THE WELFARE ECONOMICS OF INSURANCE CONTRACTS

THAT PAY OFF BY REDUCING PRICE

by

John A. Nyman

Discussion Paper No. 308, April 1999

University of Minnesota
April 22, 1999

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, MN 55455

Introduction

For the purposes of this paper, moral hazard is defined as the change in consumption that results from becoming insured when the insurance contract pays off by reducing the price of the insured commodity.[1] In the case of "health" insurance, moral hazard most commonly refers to the change in health care utilization or spending that results from becoming insured under a contract that reduces the price of health care. Early health insurance contracts typically paid off by reducing the price to zero, but modern contracts typically establish a percentage reduction in price (the coinsurance rate) or a more complex payoff scheme that includes such a percentage reduction in price in addition to other parameters. This analysis focuses on that portion of any insurance contract that pays off with any reduction in price.

The welfare implications of moral hazard have been analyzed by Pauly (1968). In what has become one of the most influential articles in the health economics literature, Pauly characterized the welfare loss as the difference between (1) the Marshallian consumer surplus for the increased consumption at the reduced price, and (2) the cost of that increased consumption, at the pre-insurance prices, assumed to reflect the marginal cost of producing it. Pauly's analysis assumes that consumption of medical care is not related to income.[2] More recent empirical evidence, however, suggests that medical care consumption is indeed related to income.[3] This paper presents the welfare implications of moral hazard under the latter assumption. It suggests that a portion of moral hazard is due to income transfers and that this portion of moral hazard should be excluded from the welfare loss calculations. Only the portion of moral hazard that is due to the pure price effect has conventional welfare loss implications.

Concomitant to making these arguments, this paper presents a taxonomy of price change

---

[1]A more general definition is that moral hazard is any change in behavior that occurs as a result of becoming insured. This definition would then encompass the concept of moral hazard qua "self protection," which is a reduction of the probability of loss (Ehrlich and Becker, 1972). This paper takes the probability of loss, or more generally, the probability of the event that triggers the payoff as given exogenously. Instead, we focus on the aspect of moral hazard that Pauly (1968) addressed: the change in consumption that results from purchasing an insurance contract that pays off by reducing price.

[2]Pauly writes, "[s]uppose there are no significant income effects on the individual's demand for medical care resulting from his payment of a lump-sum premium for insurance." (Pauly, 1968, p. 534). In other word, the loss of disposable income based on the premium payment has no effect on health care spending. He is silent, however, regarding the effect of income transfers made through insurance *payoffs* on health care spending.

[3]Manning and Marquis (1996) find an overall income elasticity of .22 using the RAND health insurance experiment data. Income elasticities close to 1 are found in studies using cross country data (e.g., Parkin, McGuire and Yule, 1987).

decompositions. In this taxonomy, a demand curve is identified that has not yet appeared in the literature. This new demand curve is used to compare the welfare loss associated with the pure price effect of insurance with the welfare loss associated with Marshallian demand that was used by Pauly.

This paper has six additional sections. In the next section, a simple health insurance example is presented to motivate the issue and show how income enters the insurance problem. This example shows that the increase in consumption that occurs when a consumer becomes insured (that is, buys an insurance contract that pays off by reducing price of health care) is related both to the effect of facing a reduced price for the insured commodity and to income transfers from those who remain healthy to those who become ill. In the following section, the conditions under which income transfers are removed from the insurance contract are discussed. Next, the taxonomy of price change decompositions is presented. In the following section, the welfare loss from moral hazard is compared with Pauly's welfare loss. Following that, the magnitude of the welfare loss is estimated, and in the conclusion, the implications for public policy are discussed.

The Problem

The following example illustrates the underlying problem of using the Marshallian demand to measure the welfare loss from moral hazard when medical care spending increases with income.

Assume a large population of 10,000 consumers, each with identical preferences and incomes of $40,000. The probability that each consumer will become ill is 1/10, meaning that 1 in every 10 of all 10,000 insured consumers becomes ill. The probability of illness is fixed exogenously and cannot be influenced by the insurer or the insured.

Without insurance, if a consumer becomes ill, he would consume $10,000 of medical care, leaving them $30,000 to spend on other goods and services. The original price of medical care is assumed to be $1 per unit, so 10,000 units are purchased. The $1 price is assumed to equal the marginal cost of producing each additional unit of medical care that the consumer would purchase, the consumer being small relative to the size of the market.

With insurance that pays off by reducing the price from $1 per unit of medical care to $0 per unit, each ill consumer who is insured consumes $20,000 worth of medical care (that is, 20,000 units) after paying a $2,000 premium. The $2,000 premium was determined by an actuarial study conducted by the insurer and is actuarially fair because, given a probability of illness of 1/10 and expenditures of $20,000, each insured consumer is expected to spend $2,000. All 10,000 insured consumers pay this premium.

The consumer's situation is described by Figure 1. Each consumer who is insured and remains healthy has disposable income of $38,000 after paying the actuarially fair premium of $2,000 [$40,000 - $2,000 = $38,000]. Because they remain healthy, it is assumed they choose not to consume any medical care, so their post-premium income position is a point on the Y axis indicating that no medical care is purchased and $38,000 is spent on other consumption.

Each person who is insured and becomes ill consumes $20,000 of medical care, paid for entirely by insurance. Of this $20,000, $2,000 represents the ill consumer's contribution to the insurance pool and $18,000 represents income transfers from those who remain healthy. That is, if the probability of illness is 1 in 10, for every 1 consumer who becomes ill, there are 9 others who each transfer $2,000 through the insurer to the consumer who becomes ill. As a result, the ill consumer now is consuming at total of $58,000 worth of both medical care and other consumption, but had only $40,000 to spend without insurance. The only way that this person can spend $58,000 with a original budget of $40,000 is because of a transfer of $18,000 in income from those 9 (out of every 10 consumers) who remain healthy. The price of obtaining medical care has not fallen, so there is no increase in real income to allow the ill consumer to consume beyond his budget. The price that someone must pay to obtain this care remains at $1 per unit, its marginal cost.

The total medical care spending of $20,000 with insurance represents $10,000 more in medical care spending than the $10,000 of medical care the consumer would have purchased without insurance. This $10,000 increase in spending represents the moral hazard. If health care demand is unrelated to income, all of this increase in consumption would be related to the price change. If, however, health care demand is related to income, a portion of this additional $10,000 in spending is explained by the $18,000 in income transfers from those who remain healthy. The increased spending on health care that occurs as a result of this income transfer should be excluded from the welfare calculations, as is typically done with income transfers, because the gain to one individual ($18,000 in this case) equals the loss to others ($2,000 by each of 9 other consumers). Only that portion of medical care spending that remains and is due to a pure price effect has welfare implications.

The amount of transfers depends on the probability of illness, an exogenously determined parameter independent of price. In this example, a .1 probability of illness resulted in a transfer of $18,000. A lower probability, say .01, would have resulted in a greater amount being transferred, and a higher probability, say .5, in smaller transfers. The amount of income being transferred determines the consumption of medical care, if consumption is responsive to income. (In Figure 1, the effect of various transfer amounts on medical care consumption is traced by the income expansion path, II'.) Therefore, for an insurance contract that pays off with any given reduction in price, there is another factor, the probability of illness--completely independent of price--that determines the amount transferred and, in turn, the amount of medical care consumed. The portion of total moral hazard that is determined by income transfers (made possible by a probability of illness less than 1) must be excluded from the welfare loss calculation.

The pure price effect of insurance (that pays off by reducing price) is the change in consumption of medical care that would occur if a consumer *who is already ill* purchased a contract from an "insurer" to reduce the price of medical care in return for an actuarially fair premium. In other words, the effect of income transfers on increasing medical care consumption is eliminated when income transfers are eliminated, and this would only occur if the probability of illness were 1. In Figure 1, this decomposition of moral hazard would correspond to point A: the point where the income expansion path II' at the new prices intersects the original budget constraint. At point A, the pure price effect of insurance is the increase in medical care consumption from 10,000 units to M_A units, and the portion of moral hazard due to income transfers is from M_A units to 20,000 units.

Income Transfers in Insurance

This section lists the various steps in building an insurance contract that pays off by reducing price. At each step, the welfare implications are evaluated and whether the contract is viable--meaning whether the contract would be purchased--is discussed. Although it is presumably possible to begin with any characteristic, this discussion begins with the purchase of a reduced price.

Purchasing a reduced price. Consider a consumer facing the following familiar utility maximization problem:

$$\begin{aligned} &\max U = U(X,Y) \\ &\text{s.t. } Y^o = PX + Y \end{aligned} \qquad (1)$$

where X is a good,,Y is the numeraire, and P is the price of X in terms of Y. Assume that $\partial X/\partial P < 0$ and $\partial X/\partial Y^o > 0$, and that P represents the marginal cost of producing X. The consumer solves this problem by consuming (X^*,Y^*) as illustrated in Figure 2.[4]

Now consider a consumer who wants to purchase a contract from a firm for a reduced price of X, cP, where $0<c<1$. The consumer pays fee R to the firm to cover it costs and the firm wants to break even. There are no transactions or administrative costs. The firm conducts a demand study to determine the fee it must charge to satisfy these conditions, given the consumer's behavior. The consumer's problem is

$$\begin{aligned} &\max U = U(X,Y) \\ &\text{s.t. } Y^o - R^c = cPX + Y. \end{aligned} \qquad (2)$$

Although R^c is taken as a given by the consumer, the study conducted by the firm ensures that the consumption bundle, (X^c, Y^c), solves the above problem and the firm's breakeven constraint, $R^c = (1-c)PX$, simultaneously. Because the constraint must satisfy both $Y^o - R^c = cPX + Y$ and $R^c = (1-c)PX$, the consumer is maximizing utility by operating on his original budget constraint, $Y^o = PX + Y$, as is shown in Figure 2.

---

[4]For ease of exposition, consumer equilibria are identified only by the X values in Figure 2, The budget constraints are generally identified by their Y intercepts.

The welfare implications of such a contract can be determined by the compensating variation for the consumer. In order to purchase this contract and still be on the original indifference curve, the consumer would need to be paid $(Y^1 - Y^2)$ where $Y^2 = Y^O - R^c$. Thus, $(Y^1 - Y^2)$ is the welfare loss from such a contract. This welfare loss reflects the inefficiency caused by purchasing too much X, that is, purchasing additional units of X that were not worth the Y that was required to obtain them. This is not a viable contract because, although the firm breaks even, the consumer's utility declines.

Income transfers. Now, consider the consumer's problem if n other consumer pay the consumer's fee. The consumer now sees his fee reduced to 0 and will purchase more X because of this transfer of income and because $\partial X/\partial Y^o > 0$. The amount transferred depends on the consumer's solution to the following problem,

$$\begin{aligned} &\max U = U(X,Y) \\ &\text{s.t. } Y^o = cPX + Y. \end{aligned} \tag{3}$$

The consumer solves this problem at (X^R, Y^R) in Figure 2, and the other n consumers pay the firm $R^R = (1\text{-}c)PX^R$ to cover the consumer's uncovered costs. The consumer's compensating variation gain is the parallel shift of budget constraints from Y^2 to Y^o, but each of the n other consumers, who share the cost of the contract evenly, must reduce their income by 1/n of R^R. Unless we know something more about the utility functions of the consumers (this assumption is relaxed shortly), we must conclude that the welfare effects of such a transfer cancel each other out--one consumer gains exactly what n consumers lose--and the net welfare effect on society is the original welfare loss $(Y^1 - Y^2)$. Also, unless the other consumers are motivated by altruism, this is not a viable contract: none of the other utility maximizers would agree to reduce their income in exchange for zero gain.

Note that in Figure 2, the new problem for which (X^R, Y^R) is the solution appears to reflect the results of an exogenous decrease in price, but no price decrease has occurred. The marginal cost of producing X is still P. The purchase of a lower price has led to a welfare loss caused by the purchase of additional units of X, but these units were not worth the Y required to purchase them. The transfers, however, are welfare neutral. They reflect an increase in income to the consumer but a decrease to others of the same aggregate magnitude. Although $X^R > X^c$, the additional X consumed has no welfare implications because it reflects purchases made that are due to income transfers.

To see that the purchases due to income transfers do not result in an inefficiency loss, consider reversing the order of the first two steps. If the order is reversed--first a transfer of sufficient income to obtain (X^R, Y^R), that is, a shift to budget line intersecting the Y axis at Y^T and then the purchase of price cP--it would be clear that the transfer of income had no inefficiency effects because each additional unit of X that is purchased would be purchased at the correct price P, a price reflecting the marginal cost of producing it. The welfare loss would occur in the second step when, after this transfer, the consumer purchases a price of cP. The welfare loss for this second step can again be evaluated by compensating variation: the amount of income required to place the consumer on the indifference curve after the income transfer. Even though this welfare loss may differ in magnitude from the original loss because preferences may change at higher income levels, it is derived from the same purchased price effect and has nothing to do with the income transfer.[5]

Income transfers with portion paid by consumer. Next, consider the consumer's problem if the consumer pays $1/(n+1)$ of the fee and the n other consumer pay the remaining $n/(n+1)$ portion. The consumer now sees his fee as R^n and will purchase less X than in the above case because of this fee. The amount transferred depends now on the consumer's new optima for the following problem,

$$\begin{aligned} &\max U = U(X,Y) \\ &\text{s.t. } Y^o - R^n = cPX + Y. \end{aligned} \tag{4}$$

The consumer pays the firm $R^n = [(1-c)PX^n]/(n+1)$ and solves this problem at (X^n, Y^n) in Figure 2. The other consumers pay the firm a total of nR^n to cover the consumer's uncovered costs. The consumer's compensating variation gain is $Y^3 - Y^1$ but each of the n other consumers, who share the remaining cost of the contract evenly, must again reduce their income by R^n. Again, unless we know something more about the utility functions of the consumers, we must conclude that the welfare effects of such a transfer are a wash and the net welfare effect of this contract on society is the original welfare loss. Similarly, unless the other consumers are motivated by altruism, this is not a viable contract: none of the other utility maximizers would agree to reduce their income in exchange for no gain with certainty.

Lottery with diminishing marginal utility of income. Consider next a lottery where all $n+1$ consumers have an equal chance, $\pi = [1/(n+1)]$, of winning a contract where the winner is paid off by a reduced price of X, cP, and the losers receive nothing. Further, assume each consumer is identical to the original consumer in budget and preferences. Finally, assume that each loser chooses only to purchase Y, but the winner maximizes utility by purchasing both X and Y. The payment, R^n, to enter this lottery is actuarially fair, meaning that it reflects the expected payout as determined by an actuarial study. Thus, each consumer expects to solve problem (4) with a probability of π, and purchase $(Y^o - R^n)$ worth of other goods with a probability of $(1-\pi)$. Expected utility is therefore,

$$\begin{aligned} EU &= \pi U[X,Y|cP,(Y^o - R^n)] + (1-\pi)U[Y|(Y=Y^o - R^n)] \\ &= \pi U(X^n,Y^n) + (1-\pi)U(Y^o - R^n). \end{aligned} \tag{5}$$

If it is assumed that the marginal utility of income is declining, then the income transfers that allow the winner of this lottery to consume (X^n,Y^n) results in a smaller increase in utility than

[5]The difference is analogous to the difference between compensating variation and equivalent variation in the welfare literature.

the aggregate losses of utility by the n losers of the lottery. This would not be a viable contract because imbedded within it are two utility losses: (1) a purchased price that decreases utility and (2) a transfer from n consumers to the 1 winner that decreases aggregate utility. Thus, a double welfare loss would occur.

Lottery with state dependent utility. Finally, if it is assumed that winning the lottery coincides with a change in the consumer's state that alters the utility function to V(X,Y), such that the same mix of goods are purchased but the value of that mix exceeds the value of the mix under the original state where preferences are U(X,Y), then it is possible that a net welfare gain would occur. Expected utility is again

$$EU = \pi V(X^n, Y^n) + (1-\pi)U(Y^o - R^n).$$

The effect of the income transfer could increase aggregate utility--the losses of income to the n losers result in a utility decrease that is smaller than the expected gain in income to the winner in this state of the world. There is still the welfare loss from purchasing the lower price, but the welfare effect of the transfers may result in a smaller net loss or a net gain. In the latter case, the contract is viable.

Insurance that pays off by reducing the price. The reader may recognize that the last case also represents a model of a actuarially fair health insurance contract, where X is medical care and the insurer pays off the consumer who becomes ill by reducing the price of care from P to cP.

The probability of illness, π, is a parameter that is exogenous to the insurer's or consumer's decision and represents the proportion of consumers in the insurance pool who become ill. For example, if the probability of illness is 2/7, this implies that for every 7 consumers in the pool, 2 will become ill on average and receive the income transfer. It also determines the number of consumers who remain healthy and transfer income to the ill. If the probability of illness is 2/7, 5 of 7 will finance the income transfer. Larger πs result in smaller transfers, and smaller πs, in larger transfers. If π is 1, no transfers would occur at all. Because the probability of illness determines the amount that is transferred, and that in turn determines the amount of medical care consumed, the probability of illness indirectly determines the amount of medical care consumed. The additional medical care consumption that is due to transfers is a component of moral hazard and is only eliminated when transfers are eliminated. This only occurs if π is 1.

Aside from this transfer, there is a loss that is due to the insurer's paying off by reducing the price. The price reduction does not occur exogenously as a result of some fundamental change in supply or demand reducing the marginal cost price, but must be purchased. The marginal cost of producing medical care remains the same and this represents the price that someone must pay to obtain this additional consumption. This, of course, is Pauly's original point. The refinement from this paper is the recognition that consumption due to transfers does not enter the welfare loss calculations. If transfers are eliminated, the effect of the price reduction is smaller because the consumer would need to pay the entire cost of this additional consumption himself. With insurance, this cost is dispersed among the many insured consumers, so that it appears that the consumer does not bear this cost. Yet the cost that insured consumers bear is the expected cost of the entire medical care spending: (1) a portion of this is the medical care spending that would have occurred without insurance, (2) a portion is the increased spending on medical care and other goods and services that is due to income transfers, and (3) a portion is the increased medical care spending due to purchasing a contract for a reduced price from that type of insurance contract. The welfare gains from insurance theoretically derive from the first two, the loss only from the last.

As defined above, moral hazard is the increased consumption of medical care that occurs when a consumer becomes insured. Thus, welfare effects from moral hazard derive from the latter two sources of spending: the transfers and the price reduction. Therefore, not only is the price effect loss smaller than previously represented, but the loss due to moral hazard is a net effect of this smaller price loss plus a potential income transfer gain represented by the increased consumption of medical care due to income transfers. Thus, moral hazard might result in a net gain.

These gains and losses are typically described differently in the literature. Typically, the welfare loss from the price effect is represented by the same welfare loss as would occur from an exogenous decrease in price and any effect of income on the quantity of medical care demanded--either through the reduction of income due to the payment of a premium or the increase in income due to the transfers from the healthy when the insurer pays off by reducing price--is ignored or deemed negligible. The gain from the income transfer is often modeled ignoring the effect of the income transfer on the purchase of medical care altogether. The gain from income spent on medical care purchases is ignored, and the welfare implications are only captured by the risk reduction in purchasing other consumption.

Before estimating the net moral hazard welfare loss, the theory for representing the welfare loss due to the purchased price effect must be developed. To evaluate this loss, a new demand curve is identified. We turn now to a discussion of this demand curve and how it fits into the existing literature on price change decompositions.

Digression on Price Change Decompositions

The implication of a price change is an issue of fundamental importance in economic theory. It has long been recognized that an exogenous price decrease causes an increase in quantity demanded that is in part due to an increase in real income. In order to isolate the pure effect of reducing the price on consumption, it is necessary to hold real income constant. The existing literature offers two alternatives for accomplishing this.

Original indifference curve held constant. Consider a consumer who is originally at equilibrium consuming bundle O in panel A of Figure 3. The price of good X falls as a result of some exogenous change in the market, the price of Y held constant. The consumer purchases bundle M in equilibrium, following that price decrease. Some of this increase in the consumption of good X is due to the increase in real income, some to the pure decrease in price.

This familiar Hicks decomposition isolates the pure price effect of a price decrease by removing sufficient income to place the consumer on his original indifference curve. Figure 3 shows this decomposition at point H. Thus, of the total change in quantity X_O to X_M, a portion, X_O to X_H is due to the pure price effect, and another portion, X_H to X_M, is due to the increase in real income.

The intuition behind the Hicksian demand curve is derived from the willingness to pay for good X in terms of good Y, represented by the slope of the indifference curve that goes through point O (Mishan, 1988). The slope changes as the consumer purchases more of good X. At point O, the consumer is willing to pay exactly the old (higher) price for a unit of X, so the Hicksian and Marshallian demand coincide at point O. As more X is purchased, the willingness to pay decreases, until at point H, the consumer's willingness to pay exactly equals the new (lower) price, P_N. The area under the Hicksian demand curve is the integral of the willingnesses to pay for the various units between X_O and X_H. Indiscrete terms, to purchase unit X_{O+1} of good X, the consumer is willing to pay ΔY_{O+1} of good Y and still be on the original indifference curve. Because for all those additional units of X, the consumer is charged exactly the new price, the difference between the willingness to pay at each successive unit and the price is a measure of the Hicksian consumer surplus. Hicksian demand is re represented by curve H in panel B of Figure 3.

"Apparent real income" held constant. An alternative approach to removing real income is identified by Friedman (1962), who attributes the decomposition to Slutsky (1915). This decomposition isolates the pure price effect by removing income sufficient to allow the consumer to purchase the original bundle of goods, point O, at the new prices, P_N. Friedman referred to this concept as holding "apparent real income" constant. Thus, if the consumer is obliged to maintain an income level sufficient to purchase original bundle O at the new prices, he will purchase good X until the willingness to pay equals the new prices, that is, at point S in panel A of Figure 3.

The intuition of the Slutsky demand can be understood by comparing it to the more familiar Hicksian demand. The Slutsky demand curve at price P_O would be represented by the willingness to pay at point O, as determined by the slope of the original indifference curve. This point would coincide with the Hicksian and Marshallian demand curves. As noted above, if the price were to fall to P_N, the willingness to pay for the X_{O+1} unit under Hicks is represented by ΔY_{O+1}. In contrast, the willingness to pay for unit X_{O+1} under Slutsky's decomposition is determined by the indifference curve that intersects budget line S at X_{O+1}. That is, in order to hold constant the income necessary to purchase the bundle O at the new prices P_N, at X_{O+1} the willingness to pay would need to be evaluated at a greater level of Y than under the Hicksian constraint. A greater level of Y would imply that the consumer would be willing to trade more Y for a unit of X, and the willingness to pay at the X_{O+1} unit of X would be greater than under the

Hicks decomposition. Similarly, the willingness to pay for all units of X purchased between X_O and X_S under the Slutsky constraint and would exceed the willingness to pay under Hicks. The last unit that the consumer would be willing to purchase is X_S. At X_S, the willingness to pay for that unit of X equals the new price.

Demand derived from the Slutsky decomposition is labeled with an S in panel B of Figure 3. The Slutsky demand lies everywhere above the Hicksian demand for price decreases from point O, and the consumer surplus is greater than the Hicksian consumer surplus.

Original budget constraint held constant. Still another decomposition is possible. This decomposition isolates the pure price effect by removing sufficient income to allow the consumer to purchase the original bundle of goods, point O, at the old prices, P_O. That is, real income is held constant by constraining the consumer to consume within his original feasibility set. Thus, the consumer is responding to the new prices but is obliged to purchase only consumption bundles along the original budget constraint.

If the price were P_O, the consumer at X_O would be willing to trade Y for that unit of X at the same rate as with the above two decompositions, so this demand curve coincides with the Marshallian, Hicksian, and Slutsky at point O. At X_{O+1}, the consumer would need to reduce Y by P_O in order to stay on the original budget constraint, but at that point, the consumer's willingness to pay for X exceeds the new price P_N he must pay. This would be true of those bundles along the original budget constraint from X_0 until X_I, where the consumer is still on his original budget constraint but is willing to pay exactly the new price for that last bundle. Additional consumption of X would cease at point I, the point on the original budget constraint where the willingness to pay for X equals the new price P_N. X_I represents the quantity that corresponds to P_N on the new demand curve.

In comparison with the Hicks constraints, this new demand would be below the Hicksian demand for price decreases. To gain an additional unit of X and stay on the original budget constraint would mean that the willingness to pay for X at X_{O+1} would need to be evaluated at a lower level of Y than the level of Y required to remain on the original indifference cure. Thus, because Y is scarcer, willingness to pay Y for that additional unit of X at X_{O+1} would be smaller than the Hicksian willingness to pay. For any level of X between X_O and X_I, the willingness to pay for X under the Hicksian constraints is greater than the willingness to pay under the new decomposition case. Thus, the demand curve derived from this new decomposition is everywhere below the Hicksian demand except for at the original bundle O. The corresponding demand curves in Figure 3 is labeled I.

In the foregoing analysis, the decomposition occurs by removing income (or adding income in the case of a price increase) to place the consumer at their original situation (variously defined) but at new prices. An alternative approach would decompose by adding income (or removing it in the case of a price increase) to place them at the new situation, but at the old prices.

Welfare Loss from Insurance

The foregoing analysis represents various decompositions of the increase in quantity demanded that results from an exogenous decrease in price. The Hicksian decomposition is useful for describing the welfare implications of an exogenous decrease in price, and the Slutsky decomposition is useful in understanding the construction of price indices (Friedman, 1962; Hirschleifer, 1976). The third decomposition is useful in describing the welfare effects of an insurance contract that pays off with a reduction in price because the consumer's problem in equation set 2 corresponds exactly to this decomposition. Thus, the consumer surplus derived from this third demand curve can be used to evaluate the welfare loss from such a contract.

The graphical analysis of the welfare loss is shown in Figure 4. Without insurance, the consumer purchases X_O at price P. With insurance that reduces price to $P = 0$, X_M is consumed.[6] Using Marshallian demand, D_M, to evaluate this increase would result in a gain that has been evaluated in the literature (Pauly, 1968) as the area under the Marshallian demand, area OX_MX_O. Because the price of X has not actually changed, this additional consumption costs $P(X_M - X_O)$ to produce, thus the welfare loss would be OMX_M. If Marshallian demand were used to estimate the moral hazard, however, it would include additional purchases that derive from income transfers. The pure welfare effect is only isolated when these income transfers are eliminated.

The pure welfare effect of moral hazard is represented by the change in quantity using demand curve D_I. At $P = 0$, the pure price effect in insurance results in a demand of X_I. Therefore, becoming insured would result in added consumption of X equal to $(X_I - X_O)$ that was worth the area under the new demand curve, area OX_IX_O. The cost of acquiring this additional X under insurance is $P_O(X_I-X_O)$. Thus, the moral hazard welfare loss equals area OIX_I. This loss is also smaller than either the loss evaluated using the demand curve attributed to Slutsky or the loss evaluated using the Hicksian utility constant demand curve.

Estimating the Welfare Loss

The decomposition of moral hazard into an income transfer effect and a pure price effect is presented above. According to this analysis, the pure price effect of insurance (that pays off by reducing price) is the change in consumption of medical care that would occur if a consumer *who is already ill* purchased a contract from an "insurer" to reduce the price of medical care in return for an actuarially fair premium. In other words, the income transfer effect on medical care consumption is eliminated when income transfers are eliminated, and this would only occur if the probability of illness were 1. In Figure 1 above, this decomposition of moral hazard would correspond to point A: the point where the income expansion path II' at the new prices intersects the original budget constraint. At point A, the pure price effect of insurance is the increase in medical care consumption from 10,000 units to M_A units, and the portion of moral hazard due to income transfers is from M_A units to 20,000 units.

---

[6] P=0 is chosen because it represents the price decrease in Pauly's original analysis (1968).

The welfare loss from health insurance that pays off by reducing the price of medical care has been estimated using estimates of Marshallian price elasticities of demand to determine the moral hazard effect and willingness to pay (Feldstein, 1973; Feldstein and Friedman, 1977; Feldman and Dowd, 1991; Manning and Marquis, 1996). Figure 5 shows the demand diagram corresponding to Figure 1, ignoring the effect of the premium payment on demand, as is conventionally done.[7] Demand for medical care is 10,000 units at price \$1 per unit. If the price were to fall to \$0 per unit, 20,000 units would be demanded. The cost of producing the additional 10,000 units at a marginal cost of \$1 per unit is area klmn (or \$10,000) whereas the value of the extra 10,000 units is area kmn (or \$5,000). The welfare loss is conventionally calculated as representing area klm (or \$5,000).

To estimate the true welfare loss would require estimates of how much medical care a consumer would demand if, *when ill*, he purchases an actuarially fair contract for a reduced price (Nyman, 1999a). In Figure 5, this would correspond to an increase from 10,000 units to M_A units, instead of from 10,000 units to 20,000 units. Data on the magnitude of this increase are not available and collecting them is beyond the scope of this theoretical exposition.

In lieu of such estimates, the Slutsky equation can be used to obtain a first approximation of the overstatement of the welfare loss. The Slutsky equation is conventionally characterized as measuring the Hicksian decomposition of the effect of a price decrease on Marshallian demand. With the Hicksian decomposition, the pure price effect would be equal to an increase from 10,000 to M_B units, however, we are interested in the welfare loss associated with an increase from 10,000 to M_A, a smaller increase.[8] Therefore, using the Slutsky equation to isolate the pure price does not completely remove the entire effect of income transfers on medical care consumption.

According to the Slutsky equation, the compensated (pure) price elasticity is,

$$\xi = \eta + \alpha\epsilon,$$

where η is the Marshallian price elasticity of demand, ϵ is the income elasticity of demand, and α is the share of household income devoted to medical care (Henderson and Quandt, 1958). According to Manning and Marquis' (1997) estimates from Rand health insurance experiment data, η is -.18 and ϵ is .22. The proportion of the household budget spent on medical care is .15,

---

[7]Accounting for the premium payment would shift inward the Marshallian demand for medical care and result in a smaller moral hazard increase. The Rand Health Insurance Experiment did not require a premium payment from its participants, so the effect of insurance (that pays off by a given price reduction) on medical care consumption would be slightly overestimated. As noted above, Pauly (1968) makes a corresponding assumption.

[8]Friedman (1962) and Hirshleifer (1976) note that the Slutsky decomposition only approximates the Hicksian decomposition (Hicks, 1946). Specifically, pure price effect estimated by the Slutsky decomposition is larger than the pure price effect described by Hicks.

estimated by the proportion of GDP devoted to personal medical care spending in 1997 (Economic Report of the President, 1998). Therefore, $\xi = .15$, compared with $\eta = .18$. This implies that using the estimates of the Marshallian price elasticity to estimate the demand response to an insurance contract overstates the pure price effect of insurance by at least 20 percent. Equivalently, the pure price effect estimates derived from the Slutsky equation are 83 percent of the estimates using the Marshallian demand. As a result, the estimates of the welfare loss using Slutsky's pure price effect are 83 percent of the welfare loss estimated using Marshallian demand.[9]

In addition to the theoretical argument that the Slutsky price effect overstates the true insurance price effect, there are two empirical arguments why this estimate of the Slutsky price effect is too large. First, the Rand income elasticity estimate reflects the increase in medical care consumption if the consumer's total annual wages or annual salaries were higher. Thus, a consumer would spend about .22 percent more on medical care if the consumer's annual income was 1 percent higher (Manning and Marquis, 1996). The response to a 1% income transfer that occurs simultaneously and is triggered by becoming ill, however, may be greater. This would imply that the portion of moral hazard that is due to income transfers would be greater, and the welfare loss would be smaller.[10] In addition, the income elasticity that should be used is

---

[9]Because the Rand experiment incorporated a hold-harmless provision, such that the cost sharing groups were initially paid an amount equal to their highest possible out-of-pocket expenditures and the free fee-for-service group was not paid anything, some might argue that the calculated price elasticity represents a compensated price elasticity. If so, then $\xi = .18$, $\eta = .21$, moral hazard from the Slutsky price effect is 86% of the Marshallian moral hazard, and the true welfare loss is 86% of the welfare loss calculated from Marshallian demand. If this were true, then estimates of the welfare loss based on Rand data would still be too large because the insurance price effect is still smaller than the Slusky price effect, reflected by the price elasticity of the compensated demand. It would also imply that the Rand price elasticity estimates, intended to be and widely interpreted as representing Marshallian price elasticities, are too small.

[10]For example, in the original example, in order to be consistent with an income elasticity of .22, a transfer of income of \$18,000 (an increase of 45 percent over the base of \$40,000) would lead to an increase in consumption of only about 1,000 units of medical care (over the base of 10,000 units consumed before the increase in income), since (1,000 units/10,000 units)/(\$18,000/\$40,000) ≈ .22. That is, for an additional \$18,000 in income transfers, an additional \$1,000 would be spent on medical care for a total spending of \$11,000 on medical care.

Alternatively, if the ill consumer received an income transfer shock of \$18,000 that coincided exactly with becoming ill and was implicitly intended to cover spending on medical care, a greater medical care consumption response might have occurred. For example, if as a result of the \$18,000 transfer when ill, an additional \$8,000 were spent on medical care for a total of \$18,000--a reasonable possibility if spending on medical care would total \$20,000 with insurance that paid off by a reduction in price--then the equivalent income elasticity would be (8,000 units/10,000 units)/(\$18,000/\$40,000) = 1.8. Although the resulting elasticity is many contingent on becoming ill. Therefore, the income elasticity should reflect the increased consumption of only ill persons, whereas Rand estimates reflect increased consumption on average by all consumers.

Second, the income share figure used reflects the share of medical care expenditures as a percentage of gross domestic product. Clearly, for an ill consumer, the proportion of medical care spending to overall spending would be larger than for the average person in the economy, ill or healthy. Therefore, the share would be larger and, for any given income elasticity, the pure price effect from the Slutsky decomposition would be smaller.

Conclusions

Intuition. The intuition of this paper can be summarized by referring again to Figure 1. In Figure 1, the income expansion path shows the various consumption bundles as π, the probability of the event that triggers the insurance payoff, changes. Small πs are associated with larger consumption of X, and larger πs are associated with smaller consumption of X. Whatever the π happens to be between 0 and 1, it will determine the size of the expenditures on good X, the share of those expenditures paid for by others, and the share paid for by the beneficiary in the form of an actuarially fair premium.

For any such insurance contract, there will be an additional cost that is independent of π. This cost would occur regardless of whether the π was large or small, and it simply derives from the mechanism by which income transfers are made in the insurance contract. Thus, any insurance contract that transfers income by reducing the price will incur an inefficiency loss that depends on preferences, the consumer's income, and the size of the price reduction, but not the probability of illness.

Welfare Effect of Transfers. The argument for isolating the pure price effect is that in welfare economics, income transfers are conventionally excluded from the analysis. This is also the conventional treatment of income transfers in cost-benefit analysis. With insurance theory, however, the welfare benefits are derived from additional assumptions regarding the effect of these transfers on utility. According to one of the two dominant theories of insurance demand, people demand insurance as a way of transferring income to states of the world where it is worth more than in the present state. If consumers purchase insurance because they view the expected gain from an insurance payoff as being worth more than the actuarially fair premium they must pay for the contract, then when the future state is revealed and the payoff is made, the aggregate loss (by all those who simply reduced their income by paying the premium and do not receive a payoff) is smaller than the payoff gain (by the consumer who incurred who received the payoff in that state).

---

times larger than the original price elasticity, a response of this magnitude (or even larger) does not seem unreasonable. A larger income elasticity would reduce the pure price effect.

If this is the case, then moral hazard has two welfare components: the welfare loss described in this paper, and a welfare gain associated with the increase in consumption due to income transfers from insurance. Thus, the welfare effect of moral hazard are likely to be even smaller than described in this paper because for any loss, the gains must be subtracted to find the net loss. If these gains are sufficiently large, they may swamp the welfare loss and result in a net gain from moral hazard.

Nyman (1999) has presented evidence that about 30 percent of the insurance premium of the household at the median of the wealth distribution is devoted to paying for expenditures that would otherwise be unaffordable. The additional consumption that occurs as a result of becoming insured for otherwise unaffordable medical expenditures also represents moral hazard. For these expenditures, the welfare gain is represented by the expected consumer surplus for medical services to which insurance provides access. Thus, for these services, there is no pure price welfare loss--the consumer could not afford to purchase a contract from the insurer to provide care at a lower price if the consumer were already ill. These expenditures, therefore, represent a pure income effect gain.

It should also be noted that the other dominant insurance theory, expected utility theory, cannot apply to moral hazard. Under the expected utility theory, insurance is purchased because of the gain from certainty. If a large loss is looming, a certain smaller loss that is actuarially equivalent to the large loss is preferred because of the convexity of the utility function. Therefore, insurance is purchased. With moral hazard, however, the moral hazard loss would not occur without insurance by definition. Therefore, without insurance there would be no loss for which to avoid the risk, and this theory cannot apply.

Policy implications. In determining the welfare loss from moral hazard, only the loss related to price effects should be counted. With insurance that pays off by lowering price, a large portion of moral hazard is due to the income transfers that derive from the probability of illness being less than 1. These income transfers are only eliminated when probability of illness is 1. Therefore, the pure price effect of insurance is isolated when we consider the increase in medical care consumed by the ill consumer who must purchase a fair contract from an "insurer" to provide care at a reduced price. Because consumers purchase less medical care when they are faced with the higher premium costs (which they would be, if the insurance contract they purchased were based on a probability of illness equal to 1), this decomposition implies a smaller welfare loss than that suggested by Pauly (1968).

The theoretical justification for much of current health policy stems from Pauly's 1968 analysis of the welfare loss from health insurance, and from subsequent empirical studies that have suggested that the size of this loss generally swamps any welfare gains from insurance. These papers provide the theoretical and empirical underpinnings that support the incorporation of cost-sharing deductibles and coinsurance into insurance policies, and the adoption of utilization review and other forms of managed care by health care institutions and health plan administrations to limit inappropriate health care spending. These papers also provide the main support for economists' traditional opposition to the income tax subsidy of health insurance premiums.

This paper presents the contrary argument that moral hazard--the increase in consumption of medical care that occurs when one becomes insured--may exist, but that the price-related welfare losses associated with it are smaller than have previously been understood. These losses may be outweighed by the welfare gains derived from the income transfers that allow the consumer to consume more medical care and other items in an ill state, or that simply allow the consumer to gain access to medical care procedures that would otherwise have been unaffordable.

From this perspective, the public policy emphasis on cost-sharing and managed care may need to be reevaluated. Prescriptions for optimal cost sharing would need to be recalculated because a given cost-sharing measure is likely to result in a smaller reduction of the moral hazard welfare loss and a larger reduction of the welfare gain. Managed care organizations using utilization review are most likely to scrutinize the necessity of those medical procedures that are the most costly. But it is precisely the expensive procedures--those which are most at risk of coverage denial--that represent the type of medical care that may make health insurance most valuable to consumers: procedures that are too expensive to afford without insurance. The tax subsidy has been criticized for encouraging the purchase of too much health insurance coverage, but it may instead represent a counterweight for the irreducible transactions costs that occur in any mechanism for transferring income to someone who is ill.[11] These policies all trace their theoretical origins to Pauly's 1968 article. This makes imperative a more complete understanding of Pauly's welfare loss and of the limits of his analysis.

[11]The regressivity of the tax subsidy is, of course, still an important welfare issue.

References

Arrow, Kenneth J. "Uncertainty and the Welfare Economics of Medical Care," American Economic Review vol. 53, 1963, pp. 941-973.

Blaug, Mark. Economic Theory in Retrospect, Fourth edition. Cambridge, UK: Cambridge University Press, 1985.

de Meza, David. "Health Insurance and the Demand for Medical Care," Journal of Health Economics vol. 2, no. 1, March 1983, pp. 47-54.

Ehrlich, Isaac and Gary S. Becker. "Market Insurance, Self-Insurance and Self-Protection," Journal of Political Economy vol. 80, 1972, pp. 623-648.

Economic Report of the President: Transmitted to Congress February 1998. Washington, DC: U.S. Government Printing, 1998.

Feldman, Roger and Bryan Dowd. "A New Estimate of the Welfare Loss of Excess Health Insurance," American Economic Review vol. 81, no. 1, March 1991, pp. 297-301.

Feldstein, Martin S. "The Welfare Loss of Excess Health Insurance," Journal of Political Economy vol. 81, March/April 1973, pp. 251-280.

Feldstein, Martin and Bernard Friedman. "Tax Subsidies, the Rational Demand for Insurance, and the Health Care Crisis," Journal of Public Economics vol. 7, 1977, pp. 155-178.

Friedman, Milton. Price Theory: A Provisional Text. Chicago: Aldine Publishing Co., 1962.

Friedman, Milton and L. J. Savage. "The Utility Analysis of Choices Involving Risk," Journal of Political Economy vol. 56, no. 4, August 1948, pp. 279-304.

Gertler, Paul, and Jonathan Gruber. "Insuring Consumption Against Illness," National Bureau of Economic Research Working Paper #6035. NBER: Cambridge, MA, 1997.

Hausman, J. A. "Exact Consumer's Surplus and Deadweight Loss," American Economic Review vol. 71, 1981, pp. 662-676.

Hirshleifer, Jack. Price Theory and Applications. Englewood Cliffs, NJ: Prentice-Hall, 1976.

Hicks, J. R. Value and Capital (2nd edition) Oxford: Clarendon Press, 1946.

Manning, Willard G. and M. Susan Marquis. "Health Insurance: The Tradeoff Between Risk Pooling and Moral Hazard," Journal of Health Economics vol. 15, no. 5, October 1996, pp. 609-640.

Mishan, E. J. Cost-Benefit Analysis: An Introduction. New York: Praeger, 1971.

Mishan, E. J. Introduction to Normative Economics. New York: Oxford, 1981.

Newhouse, Joseph P. "Medical Care Costs: How Much Welfare Loss?" Journal of Economic Perspectives vol. 6, no. 3, Summer, 1992, pp. 3-22.

Nyman, John A. "Costs, Technology, and Insurance in the Health Care Sector," Journal of Policy Analysis and Management vol. 10, no. 1, Winter 1991, pp. 106-111.

Nyman, John A. "The Value of Health Insurance: The Access Motive," Journal of Health Economics vol. 18, no. 2, April 1999b, pp. 141-152.

O'Connell, John F. Welfare Economic Theory. Boston: Auburn House, 1982.

Parkin, David, Allistar McGuire and Brian Yule. "Aggregate Health Care Expenditure and National Income: Is Health Care a Lusury Good? Journal of Health Economics vol. 6, no. 2, June 1987, pp. 111-127.

Pauly, Mark V. "The Economics of Moral Hazard: Comment," American Economic Review vol. 58, no. 3, 1968, pp. 531-537.

Rice, Thomas. "An Alternative Framework for Evaluating Welfare Losses in the Health Care Market," Journal of Health Economics vol. 11, no. 1, May 1992, pp. 86-92.

Rice, Thomas. "Can Markets Give Us the Health System We Want?" Journal of Health Politics, Policy and Law vol. 22, no. 2, April 1997, pp. 383-426.

Rice, Thomas. The Economics of Health Reconsidered. Chicago: Health Administration Press, 1998.

Rothschild, Michael and Joseph E. Stiglitz. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," Quarterly Journal of Economics vol. 90, 1976, pp. 629-649.

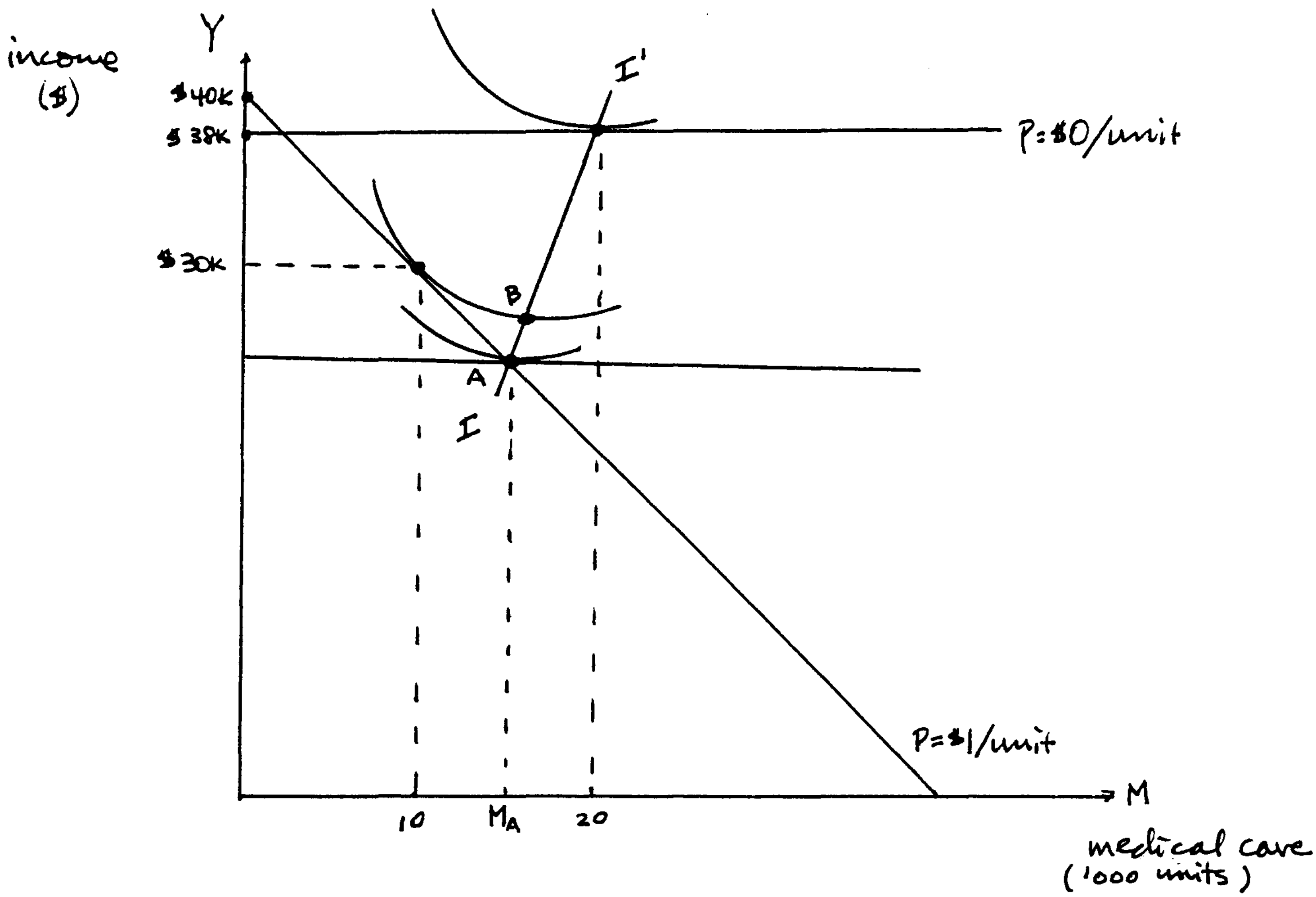
Slutsky, Eugen E. "On the Theory of the Budget of the Consumer," Translated by Olga Ragusa and reprinted in Readings in Economic Theory. Homewood, IL: Irwin, 1952.

Viscusi, W. Kip and William N. Evans. "Utility Functions that Depend on Health Status: Estimates and Economic Implications," American Economic Review vol. 80, 1990, pp. 353-374.

Von Neumann, John and Oskar Morgenstern. Theory of Games and Economic Behavior, 2nd Edition. Princeton, NJ: Princeton University Press, 1947.

Willing, Robert D. "Consumer Surplus Without Apology," American Economic Review vol. 66, no. 4, September 1976, pp. 589-597.

Zweifel, Peter and Friedrich Breyer. Health Economics. New York: Oxford, 1997.



income ($)
Y
$40K
$38K
$30K
I'
P=$0/unit
B
A
I
P=$1/unit
M
10
M_A
20
medical care ('000 units)

Figure 1

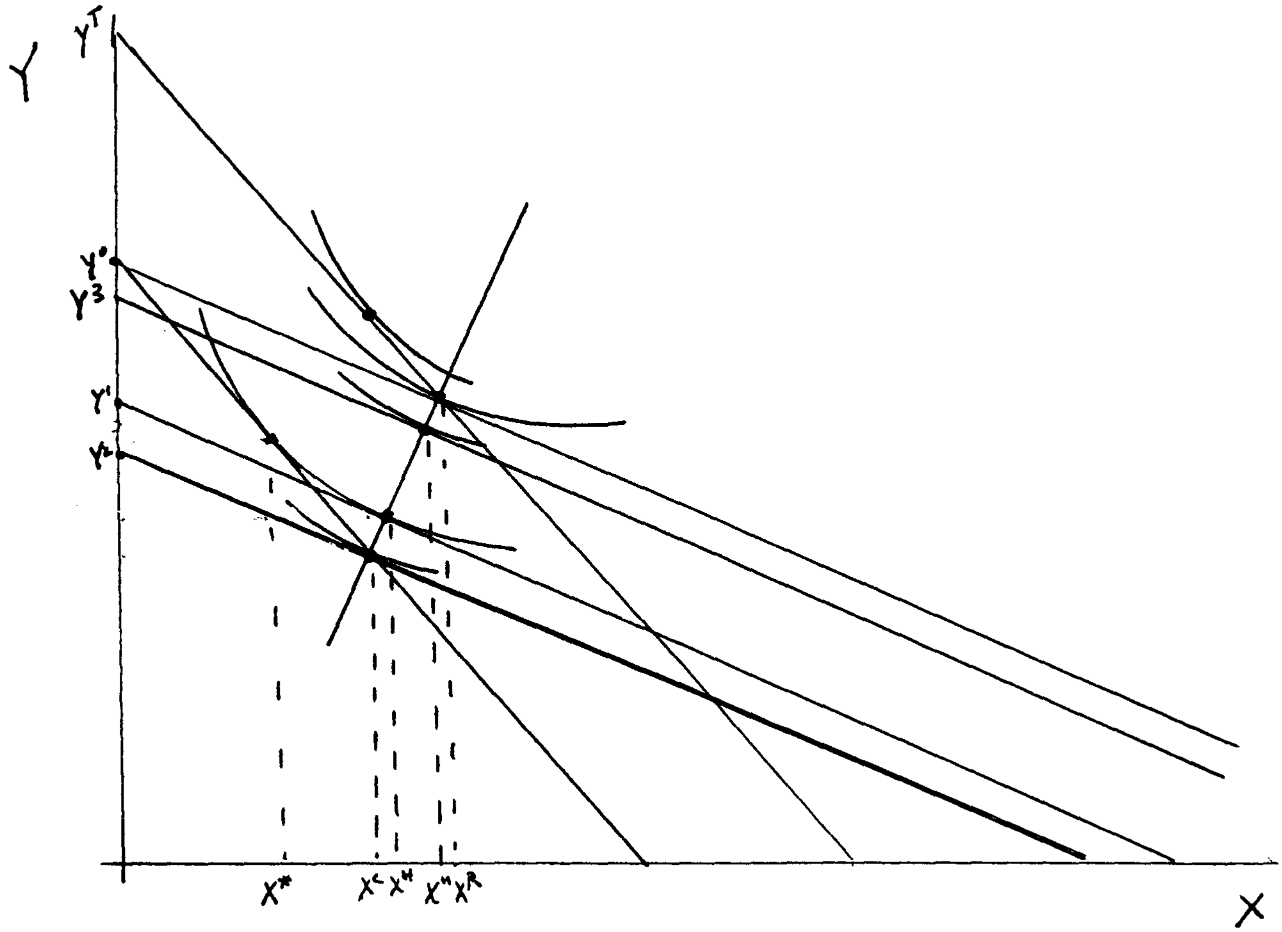
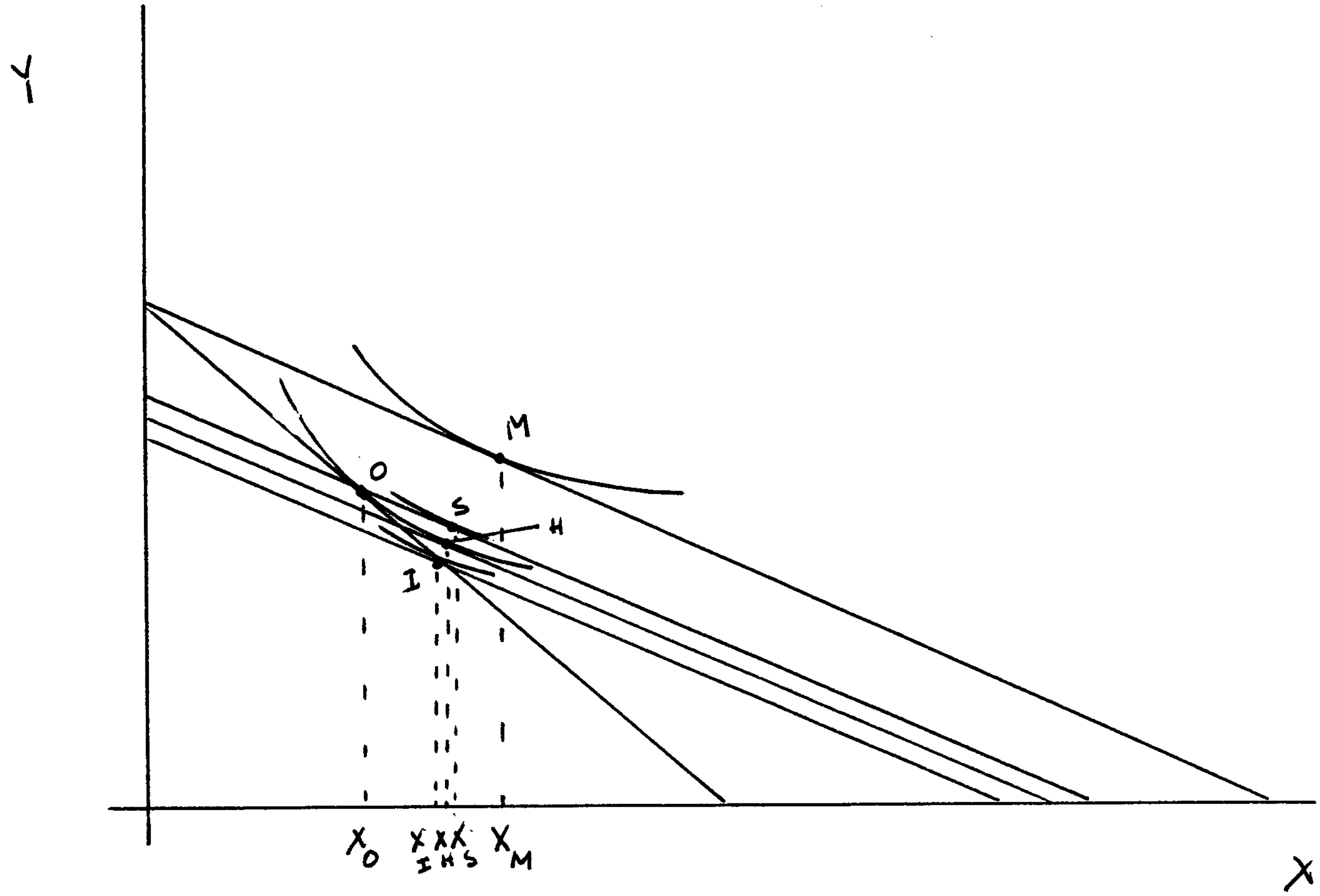
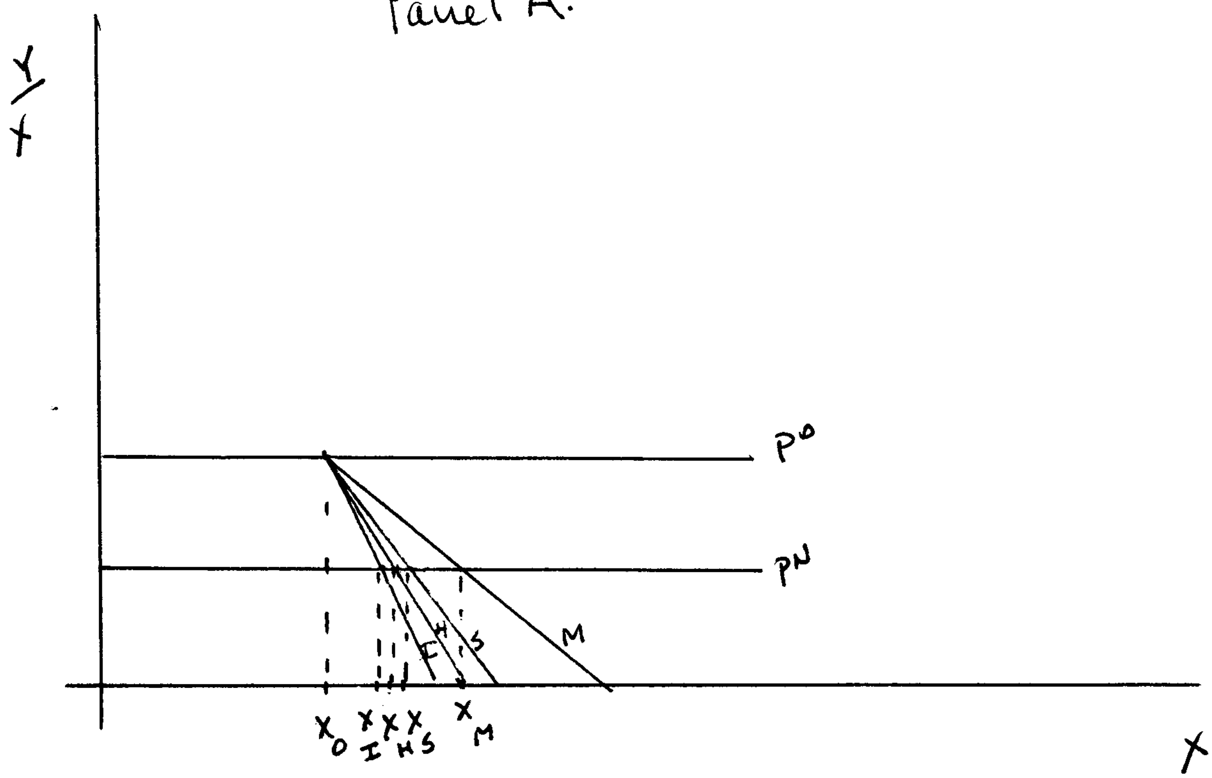


Figure 2



Panel A.



Panel B
Figure 3

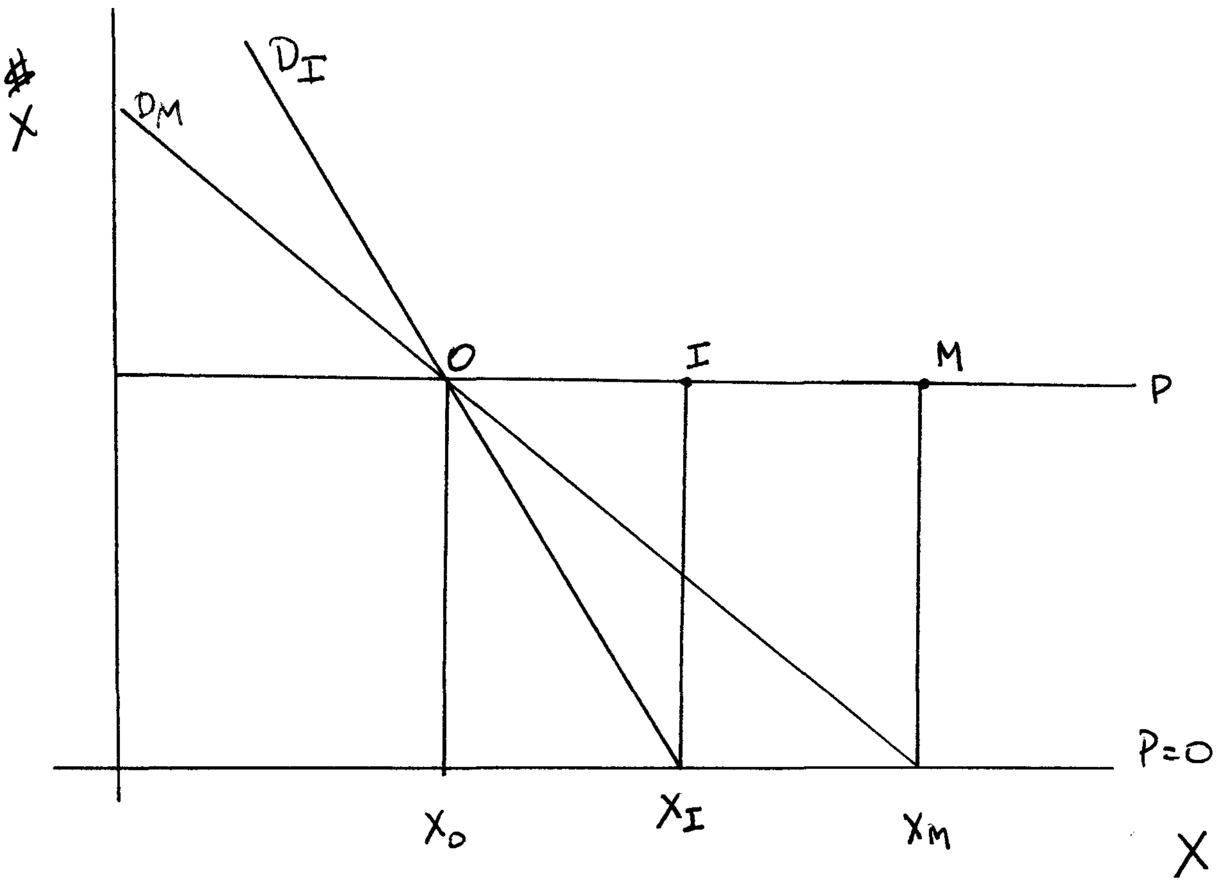
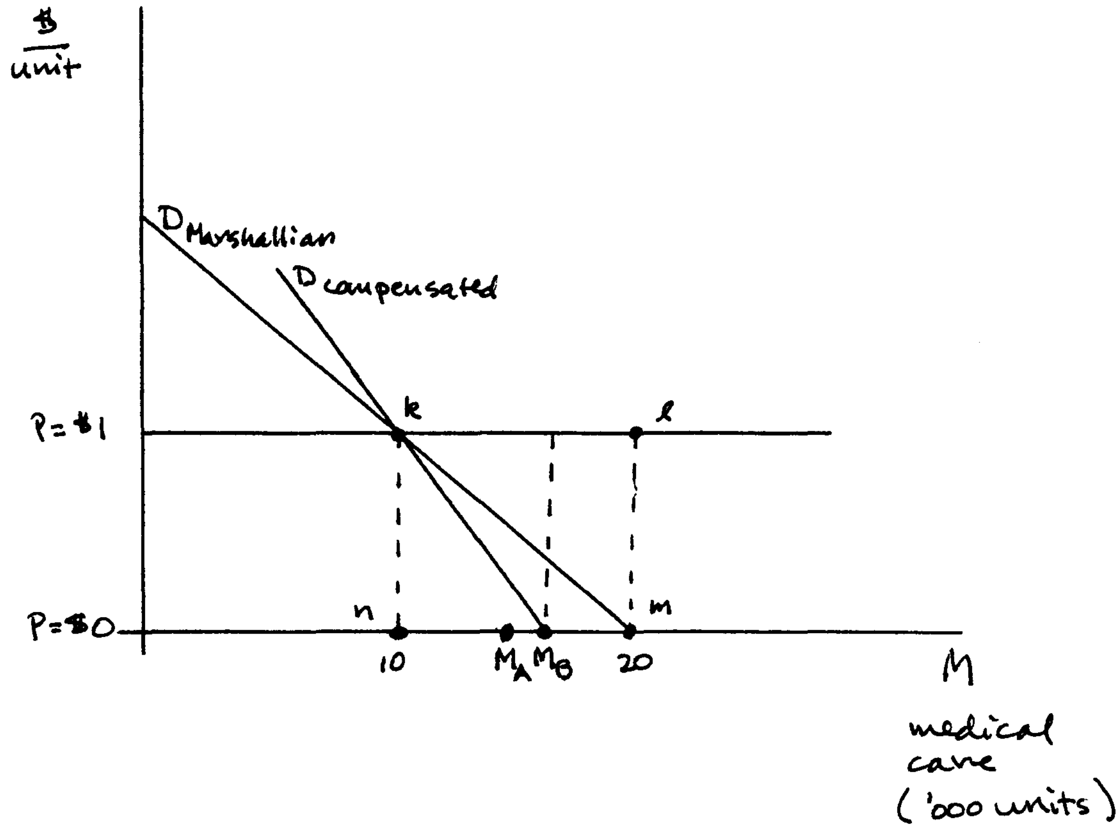


Figure 4



$/unit
D Marshallian
D compensated
k
l
P=$1
n
m
P=$0
10
M_A
M_B
20
M
medical care ('000 units)

Figure 5