I. Adverse selection in the market for insurance

We consider the market for insurance. There are many individuals indexed by \( i = 1, \ldots, N \), each facing a possible loss \( L > 0 \). Individuals differ in their "riskiness," that is, the probability of the loss \( 0 < \pi_i < 1 \) depends on the individual, but individuals are otherwise identical. Notably, they have the same initial wealth \( W_0 > L \) and the same von-Neumann-Morgenstern preferences, representable by the \textit{ex post} utility function \( u(w) = \ln(w) \). Losses occur independently across individuals.

The contractual setting is different from what we saw in class. Here, we assume that the insurance contract has a price \( P > 0 \), and that it is a take-it-or-leave-it option. That is, the choices of the individual are to either fully insure against the loss for a total price of \( P \), or to not insure at all (and not pay anything).

Note: Harder questions have a star before them. You may proceed and answer them at the end.

Demand

[0.2] 1. Define and compute the coefficient of absolute risk aversion.

**Solution:** It is \( \text{ARA}(w) = -\frac{\frac{u''(w)}{u'(w)}}{\frac{1}{w^2}} = \frac{1}{w^2} \).

[0.3] 2. Write down the two lotteries that individual \( i \) can choose from when offered an insurance contract at price \( P \). For each lottery, indicate the payoffs and the probabilities of occurrence in each state. Also indicate the expected utility under each lottery. (Hint: one of these lotteries is degenerate.)

**Solution:** Individual \( i \) can either choose \( L_0 = (W_0 - L, W_0; \pi_i, 1 - \pi_i) \) or \( L_1 = (W_0 - P; 1) \). Individual \( i \) gets expected utility \((1 - \pi_i) \ln(W_0) + \pi_i \ln(W_0 - L)\) under \( L_0 \) and \( \ln(W_0 - P) \) under \( L_1 \).

[0.5] 3. Using part 2, derive individual \( i \)'s willingness to pay for insurance, denoted \( P_i \) and defined as the highest price that will induce individual \( i \) to choose insurance over no insurance. What is \( P_i \) if \( \pi_i = 0 \)? If \( \pi_i = 1 \)? Are higher-risk individuals willing to pay more for insurance than lower-risk individuals?

**Solution:** To find the willingness to pay \( P_i \), we set the expected utility from the two lotteries equal to each other and solve for the insurance price \( P_i \): \( \ln(W_0 - P_i) = (1 - \pi_i) \ln(W_0) + \pi_i \ln(W_0 - L) \iff W_0 - P_i = \exp \left[(1 - \pi_i) \ln(W_0) + \pi_i \ln(W_0 - L)\right] \iff P_i = W_0 - \exp \left[(1 - \pi_i) \ln(W_0) + \pi_i \ln(W_0 - L)\right] \). If \( \pi_i = 0 \) then \( P_i = 0 \). This makes sense, as if there is no risk of loss the individual is willing to pay nothing for insurance. If
\textbf{Solution:} The individuals choosing insurance when it is priced at \(P\) are those with willingness-to-pay \(P_i(\pi_i)\) higher than \(P\), that is

\[
Q(P) = \mathcal{N} \Pr (P_i(\pi) \geq P) \\
= \mathcal{N} \Pr (W_0 - \exp [(1 - \pi_i) \ln(W_0) + \pi_i \ln(W_0 - L)] \geq P) \\
= \mathcal{N} \Pr (\exp [(1 - \pi_i) \ln(W_0) + \pi_i \ln(W_0 - L)] \leq W_0 - P) \\
= \mathcal{N} \Pr ((1 - \pi_i) \ln(W_0) + \pi_i \ln(W_0 - L) \leq \ln(W_0 - P)) \\
= \mathcal{N} \Pr (\pi (\ln(W_0 - L) - \ln(W_0)) \leq \ln(W_0 - P) - \ln(W_0)) \\
= \mathcal{N} \Pr \left( \pi \geq \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)} \right) \\
= \mathcal{N} \left[ 1 - \Pr \left( \pi \leq \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)} \right) \right] \\
= \mathcal{N} \left[ 1 - \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)} \right].
\]

[0.2] 6. Show that the insurance demand function in part 5 is concave in \(P\).

\textbf{Solution:} This comes directly from the concavity of the logarithm function, and can be shown formally by taking the second-order derivative of \(Q(P)\) and showing it is negative.

[0.4] 7. On a new figure (Figure 2) with insurance price \(P\) on the vertical axis and total market quantity \(Q\) on the horizontal axis, represent the insurance demand function. Indicate the value of the vertical intercepts at \(Q = 0\) and \(Q = \mathcal{N}\), and make sure your demand function has the right shape.

\textbf{Supply}

There are two insurance firms competing à la Bertrand (i.e., by choosing price \(P\) taking the other firm’s price as given). Firms have zero administrative costs of providing insurance, so that their only costs are the costs of claims filed by their customers incurring the loss \(L\). On the revenue side, insurance firms charge a flat price for each contract sold (they cannot charge higher prices to higher-risk customers, either because of regulation or because they cannot observe individuals’ riskiness). Firms are risk-neutral.

[0.2] 8. If a firm sells a contract to individual \(i\) (with loss probability \(\pi_i\)), what is the firm’s expected claims cost from that particular contract? On Figure 1, depict the expected contract cost as a function of \(\pi_i\).
Figure 2: Insurance market

Solution: The expected claims cost from a contract with individual $i$ is $(1 - \pi_i) \times 0 + \pi_i \times L = \pi_i L$.

9. Carefully explain why in equilibrium, each insurance firm must make zero expected profit and charge the same price for the insurance contract. (Hint: Assume this is not the case, and find contradictions. You are not asked to compute the equilibrium price!)

Solution: First, note that if a firm charges a high enough price, e.g. above $L$, they will get zero customers and thus zero expected profit. Therefore, firms make non-negative expected profit. Firms must charge the same price: if one firm charges a lower price than the other firm, than the low-price firm could increase its price while retaining all its customers, thereby increasing expected profit. Finally, the expected profit must be zero: if it were strictly positive, than a firm could slightly lower their price (by an $e$) and steal all its rivals' customers, thereby increasing expected profit.

Since each firm charges the same price, it is natural to assume that each firm gets half of the insured customers, and that the composition of the customer base in terms of riskiness is the same for each firm. Given part 9, in equilibrium the price of the contract must then be equal to the average expected cost of a contract in the insured population.

10. *Using your answers to part 3 and part 8, compute the total expected claims cost for the insured population as a function of the contract price $P$.
Solution: The total contract costs are the sum of the contract costs $\pi_i L$ for those individuals who have signed the contract. Those have $P_i \geq P$, that is, $\pi_i \geq \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)}$. Therefore, the total claims costs are $N \times L \times \int_{\frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)}}^{1} \pi d\pi = \frac{N L}{2} \times \left[ 1 - \left( \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)} \right)^2 \right].$

[0.3] 11. *Using part 5 and part 10, show that the average expected claims cost for the insured population (hereafter "average contract cost") is

$$AC = \frac{L}{2} \left[ 1 + \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)} \right].$$

(Hint: You will get partial credit if you explain how to combine the results of part 5 and part 10 to answer the question.)

Solution: The average contract cost is the total contract cost divided by the number of insured individuals $Q(P)$, that is, using the fact that $a^2 - b^2 = (a - b)(a + b),

$$AC(P) = \frac{\frac{N L}{2} \times \left[ 1 - \left( \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)} \right)^2 \right]}{N \left[ 1 - \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)} \right]} = \frac{L}{2} \left[ 1 + \frac{\ln(W_0) - \ln(W_0 - P)}{\ln(W_0) - \ln(W_0 - L)} \right].$$

[0.2] 12. Now using part 11 and part 5, express the average contract cost $AC$ as a function of the insured population $Q$ rather than the contract price $P$. Check that $AC(0) = L$ and that $AC(N) = \frac{L}{2}$.

Solution: Simple algebra implies that

$$AC(Q) = L \left[ 1 - \frac{Q}{2N} \right].$$

It follows that $AC(0) = L$ and that $AC(N) = \frac{L}{2}$.

[0.2] 13. On Figure 2, draw the average contract cost function $AC(Q)$.

Solution: See Figure 2.

[0.2] 14. Using Figure 2, show the equilibrium contract price. On the figure, identify the individuals who are not insured in equilibrium, if any.
Welfare analysis

It is socially desirable to insure individual $i$ whenever their willingness to pay for insurance $F_i$ exceeds the expected claim cost from the insurance contract.

[0.2] 15. Based on Figure 1, is it socially desirable to insure every individual in this population? Explain.

Solution: Yes, because for every individual the expected contract cost lies below the willingness to pay.

[0.2] 16. Does competition among insurance firms lead to a socially desirable outcome? Explain.

Solution: No, in this case it does not since some individuals who should be insured are not insured. This is because high-risk individuals adversely select into the insurance market, causing the average contract cost (and thus the insurance price) to rise to a level at which low-risk individuals no longer want to purchase insurance. If insurance companies were able to observe $\pi_i$ and to price discriminate, an efficient allocation would obtain wherein everybody is insured (but pays different prices for the insurance contract).
2. A study of total advertising effort by a cross-section of firms was conducted, by examining the monthly revenues of various media outlets, along with firms’ total sales in the same month.

Advertising by firm \( i \) was measured as the firm’s total expenditure on advertising in the current month, expressed in dollars.

Sales by firm \( i \) in the current month were also collected and expressed in dollars.

The question consisted of six parts. Grades were calculated based on 2 points per part, for a maximum of 12 points. They were then rescaled so that 4 out of 12 represented a passing grade of 2.0.

For both parts (a) and (b), the math one might do, for instance, to prove something like the fact that OLS is BLUE is completely unaffected by the circumstances introduced in these problems. A general observation worth making is that answers that included the relevant math were less likely to go astray than verbal attempts.

(a) One equation considered for estimation (using Ordinary Least Squares) was

\[
SALES_i = \beta_1 + \beta_2 ADV_i + v_i, \tag{1}
\]

with \( v_i \sim N(0, \sigma_v^2) \).

However, the observed SALES variable is measured with error:

\[
SALES_i = \text{SALES}_i^* + w_i, \tag{2}
\]

with \( w_i \sim N(0, \sigma_w^2) \), where \( \text{SALES}_i^* \) denotes the unobserved actual sales for firm \( i \).

Errors \( v_i \) and \( w_i \) are independent from each other and across firms.

State any consequences for the regression model. Would you be willing to use OLS to estimate \( \beta_1 \) and \( \beta_2 \)? What properties would you claim for your estimates? State any assumptions you require for the results you claim.

Would your answer change if \( v_i \) and \( w_i \) were correlated with each other for each firm \( i \), remaining independent across firms?
Presumably based on recalling the effects of measurement error in right-hand side variables, nearly every student concluded that OLS would be biased and inconsistent. That is a correct answer if the measurement error is in X but not if it is in Y. This is based on problem 8.8 in Dougherty's book (5th ed.). Measurement error in the dependent variable is of no real consequence for the properties of OLS.

The new error term for the regression based on an imperfectly measured SALES variable is simply the sum of the new error and the original OLS error. It can be expected to have a zero mean and a variance equal to the sum of the two variances, and to be normally distributed if the two underlying errors are uncorrelated and normally distributed.

If the two errors are correlated, it has no effect on anything but the variance of their sum.

Across all exams, there was very little in the way of analysis along the lines of writing the new equation for SALES with an error term like \( w_i = v_i + w_i \), but that probably would have revealed that OLS is essentially unaffected by the measurement error.

The measurement error increases the variance of the errors, relative to the case where \( w_i = 0 \) with certainty, but there is no other consequence for the desirable properties we typically attribute to OLS estimators. It is not really accurate to refer to this as a loss of efficiency, since an estimator based on no measurement error isn't really available.

(b) Since the advertising variable was constructed using revenue sources for media outlets in the market of interest, only large firms are included in the data set. Smaller firms' expenditures are not disclosed, because they tend to use fewer media outlets.

This means that you will observe data only from firms whose total advertising budget exceeds $100,000 per month. As a simplification, you may now assume that the SALES variable is correctly measured for those firms (i.e., \( w_i = 0 \forall i \)).
State any consequences for the regression model. Would you be willing to use OLS to estimate $\beta_1$ and $\beta_2$? What properties would you claim for your estimates? State any assumptions you require for the results you claim.

Truncation based on $X$ is not the same as truncation based on $Y$; the latter is a significant problem making OLS biased and inconsistent. With that sort of truncation, the issue is that the expected value of the error term is no longer zero; it is instead a function of $X$.

Nearly every student missed the distinction between truncating $Y$'s distribution and truncating based on $X$, and most noted that biased and inconsistent estimates would result. That is not correct.

It was occasionally suggested that the OLS estimates are not representative of what might be observed outside the range observed for $X$; that is, of course, always the case, but there is nothing about this problem that implies any different behavior outside the range observed.

It is never the case that every possible $X$ value is observed, so this point could be made about every regression and is not really relevant as an answer to this question.

For the rest of this question, you may ignore any of the estimation problems considered above.

(c) Suppose that the model now under consideration is

$$ADV_i = \beta_1 + \beta_2 SALES_i + \beta_3 SIZE_i + u_i$$
$$SALES_i = \gamma_1 + \gamma_2 ADV_i + v_i$$

The added variable $SIZE_i$ represents the size of firm $i$, measured by its total number of employees in the current month.

Demonstrate formally that OLS is not appropriate for either equation.
There is simultaneous-equations bias in this two-equation system. Most students recognized that this occurs when an $X$ variable is correlated with the error term in an equation. For instance, substituting for SALES on the right-hand side of the advertising equation demonstrates that the advertising variable is correlated with the SALES error, implying that ADV is a bad right-hand side variable in the second equation. SALES can be shown to be a bad right-hand side variable in the first equation, taking a similar approach after substituting for ADV in the second (SALES) equation.

It is the correlation of right-hand side variables with the error term for that same regression that causes OLS to be biased. Simply noting that SALES or ADV appears on both the left- and right-hand sides of an equation does not prove anything; one can always bring like terms to one side of the equation. The specific instruction was to demonstrate that OLS is not appropriate, which would require showing that the OLS estimator is biased and inconsistent, given the presence of simultaneity bias. Answers here tended to be uniformly verbal, which is acceptable when they are complete and correct.

Instrumental variables is of course the solution that comes to mind. Answers tended to lack precision here.

(d) Comment on the identification status of each equation. If feasible, indicate how you would obtain consistent estimates of the parameters, stating any assumptions you make. If it is not feasible to do so, explain why.

Most answers were vague. It was sufficient to state that SALES is exactly identified, ADV is under-identified, and SIZE represents an instrument that allows estimating $\gamma_1$ and $\gamma_2$ consistently. Most answers did not carefully distinguish the property that the instrument is uncorrelated with the error term from statements that it is uncorrelated with a particular dependent variable—such an assumption is not needed and makes no sense. The reduced-form equations pretty much guarantee that instruments are correlated with the endogenous variables—that’s why they are candidates for serving as instruments.
Indirect least-squares estimates were not necessary but (for the identified AD V equation) were acceptable. There is no instrument available for ADV in the first equation; it is under-identified.

(e) How does your answer to the previous part change if the first equation includes each firm’s lagged sales (i.e., sales of firm \(i\) in the previous month), not current-period sales? Do you require any additional assumptions? Lagged sales are predetermined at time \(t\), which breaks the simultaneity problem.

The first equation can be estimated using OLS, and then so can the second, because ADV is not correlated with \(v_i\).

One could note that lagged sales depends on \(u\) at time \(t - 1\), and that autocorrelation in the error term \(u\) plus a lagged dependent variable prevents obtaining a consistent estimator using OLS. This was not primarily an autocorrelation question, but that point is not incorrect. It is about the error term, however, not SALES specifically; as in the previous part, we of course need LSALES to be correlated with both SALES and ADV, or it would not be considered as an instrument (i.e., it would not be found in the reduced-form equations).

Simply observing that SALES and LSALES could be correlated is not really a precise statement of the problem with autocorrelation, because IV would fail if that were not the case.

(f) Finally, continuing to assume that it is lagged sales (LSALES) and not current sales that should appear in the advertising equation, you would like to test the hypothesis that lagged sales and overall firm size are not the only determinants of advertising; you believe that the number of employees who are administrative also matters.

Assuming that

\[
\text{SIZE}_i = \text{ADMIN}_i + \text{NONADMIN}_i,
\]

where \(\text{ADMIN}_i\) denotes the number of administrative employees in firm \(i\)
and $\text{NONADMIN}_i = \text{SIZE}_i - \text{ADMIN}_i$, compare your advertising equation to the two models below, and indicate how you would test the hypothesis that the share of employees in administration has no effect on advertising. Answer the question assuming that only equation (3) is available for comparison to the original equation, and then assuming that only equation (4) is available for your comparison.

$$\text{ADV}_i = \beta_1 + \beta_2 \text{LSALES}_i + \beta_3 \text{NONADMIN}_i + \beta_4 \text{ADMIN}_i + v_i. \quad (3)$$

$$\text{ADV}_i = \beta_1 + \beta_2 \text{LSALES}_i + \beta_3 \text{NONADMIN}_i + \beta_4 \text{SIZE}_i + v_i. \quad (4)$$

Is there any reason to prefer using (3) or (4)? If so, why?

First, there is no reason to prefer (3) or (4); they are identical regressions, since subtracting NONADMIN from SIZE in (4) produces (3).

An F-test using either (3) or (4) as the unrestricted model and the original equation as the restricted model would be sufficient here.

The specific hypotheses are that $\beta_3 = \beta_4$ in (3) or that $\beta_3 = 0$ in (4).
Many people blame sugar-sweetened beverages (SSBs) for growing obesity and related problems in rich and poor countries alike. Many places have imposed taxes on SSBs in an attempt to remedy these problems. Read the attached ‘Research Letter’ by Cawley, Willage and Frisvold published in the *Journal of the American Medical Association*, then address the following questions. [30 points]

a) Use a simple supply and demand graph to demonstrate the basic determinants of the pass-through of a tax on SSBs to consumers. Explain the economic factors that determine this tax pass-through rate. [4 points]

i) This rate, also called the tax incidence, depends on the relative elasticity of demand and supply. Full credit requires effective use of a simple graph and an explanation that describes how demand and supply elasticity shapes how the tax burden is divided between buyers and sellers. When supply is more elastic than demand, buyers bear most of the tax burden. When demand is more elastic than supply, producers bear most of the cost of the tax.

b) Is the fact that the location of this study is a single airport an empirical advantage or disadvantage? Explain carefully. [4 points]

i) It is an advantage for ‘internal validity’ as it creates a controlled setting in which half the airport is treated and the other is not. But it is a disadvantage for external validity since consumer behavior in airports may not generalize to other settings and to the general population. Full credit requires awareness of both sides of this trade-off with sufficient discussion to make the details of the trade-off clear.

c) What assumption(s) must hold for a difference-in-differences (DID) estimator to identify causal impacts? Use your answer to assess the authors’ use of a DID regression model to identify the causal effect of the SSB tax on retail prices. How convincing or credible do you find this use of DID? Be as specific as possible. [5 points]

i) The key assumptions here are parallel trends and no spillover effects between treatment and control groups. There are other assumptions needed for DID that are less important for this particular question. Full credit requires thoughtful argument with good support for the “how convincing” question rather than the “right” answer to this question.

d) What does it mean to cluster standard errors? Why do the authors in this case cluster standard errors at the store level? In this analysis, would you expect standard errors clustered in this way to be larger or smaller than non-clustered standard errors? Justify your reasoning. [5 points]

i) Clustering standard errors entails an adjustment to estimated standard errors for known clustering in units of observations in the data. Units of observations that are related in some way (e.g., from the same geographic area or, in this case, store across rounds of data collection) are likely to have similar residuals relative to a common model across all observations. This tends to suppress variability in residuals. If STATA
doesn't know that specific subsets of observations are related and doesn't adjust the standard errors accordingly, it will generally underestimate the magnitude of errors. The authors cluster by store since pricing decisions are likely to be similar over time in a given store location, which reduces variability in the data. Clustered standard errors are likely to be higher than unclustered standard errors.

g). The authors report a p-value of 0.002 for their February DID estimate. In the clearest and most complete terms possible, explain the statistical meaning of a p-value of 0.002 in the context of this problem. [4 points]

e). The probability of finding this result if in fact there was no relationship between the SSB tax and observed prices on the taxed side (i.e., null hypothesis) is 0.002. Could happen, but is statistically very unlikely. This leads to rejecting the null hypothesis of no effect of the SSB tax at any conventional level of statistical significance.

f). Why do you think the authors discuss 93% as the key pass-through rate finding from their study rather than their DiD estimate of 55.3%? Does this reveal a general limitation of DiD estimation or something specific to this problem? Explain. [4 points]

f). Outside of this somewhat artificial airport setting, there is not likely to be a similar strategic response by untaxed stores, which waters down the DiD estimates. This is where the external validity concerns of the context (single airport) starts to bite. They choose within taxed stores as a way to dodge this quirk of the setting and to provide an estimate that is likely to be more relevant for contexts where entire 'marketsheds' are subject to a set SSB tax.

g). Economists ask and try to answer “and then what?” questions. Suppose that the entire state of Pennsylvania adopted a 1.5 cent/oz SSB tax and that the pass-through of this tax was indeed 93%. Now, answer the "and then what?" question. Specifically, use economic concepts from consumer demand theory to predict the health effects of this SSB tax for Pennsylvania residents. [4 points]

g). Consumers will adjust their consumption of SSBs based on this price increase. Price elasticity of demand is important in this case – and if SSB consumption is subject to habit formation this could be quite high. One key question is what do consumers eat/drink instead – think substitutes and complements. Just reducing SSB consumption alone is not enough to deliver strong health benefits since consumption of other food and beverages can offset these benefits. While there are both income and substitution effects, income effects are likely small in this case.