Optimal Health Insurance for Multiple Goods and Time Periods

by

Randall P. Ellis[1], Shenyi Jiang[1], Willard G. Manning[2]

[1] Economics Department
Boston University
270 Bay State Road
Boston, MA 022l5
e-mail: ellisrp@bu.edu, and jiang@bu.edu

[2] Harris School of Public Policy Studies
The University of Chicago
1155 East 60th St., Room 193
Chicago, IL 60637
e-mail: w-manning@uchicago.edu

Abstract

This paper reexamines the efficiency-based arguments for optimal health insurance, extending the classic analysis to consider both multiple treatment goods and multiple time periods. Using a utility-based framework, we reconfirm the conventional tradeoff between risk aversion and moral hazard for insuring treatment goods. Multiple goods and multiple time periods raise issues of complementarity and of correlated losses that affect the choice of optimal insurance. Substitutes and positively correlated demands over goods or time are shown to reduce optimal cost shares on treatment. In a multiperiod model, savings complicate the optimal insurance rules but positively serially correlated errors generally imply improved coverage is desirable. Further, the presence of positively correlated uncompensated costs provide a further rationale for reducing cost sharing on the covered services. Using insurance claims data we examine the empirical relevance of the contemporaneous correlations across goods, and correlations over time using three broad aggregates of health treatment spending: inpatient, outpatient, and pharmaceuticals. Our model provides a rationale for covering pharmaceuticals more fully than is implied by static models, because it is relatively highly correlated over time, especially with high deductibles.

1. Introduction

One of the major themes in health economics since the field began has been the behavior of patients and providers in the presence of health insurance or sickness funds that cover part or all of the cost of health care. The central economic motivation for such arrangements is that risk-averse individuals can reduce their financial risk by pooling the risks through insurance that effectively shifts funds from the (*ex post*) well individual paying premiums to the (*ex post*) sick individual receiving reimbursement for health care services. A theoretical and empirical concern has been the adverse effects of moral hazard that arise from the incentives in such health plans when the marginal cost of an insured service to the consumer/patient at the point of service is less than the social costs of producing it. To the extent that patients respond to lower out-of-pocket prices of health care, health insurance will increase the amount and quality of the care purchased, generating an excess burden from the increased use. The empirical support for the law of demand applying to health care is substantial from the literature on observational studies, natural experiments, and the RAND Health Insurance Experiment (Newhouse, 1981; Newhouse et al., 1993; Zweifel and Manning, 2000).

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 model with uncertainty about health states or uncertainty about levels of health care expenditure [Arrow, 1963, 1965; Besley, 1988; Cutler and Zeckhauser, 2000; Pauly, 1968, 1974; Spence and Zeckhauser, 1971; Zeckhauser, 1970]. A number of authors have derived the optimal insurance structures based on these theories. Several 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.

Our primary interest in this paper has to do with *ex ante* choices about optimal insurance for health in markets where there are two or more health care goods – either two or more contemporaneous health care goods or health care goods in two or more periods. To keep the

model tractible, we do not consider preventive care in this analysis.[3] Our analysis of multiple health care treatment goods reconfirms the findings of Goldman and Philipson (2007) on the importance of complementarity and substitutability of health care services, but adds new insights to the literature by examining the importance of the correlation structure of errors among health care goods and over time. We also examine how savings and serially correlated health shocks affect the optimal insurance calculations.

The important insight here is that, all other things being equal, health care goods which 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 serially positively correlated shocks over time: health care treatment goods that have more positive correlations should have more generous coverage.[4] The basic logic is that if the demand for two goods or over two periods are uncertain, then their ex ante variance is larger if they are positively correlated than if there is no correlation or a negative one.[5] Risk averse individuals will prefer more generous insurance (lower coinsurance rates) to reduce their financial risk than if they ignored the correlation or treated them as independent. One very specific case of this is when some aspects of a health event are covered or compensated whereas others are not; for example consider time costs of care. In that case, the positive correlation between the uncompensated and covered loss leads to a reduction in the optimal coinsurance rate. Thus, uncompensated health losses provide a new rationale for reducing cost sharing for health care treatment goods because of the positive correlation in uncompensated care and insured care for those health events. Anoher case of continuing interest is the treatment of acute versus chronic care. Our findings would

---

[3] Optimal coverage with prevention is examined in Ellis and Manning (2007), which also explores the role of uncompensated losses in a static model with one treatment good. We also do not deal with the types of complementarity that arise in the health capital or the rational addiction models.

[4] This finding runs counter to the common experience that inpatient coverage is more generous than that for most other health services. The more generous coverage of inpatient care is motivated by the high variance of inpatient care as well as by the fact that inpatient care is less responsive to cost sharing than other services (Newhouse et al., 1993).

The new insight of our analysis is that once spending on various types of services are subject to moderate stop-losses, then the higher serial correlation of drug and outpatient spending make them more risky than inpatient spending, and hence deserving of more generous coverage.

[5] As we will show below, the details are more complicated because the optimal coinsurance also depends on the variability in demand, own and cross price effects, and (in the cases of multiple periods) the discount rate.

indicate that chronic care should have better coverage than acute, all other things equal, because such care is positively serially correlated.

The criterion used in this paper is demand-side efficiency for calculating optimal cost sharing. Since optimal insurance is a topic of considerable interest to many researchers and policymakers - both economists and non-economists - much has already been written on this topic. We do not address other rationales for insurance coverage that can be found in the health economics and public health literatures. These other important rationales for insurance and subsidies that we do not address in this paper include: correcting for externalities, such as those that can occur with communicable diseases; altruism or public good arguments for insurance coverage; corrections of informational problems (i.e. uninsured consumers make the wrong decisions); distributional concerns that may underlie some forms of social insurance (such as goals of elimination of poverty, or achieving social solidarity); or insurance so as to foster more complete coordination among health care providers. This last topic, which can be seen as a variant of the informational problems argument, is part of a new and still growing literature on disease management and drug and other therapy management (Duggan, 2005; Lichtenberg, 2001; and Newhouse, 2006). Without denying the relevance of these other arguments for insurance, we reexamine the efficiency-based arguments for insurance, and derive new results which refine our understanding of the value of generous insurance coverage from the consumer's point of view.

2. Literature and theory review

There is a substantial literature on the overall tradeoff between the welfare losses from moral hazard and the welfare gains from insuring against the riskiness of health care expenditures, starting with the seminal work by Arrow(1963, 1971, and 1976) and Zeckhauser (1970); see Cutler and Zeckhauser (2000) for a detailed review of this literature. Much of this work 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. Health care expenditures are chosen conditional on the state of the world that occurs. The common conclusion of this literature is that one should select the optimal 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. Blomqvist (1997) extends the theory to nonlinear insurance schedules.

The consensus of much of this literature is that insurance does not simply create dead-weight losses because of moral hazard. When the value of avoiding or reducing an individual's risk is included, optimal levels of cost sharing involve neither fully insured (zero out-of-pocket cost), nor being uninsured. 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., 2003).

Besley (1988) provides a multi-good extension to this literature. His model implies that goods and services that are more uncertain (variable) or less price elastic (with respect to out-of-pocket costs) should have more generous coverage – lower coinsurance rates or copayments. Our results for treatment are not inconsistent with his findings, but the results are much more complex once we also consider both complementarity and of the correlation among health care goods and overtime.

The paper in the literature that is closest to ours is a recent paper by Goldman and Philipson (2007). They model the case with multiple goods (or technologies) in which the health care treatment goods are substitutes or complements. In the case where the goods are substitutes, the insured good should have a lower level of cost sharing than otherwise under traditional rules, because lowering the cost sharing increases the use of those other goods. For complements, the cost sharing should be higher. This clearly has implications for the design of coverage of health care services where drug compliance (a form of secondary prevention) is concerned; see their discussion of several recent studies in this area. We return to the contrast with our findings below when we discuss the importance of considering covariances in conjunction with complementarity, not just complementarity by itself.

3. Model assumptions

We examine a series of models that involve two health goods within one period and one health good over two time periods. The individual's utility function is defined over health state (or health status) and the consumption of other goods (Y) and health services (X). In the one period model, consumers have income (I) and face prices P_{X_i} and P_Y, where the i in P_{X_i} indicates the ith healthcare good. In the underlying behavioral model, there is a health production function that

transforms health care X into health status. For simplicity, we ignore the possibility of death; death could be included in an expanded set of health states. To simplify the discussion, we also assume that there are no health care expenditures if the consumer or patient is healthy, but expected health care expenditures are positive when sick.[6] We assume that the moments of health care shocks do not depend on the level of cost sharing or income, considering only briefly the case where these variables might also affect the distribution of health shocks, not just consumer choices.

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 treatment. We assume premiums, π, are competitively determined and depend on the copayment rates and the demand structure, but do not vary across individuals. The units of medical care have been normalized so that the market price is 1, which we also assume to be the marginal cost and hence the efficient price. The insurance policy is a pure coinsurance plan. It has no deductible, stop-loss, or limit on the maximum expenditure or level(s) of covered services.

We first study the optimal insurance coverage for health care treatment when there are two health care treatment goods. After developing a general analytical model (with mathematical results in the Appendix), we examine a series of special cases where we do comparative statics. A key attraction of our specification is that we are able to solve for the optimal cost share as a closed form solution, and to recreate the results from the basic model that involves one health care treatment good. We also derive 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 is for a two-period model in which health care shocks in one period persist over time due to chronic conditions. In a multi-period context, if a consumer's savings react to healthcare shocks, then this changes both the cost of risk as well the optimal cost sharing rates. Positively serially correlated shocks imply that healthcare should be more generously covered (lower cost sharing) than when shocks are independent or negatively correlated across periods.

---

[6] This assumption can be relaxed to allow for positive expected health care expenditures in the healthy state. In this case, the expected expenditures on health care would be greater than zero for both the sick and the well states, but the expenditures by the well would be less than those for the sick. We would also need to assume that the demand curve for the well state does not cross that for the sick state in the range of observed out-of-pocket payments. The two demand curves could be parallel or the demand when sick could be less elastic with respect to out-of-pocket payments.

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.

4. Basic model

The sequence of steps of that we use in our full model are the following.

1. The insurer chooses the premium π, coinsurance rates c_X for health care treatment (X).
2. Nature decides on the consumer's state of illness as random health shocks θ affecting the demand for a vector of health care goods X.
3. The consumer chooses quantities X and Y to maximize utility in Period 1.
4. If a two-period model, repeat steps 2 and 3 for Period 2.

The demand for medical care services has been shown by many empirical studies to be very income inelastic for generously insured consumers. For simplicity, we assume that X is perfectly income inelastic. Hence, $\partial X / \partial I = 0$. While this income elasticity assumption is strong and unrealistic, it buys us a great deal of simplicity that enables many closed form solutions for cases with multiple health treatment goods. We make a strong assumption on the income elasticities, but make weaker assumptions about other parameters below. We avoid concern 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 spending on medical care and hence the marginal utility of income, but do not directly affect this marginal utility of income. This is a common theoretical assumption and is also assumed in many empirical studies.

4.a. One period model

Assume there are two classes of goods, one or more medical treatment goods X, one composite other good Y. The marginal benefit function of medical service X_i is assumed to be a linear demand curve that has a constant slope B_i as a function of the price regardless of the realization of the random health shock θ. For simplicity, we normalize the marginal costs of all goods to be one, and express the prices in terms of the share of this marginal cost paid by the

consumer. Because $P_Y = 1$ and $P_{X_i} = 1$, c_{X_i} then becomes the consumer price of the ith health care good X_i.

There are two possible broad states of the world facing the consumer: healthy and sick. Income available for spending on the one non-health good Y is $I - \pi$, where I is the consumer's income, and π is the insurance premium. After these normalizations, the indirect utility for a one-period model is simply

$$V(I,P,\theta) = V(I - \pi) \equiv V(J) \tag{1}$$

where J is the quantity of good Y that can be purchased after paying the insurance premium π. Expressions for the premium π are derived below.

The demand curve for each medical service is linear in price the consumer's health care price c_X, and hence

$$X = A - Bc_X \tag{2}$$

where A and B are positive constants. The quadratic indirect utility function consistent with this demand function is shown in the Appendix. Stochastic health treatment demand is introduced by letting $A = \mu_X + \theta$, where $\theta \sim F(\theta)$, with $E(\theta) = 0$. We assume that the variance of θ is a constant, and specifically 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 μ_X when the out-of-pocket price is zero.

We allow health shocks to directly cause losses in consumer utility independent of the level of medical care. These losses may be of two types. One type of loss, we denote L_X is equivalent to lowering a person's effective income. Medical conditions (such as mental illness or injuries) may reduce a person's productivity on the job or ability to obtain work. While some injuries or disabilities may be eligible for imperfect compensation through insurance programs such as disability insurance, many are not. A second type of loss is those that directly affect utility independent of a person's income, which we denote $L_\theta(\theta)$. In order to introduce risk aversion, we apply a monotonically increasing concave function $V^S(...)$ 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

$$V(I, C_X, \theta) = V^S \left[\begin{array}{l} J + \dfrac{B_1(c_{X_1} + L_{X_1})^2}{2} + \dfrac{B_2(c_{X_2} + L_{X_2})^2}{2} - \left(\mu_{X_1} + \theta_1\right)(c_{X_1} + L_{X_1}) \\ -\left(\mu_{X_2} + \theta_2\right)(c_{X_2} + L_{X_2}) - G_{12}(c_{X_1} + L_{X_1})(c_{X_2} + L_{X_2}) \end{array} \right] - L_{\theta_1}(\theta_1) - L_{\theta_2}(\theta_2)$$

where

$$J = I - \pi,$$

(3)

Using the linear demand equation for X, the insurer's break-even condition for the insurance premium is

$$\pi = (1+\delta) \left\{ \begin{array}{l} (1 - c_{X_1})\left[\mu_{X_1} - B_1(c_{X_1} + L_{X_1}) + G_{12}(c_{X_2} + L_{X_2}) \right] \\ + (1 - c_{X_2})\left[\mu_{X_2} - B_2(c_{X_2} + L_{X_2}) + G_{12}(c_{X_1} + L_{X_1}) \right] \end{array} \right\}$$

(4)

where δ is the administrative loading factor such that insurance costs proportion δ 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 shares 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.

4.a.1 *Optimal coinsurance rate on health care treatment.*

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 and 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. As before, we focus on the case of a system of linear, income inelastic demand curves expressing them in terms of the consumer's cost share, c_{X_i}, normalize marginal costs to be one, and let the consumer's prices be the shares of marginal costs paid by the consumer, c_{X_1} and c_{X_2}.

This yields the following two demand equations.

$$\begin{array}{l} X_1 = A_1 - B_1 P_{X_1} / P_Y \qquad\quad + G_{12} P_{X_2} / P_Y \\ X_2 = A_2 \qquad\quad - B_2 P_{X_2} / P_Y + G_{12} P_{X_1} / P_Y \end{array}$$

(5)

Using Roy's identity and ignoring corner solutions, it is straightforward to derive the risk neutral indirect utility function $\tilde{V}$ consistent with these linear demand functions. We introduce two types of uncompensated health losses, L_{X_i} (per unit uncompensated costs of treatment X_i) and $L_{\theta_i}(\theta_i)$ (health shocks that directly affect consumer utility). We then apply a concave function to the utility arguments (except for $L_{\theta_i}(\theta_i)$) in order to introduce risk aversion. This yields the utility function (3) above for the sick health state.

The optimal cost sharing rates c_{X_i} for health care treatment will maximize the expectation of Equation (3). Taking its partial derivative with respect to c_{X_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_I^S, evaluated around the nonstochastic arguments of the utility function. This solution can be written as

$$0 = \frac{\partial E_\theta V}{\partial c_{X_1}} = E_\theta \left\{ \begin{array}{l} \left[\left[V_I^S [J-K] - V_{II}^S [J-K] \left[\theta_1 \left(c_{X_1} + L_{X_1} \right) + \theta_2 \left(c_{X_2} + L_{X_2} \right) \right] \right] \right] \\ \left(-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1 (c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) - \theta_1 \right) \end{array} \right\}$$

where

$J = I - \pi,$

$K = (c_{X_1} + L_{X_1})\mu_{X_1} + (c_{X_2} + L_{X_2})\mu_{X_2} - B_1(c_{X_1} + L_{X_1})^2 / 2 - B_2(c_{X_2} + L_{X_2})^2 / 2 + G_{12}(c_{X_1} + L_{X_1})(c_{X_2} + L_{X_2}),$

$\pi = (1+\delta)\left\{ (1 - c_{X_1})\left[\mu_{X_1} - B_1(c_{X_1} + L_{X_1}) + G_{12}(c_{X_2} + L_{X_2}) \right] + (1 - c_{X_2})\left[\mu_{X_2} - B_2(c_{X_2} + L_{X_2}) + G_{12}(c_{X_1} + L_{X_1}) \right] \right\},$

$\dfrac{\partial \pi}{\partial c_{X_1}} = (1+\delta)\left[-\mu_{X_1} - B_1 + 2B_1 c_{X_1} + B_1 L_{X_1} + G_{12} - 2c_{X_2}G_{12} - G_{12}L_{X_2} \right]$

$\hspace{12cm}(6)$

Defining $R^A \equiv -V_{II}^S \Big/ V_I^S$, $\sigma_1^2 \equiv E(\theta_1)^2$, $\sigma_2^2 \equiv E(\theta_2)^2$, $\sigma_{12} \equiv E(\theta_1\theta_2)$, and using $E(\theta_i)=0$, we show in the Appendix that this result can be rearranged to solve for the optimal insurance rates c_1 in the full specification shown above.

$(1+\delta)\left[-\mu_{X_1} - B_1 + 2B_1 c_{X_1} + B_1 L_{X_1} + G_{12} - 2c_{X_2}G_{12} - G_{12}L_{X_2} \right]$

$\hspace{1cm} + \mu_{X_1} - B_1(c_{X_1} + L_{X_1}) + G_{12}c_{X_2} + R^A \left[\sigma_1^2 \left(c_{X_1} + L_{X_1} \right) + \sigma_{12} \left(c_{X_2} + L_{X_2} \right) \right] = 0$

$\hspace{12cm}(7)$

using the relationship for $\partial\pi/\partial c_{X_1}$ in Equation (6). We can solve for the optimal coinsurance rate c_{X_1} if simplify (see Equation A7). An alternative way to interpret these first order condition is to recognize that the elements in Equation 8 consists of two terms of interest. The first term is:

$$-(1+\delta)\left[-\mu_{X_1}-B_1+2B_1c_{X_1}+B_1L_{X_1}+G_{12}-2c_{X_2}G_{12}-G_{12}L_{X_2}\right]$$
$$-\mu_{X_1}+B_1(c_{X_1}+L_{X_1})-G_{12}c_{X_2}$$

or more simply $B_1\left(1-c_{X_1}\right)+G_{12}\left(1-c_{X_2}+L_{X_2}\right)$ if $\delta=0$, that is if there is no loading fee, insurance is actuarially fair. This term corresponds to the marginal costs of moral hazard from having a coinsurance rate less than one. The first element $B_1\left(1-c_{X_1}\right)$ corresponds to the term for the own price effect and the associated moral hazard from paying a price c_{X_1} less than the marginal cost of care (which is normalized to one here). This element is zero if either demand is perfectly inelastic ($B_1=0$) with respect to its own price or if the coinsurance rate is one. The second element $+G_{12}\left(1-c_{X_2}+L_{X_2}\right)$ corresponds to the cross-effect from a complement or substitute. It is zero if the other good is neither a complement nor a substitute, its coinsurance rate is one, and the uncompensated loss L_{X_2} is zero.

The second term of interest corresponds to the marginal losses (negative benefits) of increasing the cost sharing and therefore reducing the protection against financial risk that occurs if the coinsurance rate is less than one:

$$R^A\left[\sigma_1^2\left(c_{X_1}+L_{X_1}\right)+\sigma_{12}\left(c_{X_2}+L_{X_2}\right)\right]$$

There are no marginal gains from risk sharing if there is no variance in demand, no correlated second good, and no uncompensated loss for either health care good.

At the optimal coinsurance rate c_{X_i}, the marginal cost gains (reductions from reducing moral hazard) must just equal the marginal losses from reducing risk protection.

In this general case, the presence of an uncompensated loss on a complement or substitute shifts the marginal cost of moral hazard, reducing it if the two goods are complements $G_{12}<0$ and increasing it if they are substitutes $G_{12}>0$. The effect of uncompensated losses and on positively correlated demands increases the potential gains from risk protection, while uncompensated losses for a specific good increases the potential gains from risk protection for that good. The underlying

logic is that a weighted sum of positively correlated outcomes has greater variance (and hence risk) than the case where they are negatively correlated or uncorrelated.

We now continue to interpret the single period first order condition Equation(7) in a series of special cases below.

4.a.2 One health care good, base case.

In the case of a single good in a single period with no uncompensated costs, and no loading fee, these two terms (for moral hazard and risk bearing) simplify to the well known result from the literature on the second best optimal insurance for constant coinsurance rate plans:

$$-B_1\left(1-c_{X_1}\right)+R^A\sigma_1^2 c_{X_1}=0$$

where $B_1 \geq 0$, and the first term (the marginal costs due to moral hazard) are increasing in the price response $|B_1|$. The gains from risk pooling are increasing in the variance in health care demand. Solving for the optimal coinsurance rate yields Equation 9, where the optimal coinsurance rate c_{X_1} is increasing in the price response B_1 and decreasing in the underlying variance σ_1^2 in demand.

$$c_{X_1}^* = \frac{B_1}{B_1 + R^A\sigma_1^2} \qquad (8)$$

If the demand is perfectly inelastic ($B_1 = 0$), then the optimal coinsurance rate is zero. If there is no variance or the consumer is risk neutral ($R^A = 0$), then the coinsurance rate should be 1. The optimal coinsurance rate falls between 0 and 1, inclusive, $0 \leq c_{X_1}^* \leq 1$.

4.a.3 One health care good, with insurance loading

Starting with this base case, we next relax the assumption of no insurance loading factors. The new expression for the optimal coinsurance rate becomes

$$c_{X_1}^* = \frac{B_1 + B_1\delta + \delta\mu_{X_1}}{B_1 + 2\delta B_1 + R^A\sigma_1^2} \qquad (9)$$

As long as the insurance loading factor δ is not prohibitively large, then $0 < c_{X_1}^* < 1$, and $\frac{\partial c_{X_1}^*}{\partial \delta} > 0.$ Moreover, the mean expected level of spending (with a marginal out-of-pocket price of

zero), μ_{X_1}, enters in the numerator such that as costs go up, it is desirable to increase the cost share to reduce the inefficiency due to the insurance loading factor ($\delta > 0$).

4.a.4 Adding uncompensated health-related losses

Incorporating uncompensated health loss also affects optimal cost sharing for covered treatment goods, increasing the coverage (decreasing the cost-sharing c_{X_1}) desired that we found in our analysis of prevention versus treatment (Ellis and Manning, 2007).

Using our general model, we assume for simplicity that there is one good, and that $\delta = 0$ so that there is no insurance loading factor. But we now allow $L_{X_1} > 0$. This uncompensated loss might reflect lost worker productivity (e.g., sick days without pay), the value of time spent receiving treatment, or uncompensated health care spending (e.g., over-the-counter medicine or home care). In this case, the first order condition simplifies to:

$$-B_1\left(1-c_{X_1}\right)+R^A\sigma_1^2\left(c_{X_1}+L_{X_1}\right)=0$$

The first term is the same as the moral hazard term for the case with no uncompensated loss in Section 4.a.1. The presence of the extra uncompensated loss L_{X_1} does not change the marginal loss from moral hazard because the demand curve is linear. The loss L_{X_1} does increase the losses from risk bearing from $R^A\sigma_1^2 c_{X_1}$ to $R^A\sigma_1^2\left(c_{X_1}+L_{X_1}\right)$. This shift in the demand for risk protection without any offsetting shift in the marginal costs of moral hazard leads to an unambiguous shift to a lower optimal coinsurance rate in Equation 11:

$$c_{X_1}^* =\frac{B_1-R^A\sigma_1^2 L_{X_1}}{B_1+R^A\sigma_1^2}<\frac{B_1}{B_1+R^A\sigma_1^2} \tag{10}$$

The optimal coinsurance rate c_{X_1} is increasing in the price response B_1 and decreasing in the underlying variance σ_1^2 in demand but decreasing in the size of the uncompensated loss L_{X_1}. The denominator in Equation (10) is the same as that in Equation (9). The only difference is the shift in the numerator, which is unambiguously to lower the optimal coinsurance rate because of the increased variance in the costs.

Uncompensated health-related losses that affect the marginal utility of income reduce the optimal coinsurance rate in a straightforward way, by increasing the desirability of transferring more income into less healthy states. If insurance against these losses is incomplete, reducing cost

sharing for treatment expenses is a second best solution, and cost sharing rates should be kept low. Note that only uncompensated losses of type 1, L_{X_1}, which influence the marginal utility of income, affect optimal treatment cost sharing. This is in contrast with optimal cost sharing on preventive care, which is also influenced by uncompensated losses that directly affect utility but not income, $L_{\theta_1}(\theta_1)$.

If there are uncompensated losses, it is possible for c_X^* to be negative or to reach a corner solution where $c_X^* = 0$ for either large L_X or small B_1. Equation (10) provides an efficiency-based rationale for why full insurance can be second best optimal: the absence of complete insurance markets to transfer income into particular ill 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 is rare and productivity losses from ill health are often large. Wagstaff (2007) provides recent 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.[7]

4.a.5 Multiple health care treatment goods

An important motivation for the modeling two rather than one health care treatment good was to be able to examine the role of cross price elasticities and correlated health demands. For ease of exposition, we now assume no uncompensated health losses and no insurance loading factor, $\delta = 0.$, but explore the general case for demand parameters of the two goods. In contrast to the situation with two goods in a one period model but no uncompensated losses, there are now shifts in both the marginal costs from the dead weight loss of c < 1, and from shifts in the risk

---

[7] Although it does not follow in the case of our approximation with a constant absolute risk aversion utility function, for some utility functions even uncorrelated financial risks that do not affect health may affect optimal cost sharing. For instance, if an individual faces an uncertain income or price of food, then it may be optimal to offer more generous health treatment insurance to reduce overall financial risk. We do not explore this issue further here.

bearing. If $\delta = L_{X_1} = L_{X_2} = 0$, then the relevant first order term (Equation A7) for the first health care good simplifies to:

$$\left[-B_1\left(1-c_{X_1}\right)+G_{12}\left(1-c_{X_2}\right)\right]+R^A\left[\sigma_1^2 c_{X_1}+\sigma_{12}c_{X_2}\right]=0 \tag{11}$$

with a similar one the second health care good in Equation A.8. The first term in the first square bracket is the marginal cost of moral hazard for the first good, while the second term indicates the shift in that cost if the two goods are complements or substitutes. The second square bracket is the gain from risk protection which is increasing in the own variance and if the other good is positively correlated. Thus the two major components shift in the presence of a second health care good. If the second good is a complement, then the costs of moral hazard are reduced, If the second good is positively correlated with the first, then there are greater gains from reducing risk by reducing the coinsurance rates.

The final result in terms of the optimal coinsurance rate is ambiguous because it depends on whether the two goods are complements or substitutes and whether their demands are positively or negatively correlated. The expressions for $\dfrac{\partial c_{X_1}^*}{\partial \sigma_{12}}$ and $\dfrac{\partial c_{X_1}^*}{\partial G_{12}}$ cannot be signed for all possible values of G_{ij} and σ_{ij}, but this derivative can be unambiguously signed for the limiting case where G_{ij} and σ_{ij} approach zero. In this case, the two partial derivatives $\dfrac{\partial c_{X_1}^*}{\partial \sigma_{12}}$ and $\dfrac{\partial c_{X_1}^*}{\partial G_{12}}$ are both negative. So as the covariance of the errors between two services increases (becomes more positive), then the optimal coinsurance rate decreases. Also, when two health services become stronger gross substitutes in the sense that $G_{12} = \partial X_1/\partial c_{X_2}$ is increased, then both services should have lower cost shares. This implies that goods that are complements ($G_{12} < 0$) should have higher cost sharing relative to the case in which cross price elasticities of demand for each service are zero.

An alternative way to see the interplay of complementarity and correlation is to consider the solutions to the two first order conditions jointly. If we assume that $\delta = L_{X_1} = L_{X_2} = 0$, then the optimal coinsurance rates must satisfy

$$\left[-B_1(1-c_{X_1})+G_{12}(1-c_{X_2})\right]+R^A\left[\sigma_1^2 c_{X_1}+\sigma_{12}c_{X_2}\right]=0 \tag{12}$$

$$\left[-B_2(1-c_{X_2})+G_{12}(1-c_{X_1})\right]+R^A\left[\sigma_2^2 c_{X_2}+\sigma_{12}c_{X_1}\right]=0 \tag{13}$$

We can rearrange these into

$$c_{X_1} = \frac{B_1 - G_{12}}{B_1 + R^A \sigma_1^2} + \frac{G_{12} - R^A \sigma_{12}}{B_1 + R^A \sigma_1^2} c_{X_2} \qquad (14)$$

$$c_{X_2} = \frac{B_2 - G_{12}}{B_2 + R^A \sigma_2^2} + \frac{G_{12} - R^A \sigma_{12}}{B_2 + R^A \sigma_2^2} c_{X_1} \qquad (15)$$

If the own-price effects are larger than the cross-price effects or if the two goods are complements, then both of the two intercepts are positive. For both goods, the denominator of the intercepts are strictly positive if their demand is not perfectly inelastic, or if there is risk aversion and variance in health care demand. Thus, the intercepts would be negative only if the two health care goods were substitutes and the cross price effect dominated the own price effect, which seems implausible.[8] The slopes depend on $G_{12} - R^A \sigma_{12}$, that is, on whether the two goods are substitutes or complements and on whether their demands are positively or negatively correlated. Figures 1a, 1b and 1c present the three possible cases for G_{12} according to whether goods 1 and 2 are price neutral, substitutes or complements.

Figure 1a corresponds to where $G_{12} = 0$ and the goods are neither complements or substitutes while Figure 1b corresponds to the case where $G_{12} > 0$ and the two medical goods are substitutes. In both cases as the covariance σ_{12} becomes more positive, then Line 2 (equation (15)) pivots clockwise around the origin, while Line 1 (equation (14)) rotates counterclockwise, ensuring that both coinsurance rates fall.

Figure 1c corresponds to the case where $G_{12} < 0$, and the two medical goods are complements. The opposite comparative static occurs, with optimal coinsurance rates rising as the health care goods become more positively correlated.

In general, the optimal coinsurance rate for each of the two health care goods depends on a complex interplay of own and cross-price effects, the degree of risk aversion, and the variances

---

[8] The only plausible example that we could think of was the case where two goods were close substitutes in treating a condition where the demand for treatment was very inelastic.

and covariances in demand. As we show in the Appendix, the optimal cost sharing rate c_1^* can be written as

$$c_{X_1}^* = \frac{(G_{12} - B_2)(R^A\sigma_{12} - G_{12}) - (G_{12} - B_1)(R^A\sigma_2^2 + B_2)}{(R^A\sigma_1^2 + B_1)(R^A\sigma_2^2 + B_2) - (R^A\sigma_{12} - G_{12})^2} \tag{16}$$

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

$$\frac{\partial c_{X_1}^*}{\partial \sigma_{11}} < 0, \ \frac{\partial c_{X_1}^*}{\partial \sigma_{22}} < 0, \ \frac{\partial c_{X_1}^*}{\partial B_1} > 0, \frac{\partial c_{X_1}^*}{\partial B_2} < 0,$$

But the new insight is that introducing a positive covariance leads to lower coinsurance rates. For the case of the first good, the comparative static is:

$$\left.\frac{\partial c_{X_1}^*}{\partial \sigma_{12}}\right|_{\sigma_{12}=0, G_{12}=0} = \frac{-R^A B_2}{(R^A\sigma_1^2 + B_1)(R^A\sigma_2^2 + B_2)} < 0 \tag{17}$$

With a similar result for the second good.

4.b. Multiple periods

Our framework can also address the case of multiple periods with correlated health care demands. We focus here on the case where there are only two periods, (indexed by 1 and 2), and one health treatment good in each period where X_i is health care demand in period i.[9] We only allow one cost share, and hence $c_{X_1} = c_{X_2} = c_X$, and focus on the case where demand is the same in each period except for the health shocks θ_i, $X_i = A - B(c_X + L_{X_i}) + \theta_i$. We focus on the case where the parameters and price structure are constant over time: $I_1 = I_2 = I$, $L_{X_1} = L_{X_2} = L_X$, $\pi_1 = \pi_2 = \pi$, and $\sigma_1^2 = \sigma_2^2 = \sigma^2$. 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 $\theta_2 = \rho\theta_1 + \varepsilon_2$, where $-1 \leq \rho \leq 1$

---

[9] Later in the paper and in the Appendix, we allow for more time periods.

In a dynamic model, we need to introduce savings, which we assume to be optimally chosen. In a two-period model, net saving is decided in period 1 after θ_1 is known, and spent entirely in period 2. In general, the optimal level of savings will depend on the all of the parameters of the model. Of special interest is that savings will depend on the cost share c_X and the first period health shock θ_1, $S_1^*(c_X, \theta_1)$ 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_X can be written as follows

$$
EV^* = E_{\theta_1, \theta_2} \left\{ \begin{array}{l} \left[V^1 \left(J - K - S_1^*(\theta_1, c_X) - (c_X + L_X)\theta_1 \right) - L_\theta(\theta_1) \right] \\ + \varphi\, E_{\theta_2|\theta_1} \left\{ \left[V^2 \left(J - K + (1+r)S_1^*(\theta_1, c_X) - (c_X + L_X)\theta_2 \right) - L_\theta(\theta_2) \right] \right\} \end{array} \right\}
$$

where

$$J = I - \pi$$
$$\theta_2 = \rho\theta_1 + \varepsilon_2 \tag{18}$$
$$K = (c_X + L_X)\mu_X - \frac{B(c_X + L_X)^2}{2}$$
$$\pi = (1+\delta)\left[(1 - c_X)(\mu_X - B(c_X + L_X)) \right]$$

Except for the savings function and the introduction of discounting, φ, 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_X is derived in the Appendix. We make the following three further assumptions in deriving our solution:

- Savings is optimal so that for all θ_1, $V_I^1(...) = (1+r)\varphi\, E_{\theta_2|\theta_1}\left\{ V_I^2(...) \right\}$ 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 to their discount rate φ so that $\varphi(1+r) = 1$.

In the case of the quadratic utility function that we have used in our analysis, the optimal saving is approximated by $S_1^*(\theta_1, c_X) = \overline{S}_1 - s_1(c_X + L_X)\theta_1$ (shown in Appendix), where the expected (ex ante) savings are $\overline{S}_1 = \dfrac{\varphi(1+r) - 1}{R^A\left[1 + \varphi(1+r)^2 \right]}$ and the optimal savings (ex post) are reduced by the

proportion $s_1 = \dfrac{1-\varphi(1+r)\rho}{1+\varphi(1+r)^2}$ multiplied by the out-of-pocket health payments and uncompensated

costs in time period 1. The term s_1 is the marginal propensity to save. In particular, if $\varphi(1+r)=1$

(as assumed), $S_1^*(\theta_1, c_X)$ is reduced to a simple functional form $S_1^* = -\dfrac{1-\rho}{2+r}(c_X + L_X)\theta_1$.[10]   Under

these assumptions, when $n=2$ the optimal cost share

$$c_X^* = \frac{B - R^A L_X \sigma^2 \left[1 - s_1 \dfrac{1-\rho}{1+\varphi}\right]}{B + R^A \sigma^2 \left[1 - s_1 \dfrac{1-\rho}{1+\varphi}\right]} \tag{19}$$

If we plug in the actual function forms, then $c_X^* = \dfrac{B - R^A L_X \sigma^2 \left[1 - \dfrac{(1-\rho)^2}{(2+r)(1+\varphi)}\right]}{B + R^A \sigma^2 \left[1 - \dfrac{(1-\rho)^2}{(2+r)(1+\varphi)}\right]}$

This result is very similar to equation (11) for the case of a single period, one health care good

with an uncompensated loss. Equation 15 differs only by the addition of a new savings-related

term $\Sigma \equiv \left[1 - s_1 \dfrac{(1-\rho)}{1+\varphi}\right]$ in the numerator and denominator. This Σ term is a function of the

marginal propensity to save s_1, the correlation coefficient between period 1 and period 2 health

shocks, ρ, and the consumer discount rate φ. Note that Σ is non-increasing in s_1, and non-

decreasing in ρ and φ, and that c_X^* is decreasing in Σ.

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 $s_1 = 0$, so that savings does not respond to

health shocks. While not optimal, this is true of most government and private pensions. In this

case, $\Sigma=1$, and the one period model results remain correct even with two periods. The consumer

---

[10] 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 that the

optimal saving function still has a closed form $S_1^* = -\dfrac{\displaystyle\sum_{i=1}^{n-1}(1+r)^{i-1}(1-\rho_{n-i+1})}{\displaystyle\sum_{i=1}^{n}(1+r)^{i-1}} P_X \theta_1.$

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 φ is a number close to one, and it is convenient to consider the case where it is exactly one so that there is no discounting. A plausible value for s_1 in a two period model is that s_1 will be close to one half, so that 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 Σ 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. Capturing this would require more periods. Whether c_X^* is higher or lower than implied by the one period model depends on how consumers choose their savings, which is a function of all of the parameters of the model.

Fourth, consider the more plausible case where $0< s_1<1$, $0<\rho<1$, and $\varphi \approx 1$, and $L_X \geq 0$. If we ignore the fact that s_1^* is influenced by all of the parameters of the model, then we can show that the first order effects of an increase in ρ is to lower the optimal coinsurance rate c_X^*. Similarly, we can show that increasing the discount rate φ should decrease optimal cost sharing.

Finally, in the fully general case, the effects of each parameters on c_X^* is ambiguous in general since the optimal savings rule can change as the correlation coefficient and discount factors change.

4.c. 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 solutions for optimal cost sharing on 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 could only sign the effects of a parameter on optimal cost shares for certain parameter values. These cases have the comparative static results in parentheses.

The parameters in the first four rows of Table 1 reaffirm conventional results found in the previous literature, while the terms at the bottom reflect our 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 also are 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 new finding is that positively correlated losses should lead to more generous coverage (lower cost-sharing) than uncorrelated or negatively correlated losses.

5. Empirical Relevance.

In this section we examine the empirical magnitudes of two innovation of our model: the role of contemporaneous correlations across multiple goods and of autocorrelations over time. We use data from the MEDSTAT Marketscan data base 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, which yields a sample of 1,335,448 individuals. Besides it 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, thus allowing us to contrast pharmacy expenditure patterns with those of both inpatient and outpatient care.

Although our analytical model has been developed for the simple case of pure coinsurance with no deductibles or stop-losses, we know that all optimal insurance programs are more complex than this, typically including an upper limit on the out of pocket payments by consumers for their health care (Arrow, 1964, 1971, 1976; Buchanan et al, 1991). Since the point of our analysis is to highlight the importance of correlations of the out of pocket risk facing consumers with multiple treatment goods and multiple periods, we examine these patterns in two ways: using actual (i.e., unadjusted) spending and using spending top-coded at $2000 for each of three broad spending groups.[11] The question we address from the perspective of our model is which is riskier from the consumers point of view – being exposed to the first $2000 of risk for drug spending, $2000 of risk for outpatient spending, or $2000 of spending for inpatient care?

Table 2 summarizes key means, standard deviations and correlations from our five year sample, decomposed into three broad types of services – inpatient facility payments, all outpatient services, and pharmacy services.[12] The top half displays for covered expenses for the level of services consistent with relatively generous commercial insurance, while the bottom half displays topcoded covered charges for a hypothetical simulation in which each of the three broad services are subject to a $2000 deductible. While having such a structure of deductibles would never be part of an optimal insurance program, separate deductibles on pharmacy, inpatient care, or other specific services is not unknown, and it is useful to understanding the risk implied by partially covering specific services. The threshold of $2000 was chosen because it corresponds to typical

---

[11] Our linear approximation of demand, the assumption of a constant absolute risk aversion, and the assumption that consumers are never constrained to be able to purchase less than the optimal amount of medical care also make the most sense when financial losses are limited. By top-coding spending on each type of service at a fixed dollar amount, we make the closed form solutions and approximations of our model more reasonable.

[12] 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, and overall finds 24% of all covered charges are 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.

deductibles in many high deductible or consumer directed health care (CDHC) plans as well as the magnitudes involved in some pharmacy benefit plans. Note that in this simulation we have not included adjustments for the moral hazard response to the high deductibles.

The first two columns in the top half of Table 2 reaffirms that when there is no insurance then inpatient spending, while not the largest expected cost, is by far the most risky in a one period framework. This is less compellingly true when one looks at variation over a five year period, as shown in the final column, which examines the standard deviation of the five year sum of each type of spending. The bottom half of Table 2 reveals that expected costs and their standard deviations are substantially different when consumers are fully insured for any costs about $2000 for each of these three broad services. We see that spending below a deductible – topcoded spending – on inpatient services is substantially less risky than either outpatient or pharmacy spending. Using the one period standard deviation of each service, outpatient spending is the most risky service, while looking over a five year period, pharmacy spending has the highest standard deviation. Also of interest is that being at risk over five years for the first $2000 of pharmacy spending is almost as risky (std. dev. = $2952) as being at risk for the first $2000 of total spending on inpatient, outpatient and pharmacy costs (std. dev. =$3294).

Table 2 also shows the autocorrelation coefficients for unadjusted and topcoded spending for each of the three services and total spending, which helps us better understand why inpatient spending is less risky over a multiyear period. Autocorrelation coefficients for spending over time are consistently lowest for inpatient care, followed by outpatient care, and followed by pharmacy spending. This is even more strongly true using topcoded spending, as shown in the bottom half of the table. The autocorrelations also reveal that spending is much slower to return to normal levels following a health shock than a simple autoregressive pattern would find. Chronic conditions obviously explain this pattern. The larger correlations for pharmacy than outpatient care would suggest a larger correction for pharmacy than outpatient over the one period result.

Our theoretical model shows that contemporaneous correlations between multiple health service goods can be important. Table 3 presents the correlation matrix for spending on inpatient, outpatient, pharmacy and total spending calculated in three ways. The first correlation matrix for uses one year (2004) spending and reveals that inpatient spending is the most closely correlated with total spending, and that correlations among other services are relatively modest – the conventional view. The middle part of Table 3 reveals that if a longer time frame is used – here five years, then unadjusted spending on outpatient services becomes even more highly correlated

than inpatient spending with total health care costs. Pharmacy spending also grows in its correlation with other services. The final part of Table 3 uses the five year sum of the $2000 topcoded spending on each of these three services. These calculations show that topcoded outpatient spending and pharmacy spending are similarly correlated with total spending ($\rho=.885$ versus $\rho=.806$), hence implying similar increases in riskiness due to this correlation in our framework. We are not interpreting our findings as saying that pharmacy costs deserve equal coverage as inpatient or outpatient services, only suggesting that once longer time periods are considered (where correlations matter) and high deductibles are imposed, then the remaining risk that consumers face from pharmacy spending is much higher than is suggested by conventional one-period models that ignore contemporary and serial correlations when calculating welfare losses and choosing optimal insurance rates.

It is interesting to also use the panel dimension of our data to explore how risks are changing over time. Figure 2 displays the per capita spending on the three large groupings – inpatient, outpatient, and pharmacy, both for unadjusted spending and topcoded spending, as explained above. (Note that we use different vertical axis scales in the two figures.) The pattern over time reveals both that pharmacy spending is growing over time, and topcoded pharmacy spending is growing faster than any other category of spending. To the extent that services with larger means are more likely to generate larger losses from moral hazard, the differences would suggest greater coverage of inpatient care. Whether this is a reasonable step toward a second best optimal coinsurance rate would depend on the relative own price elasticities of these three services. To the extent that the evidence shows that inpatient services are less responsive to cost-sharing, inpatient should have a lower coinsurance rate. The evidence on the relative price elasticities of outpatient services and pharmacy is mixed.

Figure 3 repeats the analysis over time of the standard deviation of the three services, both using unadjusted spending and topcoded spending. Following on the earlier literature (especially, Besley, 1988), the larger the standard deviation (and variance), the lower the coinsurance rates should be, all other things equal. Not surprisingly, inpatient care is the most variable of the three services, and pharmacy is the least when looking at unadjusted spending, while outpatient and pharmacy have the highest standard deviations when topcoded. Matching the dramatic growth in drug spending, the standard deviation of drug spending - both unadjusted and using topcodeed spending – is growing the most rapidly. While the static results suggest higher coinsurance rates for pharmacy than for outpatient care, which would have higher coinsurance rates than inpatient

care, the topcoded models suggest that pharmacy spending is more risky than inpatient spending, and becoming increasingly so.

Needless to say, the MEDSTAT data do not have information on the range of uncompensated losses other than deductibles and copayments. 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 the degree of complementarity among them, it is difficult to determine how large the shift in coinsurance rates would be under our rules would be from those based on Besley's formulation or older approaches. But the direction is clear. By considering the correlated responses over time, pharmacy would have a lower coinsurance rate than would occur under a one-period models. Spending on inpatient services, which is less correlated than outpatient and pharmacy over time would receive the least adjustment.

6. 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 with 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, 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 over insuring 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 $\left[\theta_1\left(c_{X_1} + L_{X_1}\right) + \theta_2\left(c_{X_2} + L_{X_2}\right)\right]$ where the θ's are uncertain *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). If one of the losses is much larger than the other, holding the

variance in θ constant, then the corresponding c has to fall to reduce the overall variance and risk premium. The term that gives rise to the moral hazard is the term:[13]

$$(c_{X_1} + L_{X_1})\mu_{X_1} + (c_{X_2} + L_{X_2})\mu_{X_2} - \tfrac{1}{2}B_1(c_{X_1} + L_{X_1})^2 - \tfrac{1}{2}B_2(c_{X_2} + L_{X_2})^2$$
$$+ G_{12}(c_{X_1} + L_{X_1})(c_{X_2} + L_{X_2}),$$

which involves the sum of the copayments (the c's) and the corresponding uncompensated losses (the L's). Thus the 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 shares should be lower for positively correlated treatment goods and goods with positive cross price effects reaffirm the findings of Besley (1988) on multiple goods as well as the intuition that positively correlated amounts have greater variance which need to be partially offset by lower cost sharing.[14] The empirically significance of these results is difficult to assess, since relatively little research has focused on estimating these two parameters. They may nonetheless provide guidance on coverage for certain goods such as certain brand name drugs, specialty curative goods, or the coverage of serious chronic illnesses, which may have close or not close substitutes and complements. Our framework also provides a rationale for more generously covering services provided relatively more to family (e.g., maternity care) rather than individual contracts, on the grounds that these services are more positively correlated.

Finally, our multiperiod 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 has 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 deserves 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

---

[13] This is the term K from Appendix equation A.4, which is the deterministic part of the indirect utility function for the two health care goods.

[14] Besley (1988, 329-330) indicates that cross-elasticities and covariances jointly affect which good is more generously covered.

equivalent in terms of price responses and variability, the stronger correlation in health care over spending for the chronically ill would lead to better coverage than the standard one period model would suggest. Since so much of pharmacy use exhibits the same property, this line of argument would also lead to better pharmacy coverage.

It is worth highlighting the limitations of our study. We develop all of our models using a very specific demand structure, in which demand is linear in its arguments, have errors that are additive and have constant variance, and the demand for treatment is perfectly income inelastic.[15] In doing so, 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 have used only consumer surplus or assumed only two health care states or one health care good.

We have also repeatedly used a linear approximation of the marginal utility of income 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 is not varying too much across states of the world. Our uncompensated loss function and optimal savings function were also approximated using linear functions, although again, we believe that our results should hold as an approximation for more general nonlinear functions.

The other restrictive assumption that we have made for tractability 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 $\partial \sigma_\theta^2 / \partial c_i = 0$ 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.[16]

_____

[15] The empirical literature finds that demand is income inelastic overall, especially in the absence of adverse selection on insurance coverage (Newhouse et al, 1993). But demand for specific health treatment services may be more highly income elastic and yield different results.

[16] 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

Our models point to the importance of understanding the empirical significance of a number of demand and cost parameters. Of all of our parameters, the demand responsiveness of various health care treatment goods to cost sharing has been perhaps the most carefully studied. The Rand Health Insurance Experiment and others 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 ("drugs"). 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 *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 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 does, suggesting they may deserve greater insurance coverage than would otherwise be the case. Cross elasticities of demand for drug and outpatient care are likely to be much 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.

$X(c_X)$ 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.

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 higher correlated over time than spending on inpatient care. Jiang, Ellis and Kuo (2007) 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 serial correlations justify greater insurance coverage than is implied using a one period model. The evidence from the MEDSTAT Marketscan data suggest that these correlations are sizeable and important for both outpatient and for pharmacy, but stronger for pharmacy than outpatient. We suspect that the cause of this difference is that so much of the use of pharmacy is related to the treatment and management of 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 shares on treatment goods and increased cost shares on prevention in the absence of perfect insurance markets. It would be interesting to know how large are the adjustments needed to the conventional model results.

REFERENCES

Arrow, K.J., 1963, Uncertainty and the welfare economics of medical care, American Economic Review 53, 941-973.

Arrow, K.J., 1971, Essays in the theory of risk-bearing (Markham Pub. Co., Chicago, IL).

Arrow, K.J., 1976, Welfare analysis of changes in health coinsurance rates, in R. Rossett (eds.), The note of health insurance in the health services sector (National Bureau of Economic Research, New York).

Blomqvist A., 1997, Optimal non-linear health insurance. Journal of Health Economics 16, 303-321.

Besley, T.J., 1988. Optimal reimbursement health insurance and the theory of Ramsey taxation. Journal of Health Economics 7, 321-336

Cutler, D.M., Zeckhauser, R.J., 2000. The anatomy of health insurance. In: Culyer, A., Newhouse, P. (Eds.), Handbook of Health Economics. North Holland.

Dardanoni, V., and Wagstaff, A. 1990. Uncertainty and the demand for medical care. Journal of Health Economics 9, 23-38.

Eeckhoudt, Louis, Marchand, Maurice, Pestieau, Pierre and Piaser, Gwenael, Are differential co-payment rates appropriate in the health sector?" (October 2004). CORE Discussion Paper No. 2004/70.

Ehrlich, I. and Becker, G. S. 1972. Market insurance, self-insurance, and self-protection. Journal of Political Economics 80, 623-648.

Ellis, R.P., and Manning, W.G. 2007. Optimal Health Insurance for Prevention and Treatment. Journal of Health Economics, 26(6), 1128-1150.

Feldstein, M. and Friedman, B. 1977. Tax subsidies, the rational demand for health insurance, and the health care crisis. Journal of Public Economics 7, 155-178.

Feldman, R., and B. Dowd, 1991. A New Estimate of the Welfare Loss of Excess Health Insurance. The American Economic Review, Vol. 81(1), pp. 297-301

Feldman, R. and W.G. Manning, 1997. Une Formule Simple du Taux Optimale pour Une Police D-Assurance Maladie. In : S. Jacobzone (ed.) Economie de la Sante: Trajectoires du Futur, Insee Methodes, number 64-65, pp.133-144, French National Ministry of Health, (English Translation: A Simple Formula for the Optimal Coinsurance Rate in a Health Insurance Policy).

Goldman, D., and T.J. Philipson, 2007. Integrated insurance design in the presence of multiple medical technologies. American Economic Review 97(2), 427-32.

Jiang, S., Ellis R.P., and Kuo, T.C, 2008. Does service-level spending show evidence of selection across health plan types? Boston University working paper.

Manning, W.G., and Marquis, M.S.,1996. Health insurance: the trade-off between risk pooling and moral hazard. Journal of Health Economics 15(5), 609-639.

Newhouse, J.P., and the Health Insurance Group 1993. Free-For-All: Health Insurance, Medical Costs, and Health Outcomes: The Results of the Health Insurance Experiment, Cambridge: Harvard University Press.

Newhouse, J.P. 2006. Reconsidering the moral hazard-risk avoidance tradeoff. Journal of Health Economics 25: 1005–1014.

Nyman, J.A., 1999. The economics of moral hazard revisited, Journal of Health Economics 18 811-824.

Pauly, M.V. 1974. Overinsurance and public provision of insurance: the roles of moral hazard and adverse selection. Quarterly Journal of Economics 88: 44-62.

Pratt, J. W., 1962. Risk Aversion in the Small and the Large. Econometrica 32, 122-36.

Zeckhauser, R., 1970. Medical insurance: a case study of the tradeoff between risk spreading and appropriate incentives. Journal of Economic Theory 2, 10–26.

Zweifel, P., and Breyer, F. 1997. Health Economics. New York: Oxford University Press.

Zweifel, P.J., and W.G. Manning, 2000, Consumer Incentives in Health Care, in Newhouse, J.P., A.J. Culyer(eds.), Handbook of Health Economics Vol. 1A, Amsterdam: Elsevier, Ch. 8.

Table 1. Comparative Statics on Optimal Coinsurance Rate $c^*_{X_i}$

Effect of		on $c^*_{X_i}$
Own demand slope	$|B_i|$	+
Risk aversion parameter	R^A	-
Variance of spending on own i	σ_1^2	-
Insurance loading factor	δ	+
Uncompensated losses affecting income	L^1	-
Uncompensated losses affecting utility directly	L^2	0
Variance of spending on other good j	σ_1^2	-
Covariance of spending on i and j	σ_{ij}	(-)
Other good demand slope	$|B_j|$	-
Cross price term for other good	G_{ij}	(-)
Correlation with next period error	ρ	(-)

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

| | | 2004 | 2004 Standard deviation | Autocorrelation with spending in year | | | | | Standard deviation of 5 year sum |
		Mean		2004	2003	2002	2001	2000	
Using unadjusted spending									
2004 inpatient spending	$	871	$ 8,116	1.000	0.111	0.082	0.062	0.055	$ 16,288
2004 outpatient spending	$	2,150	$ 7,018	1.000	0.460	0.332	0.271	0.240	$ 17,702
2004 pharmacy spending	$	977	$ 2,321	1.000	0.825	0.622	0.585	0.528	$ 7,700
2004 total spending	$	3,999	$12,856	1.000	0.418	0.304	0.248	0.223	$ 31,875
Using spending topcoded at $2000									
2004 inpatient spending	$	94	$ 421	1.000	0.132	0.115	0.102	0.090	$ 1,089
2004 outpatient spending	$	844	$ 825	1.000	0.469	0.414	0.374	0.354	$ 2,870
2004 pharmacy spending	$	614	$ 739	1.000	0.872	0.783	0.637	0.586	$ 2,952
2004 total spending	$	1,108	$ 826	1.000	0.651	0.590	0.525	0.494	$ 3,294

Note: Results are based on healthcare services for MEDSTAT Marketscan data age < 65, continuously enrolled cohort 2000-2005, N= 1,335,448.

Table 3   Correlations across spending on healthcare services

3A
2004 unadjusted spending

	Inpatient	Outpatient	Pharmacy	Total
Inpatient	1.000	.278	0.129	0.807
Outpatient		1.000	0.251	0.767
Pharmacy			1.000	0.399
Total	(symmetric)			1.000

3B
Five year sum of unadjusted spending

	Inpatient	Outpatient	Pharmacy	Total
Inpatient	1.000	.395	0.215	0.782
Outpatient		1.000	0.352	0.842
Pharmacy			1.000	0.547
Total	(symmetric)			1.000

3C
Five year sum of $2000 topcoded spending

	Inpatient	Outpatient	Pharmacy	Total
Inpatient	1.000	.372	0.269	0.350
Outpatient		1.000	0.603	0.885
Pharmacy			1.000	0.806
Total	(symmetric)			1.000

Note: Numbers shown in text are correlations among healthcare services for MEDSTAT Marketscan data, age < 65, continuously enrolled cohort. N= 1,335,448
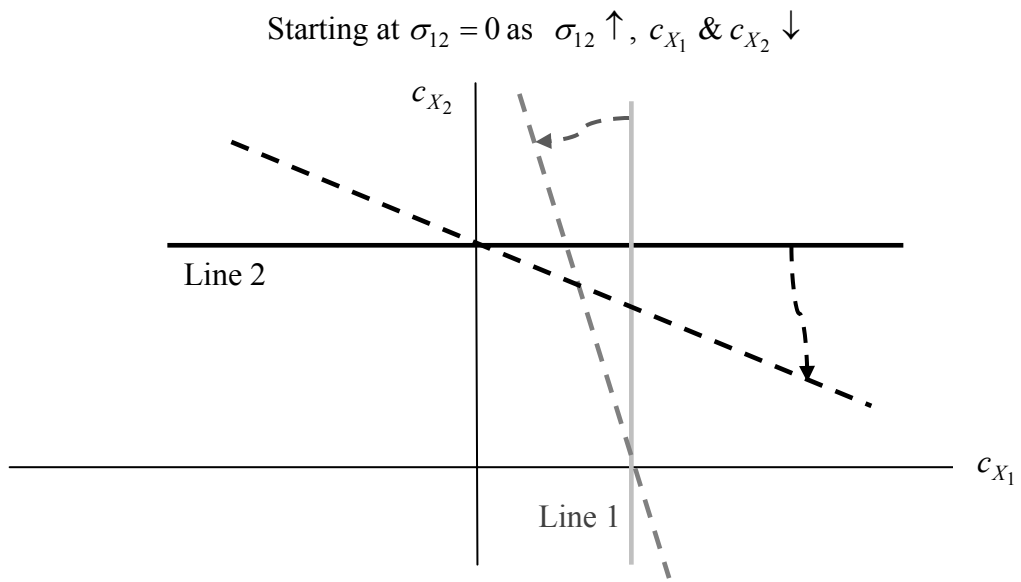
Figure 1a Optimal coinsurance rates for $G_{12} = 0$

Starting at $\sigma_{12} = 0$ as $\sigma_{12} \uparrow$, c_{X_1} & $c_{X_2} \downarrow$

c_{X_2}

Line 2

Line 1

c_{X_1}

Figure 1b Optimal coinsurance rates for $G_{12} < 0$

c_{X_2}

Starting at $\sigma_{12} = 0$ as $\sigma_{12} \uparrow \Rightarrow c_{X_1}^*$ & $c_{X_2}^* \downarrow$

Line 2

Line 1

c_{X_1}
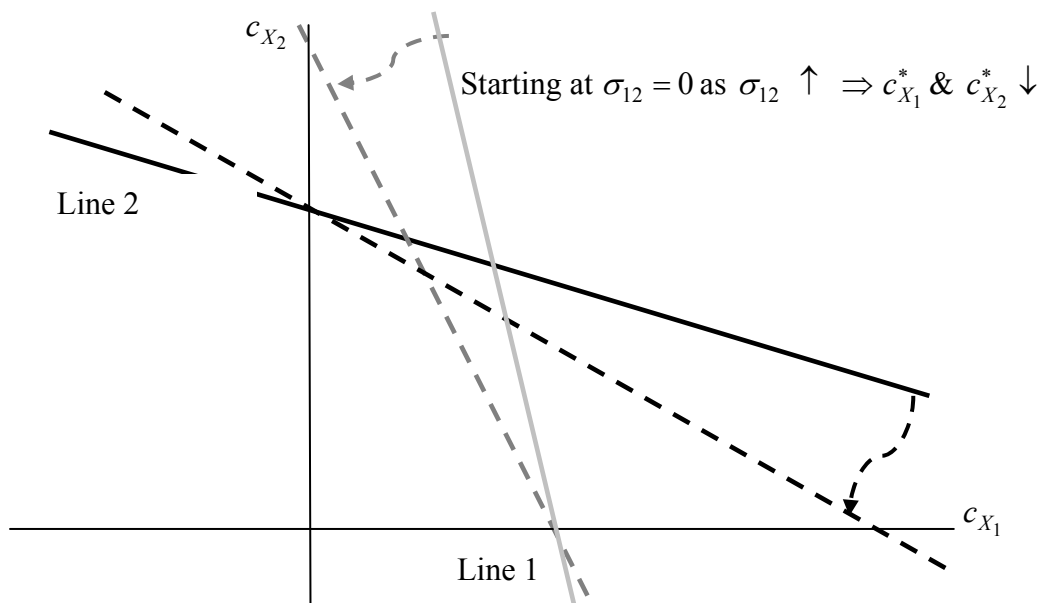
Figure 1c Optimal coinsurance rates for $G_{12} > 0$
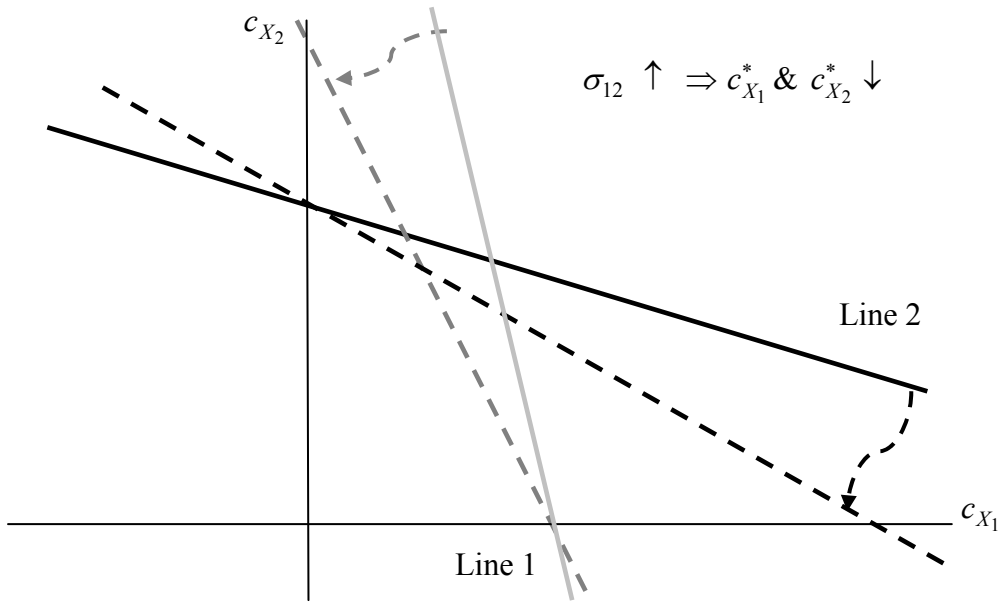
$$\sigma_{12} \uparrow \Rightarrow c^*_{X_1} \& c^*_{X_2} \downarrow$$

Notes for figures 1a-1c

$$c_{X_1} = \frac{B_1 - G_{12}}{B_1 + R^A \sigma_1^2} + \frac{G_{12} - R^A \sigma_{12}}{B_1 + R^A \sigma_1^2} c_{X_2} \ \text{(Line 1)} \quad c_{X_2} = \frac{B_2 - G_{12}}{B_2 + R^A \sigma_2^2} + \frac{G_{12} - R^A \sigma_{12}}{B_2 + R^A \sigma_2^2} c_{X_1} \quad \text{(Line 2)}$$
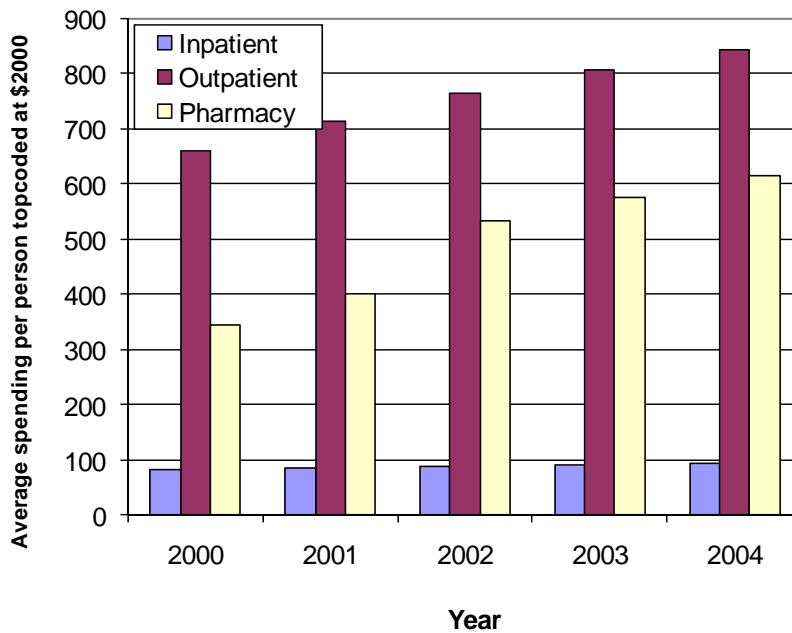
Intercepts on axes for c_{X_1} and c_{X_2} do not shift because they do not depend on σ_{12}.

Figure 2
Mean Per Capita Spending on Health care Services, 2000-2004
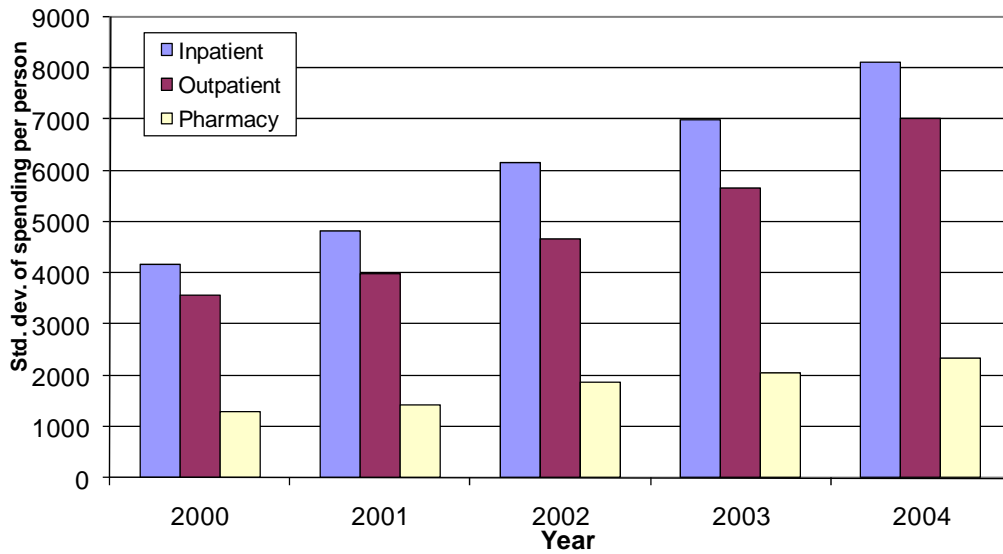
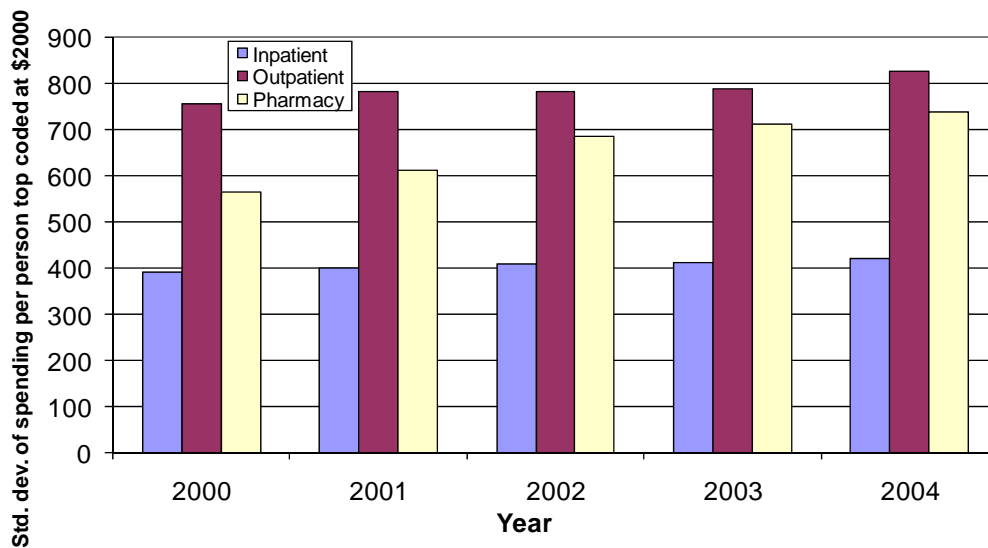2a Unadjusted spending
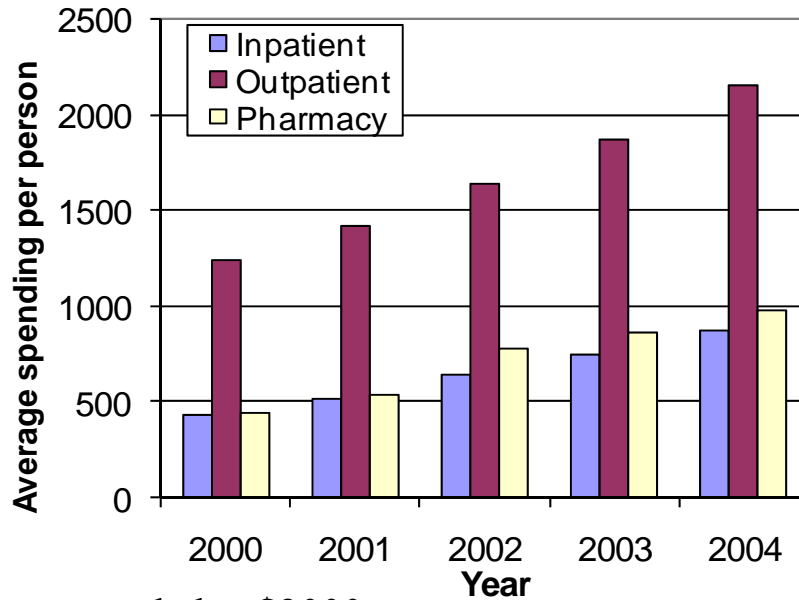
2b Spending topcoded at $2000

Note: Sample is five-year continuously enrolled cohort, aged < 65, from MEDSTAT Marketscan data 2000-2004 (N=1,335,448).

Figure 3
Std. Dev. of Per Capita Spending on Health care Services

3a Unadjusted spending

3b Spending topcoded at $2000

Note: Sample is five-year continuously enrolled cohort, aged < 65, from MEDSTAT Marketscan data 2000-2004 (N=1,335,448)

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

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

π = premium paid by consumer

P_{X_i}, P_Y = demand prices of X_i and Y

c_{X_i} = cost share rates on treatment X_i (share paid by consumer)

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

μ_{X_i}, μ_X = mean of health care spending on X_i or X when care is free

B_i = slope of demand curves when written in the form $X_i = \mu_{X_i} - B_i c_{X_i} + \theta_i$

G_{ij} = cross price effect of c_{X_j} on X_i and also c_{X_i} on X_j

δ = insurance loading factor

φ = discount rate used by consumer

r = interest rate received on savings by consumer

S_1 = savings in period 1

σ_i^2 = variance of θ_i and also variance of health care spending X_i

text. V = the consumer's utility function

R^A = absolute risk aversion constant = $-V_{II}/V_I$

L_{X_i} = per unit uncompensated costs of treatment X_i

$L_{\theta_i}(\theta_i)$ = uncompensated health losses that reduce effective income from random shock θ_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

$$
\begin{aligned}
X_1 &= A_1 - B_1 P_{X_1}/P_Y && + G_{12} P_{X_2}/P_Y \\
X_2 &= A_2 && - B_2 P_{X_2}/P_Y + G_{12} P_{X_1}/P_Y
\end{aligned}
\tag{A.1}
$$

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

$$\tilde{V}^{S} = \frac{I}{P_Y} + \frac{B_1 P_{X_1}^2}{2P_Y^2} + \frac{B_2 P_{X_2}^2}{2P_Y^2} - A_1 \frac{P_{X_1}}{P_Y} - A_2 \frac{P_{X_2}}{P_Y} - G_{12} \frac{P_{X_1} P_{X_2}}{P_Y^2} \qquad (A.2)$$

Using the normalizations $P_Y = 1$, $P_{X_1} = c_{X_1} + L_{X_1}$, $P_{X_2} = c_{X_2} + L_{X_2}$, $A_i = \mu_{X_i} + \theta_i$, letting

$J = I - \pi$, applying the concave transformation $V^S(\ldots)$, this yields an indirect utility function that

can be written as

$$V^{S} = V^{S} \left[\begin{array}{l} J + \dfrac{B_1(c_{X_1} + L_{X_1})^2}{2} + \dfrac{B_2(c_{X_2} + L_{X_2})^2}{2} - \left(\mu_{X_1} + \theta_1\right)(c_{X_1} + L_{X_1}) \\ -\left(\mu_{X_2} + \theta_2\right)(c_{X_2} + L_{X_2}) - G_{12}(c_{X_1} + L_{X_1})(c_{X_2} + L_{X_2}) \end{array} \right] - L_{\theta_1}(\theta_1) - L_{\theta_2}(\theta_2) \qquad (A.3)$$

Taking partial derivatives with respect to c_{X_1} yields

$$0 = \frac{\partial V^{S}}{\partial c_{X_1}} = E_{\theta} \left\{ \begin{array}{l} V_I^S \left[J - K - \theta_1(c_{X_1} + L_{X_1}) - \theta_2(c_{X_2} + L_{X_2}) \right] \\ \times \left(-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1(c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) - \theta_1 \right) \end{array} \right\}$$

where

$J = I - \pi$,

$K = (c_{X_1} + L_{X_1})\mu_{X_1} + (c_{X_2} + L_{X_2})\mu_{X_2} - \tfrac{1}{2}B_1(c_{X_1} + L_{X_1})^2 - \tfrac{1}{2}B_2(c_{X_2} + L_{X_2})^2 + G_{12}(c_{X_1} + L_{X_1})(c_{X_2} + L_{X_2})$,

$$\pi = (1+\delta)\left\{ \begin{array}{l} (1 - c_{X_1})\left[\mu_{X_1} - B_1(c_{X_1} + L_{X_1}) + G_{12}(c_{X_2} + L_{X_2})\right] \\ + (1 - c_{X_2})\left[\mu_{X_2} - B_2(c_{X_2} + L_{X_2}) + G_{12}(c_{X_1} + L_{X_1})\right] \end{array} \right\},$$

$$\frac{\partial \pi}{\partial c_{X_1}} = (1+\delta)\left[-\mu_{X_1} - B_1 + 2B_1 c_{X_1} + B_1 L_{X_1} + G_{12} - 2c_{X_2}G_{12} - G_{12}L_{X_2} \right]$$

$$(A.4)$$

Taking a first order approximation of V_I^S around J-K, we can write

$$\frac{\partial V^S}{\partial c_{X_1}} \approx E_\theta \left\{ \begin{array}{l} \left[\begin{array}{l} \left[V_I^S[J-K] - V_{II}^S[J-K] \left[\theta_1 \left(c_{X_1} + L_{X_1} \right) + \theta_2 \left(c_{X_2} + L_{X_2} \right) \right] \right] \\ \left(-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1(c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) - \theta_1 \right) \end{array} \right] \end{array} \right\}$$

$$= E_\theta \left\{ \begin{array}{l} V_I^S[J-K] \left[-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1(c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) - \theta_1 \right] \\ -V_{II}^S[J-K] \left[\theta_1 \left(c_{X_1} + L_{X_1} \right) + \theta_2 \left(c_{X_2} + L_{X_2} \right) \right] \\ \times \left[-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1(c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) - \theta_1 \right] \end{array} \right\} \qquad \text{(A.5)}$$

Defining $R^A \equiv -V_{II}^S \Big/ V_I^S$, $\sigma_1^2 \equiv E_{\theta_1}(\theta_1)^2$, $\sigma_2^2 \equiv E_{\theta_2}(\theta_2)^2$, $\sigma_{12} \equiv E_\theta(\theta_1\theta_2)$, using $E_\theta(\theta_i) = 0$, the first

order conditions for the maximum, once divided through by the (nonstochastic) $V_I^S(J-K)$ can be

approximated as follows.

$$0 = E_\theta \left\{ \begin{array}{l} \left[-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1(c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) - \theta_1 \right] \\ + R^A \left[\theta_1(c_{X_1} + L_{X_1}) + \theta_2(c_{X_2} + L_{X_2}) \right] \left[-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1(c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) - \theta_1 \right] \end{array} \right\}$$

$$= \left\{ \begin{array}{l} \left[-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1(c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) - E_\theta[\theta_1] \right] \\ + R^A E_\theta \left[\theta_1(c_{X_1} + L_{X_1}) + \theta_2(c_{X_2} + L_{X_2}) \right] \left[-\dfrac{\partial \pi}{\partial c_{X_1}} - \mu_{X_1} + B_1(c_{X_1} + L_{X_1}) - G_{12}(c_{X_2} + L_{X_2}) \right] \\ + R^A E_\theta \left[\theta_1(c_{X_1} + L_{X_1}) + \theta_2(c_{X_2} + L_{X_2}) \right] [-\theta_1] \end{array} \right\}$$

$$= \left[-\frac{\partial \pi}{\partial c_{X_1}} \right] - \left[\mu_{X_1} - B_1(c_{X_1} + L_{X_1}) + G_{12}(c_{X_2} + L_{X_2}) \right] - R^A \left[\sigma_1^2(c_{X_1} + L_{X_1}) + \sigma_{12}(c_{X_2} + L_{X_2}) \right]$$

$$\text{(A.6)}$$

Finally replacing $\dfrac{\partial \pi}{\partial c_{X_1}}$ and rearranging slightly, we get an expression that is linear in

c_{X_1} and . c_{X_2}. By symmetry, we get the corresponding expression for c_{X_2}, which is also presented below.

$$(1+\delta)\left[-\mu_{X_1} - B_1 + 2B_1 c_{X_1} + B_1 L_{X_1} + G_{12} - 2c_{X_2} G_{12} - G_{12} L_{X_2}\right]$$
$$+ \mu_{X_1} - B_1(c_{X_1} + L_{X_1}) + G_{12}c_{X_2} + R^A\left[\sigma_1^2\left(c_{X_1} + L_{X_1}\right) + \sigma_{12}\left(c_{X_2} + L_{X_2}\right)\right] = 0 \tag{A.7}$$

$$(1+\delta)\left[-\mu_{X_2} - B_2 + 2B_2 c_{X_2} + B_2 L_{X_2} + G_{12} - 2c_{X_1} G_{12} - G_{12} L_{X_1}\right]$$
$$+ \mu_{X_2} - B_2(c_{X_2} + L_{X_2}) + G_{12}c_{X_1} + R^A\left[\sigma_2^2\left(c_{X_2} + L_{X_2}\right) + \sigma_{12}\left(c_{X_1} + L_{X_1}\right)\right] = 0 \tag{A.8}$$

In the main text we examine a number of special cases.

A-2.1 One good, no uncompensated losses, no insurance loading costs

$$\mu_{X_2} = B_2 = G_{12} = L_{X_1} = L_{X_2} = \sigma_2^2 = \sigma_{12} = \delta = 0$$

$$\left[-\mu_{X_1} - B_1 + 2B_1 c_{X_1}\right] + \mu_{X_1} - B_1 c_{X_1} + R^A \sigma_1^2 c_{X_1} = 0$$

$$\left[-B_1 + B_1 c_{X_1}\right] + R^A \sigma_1^2 c_{X_1} = 0$$

$$c_{X_1}^* = \frac{B_1}{B_1 + R^A \sigma_1^2} \tag{A.9}$$

A-2.2 One good, no uncompensated losses, positive insurance loading costs

If $\mu_{X_2} = B_2 = G_{12} = L_{X_1} = L_{X_2} = \sigma_2^2 = \sigma_{12} = 0$, and $\delta > 0$, then the FOC

Error! Reference source not found. simplifies to

$$-(1+\delta)\left[-\mu_{X_1} - B_1 + 2B_1 c_{X_1}\right] - \mu_{X_1} + B_1 c_{X_1} - R^A \sigma_1^2 c_{X_1} = 0$$

$$c_{X_1}^* = \frac{B_1 + B_1\delta + \delta\mu_{X_1}}{B_1 + 2\delta B_1 + R^A \sigma_1^2} \tag{A.10}$$

A-2.3 One good, no insurance loading factor, uncompensated losses

If $\mu_{X_2} = B_2 = G_{12} = \sigma_2^2 = \sigma_{12} = \delta = 0$, but $L_{X_1} > 0$, then the FOC

Error! Reference source not found. simplifies to

$$\left[-\mu_{X_1} - B_1 + 2B_1 c_{X_1} \right] + \mu_{X_1} - B_1 c_{X_1} + R^A \sigma_1^2 \left(c_{X_1} + L_{X_1} \right) = 0$$

$$c_{X_1}^* = \frac{B_1 - R^A \sigma_1^2 L_{X_1}}{B_1 + R^A \sigma_1^2} \tag{A.11}$$

A-2-4 Two goods, no insurance loading factor, no uncompensated losses, general demand structure

If $\delta = L_{X_1} = L_{X_2} = 0$, then equations **Error! Reference source not found.** and

Error! Reference source not found. can be solved for c_{X_i} as

$$c_{X_1}^* = \frac{(G_{12} - B_2)(R^A \sigma_{12} - G_{12}) - (G_{12} - B_1)(R^A \sigma_2^2 + B_2)}{(R^A \sigma_1^2 + B_1)(R^A \sigma_2^2 + B_2) - (R^A \sigma_{12} - G_{12})^2} \tag{A.12}$$

It is straightforward to show from **Error! Reference source not found.** that

$$\frac{\partial c_{X_1}^*}{\partial \sigma_{11}} < 0, \ \frac{\partial c_{X_1}^*}{\partial \sigma_{22}} < 0, \ \frac{\partial c_{X_1}^*}{\partial B_1} > 0, \frac{\partial c_{X_1}^*}{\partial B_2} < 0, \tag{A.13}$$

The expressions for $\dfrac{\partial c_{X_1}^*}{\partial \sigma_{12}}$ and $\dfrac{\partial c_{X_1}^*}{\partial G_{12}}$ cannot be signed for all possible values of G_{ij} and σ_{ij}, however

these derivatives can be unambiguously signed for the limiting case where G_{ij} and σ_{ij} approach

zero. In this limiting case the two partial derivatives become

$$\left. \frac{\partial c_{X_1}^*}{\partial \sigma_{12}} \right|_{\sigma_{12}=0, G_{12}=0} = \frac{-R^A B_2}{(R^A \sigma_1^2 + B_1)(R^A \sigma_2^2 + B_2)} < 0 \tag{A.14}$$

$$\left. \frac{\partial c_{X_1}^*}{\partial G_{12}} \right|_{\sigma_{12}=0, G_{12}=0} = \frac{-R^A \sigma_2^2}{(R^A \sigma_1^2 + B_1)(R^A \sigma_2^2 + B_2)} < 0 \tag{A.15}$$

Hence in this limiting case $\dfrac{\partial c^{*}_{X_1}}{\partial \sigma_{12}}$ and $\dfrac{\partial c^{*}_{X_1}}{\partial G_{12}}$ are both negative. This means that for sufficiently small

G_{ij} and σ_{ij}, as the covariance of the errors between two services increases (becomes more positive), then the optimal coinsurance rate decreases. Also, when two health services become stronger gross

substitutes in the sense that $G_{12} = \partial X_1 \big/ \partial P_{X_2}$ is increased, then both services should have lower cost

shares, while goods that are complements should have higher cost sharing relative to the case in which cross price elasticities of demand for each service is zero.

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 is assumed to be the same in both periods, and hence so are premiums. 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 $L_{X_1} = L_{X_2} = L_X$, hence $\pi_1 = \pi_2 = \pi$. We also assume constant

variances over time $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and $\varphi(1+r) = 1$.

$$
\begin{array}{c}
Max \\
X_1, Y_1, S_1
\end{array}
EV = E_{\theta_1, \theta_2}
\left\{
\begin{array}{l}
V^1\left(X_1, Y_1, \theta_1\right) - L_{\theta_1}(\theta_1) \\
+\varphi\, E_{\theta_2|\theta_1}\left\{\left[V^2\left(X_2, Y_2, \theta_2\right) - L_{\theta_2}(\theta_2)\right]\right\}
\end{array}
\right\}
$$

(A.16)

where

$$V^1(X_1, Y_1, \theta_1) = V^S(Y_1 - \frac{(\mu_X + \theta_1 - X_1)^2}{2B})$$

$$V^2(X_2, Y_2, \theta_2) = V^S(Y_2 - \frac{(\mu_X + \theta_2 - X_2)^2}{2B})$$

$$\theta_2 = \rho\theta_1 + \varepsilon_2 \tag{A.17}$$

$$\varphi = \text{consumers discount factor}$$

$$K_1 = K_2 = (c_X + L_X)\mu_X - \frac{B(c_X + L_X)^2}{2}$$

$$\pi = (1 - c_X)(\mu_X - B(c_X + L_X))$$

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:

$$X_1^* = \mu_X + \theta_1 - B(c_X + L_X) \tag{A.18}$$

$$X_2^* = \mu_X + \theta_2 - B(c_X + L_X)$$

Savings will equilibrate the expected marginal utility of income in period 2 with the

marginal utility in period 1. Hence we have

$$V_I^1(\theta_1) = (1 + r)\varphi \, E_{\theta_2|\theta_1}\left\{\left[V_I^2(\theta_2)\right]\right\} \tag{A.19}$$

Let the optimal savings function be $S_1^*(\theta_1, c_X)$ (derived below) and write the problem using indirect
utility as

$$EV^* = E_{\theta_1, \theta_2}\left\{\begin{array}{l}\left[V^1\left(J - K - S_1^*(\theta_1, c_X) - (c_X + L_X)\theta_1\right) - L_{\theta_1}(\theta_1)\right] \\ +\varphi \, E_{\theta_2|\theta_1}\left\{\left[V^2\left(J - K + (1+r)S_1^*(\theta_1, c_X) - (c_X + L_X)\theta_2\right) - L_{\theta_2}(\theta_2)\right]\right\}\end{array}\right\}$$

where

$$J_1 = J_2 = J = I - \pi$$

$$\theta_2 = \rho\theta_1 + \varepsilon_2 \tag{A.20}$$

$$\varphi = \text{consumers discount factor}$$

$$K = (c_X + L_X)\mu_X - \frac{B(c_X + L_X)^2}{2}$$

$$\pi = (1 - c_X)(\mu_X - B(c_X + L_X))$$

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

$$\frac{\partial EV^*}{\partial c_X} = E_{\theta_1,\theta_2}\left\{\begin{array}{l}\left[\begin{array}{l}V_I^1\left(J-K-S_1^*(\theta_1,c_X)-(c_X+L_X)\theta_1\right)\\ \times\left(-\dfrac{\partial S_1^*(\theta_1,c_X)}{\partial c_X}-\dfrac{\partial \pi}{\partial c_X}-\mu_X+B(c_X+L_X)-\theta_1\right)\end{array}\right]\\ +\varphi\, E_{\theta_2|\theta_1}\left[\begin{array}{l}V_I^2\left(J-K+(1+r)S_1^*(\theta_1,c_X)-(c_X+L_X)\theta_2\right)\\ \times\left((1+r)\dfrac{\partial S_1^*(\theta_1,c_X)}{\partial c_X}-\dfrac{\partial \pi}{\partial c_X}-\mu_X+B(c_X+L_X)-\theta_2\right)\end{array}\right]\end{array}\right\} \tag{A.21}$$

By the envelope theorem, the terms involving $\dfrac{\partial S_1^*(\theta_1,c_X)}{\partial c_X}$ in the above expression will cancel out due to the assumption of optimal savings. Hence we can rewrite this as:

$$\frac{\partial EV^*}{\partial c_X} = E_{\theta_1}\left\{\begin{array}{l}\left[V_I^1\left(J-K-S_1^*(\theta_1,c_X)-(c_X+L_X)\theta_1\right)\times\left(-\dfrac{\partial \pi}{\partial c_X}-\mu_X+B(c_X+L_X)-\theta_1\right)\right]\\ +\varphi\, E_{\theta_2|\theta_1}\left[\begin{array}{l}V_I^2\left(J-K+(1+r)S_1^*(\theta_1,c_X)-(c_X+L_X)\theta_2\right)\\ \times\left(-\dfrac{\partial \pi}{\partial c_X}-\mu_X+B(c_X+L_X)-\theta_2\right)\end{array}\right]\end{array}\right\} \tag{A.22}$$

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_I^1 this is identical to the earlier specification. Taking a first order Taylor series approximation of the V_I^1 and V_I^2 functions, we can use the familiar expansions

$$0 = E_{\theta_1}\left\{\begin{array}{l}\left[1-R^A\left(-S_1^*(\theta_1,c_X)-(c_X+L_X)\theta_1\right)\right]\left(-\dfrac{\partial \pi}{\partial c_X}-\mu_X+B(c_X+L_X)-\theta_1\right)\\ +\varphi\, E_{\theta_2|\theta_1}\left\{\left[1-R^A\left((1+r)S_1^*(\theta_1,c_X)-(c_X+L_X)\theta_2\right)\right]\left(-\dfrac{\partial \pi}{\partial c_X}-\mu_X+B(c_X+L_X)-\theta_2\right)\right\}\end{array}\right\}$$

$$\frac{\partial EV^*}{\partial c_X} \approx E_{\theta_1}\left\{\begin{array}{l}\left[V_I^1(J-K)+V_{II}^1(J-K)\left(-S_1^*(\theta_1,c_X)-(c_X+L_X)\theta_1\right)\right]\left(-\dfrac{\partial \pi}{\partial c_X}-\mu_X+B(c_X+L_X)-\theta_1\right)\\ +\varphi\, E_{\theta_2|\theta_1}\left\{\begin{array}{l}\left[V_I^2(J-K)+V_{II}^2(J-K)\left((1+r)S_1^*(\theta_1,c_X)-(c_X+L_X)\theta_2\right)\right]\\ \times\left(-\dfrac{\partial \pi}{\partial c_X}-\mu_X+B(c_X+L_X)-\theta_2\right)\end{array}\right\}\end{array}\right\}$$

$$= \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) \right)$$

$$-R^A E_{\theta_1} \left\{ \left(-S_1^*(\theta_1, c_X) - (c_X + L_X)\theta_1 \right) \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) - \theta_1 \right) \right\}$$

$$+\varphi \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) \right)$$

$$-\varphi R^A E_{\theta_1, \theta_2} \left\{ \left((1+r)S_1^*(\theta_1, c_X) - (c_X + L_X)\theta_2 \right) \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) - \theta_2 \right) \right\}$$

$$= (1+\varphi) \left(-\mu_X + Bc_X - \frac{\partial \pi}{\partial c_X} + BL_X \right)$$

$$-R^A \left\{ E_{\theta_1} \left[\left(-S_1^*(\theta_1, c_X) \right) \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) - \theta_1 \right) \right] + E_{\theta_1} \left[(c_X + L_X)\theta_1^2 \right] \right\}$$

$$-\varphi R^A \left\{ E_{\theta_1, \theta_2} \left[(1+r)S_1^*(\theta_1, c_X) \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) - \theta_2 \right) \right] + E_{\theta_1, \theta_2} \left[(c_X + L_X)\theta_2^2 \right] \right\}$$

$$= (1+\varphi) \left(-\mu_X + Bc_X - \frac{\partial \pi}{\partial c_X} + BL_X \right)$$

$$-R^A \left\{ E_{\theta_1} \left[\left(-S_1^*(\theta_1, c_X) \right) \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) - \theta_1 \right) \right] + (c_X + L_X)\sigma^2 \right\}$$

$$-\varphi R^A \left\{ E_{\theta_1, \theta_2} \left[(1+r)S_1^*(\theta_1, c_X) \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) - \theta_2 \right) \right] + (c_X + L_X)\sigma^2 \right\}$$

$$= (1+\varphi) \left(-\mu_X + Bc_X - \frac{\partial \pi}{\partial c_X} + BL_X \right) - R^A(c_X + L_X)\sigma^2 (1+\varphi)$$

$$-R^A \left\{ E_{\theta_1} \begin{bmatrix} -\left(S_1^*(\theta_1, c_X) \right) \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) \right) \\ +\varphi \left((1+r)S_1^*(\theta_1, c_X) \right) \left(-\frac{\partial \pi}{\partial c_X} - \mu_X + B(c_X + L_X) \right) \end{bmatrix} \right\} \quad \text{(A.23)}$$
$$\phantom{-R^A \Big\{} + E_{\theta_1, \theta_2} \left[\left(S_1^*(\theta_1, c_X) \right) \left(\theta_1 - \varphi(1+r)\theta_2 \right) \right] \Bigg\}$$

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

$$0 = (1+\varphi)\left(-\mu_X + Bc_X - \frac{\partial \pi}{\partial c_X} + BL_X\right) - R^A\sigma^2(c_X + L_X)(1+\varphi)$$

$$- R^A E_{\theta_1,\theta_2}\left[\left(S_1^*(\theta_1,c_X)\right)(\theta_1 - \theta_2)\right] \tag{A.24}$$

As a reminder, $\frac{\partial \pi}{\partial c_X} = -\mu_X + 2Bc_X + BL_X - B$.

$$0 = (1+\varphi)(B - Bc_X) - R^A\sigma^2(c_X + L_X)(1+\varphi) - R^A E_{\theta_1,\theta_2}\left\{\left(S_1^*(\theta_1,c_X)\right)(\theta_1 - \theta_2)\right\} \tag{A.25}$$

This expression cannot be simplified further without explicit savings function form $S_1^*(\theta_1,c_X)$. As is showed later, optimal savings rule can be approximated by the following linear function

$$S_1^*(\theta_1,c_X) = \bar{S}_1 - s_1(c_X + L_X)\theta_1 \tag{A.26}$$

Where $\bar{S}_1 = \dfrac{\varphi(1+r)-1}{R^A\left[1+\varphi(1+r)^2\right]}$, $s_1 = \dfrac{\varphi(1+r)\rho - 1}{1+\varphi(1+r)^2}$,

This implies that there is an average savings level $\bar{S}_1$ but that savings is reduced by proportion s_1 for all losses (compensated or uncompensated). Since $\bar{S}_1$ will be uncorrelated with $(\theta_1 - \theta_2)$, it will drop out once expectations are taken and we can write

$$0 = (1+\varphi)(B - Bc_X) - R^A(c_X + L_X)\left(\sigma_1^2 + \varphi\sigma_2^2\right) - R^A E_{\theta_1,\theta_2}\left\{\left(\bar{S}_1 - s_1(c_X + L_X)\theta_1\right)(\theta_1 - \theta_2)\right\}$$

$$0 = (1+\varphi)(B - Bc_X) - R^A(c_X + L_X)\left(\sigma_1^2 + \varphi\sigma_2^2\right) + R^A\left[s_1(c_X + L_X)\left(\sigma^2 - \rho\sigma^2\right)\right]$$

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

$$\tag{A.27}$$

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

$$c_X^* = \frac{B - R^A L_X\sigma^2\left[1 - s_1\dfrac{1-\rho}{1+\varphi}\right]}{B + R^A\sigma^2\left[1 - s_1\dfrac{1-\rho}{1+\varphi}\right]} \tag{A.28}$$

If we plug in function form of $s_1 = \dfrac{1-\rho}{2+r}$, then $c_X^* = \dfrac{B - R^A L_X\sigma^2\left[1 - \dfrac{(1-\rho)^2}{(2+r)(1+\varphi)}\right]}{B + R^A\sigma^2\left[1 - \dfrac{(1-\rho)^2}{(2+r)(1+\varphi)}\right]}$.

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. As long as income is sufficient, 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_1, Y_1 and S_1 simultaneously in the first period.

$$\underset{S_1}{Max} \ \ EV = \left[\ V(J - K - S_1 - (c_X + L_X)\theta_1) - L_{\theta_1}(\theta_1) \right]$$

$$+ \varphi E_{\theta_2} \left[\ V(J - K + (1+r)S_1 - (c_X + L_X)\theta_2) - L_{\theta_2}(\theta_2) \right]$$

where

$$J = I - \pi$$

$$\theta_2 = \rho\theta_1 + \varepsilon_2$$

$$\varphi = \text{consumers discount factor}$$

$$K = (c_X + L_X)\mu_X - \frac{B(c_X + L_X)^2}{2}$$

$$\pi = (1 - c_X)(\mu_X - B(c_X + L_X))$$

The optimal saving function S_1^* satisfies

$$-V_I\left(J - K - S_1^* - (c_X + L_X)\theta_1\right) + \varphi(1+r)E_{\theta_2|\theta_1}\left[V_I\left(J - K + (1+r)S_1^* - (c_X + L_X)\theta_2\right)\right] = 0$$

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

$$V_I(J-K) - \left[S_1^* + (c_X + L_X)\theta_1\right]V_{II}(J-K) = \varphi(1+r)E_{\theta_2|\theta_1}\left\{\begin{matrix} V_I(J-K) \\ +\left[(1+r)S_1^* - (c_X + L_X)\theta_2\right]V_{II}(J-K) \end{matrix}\right\}$$

$$V_I(J-K) - \left[S_1^* + (c_X + L_X)\theta_1\right]V_{II}(J-K) = \varphi(1+r)V_I(J-K) + \varphi(1+r)^2 S_1^* V_{II}(J-K)$$

$$-\varphi(c_X + L_X)(1+r)V_{II}(J-K)E_{\theta_2|\theta_1}(\theta_2)$$

As is assumed $\theta_2 = \rho\theta_1 + \varepsilon_2$, $E_{\theta_2|\theta_1}(\theta_2) = \rho\theta_1$, we obtain

$$V_I(J-K) - \left[S_1^* + (c_X + L_X)\theta_1\right]V_{II}(J-K) = \varphi(1+r)V_I(J-K) + \varphi(1+r)^2 S_1^* V_{II}(J-K)$$

$$-\varphi(1+r)(c_X + L_X)V_{II}(J-K)\rho\theta_1$$

$$1 + \left[S_1^* + (c_X + L_X)\theta_1\right]R^A = \varphi(1+r) - \varphi(1+r)^2 S_1^* R^A + \varphi(c_X + L_X)(1+r)R^A\rho\theta_1$$

Rearranging the above equation, we solve the optimal saving function as following

$$S_1^* = \frac{\varphi(1+r) - 1}{R^A\left[1 + \varphi(1+r)^2\right]} - \frac{1 - \varphi(1+r)\rho}{1 + \varphi(1+r)^2}(c_X + L_X)\theta_1, \text{ which implies that}$$

$$\bar{S}_1 = \frac{\varphi(1+r)-1}{R^A\left[1+\varphi(1+r)^2\right]}, \quad s_1 = \frac{1-\varphi(1+r)\rho}{1+\varphi(1+r)^2}$$

Given our assumption $\varphi(1+r)=1$, S_1^* can be simplified further as

$$S_1^* = -\frac{1-\rho}{2+r}(c_X + L_X)\theta_1 \text{ and } \bar{S}_1 = 0, \ s_1 = \frac{1-\rho}{2+r}$$

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

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

As is shown in B-1(available upon request), if we have three periods,
$$S_1^* = -\frac{(1+r)(1-\rho_1)+(1-\rho_2)}{1+(1+r)+(1+r)^2}P_X\theta_1\,.$$

$$S_1^* = -\frac{(1+r)^2(1-\rho_1)+(1+r)(1-\rho_2)+(1-\rho_3)}{1+(1+r)+(1+r)^2+(1+r)^3}P_X\theta_1 \text{ for T=4.}$$

And $S_1^* = -\dfrac{(1+r)^3(1-\rho_1)+(1+r)^2(1-\rho_2)+(1+r)(1-\rho_3)+(1-\rho_4)}{1+(1+r)+(1+r)^2+(1+r)^3+(1+r)^4}P_X\theta_1$ when T=5 (same with the number of years of data we have).

.

From the pattern of optimal saving for multiple period models, it is easy to see that the optimal

saving result for any $T \geq 2$ as $S_1^* = -\dfrac{\sum\limits_{t=1}^{T-1}(1+r)^{t-1}(1-\rho_{T-t+1})}{\sum\limits_{t=1}^{T}(1+r)^{t-1}}P_X\theta_1\,.$