Uncertainty, Asymmetric Information, and Market Failure

The failure of the market to insure against uncertainties has created many social institutions in which the usual assumptions of the market are to some extent contradicted.

—Kenneth Arrow (1963)

Life is full of uncertainties. Unexpected events such as a fire or an illness can dramatically lower a person’s well-being. One way to gain some protection against such eventualities is to purchase insurance. In return for paying premiums to an insurance company, an individual receives benefits in the event of a loss. Federal and provincial programs, such as unemployment insurance, worker’s compensation, and medicare, replace lost income or cover losses that are consequences of events at least partly outside personal control. Public pensions ensure that individuals do not suffer drastic declines in their incomes when they retire and that they do not “outlive their savings.” These programs, collectively referred to as social insurance, represent a large proportion of government expenditures in Canada.

Although the various programs serve different functions, they often have some of these characteristics:

• Participation is compulsory.
• Eligibility and benefit levels depend, in part, on past contributions made by the worker.
• Benefit payments begin with some identifiable occurrence such as unemployment, illness, or retirement.
• The programs are not means-tested—financial distress need not be established to receive benefits.

We begin by discussing an individual’s demand for insurance and the supply of insurance by the private sector. We then examine possible causes of market failure, and whether they provide a rationale for government involvement in social insurance.
The expected utility model is the most widely used framework for analyzing decision making under uncertainty, and we use this model to explain an individual's demand for insurance.1 Suppose Jones has wealth equal to \( W \). However, if a fire burns his home down he will face a loss of \( L \) and his wealth will be reduced to \( W - L \). Let the probability of a fire be \( \pi \), where \( \pi \) is a number between 0 and 1. Therefore his expected loss from fire is \( \pi L \) and his expected wealth is:

\[
EW = \pi(W - L) + (1 - \pi)W = W - \pi L \quad (IC2.1)
\]

It is assumed that Jones evaluates his situation by calculating his expected utility. He does this by assigning a utility index number to the situation in which a loss occurs. It will be denoted by \( U(W - L) \). The utility index number assigned to the situation where no loss occurs is denoted by \( U(W) \). Since Jones is worse off if a loss occurs, \( U(W - L) \) is less than \( U(W) \). Figure IC2.1 shows the utility indices assigned by Jones under the assumption that he uses a concave utility function to evaluate these alternative outcomes. (As will be noted later, the shape of the \( U \) curve reflects the individual's attitude toward risk.) His expected utility is defined as:

\[
EU = \pi U(W - L) + (1 - \pi)U(W) \quad (IC2.2)
\]

Consequently, the expected utility level can be represented by a point on the vertical axis that lies between \( U(W) \) and \( U(W - L) \). The \( EU \) level can be determined by the following geometric procedure. Draw a straight line between the points a and b on the \( U \) curve. Draw a vertical line from the individual's expected wealth to the line ab. The length of this vertical line is equal to \( EU \).

The expected utility model assumes that Jones will decide whether or not to purchase a fire insurance policy by comparing the expected utility with the policy to his \( EU \) without

---

1 Most intermediate microeconomics textbooks contain a more detailed discussion of decision making under uncertainty. See, for example, Pindyck and Rubinfeld (1998: ch. 5) or Varian (1987: ch. 13).
insurance. Suppose that an insurance company offered Jones a full coverage insurance policy for a premium of \( P \) dollars. Under this contract, Jones pays the insurance company \( P \) dollars before it is known whether a fire will occur or not. If a fire occurs, the insurance company will pay Jones \( L \) dollars, and he is fully compensated for his loss. If a fire does not occur, Jones does not receive any payment from the insurance company. Thus, with the full coverage policy, Jones’s wealth is \( W - P \), whether or not an accident occurs, and therefore his expected utility level will be \( U(W - P) \). The expected utility model predicts that Jones will purchase the policy if his expected utility level with the policy exceeds his expected utility without insurance coverage, or in other words if:

\[
U(W - P) > \pi U(W - L) + (1 - \pi)U(W) \tag{IC2.3}
\]

Whether or not Jones will purchase the policy depends on the premium and on his attitude toward risk.

We will begin by considering whether Jones will purchase the insurance policy at what is known as an **actuarially fair premium**, \( P_{ACT} \). For a full coverage insurance policy, the actuarially fair premium is the individual’s expected loss, or:

\[
P_{ACT} = \pi L \tag{IC2.4}
\]

If the insurance company offered Jones the full coverage policy at \( P_{ACT} \), he would have his expected wealth level with certainty:

\[
W - P_{ACT} = W - \pi L = EW \tag{IC2.5}
\]

Figure IC2.1 shows that Jones would be willing to purchase this policy because the \( U(EW) \) is greater than \( EU \). Why? Because the \( U \) curve is concave. Any cord such as \( \alpha \beta \) will always lie below the \( U \) curve, and therefore \( EU \), the vertical distance to the \( \alpha \beta \) line at \( EW \), will always be less than \( U(EW) \), the vertical distance to the \( U \) curve. Individuals, such as Jones, who use a concave \( U \) function in calculating the expected utility to evaluate risky alternatives are said to be **risk averse**. Such individuals will always be willing to purchase a full coverage insurance policy at an actuarially fair premium.

If Jones had used a utility function that was a linear function of wealth, such as \( U(W) = a + bW \) where \( a \) and \( b \) are constants and \( b \) is positive, then Jones would have been indifferent about purchasing the insurance policy at the actuarially fair premium or going without insurance. (It is left to the reader to demonstrate this using a diagram.) Such individuals are said to be **risk neutral**.

On the other hand, if Jones had used a convex utility function—such that the tangent lines to the utility curve become steeper when wealth increases—then the \( \alpha \beta \) line joining the points \( U(W - L) \) and \( U(W) \) on the \( U \) curve would lie above the \( U \) curve. In this case, the \( U(EW) \) would be less than \( EU \) and Jones would not want to purchase the full coverage insurance policy at an actuarially fair premium. Such individuals are said to be **risk seekers**.

To summarize, individuals are said to be risk averse if they are willing to purchase a full coverage insurance policy at an actuarially fair premium. Such individuals use a concave utility function in evaluating alternative courses of action that result in different levels of wealth. The fundamental characteristic of the concave utility function is that the marginal
utility of wealth, which is given by the slope of the tangent to the $U$ function, decreases as wealth increases. In other words, risk-averse behaviour is implied by diminishing marginal utility of wealth. We will assume that most individuals can be characterized as risk averse because most people are willing to purchase insurance to avoid major financial losses.

The EU model predicts that a risk-averse individual will never purchase a lottery ticket if the price of the lottery ticket is equal to, or greater than, the expected prize. (See Discussion Question 1 at the end of this chapter.) In fact, many individuals who purchase insurance also participate in unfair gambles—they purchase lottery tickets at prices that exceed their actuarially fair value. This suggests that their behaviour is inconsistent with the simple model of risk aversion that we have outlined above. However, since the amounts that most individuals are prepared to bet on unfair gambles are very small compared to their total wealth or income, and because they generally behave in ways that are consistent with risk aversion, such as holding diversified portfolios of assets, when it comes to major financial decisions, we will treat individuals as if they are risk averse.

Having established that a risk-averse individual will always purchase full insurance coverage at an actuarially fair premium, we can ask: What is the maximum premium that a risk-averse individual would be willing to pay for a full coverage policy? The maximum premium, $P_{\text{MAX}}$, leaves the individual no better off than he or she would be without insurance, and it is implicitly defined by the following equation:

$$U(W - P_{\text{MAX}}) = \pi U(W - L) + (1 - \pi) U(W) \quad (IC2.6)$$

Figure IC2.1 shows Jones’s $P_{\text{MAX}}$. Note that for a risk-averse individual, $P_{\text{MAX}} > P_{\text{ACT}}$ and that $P_{\text{MAX}}$ increases as $\pi$ increases. The wealth level, $W - P_{\text{MAX}}$, which Jones views as equivalent to his expected utility without insurance, is called his certainty equivalent wealth, CEW. For a risk-averse individual, $CEW < EW$.

The consumer surplus, $G$, from purchasing a full coverage policy at an actuarially fair premium can be defined as the difference between the maximum premium that an individual would be willing to pay and the actuarially fair premium, or:

$$G = P_{\text{MAX}} - P_{\text{ACT}} = EW - CEW \quad (IC2.7)$$

In Figure IC2.1, $G$ is the horizontal distance, $\delta \gamma$, between the $U$ function and the $\alpha \beta$ line, measured at $EU$.

The magnitude of $G$ obviously depends on the individual’s degree of risk aversion—the more risk averse, the greater the concavity of the $U$ function, the larger $G$ will be. The expected utility model also predicts that the consumer surplus from actuarially fair insurance will be very small when the losses are very small, very rare, or very common. As we will see in the next section, this has implications for the types of losses that will be covered by private insurance markets.

---

2 These predictions can be demonstrated with reference to Figure IC2.1. As $\pi$ goes to zero, $EW$ approaches $W$, $EU$ approaches $U(W)$, and $\gamma$ approaches $\beta$. Therefore the horizontal distance $\delta \gamma$ shrinks to zero. Conversely, if $\pi$ approaches 1, $EW$ approaches $W - L$, $EU$ approaches $U(W - L)$, $\gamma$ approaches $\alpha$, and therefore $\delta \gamma$ shrinks to zero. The demonstration that $G$ goes to zero as $L$ goes to zero is similar, except that the point $\alpha$ also approaches $\beta$ as $L$ goes to zero, which implies that $G$ declines at a faster rate as $L$ decreases than as $\pi$ decreases.
THE SUPPLY OF INSURANCE

Having described the demand for insurance, we will now consider under what terms and conditions private firms will provide insurance coverage. Long-run equilibrium in a perfectly competitive insurance market requires that the firms earn zero economic profit. This implies that the premium for a policy must equal the expected average cost of providing the insurance coverage. The average cost of a policy can be decomposed into three components:

\[
\text{average cost per policy} = \text{expected claim per policy} + \text{administration cost per policy} + \text{insurance risk per policy}
\]

\[
= \pi L + A + R
\]

The claims that an insurance company has to pay are a major component of its costs. The expected claim loss, \(\pi L\), is what each policy adds to the firm’s expected total claims. Note that this is equal to the \(P_{ACT}\). The administration costs or “loading costs” are the costs incurred by the firm in handling claims and billing customers. This component includes the wages and salaries, rent, and other input costs of the insurance company. \(R\) is the risk premium that the shareholders of the insurance company must be paid to compensate them for the risks they are assuming in providing the insurance policy. The premium, \(P\), charged by the insurance company, will equal \(\pi L + A + R\), and therefore because resources are used up in running an insurance company and the shareholders may have to be compensated for risking their wealth by investing in the insurance company, the premium charged by the insurance company will exceed the actuarially fair premium.

Since Jones will only purchase the insurance contract if \(P_{MAX}\) exceeds \(P\), it follows he will only purchase insurance if the following condition is satisfied:

\[
P_{MAX} = G + P_{ACT} \geq P = \pi L + A + R \quad (IC2.8)
\]

or equivalently if:

\[
G \geq A + R \quad (IC2.9)
\]

That is, Jones will only buy the insurance policy if his consumer surplus from actuarially fair insurance exceeds the administration cost per policy and the risk premium that the shareholders of the insurance company must be paid.

This presents an interesting problem. If the shareholders of the insurance company are as risk averse as the individuals who would like to buy insurance, how is it possible for insurance policies to be traded? Risk-averse shareholders need to be compensated for the risk that they incur by insuring Jones, and this compensation, \(R\), may be so large that Jones would not be prepared to pay it. The answer to this question is that there are mechanisms by which \(R\) can be greatly reduced or eliminated entirely. These mechanisms are known as risk pooling and risk spreading, and we discuss them below.

Risk Pooling and Insurance

Consider the following story. Captain Ahab sends ships to the Levant. Over many years, he has observed that in one out of four years his ships are attacked by pirates off the Barbary Coast. The attacks are random events. When his ships are attacked, he suffers a financial loss.
of $L$. Ahab, after receiving word of an attack on this year’s convoy, goes to Lloyd’s Coffee House in the City of London, to console himself. Isaac Mutant also frequents the coffee house and overhears Ahab’s fulminations against the pirates. He decides to plot a histogram of the probability of Ahab’s annual losses, which is shown in Figure IC2.2, and he calculates that Ahab’s expected loss is $(1/4)L$. Captain Bligh also hears Ahab’s stories and is very sympathetic. He sends ships to Jamaica, and by a remarkable coincidence, he also suffers a loss of $L$ in one out of four years because storms sink his ships. Bligh, in an effort to cheer Ahab, tells him about his situation. Upon overhearing Bligh’s story, Isaac Mutant decides to plot the histogram of probability of the average losses incurred by Ahab and Bligh. Mutant reasons that the losses of Ahab and Bligh are independent events since they occur in different parts of the world and have different causes. Therefore, the probability that both Ahab and Bligh would have a loss in the same year is $(1/4)^2$ or $1/16$. In that case the average loss is $L$. On the other hand, the probability that neither captain has a loss in a given year is $(3/4)^2 = 9/16$. The probability that Ahab has a loss and Bligh does not is $(1/4)(3/4) = 3/16$. Similarly, the probability that Bligh has a loss but Ahab does not is $3/16$. The average loss, when only one of the captains has a loss, is $L/2$. Mutant tells the captains that if they share the losses that are incurred, they would face the probability distribution shown in Figure IC2.3. Both captains agree that the distribution of losses in Figure IC2.3 is less risky than the distribution in Figure IC2.2.

**Figure IC2.2**
Probability Distribution for Ahab’s Losses

**Figure IC2.3**
Probability Distribution of Ahab’s Share of the Total Losses
because the probability of having to pay $L$ drops from $1/4$ to $1/16$. On the other hand, there is also a lower probability of suffering no loss and a high probability of having to contribute $(1/2)L$ if one of them has a loss. They want to know what their expected contribution will be if they share in the losses. Isaac Mutant calculates that each captain’s expected contribution when they share in the losses is:

$$\text{Expected Contribution When Losses Are Shared} = \frac{9}{16} \cdot 0 + \frac{6}{16} \cdot \frac{L}{2} + \frac{1}{16} \cdot L$$

$$= \frac{L}{4} \quad (\text{IC2.10})$$

Thus, the expected contribution when the captains share their losses is exactly the same as the expected loss that each captain faces on his own. Both captains recognize that they are better off when they share their losses because their expected payment or loss is the same as when they are on their own, but by sharing in the losses they will face a situation where there is a lower probability of incurring a loss of $L$.

At this point all of the other captains in Lloyd’s Coffee House ask to see what the probability distribution of the average loss would look like if they also shared the losses with Ahab and Bligh. Mutant refuses to draw any more diagrams, but he agrees to derive another measure of the risk that each sea captain faces. There are $n$ captains in Lloyd’s Coffee House. Let $L_i$ be the loss sustained by captain $i$ where $L_i$ is a random variable that takes on values of 0 or $L$. The probability of a loss of $L$ is $\pi$. The expected loss for each captain is $\pi L$, and the variance of his loss is:

$$\text{Var}(L_i) = \pi(1 - \pi)L^2 \quad (\text{IC2.11})$$

Mutant then notes that if the $n$ sea captains form The Sea Dogs Mutual Insurance Company, they will have to contribute $S/n$ where $S$ is the sum of their losses. The expected value for $S$ is the number of captains times their expected loss, or $n \pi L$. The expected contribution that each captain will make to the insurance pool is $\pi L$. However, the contribution will vary from year to year depending on how many of the captains experience losses. If they are extremely lucky and no captains experience any losses, their contributions will be zero. However, if they are very unlucky and they all suffer losses, each will have to contribute $L$. Thus, their contribution will vary from year to year depending on the members’ losses.

What risk would the members of The Sea Dogs Mutual face? Suppose that the members’ losses are statistically independent, then Mutant shows that the variance of each member’s contribution would equal the following:

$$\text{Var}(S/n) = \frac{\pi(1 - \pi)L^2}{n} \quad (\text{IC2.12})$$

Thus, the variance of a member’s contribution becomes smaller as $n$ increases, and if the mutual insurance company is very large, its members will not face any risk. Each member’s contribution will almost always be very close to $\pi L$. The risk of a loss of $L$ is

---

3 See the appendix to this chapter for the derivation of this equation.
eliminated through risk pooling. At this point, the captains agree to form the mutual insurance company, and they select Isaac Mutant as President and CEO.

The ability of the risk-pooling mechanism to eliminate risk has been discussed within the context of a mutual insurance company, but the same mechanism can eliminate the risk that the shareholders of a stock insurance company face from very large losses, thereby reducing the $R$ component of cost to zero if the number of policy holders is sufficiently large.

The ability of the risk-pooling mechanism to reduce risk hinges on the assumption that the losses are independent events. If the losses are positively correlated, then the risk-pooling mechanism's ability to reduce risk is impaired, but not entirely eliminated. In the extreme case, where losses are perfectly correlated—if one individual has a loss, then they all suffer losses—the risk-pooling mechanism is completely ineffective in reducing risk.

The risk-pooling mechanism operates when the potential loss from a large number of different risks can be combined. Another mechanism is also available to ameliorate the risk for a unique event.

**Risk Spreading**

The costs and benefits of a given risky project can be shared or spread over a number of individuals. If the net return on the project is uncorrelated with individuals' wealth from other sources, then the total cost of risk bearing goes to zero as the number of individuals who share in the net return on the project becomes very large. This is known as the Arrow–Lind Theorem.

Suppose that Connie could invest in a research project that may lead to a desk-top nuclear fusion reactor. The cost of the research project is $C$. If the project is successful, desk-top fusion will yield a return of $S$ and Connie's wealth will be $W + S - C$. If the project fails to produce a viable desk-top fusion reactor, the project will not generate any return and Connie's wealth will be $W - C$. The probability of success is $\pi$, and therefore Connie's expected wealth, if she invests in the project, is $EW = W + \pi S - C$. To determine whether she should invest in the project, Connie compares her expected utility with the project to her expected utility without the project. If she does not invest, her expected utility is $U(W)$. It is assumed that the project would raise her expected wealth, but that it is so risky that her expected utility with the project is less than $U(W)$. This situation is shown in Figure IC2.4. In this context the horizontal distance, $k_1$, between $EW$ and Connie's $CEW$ with the project can be considered the cost of bearing the risk for this project when she is the only investor.

Suppose Connie were to share the costs and the benefits from the desk-top fusion project with another investor, Bjorn. It will be assumed for convenience that Connie and Bjorn exhibit the same degree of risk aversion and have the same wealth, $W$, if they do not invest in the project. Both individuals will contribute half of the cost and receive half of the gross return if it is successful. Therefore, each individual's wealth would be $W + (S - C)/2$ if the project is a success and $W - C/2$ if it is a failure. Figure IC2.5 shows each individual's expected utility. Note that in this case, $EU$ exceeds $U(W)$ and therefore Connie and Bjorn would be willing to invest in the desk-top fusion research project. The key point is that by sharing in the gains and the losses from the project, the investors have reduced the cost of risk bearing. In Figure IC2.5, the cost of risk bearing is the distance $k_2$. Not only is $k_2 < k_1$ and therefore the cost of risk bearing per investor reduced, but careful inspection reveals that $2k_2 < k_1$. That is, the total cost of risk bearing has been reduced. If the project were
shared by \( n \) investors, the total cost of risk bearing would be \( nk_n \), where \( k_n \) is the cost of risk bearing to any one of the individuals. Arrow and Lind (1970) showed that \( nk_n \) approaches zero as \( n \) becomes very large.

The Arrow–Lind Theorem implies that a project whose net return is uncorrelated with the return on other assets should be undertaken if the expected net return is positive. No risk premium is required by investors in such projects because the total cost of risk bearing can be completely eliminated through the risk-spreading mechanism. This mechanism allows firms with many shareholders to undertake projects that would be too risky for an individual or a small group of investors to undertake. Lloyd’s of London provides insurance coverage for a wide variety of special risks by spreading the risk among a syndicate of “names,” wealthy individuals who agree to cover losses in exchange for a share of the premiums. Governments can also use the risk-spreading mechanism to eliminate risk if the costs and benefits of a project are spread over the population through the government’s tax and expenditure system. For example, governments can sponsor basic scientific research into finding a cure for cancer without worrying that such research activity is highly risky in the sense that it may fail to produce significant results. As long as the expected benefit from the research exceeds its expected cost, then the research should be undertaken. The
riskiness of the research can be ignored because the benefits will be shared by all citizens (if they face a reduced risk of cancer) and the costs are spread over all taxpayers.

As with the risk-pooling mechanism, risk spreading cannot be used to eliminate the risk from projects where their returns are positively correlated with labour income or the returns on other assets. In this case, the project would tend to be a success when the returns on other assets are high, and therefore the marginal utility of wealth is relatively low, and it would tend to be a failure when the returns on other assets are low and the marginal utility of wealth is relatively high. Spreading the net returns from the project across many investors in this case would not eliminate the risk premium that investors would require to hold this asset in their portfolios.

**Administration Costs**

Of course, even if the risk-pooling and risk-spreading mechanisms are used by insurance and capital markets to eliminate the need for the risk premium, \( R \), the insurance contract will not be purchased if the administration cost of the policy exceeds the consumer surplus from actuarially fair insurance. Recall that \( G \) will be small if the probability of a loss is either very high or very small, or if the magnitude of the loss is small. Under these circumstances, it is likely that \( G \) will be less than \( A \), and insurance will not be provided for these risks. This does not constitute a market failure. The cost of providing the policy simply exceeds the potential gain.

For some risks, the magnitude of a loss may be a random variable—a collision with another vehicle can result in damages that range from a minor scrape to a total write-off. The gain from being insured against a small loss may be less than the cost of administering a small claim, and therefore the insured will want a **deductible** in the insurance policy so that small losses are not insured, and these administration costs are avoided. With a deductible of \( D \), only losses in excess of \( D \) are insured. The use of a deductible in an insurance contract helps to economize on administration costs, but the policy still provides protection against large losses. The individual only has partial coverage, but this should not be construed as a market failure.

**SOURCES OF MARKET FAILURE**

Under what conditions do private markets fail to provide appropriate insurance coverage? When is the provision of social insurance by government warranted?

**Inability to Pool Risks**

As we have seen, the risk-pooling mechanism is impaired when losses are positively correlated. Should governments provide social insurance when risks are positively correlated? It is useful to consider this issue within the context of unemployment insurance because unemployment is a cyclical phenomenon. When the unemployment rate is high in one industry or region, it is usually high in other industries or regions. For example, over the period 1975 to 1990, correlation coefficients between provincial unemployment rates were 0.49 for Ontario and Newfoundland, 0.86 for British Columbia and Ontario, and 0.90 for British Columbia and Newfoundland.
years in which claims were low, and this surplus would be held as a reserve to pay for claims when the unemployment rate was high. In a major recession, claims might exceed premiums and reserves, and the insurance company would have to borrow to meet its obligations. Even a very large and well-run insurance company would be unable to borrow enough to pay all the private unemployment insurance claims during a major recession, and therefore would be forced to default on its claims and declare bankruptcy. One of the advantages that most governments have in the provision of unemployment insurance is that they have a greater capacity to borrow than even the largest private corporations. Unlike the shareholder in a private insurance company, the taxpayers’ liability is not limited. Therefore, governments can provide social insurance for losses that the private sector would not cover because of the inability to pool risks.

Increasing Returns to Scale in the Provision of Insurance

We have seen that the ability to reduce risk through risk pooling increases with the number of individuals who are insured. Therefore, at least one component of an insurance company’s average cost declines as the number of policies increases, making insurance a potential natural monopoly. However, studies of returns to scale in the insurance industry suggest that for most forms of property and casualty insurance the economies of scale from risk pooling are exhausted at relatively low volumes of insurance. The property and casualty insurance industry generally has relatively low barriers to entry and relatively low concentration ratios. For example, there were more than one hundred firms in the automobile insurance industry in Alberta in 1982, and the four largest firms had 31 percent of the market (see Dahlby, 1992). On the other hand, it is often claimed that provincial health insurance has lower average costs than comparable private insurance companies operating in the United States. These cost differences may be due to economies of scale in administration, different administration procedures (U.S. hospitals have to keep detailed records of the costs incurred in treating patients in order to bill the private insurers), or the exercise of monopsony power in the purchase of inputs in providing health care.

Asymmetric Information

To this point, we have considered insurance markets where the buyer and the seller of the insurance policy have the same information concerning the probability and the magnitude of the loss. In many situations the purchaser of the insurance policy has more information about $\pi$ and $L$ than the insurance company, and this gives rise to a situation of asymmetric information—the buyer and the seller of a product do not possess the same information about the quality of the product. Asymmetric information can occur in a wide variety of markets, but it seems to be particularly important in insurance markets. Two basic types of asymmetric information problems can be distinguished. Adverse selection occurs when the insured has more accurate information concerning his or her loss probability than the insurance company. Moral hazard occurs when the insured can influence the magnitude or the probability of a loss. The implications of these phenomena for the operation of private insurance markets are considered below.

---

5 In the early 1980s, a private company began offering “executive” unemployment insurance. Unfortunately, the company entered the market just before the 1982 recession. Within six months it was unable to pay the claims on its policies and was forced into bankruptcy.
Adverse Selection

Suppose there is a large group of risk-averse individuals. Each has wealth equal to $W$ and faces a possible loss of $L$. There are two types of individuals—high risks and low risks. The loss probability is $\pi_h$ for a high-risk individual and $\pi_l$ for a low-risk individual. Let $h$ represent the fraction of the population that is in the high-risk group. The average probability of a loss is $\bar{\pi} = h\pi_h + (1 - h)\pi_l$, and therefore $0 < \pi_i < \bar{\pi} < \pi_h < 1$. In the absence of insurance, the expected utility of a low-risk individual is $EU_l$ and the expected utility of a high-risk individual is $EU_h$ shown in Figure IC2.6.

Would a private insurance industry be able to provide full insurance coverage in this market? To simplify the analysis, it is assumed that the individuals' losses are independent events and that there are no administration costs in running an insurance company. Hence $A$ and $R$ are zero, and competition in the insurance market will ensure that premiums are equal to the expected claims. If an insurance company can distinguish high-risk individuals from low-risk individuals, it would offer two full coverage policies. Members of the high-risk group would be charged a premium equal to $\pi_hL$, and members of the low-risk group would be charged a lower premium, equal to $\pi_L$.

Now suppose each individual knows his risk group, but the insurance company cannot distinguish a high-risk individual from a low-risk individual. Clearly, the insurance company cannot continue to offer two full coverage insurance policies with different premiums because everyone would claim to be a member of the low-risk group and pay $\pi_L$. The average claim per policy would be $\bar{\pi}L$, and the insurance company would lose money on each policy. Eventually, it would have to raise its premium or go out of business. Suppose the insurance company raised its premium to $\bar{\pi}L$, which would cover the expected claims if the policy were purchased by both risk groups. Would the insurance company be able to sell this policy to both risk groups? Not necessarily. Figure IC2.6 shows a situation in which the maximum premium that the low-risk individuals would pay for a full coverage insurance policy, $P_{\text{MAX}}^l$, is less than $\bar{\pi}L$. Thus, only the high-risk individuals would purchase the full coverage policy, and the insurance company would have to raise the premium to $\pi_hL$ in order for the premiums to cover the expected losses. High-risk individuals would be willing to pay this premium for full coverage, but the low-risk individuals would not because $\pi_hL > \bar{\pi}L > P_{\text{MAX}}^l$. If $P_{\text{MAX}}^l > \bar{\pi}L$, the low risks would still prefer a partial coverage policy.

\[\text{FIGURE IC2.6} \quad \text{Adverse Selection}\]
Does this mean that the low risks would not receive any insurance coverage? Again, not necessarily. Insurance companies would recognize that there are low-risk individuals who would be willing to buy some insurance coverage as long as the premium is not driven up because high-risk individuals purchase it. Therefore, insurance companies may offer a partial coverage policy with a deductible $D$ that is large enough so that the high-risk individuals prefer the full coverage policy at $\pi_h L$ (see Rothschild and Stiglitz, 1976). However, it is possible that when the administration costs of providing insurance are taken into account, the premium for the partial coverage policy may be so high, relative to the protection that it provides, that the low risks would rather remain uninsured.

To summarize, one possible outcome of adverse selection in an insurance market is a situation where full coverage policies are offered, but the premiums are very high because they reflect the loss experience of high-risk individuals. Partial coverage policies, which expose the individual to some risk, may be offered at much lower premiums, and these will be purchased by low-risk individuals. When administration costs are relatively high, the low-risk individuals may be better off with no insurance coverage.

Before considering whether government intervention is warranted when there is adverse selection in an insurance market, we will consider some other private market responses to adverse selection. One market response is to provide insurance coverage to all the members of a particular group such as the employees of a firm. Group coverage circumvents the adverse selection problem because all of the members of the group are covered, and not just those who think that they have a high probability of a loss. Thus, group health insurance or group life insurance can be provided to all of the employees of a firm at premiums that are lower than in the open market, which is subject to an adverse selection process. The problem with group coverage is that it is only available to individuals who are members of large groups that qualify for such coverage. The employees of small firms or the self-employed cannot obtain group coverage, and this may distort individuals’ decisions concerning whether to work for a large or a small firm. It also means that individuals risk losing their insurance coverage if they are laid off by their employer.

Another response to the adverse selection problem is that insurance companies may categorize individuals according to some observable characteristic—such as age, sex, or marital status—which is an imperfect indicator of an individual’s risk type. For example, private automobile insurers charge a 20-year-old male a higher premium than a 20-year-old female because on average young males have a higher claim frequency than young females. This does not mean that all 20-year-old males are high-risk drivers or that all 20-year-old females are low-risk drivers. Each group contains high and low risks, but there is a higher proportion of high-risk drivers among the 20-year-old male population than among the 20-year-old female population. While an insurance company may not be able to distinguish who is a high or low risk, it can tell males from females, and it may decide to use these imperfect indicators of loss probability to charge different premiums to males and females. The use of these imperfect indicators helps to create more homogeneous risk groups, but it raises ethical issues regarding discrimination. The Supreme Court has ruled that sex-based automobile insurance premiums do not contravene the Canadian Charter of Rights and Freedoms because the insurance industry has demonstrated that its use of these categories is based on statistical evidence that young males have higher loss probabilities than young females. However, many people feel uncomfortable with such rationalizations and wonder whether discrimination on the basis of race or religion would also be accepted if it were supported by statistical evidence. It is well known that the elderly, on average, incur higher
health care expenditures than the rest of the population. Private health insurers would charge higher premiums to the elderly and, indeed, to anyone with a history of illness. The insurance industry’s use of information obtained from genetic testing to predict individuals’ likelihood of contracting hereditary diseases raises many important ethical and economic issues. In summary, categorizing risks and charging them different premiums would reduce the scope for insurance coverage for many groups. Individuals may prefer a social insurance scheme that does not discriminate among individuals according to some observable characteristics that are correlated with expected losses.

One of the primary characteristics of social insurance programs is that they are compulsory. This means that they are not subject to the adverse selection process that arises when individuals can use their private information regarding their loss probability to choose their insurance coverage. The welfare implications of adopting compulsory social insurance in order to overcome an adverse selection problem are examined below. Suppose that the private market provides full coverage to high-risk individuals at a premium equal to \( \pi_h L + A \), which covers their expected losses and administration costs, and that low-risk individuals do not purchase any insurance coverage. When the social insurance program is introduced, it provides full coverage for the entire population. All individuals will have to contribute, either in premiums or in taxes, an amount equal to \( \bar{\pi}L + A \) to cover the expected cost of the claims and the administration costs of the program. (It is assumed that the administration cost per policy, \( A \), is independent of the amount of coverage and the same for private and public insurance.) High-risk individuals, who previously purchased full coverage from the private sector, will be better off with the social insurance scheme because their premium or contribution will decline from \( \pi_h L + A \) to \( \bar{\pi}L + A \). Low-risk individuals will be made worse off because they prefer zero coverage to full coverage at a premium of \( \pi_l L + A \). Therefore, a compulsory full coverage insurance program will not be a Pareto improvement over the private market outcome. It will make the high-risk individuals better off, but it will make the low-risk individuals worse off. It can be shown, however, that a compulsory partial coverage policy may be able to achieve a Pareto improvement over the private market equilibrium.\(^7\)

If a compulsory full coverage social insurance scheme does not represent a Pareto improvement over the private market outcome, can it be justified on other grounds? First, consider the distributional effects of a compulsory social insurance program. Even though the high risks receive full coverage with private insurance, they are less well off than the low risks who are not insured because \( U(W - (\pi_h L + A)) < EU_l \). Otherwise the low risks would purchase the full coverage policy that is offered in the market. Thus, it can be argued that compulsory full coverage insurance improves distributional equity. Furthermore, it can be shown that with the compulsory full coverage policy the high-risk individuals gain more than the low-risk individuals lose. The per capita net gain is \( h(\pi_h - \bar{\pi}) L + (1 - h)(P_{\text{MAX}}^l - \pi_l L - A) \) where the first term is positive and represents the per capita gain to the high-risk individuals from the reduction in their premiums. The second term is negative and represents the per capita loss to the low risks from being forced to contribute to a full coverage insurance policy. Their loss is the difference between the value that they place on full coverage insurance, \( P_{\text{MAX}}^l \), and the amount that they have to contribute, \( \pi_l L + A \). By adding and subtracting \( (1 - h)\pi_h L \) to the above expression, it can be shown that the net gain

\(^7\) See Dahlby (1981).
from compulsory full coverage insurance is \((1 - h)(P_{MAX}^L - \pi L - A)\). Thus, as long as the maximum premium that a low-risk individual would pay for full insurance exceeds his or her actuarially fair premium and administration cost, there is an net social gain from compulsory insurance. Therefore, as Akerlof (1970) speculated in his seminal article on adverse selection, compulsory insurance can be justified on a cost–benefit basis when the private market is afflicted with an adverse selection problem.

Moral Hazard

A moral hazard problem arises in an insurance market when an individual can influence the probability or the magnitude of a loss by undertaking some action, and that action is not observable by the insurance company.\(^8\) Actions that affect the magnitude of an individual’s loss are called self-insurance activities. Actions that affect the probability of an individual’s loss are called self-protection activities. Examples of self-insurance and self-protection activities are given in Table IC2.1. In some cases, the actions affect both the magnitude and probability of a loss.

In analyzing the effect of moral hazard on insurance markets, we will focus on self-protection activities. We begin by examining the choice of self-protection activity in the absence of insurance.

### Self-Protection Activity in the Absence of Insurance

Suppose that an individual can reduce the probability of a loss by undertaking an activity \(x\) that costs \(r\) dollars per unit. The effect of the activity on the individual’s loss probability is shown in Figure IC2.7. If \(x\) is zero the loss probability, \(\pi(0)\), is less than one. More \(x\) always reduces the loss probability, but at a decreasing rate. The expenditure on self-protection, \(\rho x\), is made before the individual knows whether a loss will occur. The individual chooses \(x\) in order to maximize expected utility:

\[
EU = \pi(x)U(W - L - \rho x) + (1 - \pi(x))U(W - \rho x)
\]

The level of self-protection activity that maximizes the individual’s expected utility occurs where the marginal benefit from additional \(x\), \(MB_x\), equals \(\rho\). The \(MB_x\) is equal to the marginal reduction in the probability of a loss, \(MRP\), which is the absolute value of the slope of the tangent to the \(\pi(x)\) curve, multiplied by a dollar measure of the gain from

---

\(^8\) We will treat adverse selection and moral hazard as separate phenomena, but in many situations both moral hazard and adverse selection will be present in the market.
avoiding the loss, $GU$. The condition for the optimal amount of self-protection activity and the components of the $MB_x$ are expressed below:

$$MB_x = MRP \cdot GU = \rho$$  \hspace{1cm} (IC2.14)

where

$$MRP = \frac{\Delta \pi}{\Delta x}$$  \hspace{1cm} (IC2.15)

$$GU = \frac{U(W - \rho x) - U(W - L - \rho x)}{EU'}$$  \hspace{1cm} (IC2.16)

The utility gain from avoiding a loss, $U(W - \rho x) - U(W - L - \rho x)$, is converted into a dollar figure by dividing it by the expected marginal utility of wealth, $EU'$. For a risk-neutral individual, $EU'$ is a constant and $GU$ is the magnitude of the loss, $L$. Therefore, $MB_x^N$ is $MRP \cdot L$. In Figure IC2.8, a risk-neutral individual would choose $x^*$. This level of loss-prevention activity...
activity would minimize \( \pi(x)L + \rho x \), the sum of the expected loss and the cost of the loss-prevention activity. In general, it will not be optimal for a risk-neutral individual to reduce \( \pi \) to zero by increasing \( x \) even if this were possible. The amount of loss-prevention activity that a risk-averse individual will undertake is more difficult to determine because \( GU \) may be greater or smaller than \( L \). In the situation portrayed in Figure IC2.8, a risk-averse individual would choose loss-prevention activity equal to \( x^{**} \), which would exceed the amount chosen by a risk-neutral individual. However, at a higher price for loss-prevention activity, a risk-averse individual may choose a level of loss-prevention activity that is less than \( x^* \).

**Self-Protection Activity with Insurance**

If an individual is risk averse, there are potential gains from purchasing insurance. We begin by considering the case where an insurance company can observe the insured’s level of loss-prevention activity. It will be assumed for simplicity that there are no administration costs and that losses are uncorrelated. Since \( x \) is observable, the insurance contract will specify the level of \( x \), along with the deductible, \( D \), and the premium, \( P \). In this context, the optimal insurance contract will have the following specification: \( D = 0, x = x^* \), and \( P = \pi(x^*)L \). In other words, the insurance industry would offer a full coverage contract and specify that the insured has to undertake the level of loss-prevention activity that minimizes the sum of the expected loss and the cost of the loss-prevention activity. The premium would be the expected loss, given the level of loss-prevention activity \( x^* \). This implies that the optimal insurance contract when \( x \) is observable may specify either more or less loss-prevention activity than a risk-averse individual would undertake in the absence of insurance. If, in the absence of insurance, a risk-averse individual would choose \( x^{**} > x^* \), then the loss probability would increase under the full coverage policy because prevention activity would decline. The increased frequency of losses would represent an improvement in the allocation of resources in the economy because we would be using the risk-pooling mechanism, instead of loss-prevention activity, to reduce risks. For example, in the absence of insurance coverage against theft, individuals might invest in very expensive security devices to protect their property. If the expenditure on security devices exceeded the amount that would minimize the expected loss plus the cost of the devices, then society would be better off with a somewhat higher loss probability, risk reduction provided by the risk-pooling mechanism, and more resources available for other productive purposes. Although it may seem paradoxical, insurance may serve a useful role in reducing loss-prevention activity and, as a consequence, increasing the frequency of losses.

The preceding analysis was based on the assumption that the insurance company could observe the insured’s level of self-protection activity. If \( x \) is not observable by the insurance company, its value cannot be specified in the insurance contract. This leads to the problem of moral hazard. If the insurance industry continued to offer full insurance contracts, the optimal level of \( x \) for the insured individual would be 0, not \( x^* \). If all of the insured individuals reduced their provision of \( x \) to 0, the premium for full coverage insurance would have to increase to \( \pi(0)L \). With this higher premium, the full insurance contract may be less attractive than no insurance.

However, the insurance industry is not restricted to full coverage contracts. It can offer partial coverage insurance contracts, which give the insured an incentive to engage in some self-protection activity. When there is a moral hazard problem, the optimal insurance contract will be a partial coverage contract that specifies \( D \) and \( P \), but not \( x \). The insured will
choose $x$ to satisfy the condition $MB_x = \rho$. A larger deductible will induce the insured to choose more loss-prevention activity (thereby reducing expected losses and allowing the insurance company to offer the policy at a lower premium), but a higher deductible will also impose a larger financial penalty on the insured if a loss occurs. With the optimal deductible, the additional gain from providing the insured with a greater incentive to engage in self-protection activity equals the additional loss in expected utility from reducing insurance coverage by increasing the deductible by one dollar.

Does moral hazard cause market failure? Is social insurance warranted when an insurance market is affected by a moral hazard problem? These issues are complex and have not been fully resolved by economists. First consider the question of whether moral hazard provides a justification for social insurance. It should be emphasized that, if a moral hazard problem exists, it will affect the public sector as much as the private sector. In general, the public sector does not have any advantage over the private sector in monitoring the loss-prevention activity of the insured. If the public sector provides social insurance, it should be a partial coverage, so that individuals have some incentive to reduce expected losses. The optimal deductible under social insurance would be determined according to the same criteria as under private insurance. Therefore, moral hazard does not provide a strong rationale for the provision of social insurance.

There are, however, other forms of government intervention that may improve the private sector’s allocation of resources when there is a moral hazard problem. Arnott and Stiglitz (1986) showed that a Pareto improvement in the allocation of a resource is possible if the government can subsidize the price of loss-prevention activity. The insured will engage in more loss-prevention activity, thereby permitting the insurance industry to increase its insurance coverage and moving the allocation of resources closer to the full insurance contracts that would be provided in the absence of the moral hazard problem. Similar arguments can be used to justify other forms of government intervention that affect individuals’ loss-prevention or self-protection activity. In terms of the examples presented in Table IC2.1, the allocation of resources in the economy may be improved by enforcing regulations that sprinkler systems be installed in some types of buildings, by subsidizing the cost of providing unemployed workers with information on job vacancies at government employment centres, by subsidizing the provision of recreation facilities, by enforcing speed limits and the wearing of seat-belts, and by taxing alcohol and cigarettes. A more complete discussion of the welfare implications of moral hazard is contained in Arnott and Stiglitz (1990).

### CONCLUSIONS

We began this chapter by noting that many important government programs—unemployment insurance, health insurance, and public pensions—provide insurance coverage. Social insurance programs have four primary characteristics: they are compulsory, financed by contributions, triggered by undesirable events, and not means-tested.

To understand the rationale for social insurance and the context in which social insurance programs operate, we have reviewed an economic model of the demand for insurance.

---

9 For example, if an unemployed worker can increase his or her probability of finding a new job by devoting more effort to searching for a job, then the optimal unemployment insurance system should not replace all of the unemployed worker’s lost earnings.
We have seen that if individuals are risk averse they are willing to pay more than their expected loss for a full coverage insurance policy. The private insurance industry can provide insurance coverage for some losses, especially if the collective risk can be reduced or eliminated through risk pooling or risk spreading. However, it may not be worthwhile to insure some risks if their administration cost is high, the loss is relatively small, or the probability of a loss is either very high or very low.

The private sector may not be able to provide insurance coverage when risks are highly correlated. In cases such as unemployment insurance, the public sector may be able to provide a form of insurance because of its superior ability to borrow in the event that there are very high losses.

Asymmetric information may also provide a rationale for social insurance. Adverse selection occurs when individuals differ in their loss probabilities, but insurance companies cannot distinguish high-risk individuals from low-risk individuals. As a consequence, many low-risk individuals may purchase little or no insurance coverage because the presence of the high risks drives up the insurance premiums. Adverse selection also gives rise to segmented insurance markets where some individuals can obtain group coverage, usually through their employer. Other individuals are assessed premiums based on observable characteristics, such as age, sex, or claims history, which are imperfect indicators of an individual’s risk group. Compulsory social insurance programs solve the adverse selection problem, but full coverage policies will generally make the low-risk individuals worse off and benefit high-risk individuals. However, this may be viewed as acceptable on equity grounds.

Moral hazard occurs when the insured can affect the magnitude or the probability of a loss, but the insurance company cannot directly monitor the individual’s actions. The optimal insurance contract will be a partial coverage contract, which provides the insured with some incentive to reduce expected losses, but that does not expose the insured to large financial penalty if a loss occurs. Moral hazard does not provide a strong rationale for compulsory social insurance, but it may influence the design of social insurance programs, and it justifies the use of regulation, taxes, and subsidies to influence the behaviour of individuals who are covered by public or private insurance.

It is possible that some social insurance programs are motivated by the perception of market failure. However, social insurance programs may also be desired as an instrument for achieving certain distributional goals. In the next chapter, we examine some of the conceptual issues concerning the design of income redistribution programs, and then consider the use of other instruments, such as in-kind transfers and social insurance, to supplement the more traditional transfer mechanisms.
• To economize on administration costs, insurance policies may contain deductibles so that small losses are not covered.

• The ability of the risk-pooling mechanism to eliminate risk is impaired if the losses are positively correlated.

• Social insurance may be a response to market failure. The main reasons why the private sector may fail to provide an appropriate level of insurance are:
  • an inability to pool risks;
  • increasing returns to scale in the provision of insurance; and
  • adverse selection.

• Adverse selection is the problem of hidden information. It occurs in an insurance market when:
  • the population is not homogeneous with respect to the probability of a loss;
  • each individual knows his or her own loss probability; and
  • an insurance company does not know an individual’s loss probability.

• With adverse selection, the bad risks drive up insurance premiums. The good risks respond by purchasing insurance policies with low coverage, i.e., high deductibles. The policies with high deductibles are less attractive to high-risk individuals, and therefore low-risk individuals can purchase them at relatively low premiums, which reflect their low-loss probability. Some low risks may choose to go without insurance coverage.

• The adverse selection problem may also cause private insurance companies to categorize individuals according to some observable characteristic, such as age, sex, or marital status, which is an imperfect indicator of an individual’s risk type.

• The adverse selection problem may justify compulsory social insurance, but it is likely that compulsory full coverage insurance would make the low-risk individuals worse off.

• Moral hazard is the problem of hidden action. A moral hazard problem arises in an insurance market when:
  • the individual can influence the probability and/or the magnitude of a loss by undertaking some action; and
  • an insurance company cannot observe the individual’s action.

• If the loss-prevention activity could be monitored by an insurance company, the optimal full coverage insurance contract would specify that the insured undertake the level of activity that would minimize the expected loss and the cost of the loss-prevention activity. In the absence of insurance, a risk-averse individual may engage in more or less loss-prevention activity than the level that minimizes the expected loss and the cost of the loss-prevention activity.

• If the insurance company cannot monitor the individual’s loss-prevention activity, the insurance company will offer a partial coverage insurance policy that provides the insured with some incentive to engage in loss-prevention activity. The level of insurance coverage will be determined by the trade-off between providing the insured with greater security from losses and with an incentive to engage in some self-protection activity that will reduce expected losses.

• Moral hazard does not in itself justify social insurance programs, but it may provide a rationale for regulations, taxes, or subsidies that may modify the insured’s behaviour and affect the probability or the magnitude of losses.

**DISCUSSION QUESTIONS**

1. Suppose that an individual, with wealth equal to $W$, is offered a lottery ticket where the prize is $1,000,000 (tax free). Suppose that the probability of winning the prize is 1/100,000 if one purchases one ticket.
   a. Calculate the actuarially fair price for this lottery ticket.
   b. Show that a risk-averse individual would never be willing to purchase a lottery ticket if its price is equal to or greater than its actuarially fair price.

2. Suppose that two individuals have the same $W$ in the absence of an accident, have the same degree of risk aversion, and face the same expected loss if an accident occurs. Individual 1 faces a larger loss than individual 2, $L_1 > L_2$, but with a lower probability such that $\pi_1 L_1 = \pi_2 L_2$. Using a diagram, show that with an actuarially fair insurance policy individual 1 obtains a larger consumer surplus than individual 2.

3. Three individuals each face a one-third probability of suffering a loss of $L$. Suppose they form a mutual insurance company and agree to share their losses. Calculate their expected contribution to total loss and plot a histogram of the probability distribution of each individual’s contribution if the losses are independent events.
4. Suppose that individuals use the following utility function to evaluate their expected utility in the absence of insurance:

\[ EU = \pi \ln(W - L) + (1 - \pi) \ln(W) \]

Each individual’s wealth, \( W \), is 100, and the magnitude of a loss, \( L \), is 30.

a. Calculate the actuarially fair insurance premium for a full coverage policy if the loss probability, \( \pi \), is 0.10.

b. Calculate the maximum premium that this individual would pay for a full coverage policy.

5. Explain why the private sector may not provide full coverage insurance at an actuarially fair premium for the following types of insurance coverage:

a. Automobile collisions.

b. Medical malpractice.

c. Home (including fire and theft).

6. The federal and provincial governments provide farmers with crop insurance. What problems or issues may explain why governments, rather than the private sector, provide crop insurance?

REFERENCES


APPENDIX

Derivation of Equation (IC2.12)

If \( X \) and \( Y \) are random variables and \( a \) and \( b \) are constants, then:

\[ \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X,Y) \]

where:

\[ \text{Cov}(X,Y) = E((X - EX)(Y - EY)) = \rho_{xy}(\text{var}(X)\text{var}(Y))^{1/2} \]

and \( \rho_{xy} \) is the correlation coefficient between \( X \) and \( Y \). For the mutual insurance company, \( \text{Cov}(L_i, L_j) = \rho_{ij}\pi(1 - \pi)L^2 \). Suppose that the individuals’ losses are statistically independent, i.e., \( \rho_{ij} = 0 \). Consequently,

\[ \text{Var}(S/n) = \sum_{i=1}^{n} \frac{(1/n)^2 \text{Var}(L_i)}{n} = \frac{(1/n)^2 n\text{Var}(L_i)}{n} = \frac{\pi(1 - \pi)L^2}{n} \]

Note that if \( n \) is large, \( \text{Var}(S/n) \) becomes small. Thus, if the mutual insurance company is large, its members will not face any risk. Uncertainty is eliminated through risk pooling.