OPTIMAL COMMODITY TAXATION
WITH MORAL HAZARD AND UNOBSERVABLE OUTCOMES

by

Gerard Russo*

Working Paper No. 93-23
September, 1993 (revised)

The optimal public insurance-taxation scheme is derived for a model with unobservable outcomes. If the government can only observe aggregate commodity expenditures, reimbursement insurance is constrained-efficient. However, two distortions accompany. First, consumers are induced to take (forego) actions which increase (decrease) the likelihood of adverse outcomes (moral hazard). Second, reimbursement insurance creates a subsidy distortion. Moral hazard calls for taxation (subsidization) of commodities which increase (decrease) the probability of adverse outcomes. The second calls for taxation (subsidization) of commodities which are complements to (substitutes for) the insured commodity. An example centered on cigarettes and medical insurance is presented.

*University of Hawaii, Department of Economics, 2424 Maile Way, Honolulu, HI 96822 USA and East-West center, Program on Population, 1777 East-West Road, Honolulu, HI 96848 USA.

*I would like to thank two anonymous referees for their helpful and detailed suggestions. I owe considerable thanks to one referee for suggesting the new title and complete reorganization of the paper. I would also like to thank Mark Satterthwaite, John Panzar, Kip Viscusi, Bobbi Wolfe, Gene Smolensky, Jim Roumasset, Sumner La Croix, Tom Getzen and Luigi Ermini for the earlier assistance. Of course any remaining errors are my own.
1. Introduction

This paper is about the use of commodity taxation to correct moral hazard when outcomes are unobservable. The unobservability of outcomes means lump-sum outcome-contingent payments are not possible. Under this assumption, optimal insurance takes the form of reimbursement rather than lump-sum indemnification and gives rise to price distortions in addition to moral hazard. The optimal tax scheme balances these distortions.

This paper extends the work of Arnott and Stiglitz (1986) to the case of reimbursement insurance. In their paper commodity taxation is viewed as a corrective mechanism for the moral hazard associated with public insurance [Arnott and Stiglitz (1986)]. In their model Arnott and Stiglitz (1986) assume outcomes are perfectly observable and as a result the optimal social plan includes lump-sum outcome-contingent payments. So called indemnity insurance contracts are designed this way. For example life insurance and automobile insurance contracts are typically written as indemnity contracts providing lump-sum cash payments in the event of death or auto accident. Health insurers, on the other hand, utilize reimbursement contracts which pay a portion of medical expenses and thus reduce the consumer’s price below marginal cost. Lump-sum indemnity insurance only results in a moral hazard problem [Pauly (1974), Arnott and Stiglitz (1988)]. Reimbursement insurance is subject to both moral hazard and subsidy distortions [Pauly (1968), Zeckhauser (1970)].

Consider automobile insurance and health insurance as illustrative examples. Auto accidents are observable at a relatively modest cost to auto insurers because the cost is subsidized by law enforcement agencies which provide information via accident reports to auto insurers. State laws typically require drivers to report accidents to the police and auto insurance
companies rarely pay without a police report. Thus, with the information on the outcome, auto insurers can make lump-sum outcome-contingent cash payments. This fits the case analyzed by Arnott and Stiglitz (1986) quite well. By contrast, health insurers do not easily observe health status independent of medical service consumption. Only when diagnosis and treatment have been rendered and expenses incurred will a typical health insurer know something is afoot and be able to provide payment (reimbursement insurance).\(^1\) With auto insurance however, the insured can acquire repair estimates from mechanics and body shops, receive a lump-sum cash payment from the insurer and spend those payments in any manner (indemnity insurance).\(^2\)

There are two main theoretical results. First, when outcomes are unobservable, the optimal linear tax/subsidy scheme implies subsidization of commodities the demand for which is positively correlated with the ex post marginal utility of income. That is, subsidize demand as a second-best solution when lump-sum outcome-contingent payments are not possible. Second, when reimbursement is optimal, the optimal tax (subsidy) depends both on moral hazard [Arnott and Stiglitz (1986)] and on the complementarity (substitutability) [Balcer (1980)] between insured and uninsured commodities.

Although analytically similar to the results presented by Arnott and Stiglitz (1986), this model renders substantially different interpretations. In their model (Arnott and Stiglitz 1986), subsidies are seen as corrections for insurance distortions. Here, the subsidy is the insurance. Any time government provides a subsidy to a commodity, the demand for which is stochastic, it is providing insurance.

When insurance takes this reimbursement form it generates a classic moral hazard problem [Pauly (1974), Shavell (1979)]. Arnott and Stiglitz (1986) derive the corrective tax for
a model with moral hazard. In addition reimbursement insurance generates a subsidy distortion [Pauly (1968)] which calls for taxation (subsidization) of complements to (substitutes for) the insured good, a result found in different forms in the literature [e.g., Diamond and Mirrlees (1971), Green and Sheshinski (1976), Sandmo (1976), Balcer (1980)]. This paper incorporates both the moral hazard effect and the subsidy effect in the optimal tax formulae derivation, and thus integrates both literatures.

2. The model

The theoretical framework is that of expected utility theory with public insurance generating moral hazard and subsidy distortions. A single representative consumer maximizes expected utility. There is no private insurance. Producer prices are assumed to be constant (linear production possibilities frontier) and equal to marginal cost. Given the representative consumer's demand functions, the social planner selects the optimal reimbursement insurance tax/subsidy scheme. The model is an extension of the work of Arnott and Stiglitz (1986) to the case of unobservable outcomes.

2.1. The consumer's problem

The representative consumer's income with outcome $i$, $Y_i$, equals the consumer's initial income level with outcome $i$, $y_i$, plus (minus) a lump-sum subsidy (tax), $x$. As the outcomes are assumed to be unobservable, the lump-sum payment (charge) is independent of outcome. This specification is critically different from that of Arnott and Stiglitz (1986) for whom outcomes are observable and lump-sum payments are outcome-contingent.
\[ Y_i = y_i + x \] (1)

\[ \rho_i = \rho_i(c) \] (2)

Equation (2) represents the probability of outcome \( i \) which depends on the consumption vector \( c \). The consumption vector \( c = (c_1, c_2, \ldots, c_K) \) is chosen ex ante and therefore is independent of the outcome.

\[ u_i = u_i(c, m_i) \] (3)

The consumption vector \( m_i = (m_{i1}, m_{i2}, \ldots, m_{ik}) \) is chosen ex post and therefore the necessity of the subscript \( i \). Utility with outcome \( i, u_i \), which is a function \( c \) and \( m_i \), is also assumed to be outcome-dependent [Cook and Graham (1977), Shavell (1978), Dionne (1982), Hey and Patel (1982)]. Furthermore, the consumer is assumed to be an expected-utility maximizer.

\[ EU = \sum_i \rho_i(c)u_i(c, m_i) \] (4)

The consumer’s budget constraint is represented by equation (5), with prices and income exogenous to consumer choice. The consumer price vectors \( q^c \) and \( q^m \), associated with consumption vectors \( c \) and \( m \), respectively, equal the respective producer price vectors, \( p^c \) and \( p^m \), plus (minus) the respective per-unit tax (subsidy) vectors, \( t^c \) and \( t^m \).
\[ Y_i = q^C \cdot c + q^M \cdot m_i \]  \hspace{1cm} (5)

\[ q^C = p^C + t^C \]  \hspace{1cm} (6)

\[ q^M = p^M + t^M \]  \hspace{1cm} (7)

Following Arnott and Stiglitz (1986), I solve the consumer's problem in a two step optimization. First, looking at outcome \( i \), I define the restricted direct utility function \( u_i \), with \( c \) as a given choice.

\[ u_i(m_i; c) \]  \hspace{1cm} (8)

Likewise, given expenditure on \( c \), I define the restricted income with outcome \( i \), the income available for purchase of \( m_i \).

\[ \bar{Y}_i = Y_i - q^C \cdot c = q^M \cdot m_i \]  \hspace{1cm} (9)

Now, I define the restricted indirect utility function \( v_i \), which depends on the prices, \( q^m \), restricted income and \( c \).
\[ v_i(c, Y_i, q^m) = \max_{m_i} \left\{ u_i(m_i; c) \mid Y_i - q^m \cdot m_i \geq 0 \right\} \]

(10)

This equation embodies the first step restricted optimization.\(^3\)

The second step entails maximizing expected restricted indirect utility by choice of the \textit{ex ante} consumption vector \( c \):

\[ \max_c \sum_i \rho_i(c) v_i(c, Y_i - q^c \cdot c, q^m) \]

(11)

The solution to this maximization problem can be expressed as

\[ c = c(q^c, q^m, Y). \]

(12)

Thus we may write

\[ \Pi_i(q^c, q^m, Y) = \rho_i(c(q^c, q^m, Y)) \]

(13)

\[ V_i(q^c, q^m, Y) = v_i(c(q^c, q^m, Y), Y_i - q^c \cdot c(q^c, q^m, Y), q^m) \]

(14)

and

\[ EV(q^c, q^m, Y) = \sum_i \Pi_i(q^c, q^m, Y) V_i(q^c, q^m, Y) \]

(15)
where $\Pi_i$ and $V_i$ are the unrestricted indirect probability of outcome $i$ and unrestricted indirect utility with outcome $i$, respectively.$^4$

2.2. Some envelope results

Before proceeding to the social planner's problem it is useful to present some envelope results which will greatly facilitate later exposition. First, define

$$\alpha_i = \frac{\partial v_i}{\partial Y_i} = \frac{\partial v_i}{\partial Y_i}$$

(16)

as the marginal utility of income with outcome $i$. Also by virtue of the envelope theorem, the partial derivatives of the indirect expected utility can be compactly written as (17)$^5$, (18)$^6$ and (19).$^7$

$$\frac{\partial EV}{\partial Y_i} = \Pi_i \alpha_i$$

(17)

$$\frac{\partial EV}{\partial q_k} = - \sum_i \Pi_i \alpha_i c_k$$

(18)

$$\frac{\partial EV}{\partial q_j} = - \sum_i \Pi_i \alpha_i m_{ij}$$

(19)
2.3. The social planner's problem

The social planner's problem is to select the combination of taxes and subsidies which maximizes the consumer's indirect expected utility, equation (15), subject to expected tax receipts equal expected transfers, equation (20). Alternatively, view this as an expected profits equal zero condition.

\[ x = t^c \cdot c + \sum_i \Pi_i t^m \cdot m_i \]  \hspace{1cm} (20)

Forming the Lagrangian for this problem

\[ \max_{t^c, t^m, x} \mathcal{L} = EV(q^c, q^m, Y) - \lambda \left[ x - t^c \cdot c - \sum_i \Pi_i t^m \cdot m_i \right] \]  \hspace{1cm} (21)

and deriving the first-order conditions for an interior solution renders equations (22), (23) and (24).

\[
\frac{\partial \mathcal{L}}{\partial t_k^c} = -c_k \sum_i \Pi_i \alpha_i \\
- \lambda \left[ -c_k - \sum_k t_k^c \frac{\partial c_k}{\partial q_k^c} - \sum_i \frac{\partial \Pi_i}{\partial q_k^c} t^m \cdot m_i - \sum_i \Pi_i \sum_j t_j^m \frac{\partial m_{ij}}{\partial q_k^c} \right] = 0
\]  \hspace{1cm} (22)

\[ k = 1, \ldots, K \]

In equation (22) \( c_k \sum_i \Pi_i \alpha_i \) can be interpreted as the direct cost of the tax rate increase, \( \lambda c_k \) can be interpreted as the benefit due to the direct revenue effect and the residual terms can
be interpreted as the benefit due to the indirect revenue effect. Equation (23) can be interpreted in a similar fashion.

$$\frac{\partial \xi}{\partial t_j} = - \sum_i \Pi_i \alpha_i m_{ij} - \lambda \left[ - \sum_i \Pi_i m_{ij} - \sum_k t_k \frac{\partial c_k}{\partial q_j} - \sum_i \frac{\partial \Pi_i}{\partial q_j} t^m \cdot m_i - \sum_i \Pi_i \sum_j t_j \frac{\partial m_{ij}}{\partial q_j} \right] = 0 \quad (23)$$

$j = 1, \ldots, J$

In equation (24) \( \sum \Pi_i \alpha_i \) can be interpreted as the direct benefit of an increase in the lump-sum transfer, \(-\lambda\) as the cost due to the direct revenue effect, and the residual terms as the cost due to the indirect revenue effect.

$$\frac{\partial \xi}{\partial x} = \sum_i \Pi_i \alpha_i - \lambda \left[ 1 - \sum_k \sum_i t_k \frac{\partial c_k}{\partial Y_i} - \sum_i \frac{\partial \Pi_i}{\partial Y_i} t^m \cdot m_i - \sum_i \Pi_i \sum_j t_j \sum_i \frac{\partial m_{ij}}{\partial Y_i} \right] = 0 \quad (24)$$

Multiplying the first order condition associated with \( x \), equation (24), by the expected demand for good \( j \) and adding the result to equation (23) gives equation (25).
\[
\sum_k t_k \left[ \frac{\partial c_k}{\partial q_j^m} + \sum_i \frac{\partial c_k}{\partial Y_i^m} \sum_i \Pi_i m_{ij} \right] + \sum_i \left[ \sum_k \frac{\partial p_i}{\partial q_j^m} + \sum_i \frac{\partial c_k}{\partial Y_i^m} \sum_i \Pi_i m_{ij} \right] t^m \cdot m_i \\
+ \sum_i \Pi_i \sum_j t_j^m \left[ \frac{\partial m_{ij}}{\partial q_j^m} + \sum_i \frac{\partial m_{ij}}{\partial Y_i^m} \sum_i \Pi_i m_{ij} \right]
\]

\[
= \sum_i \Pi_i \alpha_i m_{ij} - \sum_i \Pi_i \alpha_i \sum_i \Pi_i m_{ij}
\]

\[
(25)
\]

\(j = 1, \ldots, J\)

Equation (25) has the following interpretation. Suppose that the tax rate on good \(j\) is raised by one unit and that \(x\) at the same time is raised \(\sum_i \Pi_i m_{ij}\). There are two net effects. The first is the indirect net revenue effect, which is captured by the terms on the LHS of equation (25). The second is the insurance effect, which is captured by the term on the RHS of (25).

If there were no insurance effect, then the optimum would entail zero taxation. Furthermore, at the zero taxation point, the indirect net revenue effect is inoperative. Thus, in the neighborhood of the zero taxation point, the utility gradient or local direction of optimal taxation is determined solely by the insurance effect. Away from the zero taxation point, the utility gradient is determined by a combination of both effects. One wants to increase the tax rate on goods with a positive indirect net revenue effect and a positive insurance effect.
Multiplying equation (24) by $c_k$ and adding to equation (22) renders, with some additional substitution,\(^9\) equation (26). The RHS of (26) equals zero. Thus, no insurance effect is embodied in the choice of $c$-goods. This is a direct result of modelling $c$ as an *ex ante* choice.

\[
\sum_k t_k^c \left[ \frac{\partial c_k}{\partial q_k} c_k + \sum_i \frac{\partial c_k}{\partial Y_i} c_k \right] + \sum_i \left[ \sum_k \frac{\partial r_i}{\partial c_k} c_k + \sum_i \frac{\partial c_k}{\partial Y_i} c_k \right] t^m \cdot m_i
\]

\[+ \sum_i \Pi_i \sum_j t_j^m \left[ \frac{\partial m_{ij}}{\partial q_k} c_k + \sum_i \frac{\partial m_{ij}}{\partial Y_i} c_k \right] \]

\[c_k \sum_i \Pi_i \alpha_i = c_k \sum_i \Pi_i \alpha_i \]

(26)

\[= \frac{c_k \sum_i \Pi_i \alpha_i - c_k \sum_i \Pi_i \alpha_i}{\lambda} \]

\[k = 1, \ldots, K\]

To see the role of insurance more clearly, observe that the RHS of (25) is the covariance between the marginal utility of income and the demand for good $j$. Specifically

\[
\sum_i \Pi_i \alpha_i m_{ij} = E(\alpha m_j) = E(\alpha) E(m_j) + \text{cov}(\alpha, m_j)
\]

\[= \sum_i \Pi_i \alpha_i \cdot \sum_i \Pi_i m_{ij} + \text{cov}(\alpha, m_j).\]

(27)
Thus,

\[ \sum_i \Pi_i \alpha_i m_{ij} - \sum_i \Pi_i \alpha_i \cdot \sum_i \Pi_i m_{ij} = \text{cov}(\alpha, m_j). \] (28)

Therefore, we may write equations (25) and (26) as

\[ \sum_k t_k^c \left( \frac{\partial c_k}{\partial q_j^m} \right)_\theta + \sum_i \left[ \sum_k \frac{\partial \rho_i}{\partial c_k} \left( \frac{\partial c_k}{\partial q_j^m} \right)_\theta \right] t^m \cdot m_i + \sum_i \Pi_i \left[ \sum_j t_j^m \left( \frac{\partial m_{ij}}{\partial q_j^m} \right)_\theta \right] = \frac{\text{cov}(\alpha, m_j)}{\lambda} \]

\[ j = 1, \ldots, J \] (29)

and

\[ \sum_k t_k^c \left( \frac{\partial c_k}{\partial q_k^c} \right)_\theta + \sum_i \left[ \sum_k \frac{\partial \rho_i}{\partial c_k} \left( \frac{\partial c_k}{\partial q_k^c} \right)_\theta \right] t^m \cdot m_i + \sum_i \Pi_i \left[ \sum_j t_j^m \left( \frac{\partial m_{ij}}{\partial q_k^c} \right)_\theta \right] = 0 \]

\[ k = 1, \ldots, K \] (30)

where the subscripts \( \theta \) and \( \delta \) on the partial derivatives indicate exact (Hicksian) and approximate compensated price effects, respectively. The optimal tax scheme depends on the price effects.
and the covariance between the marginal utility of income and demand for good \( j \). The covariance has been shown to be important in the optimal reimbursement insurance scheme [Arrow (1976) and Besley (1988)]. Goods positively correlated with the marginal utility of income will tend be subsidized under this scheme and subsidization is equivalent to reimbursement-type insurance coverage. The covariance between good \( c_k \) and the marginal utility of income with outcome \( i \), \( \alpha_i \), is zero as \( c_k \) is chosen \textit{ex ante} and is therefore independent of outcome (i.e., there is no insurance effect).

2.4. \textit{Optimal taxes}

If I define the following matrices composed of compensated expected price effects

\[
C_\theta = \begin{bmatrix}
\frac{\partial c_1}{\partial q_1^c} & \frac{\partial c_2}{\partial q_1^c} & \cdots & \frac{\partial c_K}{\partial q_1^c} \\
\frac{\partial c_1}{\partial q_2^c} & \frac{\partial c_2}{\partial q_2^c} & \cdots & \frac{\partial c_K}{\partial q_2^c} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial c_1}{\partial q_K^c} & \frac{\partial c_2}{\partial q_K^c} & \cdots & \frac{\partial c_K}{\partial q_K^c}
\end{bmatrix}
\]  

(31)
\[ C_b = \begin{bmatrix}
\frac{\partial c_1}{\partial q_1^m} & \frac{\partial c_2}{\partial q_1^m} & \cdots & \frac{\partial c_K}{\partial q_1^m} \\
\frac{\partial c_1}{\partial q_2^m} & \frac{\partial c_2}{\partial q_2^m} & \cdots & \frac{\partial c_K}{\partial q_2^m} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial c_1}{\partial q_J^m} & \frac{\partial c_2}{\partial q_J^m} & \cdots & \frac{\partial c_K}{\partial q_J^m}
\end{bmatrix}
\] (32)

\[ M_b = \begin{bmatrix}
\sum_i \Pi_i \left( \frac{\partial m_{i1}}{\partial q_1^m} \right) & \sum_i \Pi_i \left( \frac{\partial m_{i2}}{\partial q_1^m} \right) & \cdots & \sum_i \Pi_i \left( \frac{\partial m_{ij}}{\partial q_1^m} \right) \\
\sum_i \Pi_i \left( \frac{\partial m_{i1}}{\partial q_2^m} \right) & \sum_i \Pi_i \left( \frac{\partial m_{i2}}{\partial q_2^m} \right) & \cdots & \sum_i \Pi_i \left( \frac{\partial m_{ij}}{\partial q_2^m} \right) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_i \Pi_i \left( \frac{\partial m_{i1}}{\partial q_J^m} \right) & \sum_i \Pi_i \left( \frac{\partial m_{i2}}{\partial q_J^m} \right) & \cdots & \sum_i \Pi_i \left( \frac{\partial m_{ij}}{\partial q_J^m} \right)
\end{bmatrix}
\] (33)
\[ M_0 = \left[ \begin{array}{ccc} \sum_i \Pi_i \left( \frac{\partial m_{11}}{\partial q_1^c} \right)_0 & \sum_i \Pi_i \left( \frac{\partial m_{12}}{\partial q_1^c} \right)_0 & \cdots & \sum_i \Pi_i \left( \frac{\partial m_{1f}}{\partial q_1^c} \right)_0 \\ \sum_i \Pi_i \left( \frac{\partial m_{11}}{\partial q_2^c} \right)_0 & \sum_i \Pi_i \left( \frac{\partial m_{12}}{\partial q_2^c} \right)_0 & \cdots & \sum_i \Pi_i \left( \frac{\partial m_{1f}}{\partial q_2^c} \right)_0 \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i \Pi_i \left( \frac{\partial m_{11}}{\partial q_K^c} \right)_0 & \sum_i \Pi_i \left( \frac{\partial m_{12}}{\partial q_K^c} \right)_0 & \cdots & \sum_i \Pi_i \left( \frac{\partial m_{1f}}{\partial q_K^c} \right)_0 \end{array} \right] \] 

(34)

and also define the gradient of \( \rho_i \),

\[ D \rho_i(c) \equiv \left[ \begin{array}{c} \frac{\partial \rho_i(c)}{\partial c_1} \\ \frac{\partial \rho_i(c)}{\partial c_2} \\ \vdots \\ \frac{\partial \rho_i(c)}{\partial c_K} \end{array} \right] \] 

(35)
the matrix of marginal impacts of \( c \) on the expected demands for \( m \)-goods (marginal insurance cost),

\[
\Psi = \sum_i D \rho_i(c) \cdot m_i = \begin{bmatrix}
\sum_i \frac{\partial \rho_i}{\partial c_1} m_{i1} & \sum_i \frac{\partial \rho_i}{\partial c_1} m_{i2} & \cdots & \sum_i \frac{\partial \rho_i}{\partial c_1} m_{ij} \\
\sum_i \frac{\partial \rho_i}{\partial c_2} m_{i1} & \sum_i \frac{\partial \rho_i}{\partial c_2} m_{i2} & \cdots & \sum_i \frac{\partial \rho_i}{\partial c_2} m_{ij} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_i \frac{\partial \rho_i}{\partial c_K} m_{i1} & \sum_i \frac{\partial \rho_i}{\partial c_K} m_{i2} & \cdots & \sum_i \frac{\partial \rho_i}{\partial c_K} m_{ij}
\end{bmatrix}
\]

(36)

and the covariance vector in these manners,

\[
COV = \begin{bmatrix}
cov(\alpha, m_1) \\
cov(\alpha, m_2) \\
\vdots \\
cov(\alpha, m_J)
\end{bmatrix}
\]

(37)

then the first order conditions can be compactly written as

\[
C_0 t^c + C_0 \Psi t^m + M_0 t^m = \frac{1}{\lambda} COV
\]

(38)

and
\[ C_0 t^c + C_0 \Psi t^m + M_0 t^m = 0. \]  

(39)

Solving for the optimal tax vectors renders

\[ t^m = \left[ M_0 - C_0 C_0^{-1} M_0 \right]^{-1} \frac{1}{\lambda} COV \]  

(40)

and

\[ t^c = -\left[ \Psi + C_0^{-1} M_0 \right] \left[ M_0 - C_0 C_0^{-1} M_0 \right]^{-1} \frac{1}{\lambda} COV. \]  

(41)

In general, the optimal insurance tax/subsidy scheme depends on a complex combination of moral hazard and price effects and is somewhat difficult to interpret. However, some intuitive characterizations can be discerned. Inspection of equation (40) reveals that the importance of the covariance between the marginal utility of income and demand for \( m \)-goods. If there are no cross effects between goods then the optimal tax/subsidy scheme calls for subsidization (taxation) of goods positively (negatively) correlated with the marginal utility of money. The degree of taxation and subsidy will depend on price sensitivity in a manner similar to a Ramsey rule. The covariance embodies the insurance function of the tax scheme, which only taxes (subsidies) on \( m \)-goods can serve directly, since the demands for \( c \)-goods are outcome invariant.
This insurance leads to excessive (insufficient) consumption of goods which increase (decrease) the probability of adverse outcomes and excessive consumption of insured goods (i.e., goods positively correlated with the marginal utility of income). This calls for differential taxation/subsidization of goods based on direct effects and indirect effects transmitted through cross-price relations.

Notice that in (41), if $\Psi = 0$, then there is no moral hazard effect and the differential taxation of $c$-goods will depend on the degree of taxation of $m$-goods (see Equation (44)), the degree price sensitivity of $c$-goods (a la Ramsey) and the expected complementarity between $c$-goods and $m$-goods. If consumption of $c$-goods impacts expected revenues from $m$-goods (i.e., $\Psi \neq 0$), then there is an additional effect on the optimal taxation of $c$-goods depending on the direction and magnitude of the incentive distortion from moral hazard.

If the marginal utility of income and demands vary with the outcome, then some subsidization (reimbursement insurance) is optimal. Under the optimal scheme the moral hazard and subsidy distortions created by reimbursement insurance will be balanced by taxes and subsidies on related commodities. To further illustrate these results, I analyze the following special case.

2.5. A special case

Consider a special case with only two possible outcomes, $i=0,1$, and three goods $c$, $m$ and $z$ $(K=1$, $J=2)$. Good $c$ is chosen ex ante, and $m$ and $z$ are chosen ex post. The consumer’s direct expected utility function is
\[ EU = \rho_0(c)u_0(c, z_0) + \rho_1(c)u_1(c, m_1, z_1) \]  

(42)

If outcome 0 occurs good \( m \) provides no utility and the demand is equal zero, \( m_0 = 0 \). If outcome 1 occurs, good \( m \) provides utility and the demand will be positive. To make the example more concrete, interpret \( m \) as curative medical care and the two possible outcomes as "healthy" (\( i=0 \)) and "ill" (\( i=1 \)). Good \( c \) might be a consumption good which increases (e.g., cigarettes) or decreases (e.g., preventive care) the probability of illness.

The economy wide resource constraint facing the social planner can be written as

\[ x = t^c c + \Pi_1 t^m m_1 + \Pi_0 t^z z_0 + \Pi_1 t^z z_1. \]  

(43)

For this special case, the optimal tax/subsidy scheme can be characterized by the following expressions.
\[ t^m = \]
\[
\frac{1}{\lambda |A|} \left[ \begin{array}{c}
\text{cov}(\alpha, m) \left[ \left( \Pi_0 \left( \frac{\partial z_0}{\partial q^z} \right) + \Pi_1 \left( \frac{\partial z_1}{\partial q^z} \right) \right) \left( \frac{\partial c}{\partial q^c} \right) - \left( \Pi_0 \left( \frac{\partial z_0}{\partial q^c} \right) + \Pi_1 \left( \frac{\partial z_1}{\partial q^c} \right) \right) \left( \frac{\partial c}{\partial q^z} \right) \right] \\
- \text{cov}(\alpha, z) \left[ \left( \Pi_0 \left( \frac{\partial z_0}{\partial q^m} \right) + \Pi_1 \left( \frac{\partial z_1}{\partial q^m} \right) \right) \left( \frac{\partial c}{\partial q^c} \right) - \left( \Pi_0 \left( \frac{\partial z_0}{\partial q^c} \right) + \Pi_1 \left( \frac{\partial z_1}{\partial q^c} \right) \right) \left( \frac{\partial c}{\partial q^m} \right) \right] \end{array} \right] \\
\right]
\]

\[ t^z = \]
\[
\frac{1}{\lambda |A|} \left[ \begin{array}{c}
\text{cov}(\alpha, z) \left[ \left( \Pi_1 \left( \frac{\partial m_1}{\partial q^m} \right) \left( \frac{\partial c}{\partial q^c} \right) - \Pi_1 \left( \frac{\partial m_1}{\partial q^c} \right) \left( \frac{\partial c}{\partial q^m} \right) \right) \right] \\
- \text{cov}(\alpha, m) \left[ \left( \Pi_1 \left( \frac{\partial m_1}{\partial q^z} \right) \left( \frac{\partial c}{\partial q^c} \right) - \Pi_1 \left( \frac{\partial m_1}{\partial q^c} \right) \left( \frac{\partial c}{\partial q^z} \right) \right) \right] \end{array} \right] \\
\right]
\]
\( t^c = \)
\[-t^m \left[ \frac{\partial \rho_1}{\partial c} m_1 \right] - t^z \left[ \frac{\partial \rho_0}{\partial c} z_0 + \frac{\partial \rho_1}{\partial c} z_1 \right] \]
\[+ \frac{1}{\lambda} \frac{1}{|A|} \]
\[\times \left\{ -\text{cov}(\alpha, m) \left[ \Pi_0 \left( \frac{\partial z_0}{\partial q^z} \right)_0 + \Pi_1 \left( \frac{\partial z_1}{\partial q^z} \right)_0 \right] \Pi_1 \left( \frac{\partial m_1}{\partial q^z} \right)_0 - \left[ \Pi_0 \left( \frac{\partial z_0}{\partial q^c} \right)_0 + \Pi_1 \left( \frac{\partial z_1}{\partial q^c} \right)_0 \right] \Pi_1 \left( \frac{\partial m_1}{\partial q^c} \right)_0 \right\} \]
\[\times \left\{ -\text{cov}(\alpha, z) \left[ \Pi_0 \left( \frac{\partial z_0}{\partial q^c} \right)_0 + \Pi_1 \left( \frac{\partial z_1}{\partial q^c} \right)_0 \right] \Pi_1 \left( \frac{\partial m_1}{\partial q^c} \right)_0 - \left[ \Pi_0 \left( \frac{\partial z_0}{\partial q^m} \right)_0 + \Pi_1 \left( \frac{\partial z_1}{\partial q^m} \right)_0 \right] \Pi_1 \left( \frac{\partial m_1}{\partial q^m} \right)_0 \right\} \] (46)

where \( A \) is the matrix of expected compensated price effects.

\[
A = \begin{bmatrix}
\Pi_0 \left( \frac{\partial z_0}{\partial q^z} \right)_0 + \Pi_1 \left( \frac{\partial z_1}{\partial q^z} \right)_0 & \Pi_1 \left( \frac{\partial m_1}{\partial q^z} \right)_0 & \left( \frac{\partial \rho_0}{\partial q^z} \right)_0 \\
\Pi_0 \left( \frac{\partial z_0}{\partial q^m} \right)_0 + \Pi_1 \left( \frac{\partial z_1}{\partial q^m} \right)_0 & \Pi_1 \left( \frac{\partial m_1}{\partial q^m} \right)_0 & \left( \frac{\partial \rho_0}{\partial q^m} \right)_0 \\
\Pi_0 \left( \frac{\partial z_0}{\partial q^c} \right)_0 + \Pi_1 \left( \frac{\partial z_1}{\partial q^c} \right)_0 & \Pi_1 \left( \frac{\partial m_1}{\partial q^c} \right)_0 & \left( \frac{\partial \rho_0}{\partial q^c} \right)_0
\end{bmatrix} \] (47)
The compensated own-price effects are negative in expectation. The denominator in the tax formulae is the determinant of the Slutsky matrix and is negative in sign. The optimal scheme, then, depends on the covariances between marginal utility of income and $z$ and $m$ respectively, the cross-price effects between $c$, $z$ and $m$, and the impact of $c$ on $\rho_1$, the probability of outcome 1, and on $\rho_0$, the probability of outcome 0. Because $m_0 = 0$, we know that $\alpha_1 > \alpha_0$ implies $\text{cov}(\alpha,m) > 0$ and $\alpha_1 < \alpha_0$ implies $\text{cov}(\alpha,m) < 0$. For $z$ we know that if $\alpha_1 > \alpha_0$ and $z_t > z_0$, then $\text{cov}(\alpha,z) > 0$ and so on as represented below.

\[
\begin{array}{c|c|c}
\alpha_1 > \alpha_0 & \alpha_1 < \alpha_0 \\
\hline
z_t > z_0 & \text{cov}(\alpha,z) > 0 & \text{cov}(\alpha,z) < 0 \\
\hline
z_t < z_0 & \text{cov}(\alpha,z) < 0 & \text{cov}(\alpha,z) > 0 \\
\end{array}
\]

Inspection (44) reveals that if $\text{cov}(\alpha,m) > 0$ and $\text{cov}(\alpha,z) = 0$, then a subsidy for $m$ is advised (reimbursement insurance). If $\text{cov}(\alpha,m) < 0$ and $\text{cov}(\alpha,z) = 0$, then a tax on $m$ is the appropriate policy.

In the circumstances when there is an insurance effect for $z$, the analysis becomes more complex with ambiguous results and many possible subcases. As examples, I explore several of these subcases of $\rho^m$. Similar results apply to $\rho^z$. Cross-effects, in general, will render these results ambiguous. However, suppose $\text{cov}(\alpha,m) > 0$, $\text{cov}(\alpha,z) > 0$, $m$ and $z$ are expected substitutes, and $m$ and $z$ are independent of $c$, then a subsidy on $m$ ($\rho^m < 0$) is optimal. A subsidy to $m$ ($\rho^m < 0$) is also the optimal policy, if we suppose $\text{cov}(\alpha,m) > 0$, $\text{cov}(\alpha,z) < 0$, $m$ and $z$ are expected complements, and $m$ and $z$ are independent of $c$. 
In these two subcases, the optimal plan involves a subsidy to good $m$ (medical care), $r'' < 0$, which can be interpreted as an example is reimbursement health insurance. If $\alpha_0 < \alpha_1$, then $\text{cov}(\alpha, m) > 0$ and the consumer demands insurance against outcome 1 ("illness"). As outcome-contingent lump-sum payments are impossible, the smoothing of income across outcomes is accomplished with a subsidy. A tax serves the smoothing functions if the covariance is negative.

Given that taxes/subsidies to $z$ and $m$ are optimal, some taxation of (subsidization to) consumption of $c$ will be optimal. Inspection of equation (46) reveals that tax (subsidy) rate is partly governed by the marginal cost of moral hazard, the terms associated with rho. Given that there is subsidy for $m$ ($r'' < 0$) the formula (eq. (46)) tends to tax (subsidize) good $c$ if consumption increases (decreases) the probability of outcome 1. This is basically the case covered by Arnott and Stiglitz (1986) in their treatment of moral hazard. However, because this is reimbursement insurance, the optimal tax formulae for $r'$, equation (46) involves a cross-price effects which add ambiguity to the results.

If the consumption of good $c$ increases (decreases) the probability of the insured outcome, then moral hazard implies a tax on (subsidy to) good $c$. In the case of cigarettes, it is reasonable to assume that increased consumption will increase the probability of illness. This is a type of moral hazard and by itself implies taxation of cigarettes. If cigarettes and medical care are complements, then the optimal plan is to tax cigarettes. However, if cigarettes and medical care are sufficiently strong substitutes, then a subsidy to cigarette consumption may be optimal. Although our intuition is that cigarettes and medical care are complements, it is largely an unresolved empirical issue.
U.S. national health policy is looking at cigarette taxes as a source of financing for health insurance and as a method for increasing efficiency in the health sector. The results here imply the efficiency role will be fulfilled if cigarettes and health services are complements.

3. SUMMARY

Risk averse consumers demand insurance if the marginal utility of income varies across outcomes. If outcomes are unobservable, outcomes are directly uninsurable. Indirect public insurance can be provided by subsidizing commodities associated with outcomes for which there is a fundamental demand for insurance. This is the first main result. It is a result which is consistent with and explanatory of reimbursement insurance, the most common form of health insurance.

Given that it is optimal to provide reimbursement insurance, it is also optimal to implement other commodity taxes/subsidies to offset the distortion created by the insurance. Publicly provided reimbursement insurance creates two distortions, moral hazard and a subsidy distortion. Moral hazard calls for the taxation (subsidization) of consumption goods which increase (decrease) the probability of insurance related outcomes. The second distortion calls for the taxation (subsidization) of complements to (substitutes for) the insured commodity. This is the second main result.

For Arnott and Stiglitz (1986), the direction of the moral hazard effect is sufficient to determine commodity taxation of subsidization. This paper shows that for reimbursement insurance cross-price effects must be considered in addition to moral hazard.

I illustrate the model with an example centered on cigarettes, medical care and health
insurance. If cigarettes and medical care are sufficiently strong substitutes a subsidy to cigarettes may be optimal in the presence of health insurance distortions. The model implies that health consequences alone are insufficient to determine the optimal tax policy. Price relationships must also be examined.
REFERENCES


NOTES

1. There are exceptions of course as some cancer policies do provide lump-sum payments irrespective of treatment. However, reimbursement insurance (or service benefit) is used by public insurers and is also the dominant form of contract in private health insurance markets.

2. Life insurance contracts are also typically written to provide a lump-sum cash indemnity. The life insurer can observe death at a relatively modest cost. Subsequent to the confirmation of this outcome, the insurer provides a lump-sum cash payment. This is efficient in that such a payment does not distort the price of funeral services or other commodities that might be purchased by the beneficiaries. In contrast to life insurance, health insurance is quite distorting in this regard.

3. Assuming an interior solution, the first step optimization can be characterized by the first order conditions which follow:

\[
\frac{\partial u_i}{\partial m_{ij}} - \alpha_i q_j^m = 0 \quad j = 1, \ldots J
\]

\[
\bar{Y}_i - q^m \cdot m_i = 0
\]

Implicitly solving the first order conditions renders the restricted demand functions with outcome \( i \).

\( m_i(c, \bar{Y}_i, q^m) \)

Substituting the demands into the restricted direct utility function, equation (8), renders the restricted indirect utility function defined by (11).

\( v_i(c, \bar{Y}_i, q^m) = u_i(m_i(c, \bar{Y}_i, q^m), c) \)

4. I assume the representative consumer is an expected utility maximizer.

\[
\max_c \sum_i \rho_i(c) v_i(c, Y_i - q^c \cdot c; q^m)
\]

Further, assuming an interior solution, the first order conditions for the second step optimization are

\[
\sum_i \left[ \frac{\partial \rho_i}{\partial \epsilon} v_i + \rho_i \left( \frac{\partial v_i}{\partial \epsilon} - \frac{\partial v_i}{\partial \bar{Y}_i} q^c\right) \right] = 0 \quad k = 1, \ldots K.
\]

Solving implicitly gives the demand vector \( c \):

\( c(q^c, q^m, Y) \)

Substitution renders the demand vectors \( m_i, i = 1, \ldots J \).

\( m_i(q^c, q^m, Y) = m_i(c(q^c, q^m, Y), Y_i - q^c \cdot c(q^c, q^m, Y), q^m) \)

5.

\[
\frac{\partial EV}{\partial Y_i} = \sum_i \left[ \frac{\partial \Pi_i}{\partial Y_i} v_i + \Pi_i \frac{\partial v_i}{\partial Y_i} \right] = \sum_i \sum_k \left[ \frac{\partial \rho_i}{\partial \epsilon} v_i + \rho_i \left( \frac{\partial v_i}{\partial \epsilon} - \frac{\partial v_i}{\partial \bar{Y}_i} q^c\right) \right] \frac{\partial c_k}{\partial Y_i} \Pi_i \alpha_i = \Pi_i \alpha_i
\]
6. 
\[
\frac{\partial EV}{\partial q_k} = \sum_i \left( \frac{\partial \Pi_i}{\partial q_k} V_i + \Pi_i \frac{\partial V_i}{\partial q_k} \right) = \sum_k \sum_i \left( \frac{\partial p_i}{\partial c_k} v_i + \rho_i \frac{\partial v_i}{\partial c_k} - \frac{\partial v_i}{\partial q_k} c_k \right) \frac{\partial c_k}{\partial q_k} = \sum_i \Pi_i \alpha_i c_k = - \sum_i \Pi_i \alpha_i c_k
\]

7. 
\[
\frac{\partial EV}{\partial q_j} = \sum_i \left( \frac{\partial \Pi_i}{\partial q_j} V_i + \Pi_i \frac{\partial V_i}{\partial q_j} \right) = \sum_k \sum_i \left( \frac{\partial p_i}{\partial c_k} v_i + \rho_i \frac{\partial v_i}{\partial c_k} - \frac{\partial v_i}{\partial q_j} c_k \right) \frac{\partial c_k}{\partial q_j} = \sum_i \Pi_i \alpha_i m_{ij} = - \sum_i \Pi_i \alpha_i m_{ij}
\]

8. Note that
\[
\frac{\partial \Pi_i}{\partial q_j} = \sum_k \frac{\partial p_i}{\partial c_k} \frac{\partial c_k}{\partial q_j}
\]
and
\[
\frac{\partial \Pi_i}{\partial Y_i} = \sum_k \frac{\partial p_i}{\partial c_k} \frac{\partial c_k}{\partial Y_i}
\]

9. 
\[
\frac{\partial \Pi_i}{\partial q_k} = \sum_k \frac{\partial p_i}{\partial c_k} \frac{\partial c_k}{\partial q_k}
\]

10. Exact compensation
\[
\left( \frac{\partial c_k}{\partial q_k} \right)_b = \frac{\partial c_k}{\partial q_i} + \sum_i \frac{\partial c_k}{\partial Y_i} c_k
\]
\[
\left( \frac{\partial m_{ij}}{\partial q_k} \right)_b = \frac{\partial m_{ij}}{\partial q_i} + \sum_i \frac{\partial m_{ij}}{\partial Y_i} c_k
\]
holds both expected utility and utility with each outcome constant. Approximate compensation
\[
\left( \frac{\partial m_{ij}}{\partial q_j} \right)_b = \frac{\partial m_{ij}}{\partial q_j} + \sum_i \frac{\partial m_{ij}}{\partial Y_i} \sum_i \Pi_i m_{ij}
\]
\[
\left( \frac{\partial c_k}{\partial q_j} \right)_b = \frac{\partial c_k}{\partial q_j} + \sum_i \frac{\partial c_k}{\partial Y_i} \sum_i \Pi_i m_{ij}
\]
weights the income effect with expected demand rather than actual demand and thus over-compensates low demand outcomes and under-compensates high demand outcomes.