Optimal Health Insurance for Multiple Goods and Time Periods

Randall P. Ellis\textsuperscript{a,*}, Shenyi Jiang\textsuperscript{b}, Willard G. Manning\textsuperscript{c}

\textsuperscript{a} Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215, USA. <ellisrp@bu.edu>
\textsuperscript{b} Hanqing Advanced Institute of Economics and Finance, School of Finance, Renmin University of China, Beijing, P.R. China. <syijiang@gmail.com>
\textsuperscript{c} Harris School of Public Policy Studies, The University of Chicago, Chicago IL USA. <w-manning@uchicago.edu>

Acknowledgements: We have benefited from comments during presentations at Boston University, Harvard and Renmin University, The University of Chicago, the AEA annual meeting, and the American Society of Health Economists conference. We also thank David Bradford, Kate Bundorf, Albert Ma, Tom McGuire, David Meltzer, Joe Newhouse, Kosali Simon, and Frank Sloan for their useful comments. The opinions expressed are those of the authors, and not those of Boston University, Renmin University, or The University of Chicago.
Abstract

This paper reexamines the efficiency-based arguments for second-best optimal health insurance while considering multiple treatment goods and multiple time periods. We show how correlated shocks across health care goods and time should affect optimal insurance plans: complementarity and substitutability of multiple goods and correlated demand shocks interact to affect the optimal cost sharing: substitutes and positively correlated demand shocks across health care goods reduce optimal patient cost sharing. Correlation over time will decrease coinsurance if positive or increase it if negative. Positively correlated uncompensated losses also reduce optimal cost sharing on covered health services. Our results provide a rationale for covering pharmaceuticals and outpatient spending more fully than is implied by static, one health care good or one period models.

1. Introduction

A central theme in health economics is the behavior of patients and providers in the presence of partial health insurance that covers some but not all of the cost of health care. The economic motivation for such coverage is that risk-averse individuals can reduce their financial risk by pooling risks with others through insurance that effectively shifts funds from healthy to sick states of the world to reduce the financial risk from uncertain health. Partial, rather than complete, insurance can be desirable to offset the adverse effects of moral hazard when the marginal payment paid by the consumer/patient for a service is less than the social cost of producing that service. Since patients respond to lower out-of-pocket prices of health care, health insurance increases the amount of the care that they purchase, generating a deadweight loss from the increased use. There is a great deal of empirical support for the law of demand applying to health care in the literature, coming from observational studies, natural experiments, and randomized experiments (Newhouse et al., 1981; Manning et al., 1987; Zweifel and Manning, 2000, Baicker and Finkelstein, 2011).

We focus here on the normative issues of what the second-best coinsurance policies should have over multiple treatment goods and multiple periods. This topic has taken on new importance in the recent American debate on health care reform, where the demand-side cost sharing has reemerged as a central cost containment strategy. Cost sharing plays a central role in the health insurance exchanges, which promotes alternative plan tiers (platinum, gold, silver, bronze) that differ primarily in their covered benefits and cost sharing. This issue is also central in Value-Based Insurance, which promotes higher coverage for more effective services.
(Fendrick et al, 2001; Choudhry et al, 2010). Key issues in this policy discussion are what “essential benefits” should be covered, and how to justify higher or lower consumer cost sharing. In an earlier paper we examined optimal insurance where the tradeoff is between one treatment good and one prevention service (Ellis and Manning, 2007); in this paper we focus on the new issues that arise from more than one dimension of uncertainty coming from multiple treatment goods and multiple time periods.

Although we are fully aware of the new insights to insurance from behavioral economics (e.g., Kunreuther, Pauly and McMorrow, 2012), we nonetheless adopt a neoclassical approach and derive results based on a model of expected utility maximization in which health care demand curves reflect true marginal benefits, consumers are risk averse, and only demand side cost sharing is available to influence costs. Even if consumers make the wrong choices because of misperceptions, inertia, or for other reasons, it can still be of interest to model optimal insurance more formally as a gold standard with which to contrast imperfect choices. Even under alternative assumptions, there will always be a tradeoff between deadweight losses from moral hazard and welfare gains from the risk protection offered by incomplete insurance.

Our work is not the first to carefully consider multiple health treatment goods (Goldman and Philipson, 2007), but to our knowledge it is the first to examine the issues that arise from multiple, correlated sources of uncertainty, whether the correlations arise due to correlated demands across different health care goods at a point in time, correlated demand shocks over time, or the correlated losses that arise from uncompensated losses (e.g. job loss) that accompany the treatment cost shocks that can be covered by the insurance plan. We derive four new findings about second-best optimal insurance:

1. All other things being equal, positive contemporaneous covariances among goods justify lower coinsurance rates, while negative correlations justify higher coinsurance rates. This finding moderates previous findings on the role of complements vs. substitutes.

2. Serial correlation over multiple periods matters because the positive correlations (such as with chronic care or prescriptions filled) justify lower coinsurance rates. Static models of only one period miss the implied accumulation of risk over time.

3. Health-related losses that are not fully compensated by insurance, such as lost income or time costs of treatment, also justify lower coinsurance rates.

4. Savings behavior moderates optimal coinsurance rates and depends on the nature of serial correlation.
In the following sections of the paper, we provide a systematic review of the existing literature, a detailed description of our formal microeconomic model where two of the limiting cases are already in the literature, some empirical estimates from a set of large employment based insurance plans, and finally a set of normative conclusions. As we proceed, we show how more traditional rules for coinsurance rates change as a result of our extensions beyond the one good, one period model.

2. Literature Review

Much of the economic literature on optimal health insurance focuses on “the fundamental tradeoff of risk spreading and appropriate incentives” (Cutler and Zeckhauser, 2000, p. 576). Specifically, it examines either the dead weight losses from moral hazard, the tradeoff between moral hazard and the gains from insuring against financial risk, or the differential coverage of multiple goods with varying degrees of risk. Much of this work employs a one-period, one health care good model with uncertainty about health states or uncertainty about levels of health care expenditure [Arrow, 1963; Besley, 1988; Cutler and Zeckhauser, 2000; Dardanoni and Wagstaff, 1990; Pauly, 1968; Spence and Zeckhauser, 1971; Zeckhauser, 1970]. Papers that derive the optimal insurance contracts using this framework have employed variants of the tradeoff between the risk premium (as reflected by the Arrow-Pratt approximation) and the deadweight loss from moral hazard (as reflected in Harberger loss or related measures) or the compensating variation (Manning and Marquis, 1996). See Feldstein (1973), Feldstein and Freedman (1977), Buchanan and Keeler (1991), Manning and Marquis (1996), Newhouse et al. (1993), Feldman and Manning (1997) for other theoretical and empirically-based examinations of optimal insurance. Blomqvist (1997) extends the theory to nonlinear insurance schedules.

Almost all of this literature has been based on either a one-period model or a two-period model where the consumer selects the coinsurance before knowing his or her realized state of health, but chooses her health care expenditures conditional on the state of the world that occurs. The common result in this literature is that one should select coverage in a plan with a constant copayment or coinsurance rate such that the marginal gains from risk reduction from a change in the coinsurance (copayment) rate just equal the marginal costs of increasing moral hazard. The consensus of the empirical strain of this literature is that optimal levels of cost sharing usually involve neither full insurance (zero out-of-pocket cost), nor no insurance. Depending on the
formal model approach and the data employed, optimal coinsurance rates range from the 50-60 percent range (Feldstein and Friedman, 1977; Manning and Marquis, 1996) down to values that are in the mid 20 percent range or lower, possibly with a deductible and/or stop-loss (Blomqvist, 1997; Buchanan et al., 1991; Feldman and Dowd, 1991; Feldman and Manning, 1997; Newhouse et al., 1993).

In this paper we reconsider the optimal insurance for health care in markets with one nonhealth good and two or more health care goods. We do this in the context of both one and multi-period models. We are aware of only three papers that have considered multiple health goods. Besley (1988) provides a multi-good extension to this literature in which demands for health care goods are stochastic, but does not model either complementarity or substitutability between goods or consider correlated shocks in demand. Goldman and Philipson (2007) model two goods in one period to show how complementarity and substitutability of health care services affects optimal cost sharing in an expected utility format, but do not model multiple dimensions of uncertainty. Ellis and Manning (2007) model the case of one treatment good and one prevention good to highlight how optimal insurance rules differ for the two types of goods, but they include only one health demand shock and do not model multiple treatment goods or correlations over time.

To keep the model tractible, we take a step backwards from Ellis and Manning and do not examine the preventive effects of health care spending. In doing so we also differentiate our work from Hall and Jones (2007) in which health care affects mortality and life expectancy, but not utility, directly. Using the same utility structure as Ellis and Manning, we examine how optimal cost sharing is affected by the correlation structure of random shocks affecting demand for health care both across goods and over time. We also examine how savings and uncompensated health losses affect the optimal insurance calculations.

The important insight here is that, all other things being equal, health care goods that are positively correlated should be more generously insured than those that are negatively correlated or uncorrelated. This holds both for contemporaneously correlated health care treatment goods and shocks that are serially correlated over time. The basic logic is that if the demand for two goods or over two periods is uncertain, then the ex ante combined variance is larger if they are
positively correlated than if there is no or negative correlation.\footnote{As shown below, the details are more complicated because optimal coinsurance also depends on the variability in demand, own and cross price effects, correlated error shocks, and (in the cases of multiple periods) the discount rate.} Risk averse individuals will prefer more generous insurance (lower coinsurance rates) to reduce their financial risk when the goods are correlated than if they ignored the correlation or treated the goods as independent.

We first study the optimal demand-side insurance coverage for health care treatment when there are two health care treatment goods. After developing a general analytical model as our “base case,” we do comparative statics on a series of special cases. A key attraction of our specification is that we not only solve for the optimal cost sharing rate as a closed-form solution to recreate the results from the basic model but also address a richer class of economic behavior. These include new results involving uncompensated health care losses, correlated health care shocks, and cross price elasticities of demand with multiple goods.

Our second set of analytical results considers a two-period model in which health care shocks persist over time due to chronic conditions. In a multi-period context, if a consumer’s savings react to healthcare shocks, then both the cost of risk and the optimal cost sharing rates change. Positively serially correlated shocks (e.g. chronic illness) imply that healthcare should be more generously covered (lower cost sharing) than when shocks are independent or negatively correlated across periods. Our final set of analytical results considers the two good, two period model, with both contemporaneous and serial correlations, which gives insights into optimal coverage for acute versus chronic conditions.

The concluding section of the paper discusses a few empirical results that have a bearing on our analytical findings. We briefly discuss empirical estimates of the variance of three broad sets of services and the magnitudes of contemporaneous and intertemporal correlations that shed light on the empirical relevance of our findings.

3. **Model**

Following Arrow (1963), Pauly (1968), and Zeckhauser (1970), we focus only on the demand-side while examining optimal cost sharing. We consider risk aversion, moral hazard, insurance loading costs and uncompensated losses without attempting to model other concerns that influence optimal insurance coverage such as the following: correcting for externalities, such as those that can occur with communicable diseases (Hofmann, 2007); altruism or public good
arguments for insurance coverage (Coate, 1995; Rask and Rask, 2000); distributional concerns including goals like the elimination of poverty or achieving social solidarity (Andrulis, 1998; Atim, 1999; Maarse and Paulus, 2003; Sin et al., 2003); corrections of informational problems (i.e. consumers make the wrong decisions) (Doherty and Thistle, 1996; Choudhry et al, 2010); or insurance that fosters more complete coordination among health care providers (Duggan, 2005; Lichtenberg, 2001; and Newhouse, 2006). Without denying the relevance of these other factors affecting optimal insurance, we set them aside and reexamine the efficiency-based arguments for insurance with multiple goods and periods and derive new results which refine our understanding of the value of generous insurance coverage from the consumer’s point of view.

We assume the individual’s utility function involves one non-health consumption good, \( Y \), and two health treatment goods, \( X = \{X_1, X_2\} \), first within one period and then over two time periods.\(^2\) Prices of the three goods are \( P_y, P_1, \) and \( P_2 \), and the consumer’s income in each period is \( I \). In the underlying behavioral model, there is a health production function that transforms health care \( X \) into health status.

Following much of the literature (for example, Cutler and Zeckhauser, 2000), we examine only health insurance plans with a constant coinsurance rate \( 0 \leq c \leq 1 \) for both health treatment goods. We briefly introduce loading costs of insurance, which are assumed to be a constant proportion \( \delta \) of insurance payments. Insurance premiums, \( \pi \), are competitively determined and depend on the copayment rates and the demand structure, but do not vary across individuals. The insurance policy is thus a pure coinsurance plan with no deductible, stop-loss, or limit on the maximum expenditure or level(s) of covered services.

We assume the following sequence of steps in our full model:

1. The insurer chooses the premium \( \pi \) and coinsurance rates \( \{c_1, c_2\} \) for health care treatment for goods \( X_1 \) and \( X_2 \), respectively.
2. Nature decides on the consumer’s state of illness as a vector of random health shocks, \( \{\theta_i\} \), that affect the demand for the vector of health care goods, \( X \).
3. The consumer chooses quantities \( X \) and \( Y \) to maximize period 1 utility.
4. If a two-period model, repeat steps 2 and 3 for Period 2.

\(^2\)In an earlier version of the paper, we examined more general notation that allowed for an arbitrary number of health care goods, but we found that all of the relevant intuition is obtained using only two health care goods.
Conditional on the available technology, the demand for medical care services has been shown by many empirical studies to be highly income inelastic, with $\mu_I = .1$ or $.2$, (Chernew and Newhouse, 2012). For simplicity, we assume that $X$ is perfectly income inelastic: $\partial X / \partial I = 0$.

While this income elasticity assumption is strong, it buys us a great deal of simplicity that enables many closed form solutions for cases with multiple health treatment goods. While we make this strong assumption about income elasticities, we make weaker assumptions about other parameters below. We avoid concerns about corner solutions by further assuming that income is always sufficient to pay for at least some of all other goods $Y$ after paying for $X$.

Utility in every period is separable in health status and the utility of consumption. A corollary of this is that health care shocks do not have any effect on the marginal utility of income, other than through their effect on medical expenditures. Health shocks affect the marginal utility of income through spending on medical care, but do not directly affect the marginal utility of income. This is a common theoretical assumption, used in many empirical studies (e.g., Arrow, 1963; Zeckhauser, 1970; Manning and Marquis, 1996).

4. Single period model with two health goods

Demand curves for each of two goods are assumed to be linear.

$$
X_1 = A_1 - B_1 P_1 / P_Y + G_{12} P_2 / P_Y
$$

$$
X_2 = A_2 - B_2 P_2 / P_Y + G_{12} P_1 / P_Y
$$

For simplicity, we normalize the marginal costs of all goods to be one, and express all prices in terms of the share of this marginal cost paid by the consumer. Hence $P_Y = 1$ and $P_i = c_i$, and the cost share is the consumer price of the $i^{th}$ health care good $X_i$. Note the presence of the (symmetric) cross price coefficient, $G_{12}$, which will play a key role later. As in Ellis and Manning (2007), we also incorporate the idea that illness may involve other uninsured (uncompensated) losses. These losses can be of two types: uncompensated out-of-pocket

---

3 Across countries or over time, the income elasticity is greater than one (Newhouse, 1992; Hall and Jones, 2012). Most likely because the available technology grows with income, but for an efficiency analysis of optimal cost sharing, the relevant elasticity is the short run income elasticity, which is very low. Holding social, demographic, environmental and supply side attributes constant, low and high income individuals demand very similar levels of health care even across dramatically different income levels in generous health plans.
costs (which we assume are proportional to covered costs), and \( L \), the uncompensated health shock losses, \( L_\theta(\theta) \).

The marginal benefit function for each medical service \( X_i \) is assumed to be a linear demand curve with a constant slope \( B_i \) on the price regardless of the realization of the random health shock \( \theta \). Stochastic health treatment demand is introduced by letting \( A_i = \mu_i + \theta_i \), where \( \theta \sim F(\theta) \), with \( E(\theta) = 0 \).

Assumptions about the variance/covariance vector of \( \theta \) are made below but throughout we assume that the distribution does not depend on the out-of-pocket price or income. This corresponds to the horizontal intercept of the demand curves having a mean of \( \mu_i \) when the full consumer price \( (c_i + L_i) \) is zero. Using this convention, a single consumer’s demand curve for each medical service \( X_i \) can be written as

\[
X_i = \mu_i - B_i(c_i + L_i) + \theta_i
\]

In order to introduce risk aversion, we apply a monotonically increasing concave function \( V(\ldots) \) to the indirect utility function consistent with the demand function in Equation (2). Using this notation, we write the one period indirect utility function with two treatment good as

\[
V(I, C_X, \theta) = V \left[ J + B_i(c_i + L_i)^2 + B_2(c_2 + L_2)^2 - (\mu_i + \theta_i)(c_i + L_i) \right] - L_{\hat{\theta}_1}(\theta_1) - L_{\hat{\theta}_2}(\theta_2)
\]

where \( J = I - \pi \).

That this utility function corresponds to the above demand functions (including their symmetric cross price terms \( G_{12} \)) can be verified by applying Roy’s identity. Using the linear demand equation for \( X \), the insurer’s break-even condition for the insurance premium is

\[
\pi = (1 + \delta) \left\{ (1 - c_1) \left[ \mu_i - B_i(c_i + L_i) + G_{12}(c_2 + L_2) \right] \right\} + (1 - c_2) \left[ \mu_2 - B_2(c_2 + L_2) + G_{12}(c_1 + L_1) \right]
\]

where \( \delta \) is the administrative loading factor such that insurance costs proportion \( \delta \) more than actuarially fair insurance. In a two-period model, we assume that the same premium is charged in both periods.
While we have used the somewhat restrictive assumptions of linear demand, additive errors, and zero income effects, this specification has two attractive features. The error terms only interact with cost sharing \( c_i \) in a simple multiplicative form. This facilitates introducing multiple goods and multiple periods. The linear specification also allows us to consider cross price elasticities in a natural way.

We now turn to the social planner’s problem of choosing the optimal coinsurance rates when there are two health care treatment goods \((X_1, X_2)\) and a composite all-other-goods commodity, \(Y\). We develop the model using a general specification and then derive various cases of interest as special cases. The optimal cost sharing rates \( c_i \) for health care treatment will maximize the expectation of Equation (3). Taking its partial derivative with respect to \( c_i \) and setting equal to zero yields an equation that characterizes the social optimum. Since this expression will not in general have a simple closed form solution, we take a Taylor series approximation of the partial derivative \( V' \), evaluated around the nonstochastic arguments of the utility function. This solution can be written as

\[
0 = \frac{\partial E_\theta V}{\partial c_1} = E_\theta \left\{ V_I \left[ J - K \right] - V_II \left[ J - K \right] \left[ \theta_1 (c_1 + L_1) + \theta_2 (c_2 + L_2) \right] \right\}
\]

where

\[
J = I - \pi,
\]

\[
K = (c_1 + L_1) \mu_1 + (c_2 + L_2) \mu_2 - B_1 (c_1 + L_1)^2 / 2 - B_2 (c_2 + L_2)^2 / 2
\]

\[+ G_{12} (c_1 + L_1)(c_2 + L_2),\]

\[
\pi = (1 + \delta) \left\{ (1 - c_1) \left[ \mu_1 - B_1 (c_1 + L_1) + G_{12} (c_2 + L_2) \right] + (1 - c_2) \left[ \mu_2 - B_2 (c_2 + L_2) + G_{12} (c_1 + L_1) \right] \right\},
\]

\[\frac{\partial \pi}{\partial c_1} = (1 + \delta) \left[ -\mu_1 - B_1 + 2 B_1 c_1 + B_1 L_1 + G_{12} - 2 c_2 G_{12} - G_{12} L_2 \right].\]

Defining the absolute risk aversion parameter \( R^A \equiv -V'' / V', \sigma_1^2 \equiv E(\theta_1)^2, \sigma_2^2 \equiv E(\theta_2)^2, \) using \( \sigma_{12} \equiv E(\theta_1 \theta_2), \) and using \( E(\theta_1) = E(\theta_2) = 0, \) we show in the appendix that this result can
be rearranged to obtain first order conditions characterizing the optimal coinsurance rates $c_1$ and $c_2$ as follows.

\begin{align}
(1 + \delta)[ -\mu_1 - B_1 + 2B_1c_1 + B_1L_1 + G_{12} - 2c_2G_{12} - G_{12}L_2 ] + \mu_1 - B_1(c_1 + L_1) + G_{12}c_2 \\
+ R^4 \left[ \sigma_1^2 (c_1 + L_1) + \sigma_{12} (c_2 + L_2) \right] &= 0 \\
\end{align}

\begin{align}
(1 + \delta)[ -\mu_2 - B_2 + 2B_2c_2 + B_2L_2 + G_{12} - 2c_1G_{12} - G_{12}L_1 ] + \mu_2 - B_2(c_2 + L_2) + G_{12}c_1 \\
+ R^4 \left[ \sigma_2^2 (c_2 + L_2) + \sigma_{12} (c_1 + L_1) \right] &= 0
\end{align}

We consider a variety of ways of interpreting these two equations below. In broad terms, the first line of each equation gives the marginal cost of increasing the cost-sharing rate, while the second line, which involves $R^4$, gives the marginal benefit in terms of reduced cost of risk.

4.1 One health care good, base case

We first examine a Base Case of a single good in a single period with no uncompensated costs and no loading factor, which using our notation corresponds to

$$L_{x_1} = L_{x_2} = \delta = \mu_{x_2} = \sigma_2 = \sigma_{12} = B_2 = G_{12} = 0.$$ This yields the well-known result from the literature characterizing the second-best optimal insurance contract for constant coinsurance rate plans when there is a simple tradeoff between moral hazard and the cost of risk:

\begin{align}
-B_1 (1-c_1) + R^4 \sigma_1^2 c_1^* &= 0
\end{align}

where $B_1 \geq 0$ and the first term (the marginal costs due to moral hazard) is increasing in the price response $|B_1|$. The gains from risk pooling are increasing in the variance of health care demand. Solving for the optimal coinsurance rate yields Equation (9), where the optimal base case coinsurance rate, $c_{X_1}$ is increasing in the price response, $B_1$, and decreasing in the underlying variance in demand, $\sigma_1^2$.

\begin{align}
c_1^* &= \frac{B_1}{B_1 + R^4 \sigma_1^2} \equiv c_{X_1}^{BaseCase}
\end{align}

Intuitively, there is a tradeoff in changing $c$ to increase risk bearing as the patient pays more of the price, and to reduce moral hazard. The optimal coinsurance rate is the ratio of the loss from moral hazard to the net loss from changing risk bearing plus moral hazard. If demand
is perfectly inelastic \((B_1 = 0)\) then the optimal coinsurance rate is zero if there is any risk
\((\sigma^2_i > 0)\) at all. If there is no variance or the consumer is risk neutral \((R^d = 0)\), then the optimal coinsurance rate should be 1. Here, the optimal coinsurance rate lies between 0 and 1:
\[
0 < c^*_1 \text{BaseCase} < 1.
\]

**4.2 One health care good, with insurance loading \((\delta > 0)\)**

We next relax the assumption of no insurance loading factor, allowing for administrative costs and profits. The optimal coinsurance rate expression becomes

\[
c^*_i = \frac{B_i + (B_i \delta + \delta \mu_i)}{B_i + R^d \sigma^2_i + (2\delta B_i)}
\]

As long as the insurance loading factor, \(\delta\), is not prohibitively large, then \(0 < c^*_i < 1\), and
\[
\frac{\partial c^*_i}{\partial \delta} > 0.
\]

Moreover, the mean expected level of spending (with a marginal out-of-pocket price of zero), \(\mu_{X_1}\), enters in the numerator such that as costs go up, it is desirable to increase the cost sharing rate to reduce the inefficiency due to the insurance loading factor \((\delta > 0)\). Having a positive loading factor increases the costs of moral hazard relative to the gains from reducing risk. We leave the loading factor \((\delta = 0)\) for the remainder of our results.

**4.3 Adding uncompensated health-related losses in the one good case \((L_i > 0)\)**

Incorporating uncompensated health loss so that \(L_i > 0\) also affects optimal cost sharing for covered treatment goods, increasing the desired coverage (decreasing the cost sharing rate \(c^*_i\)) that we found in our analysis of prevention versus treatment (Ellis and Manning, 2007). The expression for the optimal cost sharing rate on health good \(X_i\) becomes

\[
c^*_i = c^*_{i \text{BaseCase}} - \frac{R^d \sigma^2_i L_i}{B_i + R^d \sigma^2_i}
\]

The optimal coinsurance rate \(c^*_i\) is decreasing in the size of the uncompensated loss \(L_i\), because all of the terms after the minus sign are positive. With uncompensated losses, it is also possible for \(c^*_i\) to be negative or to reach a corner solution where \(c^*_i = 0\) for either large \(L_i\) or
small $B_i$. Equation (11) provides an efficiency-based rationale for why full insurance can be second-best optimal: the absence of complete insurance markets to fully transfer income into high-cost health states of the world means that coinsurance rates are set at or closer to zero than they would have been if the alternative insurance markets were complete and consumers were able to insure against all health care losses. There are many health services and conditions which have substantial uncompensated health care related losses. This is particularly true in developing countries where disability and unemployment insurance are rare and productivity losses from ill health are often large. Wagstaff (2007) provides documentation of the large magnitudes of income losses from illness in Vietnam. Thus, incomplete insurance markets provide a rationale for more generous insurance coverage of health care treatment, even when welfare losses due to moral hazard and insurance loading may be important.

4.4 Multiple health care treatment goods

\[
\mu_{x_2} > 0, \sigma^2_2 > 0, \sigma_{12} > 0, B_2 > 0, L_1 = L_2 = \delta = 0
\]

An important motivation in this paper for modeling two rather than one health care treatment goods is to be able to examine the role of cross price elasticities and correlated health shocks. For ease of exposition, we now assume there are no uncompensated health losses and no insurance loading factor, $\delta = 0$, but we explore the general case for demand parameters with two goods. In contrast to the situation with two goods in a one period model with no uncompensated losses, there are now shifts in both the marginal costs from the dead weight loss of $c < 1$ and shifts in the risk bearing. If $\delta = L_{x_1} = L_{x_2} = 0$, then (6) and (7) above can be solved for the optimal cost share for health good 1 and simplified to:

\[
(12) \\ c^*_{x_1} = c^*_1 \text{Base Case} - \frac{G_{12}}{B_1 + R^4 \sigma_1^2} (1 - c^*_2) - \frac{R^4 \sigma_{12}}{B_1 + R^4 \sigma_1^2} c^*_2
\]

with a similar expression for the second health care good. Since all of the terms multiplying $\sigma_{12}$ and $G_{12}$ are non-negative, the negative sign means that the first order effect of increasing either term is to reduce the optimal cost share $c^*_{x_1}$ relative to the base case. Because

\[
\frac{\partial c^*_{x_1}}{\partial G_{12}} = \frac{c^*_2 - 1}{B_1 + R^4 \sigma_2^2} < 0,
\]

reducing $c^*_2$ reinforces the fact that $\frac{\partial c^*_1}{\partial G_{12}} < 0$. Hence, when two health services become stronger gross substitutes in the sense that $G_{12} = \partial X_1 / \partial c_2$ is increased, then both
services should have lower cost sharing rates. This finding replicates the finding of Goldman and Philipson (2007) that as goods become stronger complements \( G_{12} < 0 \) they should have higher cost sharing\(^4\).

The implications of the covariance between health goods on optimal insurance are more complex. Because of the second order effect of \( \sigma_{12} \) on \( c_i^* \) we cannot unambiguously sign \( \partial c_i^* / \partial \sigma_{12} \) for all possible values of \( G_{12} \) and \( \sigma_{12} \). We can evaluate this term for certain special cases. For example, the partial derivative can be signed as negative for the limiting case where \( \sigma_{12} \) approaches zero. Another set of special cases correspond to when \( G_{12} \) takes on certain values, which is easiest to see graphically. Figures 1a-c show the effect of changes in the optimal coinsurance rate for three ranges of \( G_{12} \) as the covariance term \( \sigma_{12} \) increases from zero to a positive level, where in each figure the solid line corresponds to the equations (6) and (7) with \( \sigma_{12} = 0 \). The figures examine \( \partial c_i^* / \partial \sigma_{12} \) for the three possible cases for \( G_{12} \) according to whether goods 1 and 2 are price neutral \( (G_{12} = 0) \), complements \( (G_{12} < 0) \) or substitutes \( (G_{12} > 0) \). In the first two cases, the optimal coinsurance rate for both health care goods will tend to fall as the covariance of the goods increases. In each case, as the covariance \( \sigma_{12} \) becomes more positive, Line 2 (Equation (7)) pivots clockwise around the origin, while Line 1 (Equation (6)) rotates counterclockwise, ensuring that both coinsurance rates fall.

However, when the two goods are substitutes \( (G_{12} > 0) \), it is possible for one line to rotate further than the other. This depends on the magnitudes of \( B_i \) and \( G_{12} \). A good with a higher price response \( B_i \) or a greater variance \( \sigma_i^2 \) will rotate less. In this situation, it is possible for one good to have its coinsurance rate fall while the other stays the same or increases. Figure 1c illustrates the case where the two goods are substitutes, \( G_{12} > 0 \), and \( \partial c_1^* / \partial \sigma_{12} < 0 \) but \( \partial c_2^* / \partial \sigma_{12} > 0 \). The opposite comparative static could occur if Line 1 swings substantially and Line 2 moves a little, with the

\(^4\)Our approach separates complementarity/substitutability and correlation.
optimal coinsurance rate falling for health care good 1 and possibly for good 2 as the health care goods become more positively correlated.

It is straightforward to show that certain elements of the conventional results for insurance still apply, even if the overall level depends on complementarity in demand or covariance information. It is shown in the appendix that

\[ \frac{\partial c^*_1}{\partial \sigma^2_1} < 0, \quad \frac{\partial c^*_1}{\partial B_1} < 0, \quad \frac{\partial c^*_1}{\partial B_2} > 0. \]

5. **Multiple period model**

Our framework can also address the case of multiple periods with correlated health care demands. We focus here on the case where there are no uncompensated costs of illness, so that \( L_i = 0 \), there are only two periods (indexed by 1 and 2), and there is one health treatment good in each period where \( X_i \) is health care demand in period \( i \). We only allow one cost share, hence \( c_1 = c_2 = c \), and focus on the case where demand is the same in each period except for the health shocks \( \theta_i, X_i = A - Bc + \theta_i \). We focus on the case where the parameters and price structure are constant over time: \( I_1 = I_2 = I, \pi_1 = \pi_2 = \pi, \) and \( \sigma^2_1 = \sigma^2_2 = \sigma^2 \). To allow for the possibility that the health care demands in the two periods are correlated, we assume that the second period health shock is \( \theta_2 = \rho \theta_1 + \epsilon_2 \), where \(-1 \leq \rho \leq 1\)

In a dynamic model, we need to introduce savings, which we assume to be optimally chosen. In our two-period model, net saving is decided in period 1 after \( \theta_1 \) is known and spent entirely in period 2. In general, the optimal level of savings will depend on all of the parameters of the model. Of special interest is that savings will depend on the cost sharing rate \( c \) and the first period health shock \( \theta_1, S^*_1(\theta_1, c) \) with \( \partial S^*_1 / \partial \theta_1 < 0 \). In the appendix, we show that the objective function to be maximized through the choice of \( c \) can be written as follows

---

Note: In a footnote and the Appendix, we allow for more time periods.
\[
EV = E_{\theta_1, \theta_2} \left\{ \left[ V^1 \left( J - K - S_1^*(\theta_1, c) - c\theta_1 \right) - L^1(\theta_1) \right] + \phi E_{\theta_2 | \theta_1} \left\{ \left[ V^2 \left( J - K + (1 + r)S_2^*(\theta_1, c) - c\theta_2 \right) - L^2(\theta_2) \right] \right\} \right\}
\]

where

\[ J = 1 - \pi \]

(13)

\[ \theta_2 = \rho \theta_1 + \epsilon_2 \]

\[ K = (c + L)\mu - \frac{B(c)^2}{2} \]

\[ \pi = (1 + \delta) \left[ (1 - c)(\mu - B(c)) \right] \]

Except for the savings function and the introduction of discounting, \( \phi \), this formulation is very similar in structure to what was used above for the case with one period with multiple states of the world. The solution for the optimal choice of c is derived in the appendix. We make the following three further assumptions in deriving our solution:

- Savings is optimal so that for all \( \theta_i \), \( V^i_1(\ldots) = (1 + r)\phi E_{\theta_2 | \theta_1} \left\{ V^2_i(\ldots) \right\} \) where r is the interest rate.
- The utility function can be approximated using a second order approximation with constant relative risk aversion.
- Consumers can earn a return on savings \( (1 + r) \) that is the inverse of their discount rate \( \phi \), so that \( \phi(1 + r) = 1 \).

In the case of the quadratic utility function that we have used in our analysis, the optimal savings is approximated by \( S_1^*(\theta_1, c) = S_i - s_i c \theta_i \) (shown in appendix), where the expected (ex ante) savings are

\[ S_i = \frac{\phi(1 + r) - 1}{R^i \left[ 1 + \phi(1 + r)^2 \right]} \]

and the optimal (ex post) savings are reduced by the proportion

\[ s_i = \frac{1 - \phi(1 + r)^2}{1 + \phi(1 + r)^2} \]

multiplied by the out-of-pocket health payments and uncompensated costs in time period 1. The term \( s_i \) is the marginal propensity to save. In particular, if \( \phi(1 + r) = 1 \)
(as assumed), then \( S_i^*(\theta_i, c_x) \) is reduced to a simple functional form \( S_i^* = -\frac{1 - \rho}{2 + r} c_x \theta_i \). Under these assumptions, when \( n = 2 \) the optimal cost share is

\[
(14) \ c^* = \frac{B}{B + R^4 \sigma^2 \left[ 1 - \frac{1 - \rho}{1 + \varphi} \right]} = \frac{B}{B + R^4 \sigma^2 \left[ 1 - \frac{(1 - \rho)^2}{(2 + r)(1 + \varphi)} \right]}
\]

This result is very similar to the Base Case equation (9) for the case of a single period and one health care good without any uncompensated losses \( L_x = 0 \). Equation (14) differs from the Base Case by the addition of a new savings-related term, \( \Sigma = \left[ 1 - \frac{(1 - \rho)^2}{(2 + r)(1 + \varphi)} \right] \) in the denominator. This \( \Sigma \) term is a function of the correlation coefficient between period 1 and period 2 health shocks, \( \rho \), the interest rate, \( r \), and the consumer discount rate, \( \varphi \). Note that \( \Sigma \) is non-decreasing in \( \rho \), \( r \), and \( \varphi \), and that \( c_x^* \) is decreasing in \( \Sigma \).

As before, we interpret the optimal cost sharing result for a variety of special cases. First, consider the case where the marginal propensity to save is zero, \( s_i = 0 \) so that savings does not respond to health shocks. In this case, \( \Sigma = 1 \) and the one period model results remain correct even with two periods. The consumer must absorb all health shocks fully in the first period, so there is no difference between the static and dynamic choices of \( c_x^* \).

Second, consider the case where the period 1 and period 2 shocks are perfectly correlated, so that \( \rho = 1 \). Once again \( \Sigma = 1 \) and the one period model results hold. Although savings is possible, there are none because the consumer knows exactly what the shock will be in period 2. There is no diversification across periods in the burden of health shocks. In this limiting case, insurance should be the same as with no savings.

Third, consider the case where health shocks are uncorrelated over time, so that \( \rho = 0 \). The discount rate \( \varphi \) is a number close to one, and it is convenient to consider the case where \( \varphi = 1 \) and \( r = 0 \) so that there is no interest or discounting. The plausible result in this two period

---

6 In the more general case of any \( n > 2 \) periods, where the autocorrelation terms are allowed to have an arbitrary pattern rather than a first order autocorrelation, we show in the appendix (A45) that the optimal saving function still has a closed form.
model is that \( s_1 \) will be close to one half, and half of the burden of a health shock is born in period 1 and half is deferred to period 2. Since the cost of risk goes up with the square of the deviation from certainty, the savings reduces riskiness in the first period to one-quarter of the one period value and, hence, the cost of first period risk (proportional to the variance) is reduced to one quarter of the one period model results. Since this burden is shared between two periods, the net reduction in risk is by one half of the variance. The reason that \( \Sigma \) only declines to 0.75 is that in a two period model there is no opportunity to reduce the burden of shocks in the second period. So while savings can reduce the burden of first period shocks to one quarter of their uncertainty cost, savings cannot reduce the burden of second period shocks.

This result with no discounting or interest rates closely approximates the result with discounting, since the two terms \((2 + r)(1 + \varphi)\) will be approximately 4 if consumers use the market interest rate for discounting.

6. Two-periods with two-goods

So far, we have addressed the cases of multiple treatment goods and multiple periods separately. It is also important to consider the case of multiple treatment goods with more than one period because this allows us to address issues such as the differential coverage of acute and chronic health care. We focus on the case where there are two goods and two periods.

The consumption of two goods in two periods is indexed by \( \{ (X_1^1, X_1^2), (X_2^1, X_2^2) \} \) with subscripts indicating goods, and superscripts time periods. Following the notation in the one-good multiple period case, we assume \( c_{X_i^1} = c_{X_i^2} = c_i, i = 1, 2, I_1 = I_2 = I \) and \( \pi_1 = \pi_2 = \pi \).

To make this more concrete, we consider \( X_1 \) as a medical treatment good for an acute condition and \( X_2 \) as a medical treatment good for a chronic condition. Then, we would expect that the amount of good one in the first period is not correlated with its own future amount, while the amounts of good two in both periods are positively correlated, where \( \theta_2^2 = \rho \theta_2^1 + \varepsilon_2, 0 < \rho \leq 1 \). We assume loading fees and other uncompensated costs are zero: \( \delta = L_{X_1} = L_{X_2} = 0 \). We also assume that \( E(\theta_{X_1^1}) = E(\theta_{X_1^2}) = E(\theta_{X_2^1}) = E(\theta_{X_2^2}) = 0 \) and
The objective function and first order conditions characterizing the optimum are shown in the appendix. The main new result of interest relates to how consumers optimally change savings in response to acute rather than chronic conditions. If discounting exactly equals 1+r, so that \( \phi(1+r) = 1 \). Then, we show in the appendix that the optimal saving function is approximated by

\[
S^*_1(\theta^1_1, \theta^1_2, c_1, c_2) = -\frac{1}{2+r} c_1 \theta^1_1 - \frac{1 - \rho}{2+r} c_1 \theta^1_2,
\]

since \( s_1 = \frac{1}{2+r} \) and \( s_2 = \frac{1 - \rho}{2+r} \). The expected (ex ante) savings \( \bar{S}_1 \) are the same with the two-period one-good case. The marginal propensity to save \( s_2 \) is lower than \( s_1 \) because good 2 is chronic care and patients who receive a large shock for chronic care anticipate a correlated shock the following period and, hence, do not adjust their savings as much as for an acute care shock.

Results for the general case are presented in the appendix. For the special case where demand shocks have zero covariance in each period and the goods are neither complements nor substitutes, so that \( \sigma_{12} = 0 \), \( G_{12} = 0 \), the two-good, two-period case reduces to the one-good two-period case. The optimal cost sharing is approximated by

\[
c^*_1 = \frac{B_1}{B_1 + R^A \sigma^2_1 \left[ 1 - \frac{1}{(2+r)(1+\phi)} \right]},
\]

\[
c^*_2 = \frac{B_2}{B_2 + R^A \sigma^2_2 \left[ 1 - \frac{1 - \rho}{(2+r)(1+\phi)} \right]},
\]

The optimal cost sharing rates are identical to equation (14) with \( \rho = 0 \) for the acute care. So the correlation of spending for chronic care in two periods does not affect the optimal cost sharing of the acute case. Thus, our findings indicate that chronic care should have better...
coverage than acute, all other things equal, because chronic care is positively serially correlated, while acute care is uncorrelated over time.

7. Theoretical Summary

We have extended the theoretical literature on efficiency-based models of optimal insurance to address issues that arise from correlated sources of uncertainty, whether the source of the correlation is due to correlated demands across different health care goods at a point in time, correlated demands over time, or the correlated losses that arise from uncompensated losses that accompany the losses covered by the insurance plan. By using a quadratic indirect utility function and, hence a linear demand specification with zero income effects on the demand for health care treatment, we have been able to derive closed form expressions characterizing optimal cost sharing for health care treatment when there are multiple health care goods or where there are multiple time periods.

Table 1 summarizes the comparative statics findings in this paper for the various parameters considered for health care treatment goods and multiple time periods. In some of the cases that we have considered, we can only sign the effects of a parameter on optimal cost sharing rates for certain parameter values. These cases have the comparative static results in parentheses.

The first five rows of Table 1 reaffirm conventional results found in the previous literature, while the bottom six rows reflect new results that extend the previous literature. It is well established that optimal cost sharing on health care treatment should be higher as demand becomes more elastic, consumers become less risk averse, or the variance of spending decreases. Our findings are consistent with the findings from Besley (1988) and others. Our theoretical findings are also consistent with those of Goldman and Philipson (2007) on complements and substitutes – that all other things equal, cost sharing should be higher for complements than substitutes. Our main new finding is that positively correlated losses across health care goods or over time should lead to more generous coverage (lower cost-sharing) than uncorrelated or negatively correlated losses.
8. **Empirical Relevance**

In this section, we briefly examine the empirical magnitudes of the two innovations of our model: the role of contemporaneous correlations across multiple goods and of serial correlation over time. We use data from the Thomson-Reuters MarketScan database from the 2000-2004 period on a population of non-elderly (age <65) enrollees in employment-based commercial plans. We have restricted the sample to FFS, HMO, PPO and POS plans which covered outpatient pharmacy services in addition to outpatient physician and inpatient services. We included only those individuals who were continuously enrolled for the full five year period This yields a sample of 1,335,448 individuals. Besides size, these data have two major advantages. The first is that the enrollees are followed for several years, allowing us to study correlations by type of health care over time. Second, all of the enrollees had pharmacy coverage, allowing us to contrast pharmacy expenditure patterns with those of both inpatient and outpatient care.

Table 2 summarizes key means, standard deviations and correlations from our five year sample, decomposed into three broad types of services – inpatient (facility, not inpatient physician) services, outpatient services, and pharmacy services.\(^7\) The first two columns in the top half of Table 2 reaffirm that inpatient spending, while not the largest expected cost, is by far the most risky in a one year framework. Table 2 also shows that the serial correlation coefficients for spending for each of the three services differ meaningfully, with pharmaceutical spending having the highest serial correlations and inpatient spending the lowest. The serial correlations also reveal that following a health shock spending returns to normal levels much more slowly than a simple autoregressive, AR(1), pattern would suggest. Chronic conditions obviously explain this pattern. The larger correlations for pharmacy versus outpatient care and for outpatient versus inpatient care suggest a larger correction for pharmacy than outpatient from the usual results for a one period model. The final column summarizes the implications of using multiple years of spending to calculate financial risk by presenting the standard deviation of five year sums of spending rather than one year spending. Whereas inpatient spending has nearly four times as

\(^7\)Some researchers may expect a higher proportion of spending to be on inpatient care. The MEDSTAT commercial claims do not include Medicaid or Medicare enrollees, who have higher hospitalization rates. MEDSTAT claims have 24% of all covered charges for inpatient care in 2004. Our sampling frame of using only people with five consecutive years of insurance coverage has somewhat lower proportion of 22 percent of spending in inpatient services.
much variation as pharmacy using a one year horizon, it is less than twice as variable if a five year horizon is used.

Our theoretical model shows that contemporaneous correlations between multiple health care goods can also be important. Table 3 presents two correlation matrices for spending on inpatient, outpatient, pharmacy and total spending. The top half is for one year (2004) while the bottom half corresponds to correlations of five year sums of each of the three components and total spending. The top correlation matrix shows that one-year inpatient spending is the most closely correlated with total spending and that correlations among other services are relatively modest – the conventional view. Taking into account the autocorrelation effect, we generate the bottom half of the table. It reveals that the pattern for inpatient spending is weaker using five year total spending. Outpatient spending becomes the category most closely correlated with five year total spending, and pharmacy spending has a significantly higher correlation with total spending when compared to the one-period case. Considering the contemporaneous correlation and the autocorrelation, our multi-period model suggests better coverage for outpatient and pharmacy spending than suggested by conventional rules for insurance coverage.

Needless to say, the MarketScan data do not have information on the range of uncompensated losses. Thus we are unable to comment on the magnitude of the correction for correlated uncompensated losses. Moreover, in the absence of estimates of the underlying demand elasticities for these three services, or, even more challenging, estimates of the degree of complementarity among them, it is difficult to determine how different the coinsurance rates under our rules would be from those based on Besley’s formulation or older approaches. But the direction is clear. Because of their higher correlation over time, pharmaceuticals should have a lower coinsurance rate than would occur under the traditional rules for one-period models. Spending on inpatient services, which is less correlated than outpatient and pharmaceuticals over time, would change the least.

9. Discussion

The new results that we find most interesting are those that (1) focus on the roles of uncompensated losses differentially over health care goods and time periods, and (2) those that address the role of correlations across goods and time. As in our earlier work (Ellis and Manning, 2007), uncompensated health losses that are related to insured services should
influence the level of cost sharing for correlated health care goods. These losses provide a rationale for both reducing out-of-pocket costs for those goods which tend to have larger uncompensated losses, such as time lost due to hospitalization and recovery, time lost by going for a physician visit, or copayments. The intuition is clear. If consumers face uncertain income losses which are correlated with health care spending shocks on certain treatment goods, then overinsuring those treatment goods is a second best response to reduce this combined risk from the compensated and uncompensated elements. In the tradeoff of moral hazard against the risk premium, the key term in uncertainty in the demand for health goods in the single period, two-good model is $\left[ \theta_1 (c_1 + L_1) + \theta_2 (c_2 + L_2) \right]$ where the $\theta$’s are uncertain \textit{ex ante}. The risk premium depends on the variance of this whole expression, which in turn depends on the size of the uncompensated losses (the L’s) compared to the out-of-pocket copayments (the c’s). Uncompensated losses increase both the benefits from risk reduction (the variance term from above) and the costs of insurance (the demand / moral hazard term.).

Our finding that optimal treatment cost sharing should be lower for positively correlated treatment goods and goods with positive cross-price effects augment the findings of Besley (1988) about multiple goods as well as the intuition that positively correlated variables have greater variance which needs to be partially offset by lower cost sharing.\footnote{Besley (1988, 329-330) indicates that cross-elasticities and covariances jointly affect which good is more generously covered.} The empirical significance of these results is difficult to assess, since relatively little research has focused on estimating these parameters. They may nonetheless provide guidance on coverage for goods such as brand name drugs, specialty curative goods, or the coverage of serious chronic illnesses, which may be strong substitutes and complements. Our framework also provides a rationale for more generously covering services provided relatively more to families (e.g., maternity care) rather than singles, on the grounds that these services are positively correlated.

Finally, our multi-period model shows the key role that savings decisions and correlated errors play in setting optimal cost sharing. We are not aware of any papers in the health economics literature that have emphasized this topic, although there is a sizeable literature on how large uncovered health losses can lead to dissaving and bankruptcy. While there is a literature on how consumers respond to health spending shocks, the implications for optimal health insurance design deserve reexamination. Expensive, chronic conditions, which exhibit
strong positive serial correlations over time for certain health care services, provide an economic rationale for more generous insurance coverage because consumers cannot use intertemporal savings to reduce the burdens of such spending. Thus, in a world where chronic and acute care look to be otherwise equivalent in terms of price responses and variability, the stronger correlation in health care spending for the chronically ill would lead to better coverage than the standard one period model would suggest. Since so much of pharmaceutical use exhibits the same property, this line of argument would also lead to better pharmaceutical coverage.

It is worth highlighting the limitations of our study. As is common in this literature, we develop all of our models using a linear demand structure, with additive errors that have constant variance, and our demand for health care treatment is perfectly income inelastic.\footnote{The empirical literature finds that demand is income inelastic overall, especially in the absence of adverse selection on insurance coverage (Manning et al, 1987; Newhouse et al, 1993). But demand for specific health treatment services may be more highly income elastic and yield different results.} By making these simplifications, we assume away most income effects or corner solutions, which are particularly relevant in any equity discussions of optimal health insurance. In our model, subsidizing health care does not affect relative incomes, although it does affect those with poor health. We recognize that these are relatively restrictive assumptions, although our models remain more general than many others that use only consumer surplus, or assume only two health states or one health care good.

We have repeatedly used a second order approximation of the utility function, which is consistent with approximating the utility function with a constant absolute risk aversion function. We are not especially troubled by this assumption because our results should hold as an approximation for any arbitrary function, as long as the absolute risk aversion parameter does not vary too much across states of the world. Our uncompensated loss function and optimal savings function were also approximated using linear functions, although again our results should hold as an approximation for more general nonlinear functions.

The other restrictive assumption that we have made for tractability’s sake is that the variance in health care expenditure is a constant, conditional on the health state. Specifically we have assumed that the variance and the other higher order moments in healthcare treatment do not depend on the level of cost sharing, that is \( \frac{\partial \sigma_i^2}{\partial c_i} = 0 \), or other observable factors in the demand function. An extension of the current work would allow for the common observation...
that the variability in health care expenditures is an increasing function of the mean or expected
value of expenditures given the covariates in the model.\footnote{When the variance becomes an increasing function of the mean, we pick up an extra term in the cost-of-risk part of the first order condition that did not exist if demand for health care treatment \(X(c_t)\) had constant variance, conditional on health status. As we increase the coinsurance rate, we have the usual increase in the term related to the variance of out-of-pocket expenses. But because increasing coinsurance also decreases the mean, it also decreases the variance. Thus, we have a partially offsetting term to include in the cost of risk. The magnitude of this reduction depends on how price responsive demand is. As long as the demand for health care treatment is inelastic with respect to cost sharing, the qualitative pattern described earlier in this section prevails. See Feldman and Manning (1997) for such an extension to the basic model that we considered in Ellis and Manning (2007) and in this paper, except that it allowed for a constant coefficient of variation property for health care expenditures instead of a constant variance assumption.}

Our models highlight the empirical significance of demand and cost parameters. Of all of our parameters, the cost sharing demand responsiveness of various health care treatment goods has been studied the most. The RAND Health Insurance Experiment and other studies have established that spending on inpatient care is less responsive to cost sharing than spending on outpatient care, which is less responsive than spending on pharmaceuticals. The evidence on demand responsiveness of preventive care is more ambiguous, but it appears to be at least as responsive to insurance coverage as outpatient treatment. Hence, the \textit{ex ante} moral hazard problem that we model seems to be very real, and insurance coverage of preventive care is justified due to the pecuniary externality of cost savings from reduced health premiums.

The variances and means of spending on different types of treatment goods are also well understood. Inpatient spending is much more variable than outpatient spending and drug spending, justifying greater coverage for inpatient care than other health services. These conclusions follow from the previous literature as well as our framework.

Less well studied are the implications of cross price effects, contemporaneous correlations, and serial correlations over time of specific treatment goods. Drug spending and outpatient spending have higher contemporaneous covariances with other types of spending than inpatient care, suggesting they may deserve greater insurance coverage than would otherwise be the case. Cross elasticities of demand for drug and outpatient care are higher than for inpatient care, justifying greater coverage. Some evidence on this is provided in Meyerhoefer and
Zuvekas (2006) who demonstrate that cross price elasticities between non-mental health drugs and spending on treatment for physical health are moderately large and statistically significant.

Relatively little work has explored the intertemporal correlations of specific treatment services. We do know that spending on drugs and outpatient care is much more highly correlated over time than spending on inpatient care. Ellis, Jiang and Kuo (2013) provide recent evidence on this issue in examining more than 30 different medical services defined by type of service, provider specialty and place of service. In our framework, high positive serial correlations justify greater insurance coverage than is implied by a one period model. The evidence from the Thomson-Reuters MarketScan data suggests that these correlations are sizeable and important for both outpatient and for pharmacy, but stronger for pharmacy than outpatient. We suspect that this is because pharmacy spending is more closely related to spending on chronic illnesses.

Perhaps the area most in need of empirical work is to document the magnitude of uncompensated health losses that are correlated with health care spending. Significant uncompensated costs provide a rationale for zero or even negative cost sharing on treatment goods. It would be interesting to know how much the conventional model results must be adjusted given our findings.
REFERENCES


Coate S. Altruism, the Samaritan’s dilemma, and government transfer policy. The American Economic Review 1995; 85; 46-57.


Marquis S.M., Manning W.G. Lifetime costs and compensation for injuries. Inquiry 1999; 36; 244-254.


Table 1. Comparative Statics on Optimal Coinsurance Rate $c_i^*$

<table>
<thead>
<tr>
<th>Effect of</th>
<th>on $c_i^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own demand slope</td>
<td>$</td>
</tr>
<tr>
<td>Other good demand slope</td>
<td>$</td>
</tr>
<tr>
<td>Risk aversion parameter</td>
<td>$R^A$</td>
</tr>
<tr>
<td>Variance of spending on own i</td>
<td>$\sigma^2_i$</td>
</tr>
<tr>
<td>Insurance loading factor</td>
<td>$\delta$</td>
</tr>
<tr>
<td>Uncompensated losses affecting income</td>
<td>$L_i$</td>
</tr>
<tr>
<td>Uncompensated losses affecting utility directly</td>
<td>$L(\theta)$</td>
</tr>
<tr>
<td>Variance of spending on other good j</td>
<td>$\sigma^2_j$</td>
</tr>
<tr>
<td>Covariance of spending on i and j</td>
<td>$\sigma_{ij}$</td>
</tr>
<tr>
<td>Cross price term for other good</td>
<td>$G_{ij}$</td>
</tr>
<tr>
<td>Correlation with next period error</td>
<td>$\rho$</td>
</tr>
</tbody>
</table>

Note: results in parentheses only hold for specific values of key parameters.
Table 2
Means, standard deviations, five year autocorrelation rates of inpatient, outpatient and pharmacy spending

<table>
<thead>
<tr>
<th></th>
<th>2004 Mean</th>
<th>2004 Standard deviation</th>
<th>Autocorrelation with spending in year</th>
<th>Standard deviation of 5 year sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using unadjusted spending</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004 inpatient spending</td>
<td>$871</td>
<td>$8,116</td>
<td>1.000</td>
<td>0.111</td>
</tr>
<tr>
<td>2004 outpatient spending</td>
<td>$2,150</td>
<td>$7,018</td>
<td>1.000</td>
<td>0.460</td>
</tr>
<tr>
<td>2004 pharmacy spending</td>
<td>$977</td>
<td>$2,321</td>
<td>1.000</td>
<td>0.825</td>
</tr>
<tr>
<td>2004 total spending</td>
<td>$3,999</td>
<td>$12,856</td>
<td>1.000</td>
<td>0.418</td>
</tr>
</tbody>
</table>

Note: Results are based on healthcare services for MEDSTAT Marketscan data for individuals aged< 65, who were continuously enrolled during 2000-2004. N= 1,335,448.
### Table 3
Correlations across spending on healthcare services

#### 2004 unadjusted spending

<table>
<thead>
<tr>
<th></th>
<th>Inpatient</th>
<th>Outpatient</th>
<th>Pharmacy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inpatient</td>
<td>1.000</td>
<td>0.278</td>
<td>0.129</td>
<td>0.807</td>
</tr>
<tr>
<td>Outpatient</td>
<td>1.000</td>
<td>0.251</td>
<td></td>
<td>0.767</td>
</tr>
<tr>
<td>Pharmacy</td>
<td></td>
<td>1.000</td>
<td>0.399</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>(symmetric)</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

#### 2000-2004 Five year total spending

<table>
<thead>
<tr>
<th></th>
<th>Inpatient</th>
<th>Outpatient</th>
<th>Pharmacy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inpatient</td>
<td>1.000</td>
<td>0.395</td>
<td>0.206</td>
<td>0.793</td>
</tr>
<tr>
<td>Outpatient</td>
<td>1.000</td>
<td>0.346</td>
<td></td>
<td>0.836</td>
</tr>
<tr>
<td>Pharmacy</td>
<td></td>
<td>1.000</td>
<td>0.529</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>(symmetric)</td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Results are based on healthcare services for MEDSTAT Marketscan data for individuals aged< 65 in 2004, who were continuously enrolled during 2000-2004. N= 1,482,232.
Figure 1a Optimal Coinsurance Rates for Neutral Goods: $G_{12}=0$

Starting at $\sigma_{12} = 0$, as $\sigma_{12} \uparrow$, $c_1$ and/or $c_2$ ↓

Figure 1b Optimal Coinsurance Rates for Complements: $G_{12} < 0$

Starting at $\sigma_{12} = 0$, as $\sigma_{12} \uparrow$, $c_1$ and/or $c_2$ ↓
Figure 1c Optimal Coinsurance Rates for Substitutes: $G_{12} > 0$

Starting at $\sigma_{12} = 0$, as $\sigma_{12} \uparrow$, $c_1$ \& $c_2$ \downarrow

as

\[ c_2 \]

\[ \text{Line 2} \]

\[ \text{Line 1} \]

\[ c_1 \]

Notes for figures 1a-c

\[ c_1 = \frac{B_1 - G_{12}}{B_1 + R^4 \sigma_i^2} + \frac{G_{12} - R^4 \sigma_{12}^2}{B_1 + R^4 \sigma_i^2} c_2 \quad \text{(Line 1)} \]

\[ c_2 = \frac{B_2 - G_{12}}{B_2 + R^4 \sigma_2^2} + \frac{G_{12} - R^4 \sigma_{12}^2}{B_2 + R^4 \sigma_2^2} c_1 \quad \text{(Line 2)} \]

Intercepts on axes for $c_1$ and $c_2$ do not shift because they do not depend on $\sigma_{12}$. 

- 34 -
Appendix

This appendix derives selected analytical results in the main paper. For convenience, Table A-1 presents our notation. Equation numbers shown without an A suffix correspond to numbering in the main text.

Table A-1 Notation

$X = \{X_i\}$ = quantity of health care treatment of service $i$

$Y$ = quantity of all other consumption goods

$I$ = consumers total income

$J$ = disposable income after premiums and prevention spending

$\pi$ = premium paid by consumer

$P_i, P_y$ = demand prices of $X_i$ and $Y$

$c_i$ = cost sharing rates on treatment $X_i$ (share paid by consumer)

$\theta = \{\theta_i\}$ = random shocks affecting health and demand for $X_i$

$\mu_i$ = mean of health care spending on $X_i$

$B_i$ = slope of demand curves when written in the form $X_i = \mu_i - B_i c_i + \theta_i$

$G_{ij}$ = cross price effect of $c_j$ on $X_i$ and also $c_i$ on $X_j$

$\delta$ = insurance loading factor

$\phi$ = discount rate used by consumer

$r$ = interest rate received on savings by consumer

$S_i$ = savings in period 1

$\sigma_i^2$ = variance of $\theta_i$ and also variance of health care spending $X_i$

$\sigma_{ij}$ = covariance of $\theta_i$ and $\theta_j$

$V$ = the consumer’s utility function

$R^i$ = absolute risk aversion constant $= -V''/V_i$

$L_i$ = per unit uncompensated costs of treatment $X_i$

$L_i(\theta_i)$ = uncompensated health losses that reduce effective income from random shocks $\theta_i$
**A-2. Optimal cost sharing rates for health care treatment**

Assume there is no preventive good, and that treatment goods $X_1$ and $X_2$ have linear demand curves of the form

\[
\begin{align*}
X_1 &= A_1 - B_1 P_1 / P_y + G_{12} P_2 / P_y \\
X_2 &= A_2 - B_2 P_2 / P_y + G_{12} P_1 / P_y
\end{align*}
\]

These demands are consistent with a risk neutral indirect utility function of

\[
\begin{align*}
V^s &= \frac{I}{P_y} + \frac{B_1 P_1^2}{2P_y^2} + \frac{B_2 P_2^2}{2P_y^2} - A_1 \frac{P_1}{P_y} - A_2 \frac{P_2}{P_y} - G_{12} \frac{P_1 P_2}{P_y^2}
\end{align*}
\]

Using the normalizations $P_y = 1$, $P_1 = c_1 + L_1$, $P_2 = c_2 + L_2$, $A_i = \mu_i + \theta_i$, letting $J = I - \pi$, applying the concave transformation $V^s(\ldots)$, this yields an indirect utility function that can be written as

\[
\begin{align*}
V^s &= V^s \left[ J + \frac{B_1 (c_1 + L_1)^2}{2} + \frac{B_2 (c_2 + L_2)^2}{2} - (\mu_1 + \theta_1)(c_1 + L_1) \\
&\quad - (\mu_2 + \theta_2)(c_2 + L_2) - G_{12}(c_1 + L_1)(c_2 + L_2) \right] - L_1(\theta_1) - L_2(\theta_2)
\end{align*}
\]

Taking partial derivatives with respect to $c_1$ yields

\[
0 = \frac{\partial V^s}{\partial c_1} = E_\theta \left\{ V^s \left[ J - K - \theta_1(c_1 + L_1) - \theta_2(c_2 + L_{X_2}) \right] \right\} \times \left( -\frac{\partial \pi}{\partial c_1} - \mu_1 + B_1(c_1 + L_1) - G_{12}(c_{X_2} + L_{X_2}) - \theta_1 \right)
\]

where

\[
\begin{align*}
J &= I - \pi, \\
K &= (c_1 + L_1)\mu_1 + (c_2 + L_2)\mu_2 - \frac{1}{2} B_1 (c_1 + L_1)^2 - \frac{1}{2} B_2 (c_2 + L_2)^2 + G_{12}(c_1 + L_1)(c_2 + L_2), \\
\pi &= (1 + \delta) \left\{ (1-c_1) \left[ \mu_1 - B_1(c_1 + L_1) + G_{12}(c_2 + L_2) \right] + (1-c_2) \left[ \mu_2 - B_2(c_2 + L_2) + G_{12}(c_1 + L_1) \right] \right\}, \\
\frac{\partial \pi}{\partial c_1} &= (1 + \delta) \left[ -\mu_1 - B_1 + 2B_1 c_1 + B_1 L_1 + G_{12} - 2c_2 G_{12} - G_{12}L_2 \right]
\end{align*}
\]

Taking a first order approximation of $V^s_1$ around J-K, we can write
\[ \frac{\partial V^s}{\partial c_1} \approx E_\theta \left\{ \left[ V^s_i [J - K] - V^s_{ii} [J - K] \left[ \theta_i (c_i + L_i) + \theta_2 (c_2 + L_2) \right] \right] \right\} \]
\[ = E_\theta \left\{ V^s_i [J - K] \left[ -\frac{\partial \pi}{\partial c_1} - \mu_{x_1} + B_i (c_i + L_i) - G_{i2} (c_2 + L_2) - \theta_1 \right] \right\} \]
\[ = E_\theta \left\{ -V^s_{ii} [J - K] \left[ \theta_i (c_i + L_i) + \theta_2 (c_2 + L_2) \right] \right\} \times \left[ -\frac{\partial \pi}{\partial c_1} - \mu_{i} + B_i (c_i + L_i) - G_{i2} (c_2 + L_2) - \theta_1 \right] \]
\[ \text{(A5)} \]

Defining \( R^4 \equiv \frac{-V^s_{ii}}{V^s_i} \cdot \sigma_1^2 \equiv E_\theta (\theta_i)^2, \sigma_2^2 \equiv E_\theta (\theta_2)^2, \sigma_{12} \equiv E_\theta (\theta_i \theta_2) \), using \( E_\theta (\theta) = 0 \), the first order conditions for the maximum, once divided through by the (nonstochastic) \( V^s_i (J - K) \) can be approximated as follows.

\[ 0 = E_\theta \left\{ \left[ -\frac{\partial \pi}{\partial c_1} - \mu_{x_1} + B_i (c_i + L_i) - G_{i2} (c_2 + L_2) - \theta_1 \right] \right\} \]
\[ + R^4 \left[ \theta_i (c_i + L_i) + \theta_2 (c_2 + L_2) \right] \left[ -\frac{\partial \pi}{\partial c_1} - \mu_{i} + B_i (c_i + L_i) - G_{i2} (c_2 + L_2) - \theta_1 \right] \]
\[ = \left[ -\frac{\partial \pi}{\partial c_1} \right] \left[ \mu_i - B_i (c_i + L_i) + G_{i2} (c_2 + L_2) \right] - R^4 \left[ \sigma_1^2 (c_i + L_i) + \sigma_{12} (c_2 + L_2) \right] \]
\[ \text{(A6)} \]

Finally replacing \( \frac{\partial \pi}{\partial c_1} \) and rearranging slightly, we get an expression that is linear in \( c_1 \) and \( c_2 \). By symmetry, we get the corresponding expression for \( c_2 \), which is also presented below.

\[ (1 + \delta) \left[ \left[ \mu_i - B_i + 2B_i c_i + B_i L_i + G_{i2} - 2c_2 G_{i2} - G_{i2} L_2 \right] \right. \]
\[ + \left. \left[ \mu_i - B_i (c_i + L_i) + G_{i2} + R^4 \left[ \sigma_1^2 (c_i + L_i) + \sigma_{12} (c_2 + L_2) \right] \right] = 0 \]
\[ \text{(A7)} \]

\[ (1 + \delta) \left[ \left[ \mu_i - B_i + 2B_i c_2 + B_i L_2 + G_{i2} - 2c_2 G_{i2} - G_{i2} L_1 \right] \right. \]
\[ + \left. \left[ \mu_i - B_i (c_i + L_i) + G_{i2} + R^4 \left[ \sigma_1^2 (c_i + L_i) + \sigma_{12} (c_2 + L_2) \right] \right] = 0 \]
\[ \text{(A8)} \]

In the main text we examine a number of special cases.
A-2.1 One good, no uncompensated losses, no insurance loading costs
\[ \mu_2 = B_2 = G_{12} = L_1 = L_2 = \sigma_2^2 = \sigma_{12} = \delta = 0 \]
(A9) \[ \left[-\mu_i - B_1 + 2B_1c_i\right] + \mu_i - B_1c_i + R^4\sigma_1^2c_i = 0 \]
(A10) \[ c_i^* = \frac{B_1}{B_1 + R^4\sigma_1^2} \]

A-2.2 One good, no uncompensated losses, positive insurance loading costs
If \( \mu_2 = B_2 = G_{12} = L_1 = L_2 = \sigma_2^2 = \sigma_{12} = \delta > 0 \), and \( \delta > 0 \), then the FOC (A7) simplifies to
(A11) \[ -(1+\delta)\left[-\mu_i - B_1 + 2B_1c_i\right] - \mu_i + B_1c_i - R^4\sigma_1^2c_i = 0 \]
(A12) \[ c_i^* = \frac{B_1 + B_1\delta + \delta\mu_i}{B_1 + 2\delta B_1 + R^4\sigma_1^2} \]

A-2.3 One good, no insurance loading factor, uncompensated losses
If \( \mu_2 = B_2 = G_{12} = L_2 = \sigma_2^2 = \sigma_{12} = \delta > 0 \), but \( L_1 > 0 \), then the FOC (A7) simplifies to
(A13) \[ \left[-\mu_i - B_1 + 2B_1c_i\right] + \mu_i - B_1c_i + R^4\sigma_1^2(c_i + L_1) = 0 \]
(A14) \[ c_i^* = \frac{B_1 - R^4\sigma_1^2L_1}{B_1 + R^4\sigma_1^2} \]

A-2.4 Two goods, no insurance loading factor, no uncompensated losses, general demand structure
If \( \delta = L_1 = L_2 = 0 \), then equations (A7) And (A8) can be solved for \( c_i \) as
(A15) \[ c_i^* = \frac{(G_{12} - B_2)(R^4\sigma_{12}^2 - C_{12}) - (G_{12} - B_2)(R^4\sigma_1^2^2 + B_2)}{(R^4\sigma_1^2^2 + B_1)(R^4\sigma_2^2^2 + B_2) - (R^4\sigma_1^2 - G_{12})^2} \]
It is straightforward to show from (A15) that
(A16) \[ \frac{\partial c_i^*}{\partial \sigma_1^2} < 0, \quad \frac{\partial c_i^*}{\partial \sigma_{12}^2} < 0, \quad \frac{\partial c_i^*}{\partial B_1} > 0, \quad \frac{\partial c_i^*}{\partial B_2} < 0, \]
The expressions for \( \frac{\partial c_i^*}{\partial \sigma_{12}} \) and \( \frac{\partial c_i^*}{\partial G_{12}} \) cannot be signed for all possible values of \( G_{12} \) and \( \sigma_{12} \), however these derivatives can be unambiguously signed for the limiting case where \( G_{12} \) and \( \sigma_{12} \) approach zero. In this limiting case the two partial derivatives become
Hence in this limiting case \( \frac{\partial c^*_1}{\partial \sigma_{12}} \) and \( \frac{\partial c^*_1}{\partial G_{12}} \) are both negative. This means that for sufficiently small \( G_{12} \) and \( \sigma_{12} \), as the covariance of the errors between two services increases (becomes more positive), then the optimal coinsurance rate decreases. The expression \( \frac{\partial c^*_1}{\partial G_{12}} \) can also be signed as negative for an arbitrary value of \( \sigma_{12} \). To show the sign of \( \frac{\partial c^*_1}{\partial G_{12}} \) when \( \sigma_{12}=0 \), use:

\[
\frac{\partial c^*_1}{\partial G_{12}} = 1 - \frac{B_2 + R^4 \sigma^2_2}{(R^4 \sigma^2_1 + B_1)(R^4 \sigma^2_2 + B_2) - (G_{12})^2}
\]

So \( c^*_1 = -\frac{\Phi(G_{12})}{\Psi(G_{12})} \), where \( \Phi \) and \( \Psi \) are as implicitly defined above. Now since \( c^*_1 < 1 \) then \( \Phi(G_{12}) > 0 \). Also, \( \Psi' < 0 \), \( \Phi' > 0 \). Hence for any \( G_{12} \), when \( \sigma_{12}=0 \), \( \frac{\partial c^*_1}{\partial G_{12}} < 0 \).

### A-3. Two period model

It is well known that there are close parallels between models with multiple states of the world and models with multiple periods. One key difference is that there is the possibility of correlated outcomes in different periods, either because health shocks are serially correlated or because the two periods are linked by savings. We examine here a two-period model with one health care treatment good in each period, allowing both savings and correlated errors. The demand structure and premiums are assumed to be the same in both periods. We use the direct utility function which is the dual to the indirect utility function used thus far. We also focus on the case where \( \delta = 0 \), \( I_1 = I_2 = I \) and \( L_1 = L_1 = L \), hence \( \pi_1 = \pi_2 = \pi \). We also assume constant variances over time \( \sigma^2_1 = \sigma^2_2 = \sigma^2 \) and \( \varphi(1+r) = 1 \).
\[
\begin{align*}
\text{Max } EV &= E_{\theta_1, \theta_2} \left[ V^1(X_1, Y_1, \theta_1) - L_{\theta_1}(\theta_1) \right. \\
&\quad \left. + \phi E_{\theta_2|\theta_1} \left[ V^2(X_2, Y_2, \theta_2) - L_{\theta_2}(\theta_2) \right] \right]
\end{align*}
\]

where

\[
\begin{align*}
V^1(X_1, Y_1, \theta_1) &= V^S(Y_1 - \frac{(\mu_x + \theta_1 - X_1)^2}{2B}) \\
V^2(X_2, Y_2, \theta_2) &= V^S(Y_2 - \frac{(\mu_x + \theta_2 - X_2)^2}{2B})
\end{align*}
\]

\[\theta_2 =\rho \theta_1 + \epsilon_2\]
\[\phi = \text{consumers discount factor}\]
\[K_1 = K_2 = (c_x + L_x) \mu_x - \frac{B(c_x + L_x)^2}{2}\]
\[\pi = (1 - c_x) \left( \mu_x - B(c_x + L_x) \right)\]

As long as income is sufficient to always buy the optimal amount of \(X_i\), then the same amount of \(X_i\) will be purchased as in the one period case.

Hence we can use:

(A20) \[X_1^* = \mu_x + \theta_1 - B(c_x + L_x)\]
\[X_2^* = \mu_x + \theta_2 - B(c_x + L_x)\]

Savings will equilibrate the expected marginal utility of income in period 2 with the marginal utility in period 1. Hence we have

(A21) \[V_i^1(\theta_i) = (1 + r) \phi E_{\theta_2|\theta_1} \left[ V_i^2(\theta_2) \right] \]

Let the optimal savings function be \(S_i^*(\theta_i, c_x)\) (derived below) and write the problem using indirect utility as
Except for the savings function, this formulation is very similar in structure to that used for multiple states of the world. Differentiating with regard to $c_x$ yields

$$\left( A23 \right) \frac{\partial E V^*}{\partial c_x} = \frac{\partial E V^*}{\partial \theta_1, \theta_2} \left\{ \begin{array}{c} V^1 \left( J - K - S^*_1(\theta_1, c_x) - (c_x + L_x) \theta_1 \right) \\
\times \left( \frac{-\partial S^*_1(\theta_1, c_x)}{\partial c_x} - \frac{\partial \pi}{\partial c_x} - \mu_x + B(c_x + L_x) - \theta_1 \right) \\
+ \phi \frac{\partial E V^*}{\partial \theta_1} \left( \frac{V^2 \left( J - K + (1+r)S^*_2(\theta_2, c_x) - (c_x + L_x) \theta_2 \right)}{(1+r) \frac{\partial S^*_2(\theta_2, c_x)}{\partial c_x} - \frac{\partial \pi}{\partial c_x} - \mu_x + B(c_x + L_x) - \theta_2} \right) \end{array} \right\}$$

By the envelope theorem, the terms involving $\frac{\partial S^*_1(\theta_1, c_x)}{\partial c_x}$ in the above expression will cancel out due to the assumption of optimal savings. Hence we can rewrite this as:

$$\left( A24 \right) \frac{\partial E V^*}{\partial c_x} = \frac{\partial E V^*}{\partial \theta_1} \left\{ \begin{array}{c} V^1 \left( J - K - S^*_1(\theta_1, c_x) - (c_x + L_x) \theta_1 \right) \\
\times \left( -\frac{\partial \pi}{\partial c_x} - \mu_x + B(c_x + L_x) - \theta_1 \right) \end{array} \right\}$$

Except for the fact that two different pieces of the utility function are used, and the appearance of the savings function as an argument of the $V^1_i$ this is identical to the earlier specification. Taking a first order Taylor series approximation of the $V^1_i$ and $V^2_i$ functions, as before, we obtain the similar expansions.
If consumers use the same discount rate as that is implied by their real interest rate on savings, then \( \varphi(1 + r) = 1 \) and the top expression in brackets will be zero. With this assumption we can further simplify

\[
0 = (1 + \varphi) \left( -\mu_x + Bc_x - \frac{\partial \pi}{\partial c_x} + BL_x \right) - R^4(c_x + L_x)\sigma^2(1 + \varphi)
\]

\[
= (1 + \varphi) \left( -\mu_x + Bc_x - \frac{\partial \pi}{\partial c_x} + BL_x \right) - R^4(c_x + L_x)\sigma^2(1 + \varphi)
\]

\[
= (1 + \varphi) \left( -\mu_x + Bc_x - \frac{\partial \pi}{\partial c_x} + BL_x \right) - R^4(c_x + L_x)\sigma^2(1 + \varphi)
\]

\[
(\text{A27})
\]

As a reminder, \( \frac{\partial \pi}{\partial c_x} = -\mu_x + 2Bc_x + BL_x - B \).

\[
(\text{A28})
\]
This expression cannot be simplified further without explicit savings function form $S_t^*(\theta, c_X)$. As is showed later, optimal savings rule can be approximated by the following linear function

(A29) $S_t^*(\theta, c_X) = \bar{S}_t - s_1(c_X + L_X)\theta$

Where $\bar{S}_t = \frac{\varphi(1+r)-1}{R^4[1+\varphi(1+r)^2]}$, $s_1 = \frac{\varphi(1+r)\rho - 1}{1+\varphi(1+r)^2}$.

This implies that there is an average savings level $\bar{S}_t$ but that savings is reduced by proportion $s_1$ for all losses (compensated or uncompensated). Since $\bar{S}_t$ will be uncorrelated with $(\theta_1 - \theta_2)$, it will drop out once expectations are taken and we can write

(A30) $0 = (1+\varphi)B(1-c_X) - R^4\sigma^2(c_X + L_X)[(1+\varphi) - s_1(1-\rho)]$

Rearranging yields the following condition for optimal cost sharing with multiple periods,

(A31) $c_x^* = \frac{B - R^4L_X\sigma^2\left[1 - s_1\frac{1-\rho}{1+\varphi}\right]}{B + R^4\sigma^2\left[1 - s_1\frac{1-\rho}{1+\varphi}\right]}$

If we plug in function form of $s_1 = \frac{1-\rho}{2+r}$, then

(A32) $c_x^* = \frac{B - R^4L_X\sigma^2\left[1 - \frac{(1-\rho)^2}{(2+r)(1+\varphi)}\right]}{B + R^4\sigma^2\left[1 - \frac{(1-\rho)^2}{(2+r)(1+\varphi)}\right]}$

Now we turn to deriving optimal saving function. As our model set up, saving is determined after health shock is revealed in the 1st period. Assuming individuals are not budget constrained, the optimal amount of $X_i$ will be purchased in both periods and will not be affected by the optimal saving decision, however saving does affect the amount of $Y_i$. Due to this feature, we solve our optimal saving function after taking optimal choices of $X_i$ as given. This approach brings us the same solution for the optimal saving function as is in the real decision process involving choosing $X_i$, $Y_i$ and $S_i$ simultaneously in the first period.
\[
\max_{S_t} \quad EV = \left[ V(J - K - S_t - (c_X + L_X)\theta_t) - L_{\theta_t}(\theta_t) \right] \\
+ \varphi E_{\theta_t} \left[ V(J - K + (1 + r)S_t - (c_X + L_X)\theta_t) - L_{\varphi \theta_t}(\theta_t) \right]
\]

where
\[
J = I - \pi \\
\theta_t = \rho \theta_t + \varepsilon_t \\
\varphi = \text{consumers discount factor}
\]
\[
K = (c_X + L_X)\mu_x - \frac{B(c_X + L_X)^2}{2} \\
\pi = (1 - c_X)(\mu_x - B(c_X + L_x))
\]

The optimal saving function \( S_t^* \) satisfies
\[
V_t \left( J - K - S_t^* - (c_X + L_X)\theta_t \right) + \varphi(1 + r)E_{\theta_t|\theta} \left[ V_t \left( J - K + (1 + r)S_t^* - (c_X + L_X)\theta_t \right) \right] = 0
\]

Taking a first order Taylor series approximation of \( V_t \) function at \( J - K \),
\[
V_t \left( J - K \right) \left[ S_t^* + (c_X + L_X)\theta_t \right] V_{\theta_t} \left( J - K \right) = \varphi(1 + r)E_{\theta_t|\theta} \left\{ V_t \left( J - K \right) + \left[ (1 + r)S_t^* - (c_X + L_X)\theta_t \right] V_{\theta_t} \left( J - K \right) \right\}
\]
\[
V_t \left( J - K \right) \left[ S_t^* + (c_X + L_X)\theta_t \right] V_{\theta_t} \left( J - K \right) = \varphi(1 + r)V_t \left( J - K \right) + \varphi(1 + r)^2 S_t^* V_{\theta_t} \left( J - K \right)
\]
\[
- \varphi(c_X + L_X)(1 + r)V_{\theta_t} \left( J - K \right) \right) E_{\theta_t|\theta} \left( \theta_t \right)
\]

As is assumed \( \theta_t = \rho \theta_t + \varepsilon_t, \) we obtain
\[
V_t \left( J - K \right) \left[ S_t^* + (c_X + L_X)\theta_t \right] V_{\theta_t} \left( J - K \right) = \varphi(1 + r)V_t \left( J - K \right) + \varphi(1 + r)^2 S_t^* V_{\theta_t} \left( J - K \right)
\]
\[
- \varphi(1 + r)(c_X + L_X)V_{\theta_t} \left( J - K \right) \right) \rho \theta_t
\]
\[
1 + \left[ S_t^* + (c_X + L_X)\theta_t \right] R^4 = \varphi(1 + r) - \varphi(1 + r)^2 S_t^* R^4 + \varphi(c_X + L_X)(1 + r)R^4 \rho \theta_t
\]

Rearranging the above equation, we solve the optimal saving function as following
\[
S_t^* = \frac{\varphi(1 + r) - 1}{R^4 \left[ 1 + \varphi(1 + r)^2 \right]} - \frac{1 - \varphi(1 + r) \rho}{1 + \varphi(1 + r)^2} (c_X + L_X)\theta_t, \text{ which implies that}
\]
\[
S_t^* = \frac{\varphi(1 + r) - 1}{R^4 \left[ 1 + \varphi(1 + r)^2 \right]}, \quad S_t = \frac{1 - \varphi(1 + r) \rho}{1 + \varphi(1 + r)^2}
\]

Given our assumption \( \varphi(1 + r) = 1 \), \( S_t^* \) can be simplified further as
\[ (A41) \, S^*_i = -\frac{1-\rho}{2+r} (c_i + L_i) \theta_i \text{ and } S^*_i = 0, \quad s_i = \frac{1-\rho}{2+r} \]

**A-4 Optimal saving for model with multiple periods** $ T \geq 2$

Now we allow more general cases. We assume that $E_{\theta_i} (\theta_{i+1}) = \rho, \theta_i$, for any $t$, without making any restrictions on the stochastic process for $\theta_i$.

For the case with three periods it can be shown that,

\[ (A42) \, S^*_i = -\frac{(1+r)(1-\rho_1) + (1-\rho_2)}{1+(1+r)+(1+r)^2} P_x \theta_i . \]

\[ (A43) \, S^*_i = -\frac{(1+r)^2 (1-\rho_1) + (1+r)(1-\rho_2) + (1-\rho_3)}{1+(1+r)+(1+r)^2 + (1+r)^3} P_x \theta_i \text{ for } T=4. \]

And

\[ (A44) \, S^*_i = -\frac{(1+r)^3 (1-\rho_1) + (1+r)^2 (1-\rho_2) + (1+r)(1-\rho_3) + (1-\rho_4)}{1+(1+r)+(1+r)^2 + (1+r)^3 + (1+r)^4} P_x \theta_i \text{ when } T=5 \text{ (same with the number of years of data we have).} \]

By induction, it is easy to see that the optimal saving result for any $T \geq 2$ is

\[ (A45) \, S^*_i = -\frac{\sum_{t=1}^{T-1} (1+r)^{-1} (1-\rho_{T-t+1})}{\sum_{t=1}^{T} (1+r)^{-1}} P_x \theta_i . \]

**A-5 Optimal saving for model with two periods and two goods**

The consumption of two goods in two periods is indexed by subscripts for goods and superscripts for periods $\{ (X^1_i, X^2_i) \}$. Following the notation for the multiple period case, we assume $c_{X^1_1} = c_{X^2_1} = c_i, i=1,2, I_1 = I_2 = I, L_{X^1_1} = L_{X^2_1} = L_{X^1_2} = L_{X^2_2} = L_2$ and

$\pi_1 = \pi_2 = \pi$.

To make the case more insightful, we might consider $X_1$ as an acute medical treatment good and $X_2$ as a chronic medical treatment good. So good one in the first period is not correlated with its
own future while good two in both periods are positively correlated, where
\[ \theta_2^2 = \rho \theta_2^1 + \epsilon_2, \quad 0 < \rho \leq 1. \]

We also assume that \( E(\theta_1^i) = E(\theta_1^j) = E(\theta_2^i) = E(\theta_2^j) = 0 \) and

\[
\text{VAR} \begin{pmatrix}
\theta_1^i \\
\theta_2^j \\
\theta_1^j \\
\theta_2^i
\end{pmatrix} = 
\begin{pmatrix}
\sigma_1^2 & \sigma_{12} & 0 & 0 \\
\sigma_{12} & \sigma_2^2 & 0 & \rho \sigma_2^2 \\
0 & 0 & \sigma_1^2 & \sigma_{12} \\
0 & \rho \sigma_2^2 & \sigma_{12} & \sigma_2^2
\end{pmatrix}
\]

As is assumed that income is sufficient to support the optimal amounts of medical service, the
demands of two treatment goods in period 1 are

\[
\begin{align*}
X_1^1 &= \mu_1 + \theta_1^1 - B_1(c_1 + L_1) + G_{12}(c_2 + L_2) \\
X_1^2 &= \mu_2 + \theta_1^1 - B_2(c_2 + L_2) + G_{12}(c_1 + L_1)
\end{align*}
\]

Similarly, the demands in period 2 are

\[
\begin{align*}
X_2^1 &= \mu_1 + \theta_2^2 - B_1(c_1 + L_1) + G_{12}(c_2 + L_2) \\
X_2^2 &= \mu_2 + \theta_2^2 - B_2(c_2 + L_2) + G_{12}(c_1 + L_1)
\end{align*}
\]

If we let the optimal savings function be \( S^*(\theta_1^i, \theta_2^j, c_1, c_2) \), then we can write the objective
function using the expected indirect utility as
The optimal saving function $S^*_i$ satisfies

$$-(1-c_i)[\mu_i - B_i(c_1 + L_1) + G_{1i}(c_2 + L_2)] + \frac{[\mu_i - B_i(c_1 + L_1) + G_{1i}(c_2 + L_2)]}{H_i(J-K)} = 0 \quad \text{(A49)}$$

Taking a first order Taylor series approximation of $V_i$ function at $J-K$,

$$V_i(J-K) = V_i(J-K) + \left[ (1+r)S^*_i - (c_i + L_1)\theta^i_1 - (c_2 + L_2)\theta^i_2 \right]V_{ii}(J-K) \quad \text{(A50)}$$

$$V_i(J-K) = V_i(J-K) + \left[ (1+r)S^*_i - (c_i + L_1)\theta^i_1 - (c_2 + L_2)\theta^i_2 \right]V_{ii}(J-K) \quad \text{(A51)}$$

Having assumed $\theta^2_2 = \rho \theta^1_2 + \epsilon_2$, $0<\rho \leq 1$, $E_{\theta^1_2, \theta^2_2 | \theta^i_1, \theta^i_2}(\theta^2_2) = \rho \theta^1_2$, we obtain

$$V_i(J-K) = V_i(J-K) + \left[ (1+r)S^*_i - (c_i + L_1)\theta^i_1 - (c_2 + L_2)\theta^i_2 \right]V_{ii}(J-K) \quad \text{(A52)}$$

$$V_i(J-K) = V_i(J-K) + \left[ (1+r)S^*_i - (c_i + L_1)\theta^i_1 - (c_2 + L_2)\theta^i_2 \right]R^4 = \phi(1+r) - \phi(1+r)^2 S^*_i R^4 + \phi(c_2 + L_2)(1+r)^4 \rho \theta^1_2 \quad \text{(A53)}$$
Rearranging the above equation, we solve the optimal saving function as following

\[
S_i^*(\theta_1, \theta_2, c_1, c_2) = \bar{S} - s_i(c_1 + L_1)\theta_1 - s_2(c_2 + L_2)\theta_2^i,
\]

where the savings and overall savings rates are

\[
s_1 = \frac{1}{1 + \rho(1+r)^2}, \quad s_2 = \frac{1 - \rho(1+r)^2}{1 + \rho(1+r)^2}.
\]

The expected (ex ante) savings \(\bar{S}\) are same with the two-period one-good case. The marginal propensity to save \(s_2\) is lower than \(s_1\) because good 2 is chronic care and patients reduce the uncertainty of expected spending of good 2 by treating it better in period 1. In particular, if \(\rho(1+r) = 1\), \(S_i^*(\theta_1, \theta_2, c_1, c_2)\) is reduced to a simple functional form

\[
S_i^* = \frac{1}{2 + r} (c_1 + L_1)\theta_1^i - \frac{1 - \rho}{2 + r} (c_2 + L_2)\theta_2^i.
\]

Differentiating with regard to \(c_i\) yields

\[
(A54) \frac{\partial EV^*}{\partial c_1} = E_{\theta_1, \theta_2, \rho} \left[ V_1 \times \left( -\frac{\partial S_i^*(\theta_1, \theta_2, c_1, c_2)}{\partial c_1} - \frac{\partial \pi}{\partial c_1} \right) \right] + \phi E_{\theta_1, \theta_2, \rho} \left[ V_2 \times \left( (1+r) - \frac{\partial S_i^*(\theta_1, \theta_2, c_1, c_2)}{\partial c_1} - \frac{\partial \pi}{\partial c_1} \right) \right] + \phi \frac{\partial S_i^*(\theta_1, \theta_2, c_1, c_2)}{\partial c_2}
\]

\[
(A55) \frac{\partial EV^*}{\partial c_2} = E_{\theta_1, \theta_2, \rho} \left[ V_1 \times \left( -\frac{\partial S_i^*(\theta_1, \theta_2, c_1, c_2)}{\partial c_2} - \frac{\partial \pi}{\partial c_2} \right) \right] + \phi E_{\theta_1, \theta_2, \rho} \left[ V_2 \times \left( (1+r) - \frac{\partial S_i^*(\theta_1, \theta_2, c_1, c_2)}{\partial c_2} - \frac{\partial \pi}{\partial c_2} \right) \right] + \phi \frac{\partial S_i^*(\theta_1, \theta_2, c_1, c_2)}{\partial c_2}
\]

By the envelope theorem, the terms involving \(\frac{\partial S_i^*(\theta_1, \theta_2, c_1, c_2)}{\partial c_i}\) in the above expression will cancel out due to the assumption of optimal savings.

The optimal cost sharing rates \(c_1^*\) and \(c_2^*\) must satisfy

\[
(A56) (1 + \phi) \left[ B_1 (1 - c_1^*) - G_{12} (1 - c_2^*) \right] - R^A \sigma_2 \left[ (1 + \phi) - s_1 \right] - R^A \sigma_{12} (c_2^* + L_2) \left[ (1 + \phi) - s_2 \right] = 0
\]
(A57)

\[(1 + \varphi)\left[ B_2 \left( 1 - c_2^* \right) - G_{12} (1 - c_1^*) \right] - R^4 \sigma_2^2 \left( c_2^* + L_2 \right) \left[ (1 + \varphi) - s_2 (1 - \rho) \right] - R^4 \sigma_{12} \left( c_1^* + L_1 \right) \left[ (1 + \varphi) - s_1 \right] = 0 \]

If we simplify the problem by assuming that there are no uncompensated cost related to health care (by assuming \( \delta = L_1 = L_2 = 0 \)), the optimal conditions become

(A58) \[ c_1^* = \frac{(1 + \varphi)(B_1 - G_{12})}{(1 + \varphi)(B_1 + R^4 \sigma_1^2) - R^4 \sigma_1^2 s_1} + \frac{(1 + \varphi)(G_{12} - R^4 \sigma_{12}) + R^4 \sigma_{12} s_2}{(1 + \varphi)(B_1 + R^4 \sigma_1^2) - R^4 \sigma_1^2 s_1} c_2^* \]

(A59) \[ c_2^* = \frac{(1 + \varphi)(B_2 - G_{12})}{(1 + \varphi)(B_2 + R^4 \sigma_2^2) - R^4 \sigma_2^2 s_2 (1 - \rho)} + \frac{(1 + \varphi)(G_{12} - R^4 \sigma_{12}) + R^4 \sigma_{12} s_1}{(1 + \varphi)(B_2 + R^4 \sigma_2^2) - R^4 \sigma_2^2 s_2 (1 - \rho)} c_1^* \]