Rising Health Care Expenditures: A Demand-Side Analysis

Babu Nahata, Krzysztof Ostaszewski, and Prasanna Sahoo*

Abstract: In this paper we consider a utility-maximization model for health care. On the basis of the equilibrium conditions derived for patients and the providers of the medical service, we evaluate the importance of cost-sharing between the patients and the third party and provide an explanation for the rising medical expenditures. We effectively assume that some form of third-party payer is always involved in the health care market and this involvement has significant consequences for the incentives of both consumers and providers of health care. The proposed demand specification explains why the empirical estimates of the price elasticity of demand for medical services could exhibit a wide range. We analyze how medical insurance can result in a market failure and evaluate ideas that can correct some of the distortions in resource allocation for medical services. Some guidelines also emerge for a national health insurance policy. [Keywords: health savings accounts; national health insurance policy; health care demand]

INTRODUCTION

Both public and private expenditures on health care are rising faster than any other sector of the U.S. economy. As the Center for Medicare and Medicaid Services (2001) states:

“National health expenditures are projected to reach $3.1 trillion in 2012, growing at a average annual rate of 7.3 percent during the forecast period 2002–2012. As a share of gross domestic product (GDP), health spending is projected to reach 17.7 percent by 2012, up from its 2001 level of 14.1 percent. This projection represents a 0.4 percent-

*Babu Nahata (nahata@louisville.edu) is on the faculty in the Department of Economics at the University of Louisville. Krzysztof Ostaszewski (krzysio@ilstu.edu) is Professor and Director of the Actuarial Program in the Department of Mathematics at Illinois State University. Prasanna Sahoo (sahoo@louisville.edu) is on the faculty in the Department of Mathematics at the University of Louisville.
age point increase in GDP share by 2011 compared with last year’s forecast.”

Moreover, it is anticipated that this trend will continue unless significant policy measures are enacted (Glied, 2003). Because health care is essential, and the economy is unable to sustain this rapid growth in expenditures, a public debate rages regarding the measures that can contain the growth. Following the 1992 presidential election, a strong sentiment both among the voters and among some candidates existed to enact a national health insurance. Similarly, recent legislation providing prescription drug coverage for Medicare participants followed the 2002 mid-term election.

Because of diverse insurance contracts, multitudes of payment scales, and imperfect markets for medical services, no single factor can account for rising expenditures. As a result, many plausible explanations have been put forward for this rapid growth (Newhouse, 1981, 1988). Feldstein (1971) has argued that the spread of health care insurance has created a demand for a higher quality of services and, at the same time, has increased the quantity of medical services. Depending on the coverage of an insurance policy, the medical costs are shared by the patient and the third party. The way in which cost sharing (percentage paid out-of-pocket) between the patient and the insurance company (third party) affects the expenditures on medical services was examined by Manning, Newhouse, Duan, Keeler, Leibowitz, and Marquis (1987) using experimental data. They conclude that “A catastrophic insurance plan reduces expenditures 31 percent relative to zero out-of-pocket price” (p. 251).

Glied (2003) provides an insightful review of the evolution in the health care cost trends between 1993 and 1999, and from 1999 on. She points out that in the period 1993–1999, health care expenditures, as a share of national income, declined for the first time since the Great Depression, per capita costs having grown at 1.9% annual rate. On the other hand, she also indicates that the upward trend in the health expenditures’ share of national income has resumed since 1999. She points out three factors as essential in explaining the rise of health care costs:

- Short term: the structure of the insurance contracts, with specific emphasis on the share of the expense covered by the third party,
- Medium term: the health insurance industry underwriting cycle, and

According to Glied (2003), “The mid-1990s were an anomalous period in the history of U.S. health expenditures.” She lists the following combination of unusual factors:
• Short term: The share of the third-party payments had a slight increase, especially after 1996, and the third parties (insurance companies, employers, and Medicare) instituted aggressive measures to control utilization (through Health Maintenance Organizations, various forms of managed care, and reductions in Medicare payments),

• Medium term: Health insurance firms went through a competitive phase of a cycle, where some operating losses could also be offset by investment returns, and

• Long term: New technologies emerged mostly in the pharmaceutical areas (this conclusion provides a new perspective on the current debate on the addition of prescription drug coverage to Medicare).

We believe that we can provide a new perspective on the conclusions of the work of Glied (2003) and the general literature concerning rising health care expenditures. The key idea of our perspective is that a low level of consumer participation in purchase decisions affects not only the behavior of the consumer (which is the standard moral hazard argument), but also the behavior of the health care provider, for whom price increases become the natural profit-maximizing route.

In this paper we analyze how the amount of out-of-pocket expenses (coinsurance rate) affects the demand for medical services, and we also provide a plausible explanation for the conclusions reached by Manning et al. (1987), based on their study using a randomized experiment. We effectively assume that some form of third-party payer is always involved in the health care market and this involvement has significant consequences of the incentives of both consumers and providers of health care. We believe that this perspective also applies to the uninsured, who nearly always default into state-supported care when faced with significant health care expenditures.

It is generally argued—and the argument has been supported by several empirical studies (see Manning et al., 1987)—that the demand for medical service is quite inelastic. Thus any increase in the price of medical services not only can be passed on to consumers, but also results in an increase in the expenditures. However, the literature exhibits substantial disagreement, by a factor of as much as 20, about the price elasticity of demand (the elasticity estimates at the mean vary from around –0.1 to –2.1; Manning et al., 1987, p. 251). On the basis of our model in Section I, we specify a demand function that shows that, depending on the shares of patients and the third party, the elasticity can vary widely.

Another reason for rising expenditures has been attributed to market failure (Manning et al., 1987). The market failure argument is based on how
employer-paid health insurance premiums are treated for income tax purposes. Since these employer-paid premiums are not considered as income and are not taxed, the employee has no incentive to economize on medical expenditures. One remedy put forth has been the Archer Medical Spending Account, or Archer MSA, authorized by The Health Insurance Portability and Accountability Act of 1996, but it has been severely limited in scope. The Medicare Prescription Drug Improvement and Modernization Act of 2003 introduced health savings accounts, a dramatic new development in the design of national health policy, which we will discuss later, that also attempts to address the problem of the perceived income and incentives of consumers.

Another frequently stated argument has been that rapid technological advances (such as medical imaging equipment, or new varieties of prescription drugs) in the delivery of medical services have led to cost increases, and because of insurance structure and an imperfect market these innovations have caused both an increase in the price of and expenditures on medical care (Newhouse, 1981, p. 7). The main reason for this explanation is that “insurance as it is structured induces more rapid technological change than would be observed in a competitive market. Assuming that optimality prevails everywhere else, there are too many resources devoted to product-enhancing technological change” (Newhouse, 1981, p. 7). Manning et al. (1987, p. 269) note that technological changes have brought many new medical products and procedures. They further argue that insurance might have played an important role in inducing technological changes, but experimental data cannot answer this question affirmatively. On the other hand, Glied (2003) argues that technological advances give rise, in the long run, to increasing health care expenditures, making this process a self-perpetuating cycle.

It is interesting to note that in many of these explanations the role of insurance appears to be a key factor. The main focus of this paper then is to examine different explanations by focusing primarily on the role played by insurance in the cost-sharing between the insurer and the insured. Section I develops equilibrium conditions based on utility-maximization, which incorporates the cost-sharing between an insurer and the insured. Section II proposes a plausible demand function that may be appropriate for medical services. Section III evaluates a recent market-based solution. Section IV summarizes the main conclusions.

**MODEL FOR HEALTH CARE DEMAND**

We consider a simple model similar to one proposed by Farley (1986). All patients have an identical utility function that depends on the consump-
tion of a medical service or procedure and a composite “numeraire,” or income. We assume that the utility function is additively separable in medical service and income.\(^1\) The utility of consuming \(q\) units of medical service is described by the relationship \(H(q)\). The marginal utility is assumed to be non-negative. The utility of consuming other goods is \(V(I - T)\), where \(I\) represents the total income and \(T\) is the total out-of-pocket expenditures on medical services. We cannot assume that \(T\), unlike the expenditures on many other goods, is very small compared to total income, \(I\). Indeed, the fact that the anticipated medical expenditures could be a large portion of income is one of the main reasons that a third-party payment system exists. Thus \((I - T)\) is the total income available for consumption of all other goods. The consumers take the price \(p\) for the medical service as given and choose the quantity \(q\) to maximize their total utility, \(U = V(I - T) + H(q)\).\(^2\) Let \(pq\) represent the total revenue of the provider of the medical service. Define \(K\) as the fraction of the provider’s revenue paid by the third party, and let us assume that \(K = pq/I\) (this is a simplified version of the assumption that the third-party subsidy increases with the portion of consumer’s income absorbed by medical expenses, exactly the model existing in the U.S. economy).\(^3\) Therefore, \(T = (1 - K)pq\). Note that our model does not allow for expenditures to exceed income, because all health care expenditures are income of the consumer in this model—in other words, if a third party pays one’s health expenditures, such payments become one’s income. Consumers face the following standard utility maximization problem.

\[
\text{Maximize} \quad U = V(I - T) + H(q) \quad (1)
\]

subject to: \(I = I_0\).

Substituting for \(T\) in Equation (1), we get

\[
U = V\left(I - \left(1 - \frac{pq}{I}\right)pq\right) + H(q) \quad (1')
\]

It should be noted that in the above, \(p\) is a function of \(q\), and the chain rule for multivariate functions must be used in differentiation. Differentiation with respect to \(q\) results in the following first-order condition for utility maximization (note, importantly, that this differentiation requires the use of the chain rule on the right-hand side, and assumes that \(I\) is not a function of \(q\):
The standard notation for the expression \( (pq)' \) is \( MR \), denoting the marginal revenue of the health care providers, while \( V'(I-(1-pq/I)\cdot pq) \cdot (pq) \cdot (1-pq/I) \cdot (-pq)' + H'(q) = \) is the marginal utility of other goods, and \( H'(q) = MU^H \) is the marginal utility of the health care consumption. Since in equilibrium, \( U'(q) = 0 \), we arrive at the following equilibrium condition for the health care market with third-party subsidies not perceived as income:

\[
\frac{MU^H}{MU^{OTH}} = MR \cdot ((1 - K) - K) \tag{3}
\]

Equation (3) gives the utility-maximizing equilibrium of a consumer in terms of the provider’s marginal revenue (\( MR \)), and the consumer’s share and the share of the third party (in this model) in health care expenditures. The left-hand side of the equation represents the ratio of the marginal utilities of medical service and the income available for consuming all other goods, and it is always positive. Thus the provider’s marginal revenue will be negative (positive) depending upon whether the patient’s share is smaller (larger) than the third party’s share. When marginal revenue is negative, the elasticity of demand (absolute value) for medical service is less than one (inelastic). In reality, many employer-paid health insurance plans are comprehensive and the third party covers most of the costs for medical service. Thus, the patient’s share is small compared to what the third party pays to the provider (this happens when \( K > 0.5 \)). Thus, the provider’s marginal revenue for these groups of patients is negative and
the demand is inelastic. With inelastic demand, an increase in price results in higher revenue (or the expenditures on health care would increase). Alternatively, for those patients who pay more out-of-pocket because of higher deductibles, or have a catastrophic insurance plan, our model results in an elastic demand. Let \( S = \frac{T}{I} \) represent the fraction of income the consumers spend on health care. In this model, \( S = K(1 - K) \). The reason many employers provide medical coverage and consumers purchase health insurance themselves is to ensure a cap on the portion of income used for medical expenses. In the proposed model, consumers’ maximum share \( (S^*) \) is 0.25, which occurs when \( K = 0.5 \). In other words, even when the third party and the consumer share equally (each provides 50 percent of the provider’s revenue), consumers’ share of income on medical services cannot exceed 25 percent. Generally, for many types of medical services (e.g., major surgeries, cancer treatments) the third party contributes a major share of the provider’s revenue, and \( K \) is significantly more than 50 percent. In this case, the provider’s revenue (or expenditures on medical services) would increase with an increase in the price. Thus, we have an explanation for rising medical expenditures for certain types of medical services and procedures.

For certain types of medical services, such as routine office visits, eye examinations, glasses, and prescriptions, consumers provide a major portion of the provider’s revenue because either these services are not covered under the medical plans or the consumers pay out-of-pocket to meet the deductibles requirements. On the basis of consumers’ ability to pay, doctors routinely practice price discrimination, charging different prices for the same procedure (Kessel, 1958). Since many of these services have a relatively elastic demand, lowering the price or practicing price discrimination will result in an increase in the provider’s revenue. Notably, we observe nonlinear pricing (e.g., two-for-one glasses sales, discounts on glasses when paying for eye examinations, coupons and rebates on contact lenses, or discounts on the first visit to a dentist) in exactly those parts of health care demand where the consumer’s large share in expenditures results in elastic demand.

**DEMAND FUNCTION FOR MEDICAL SERVICE**

According to the equilibrium conditions for consumer utility-maximization, it is clear that the out-of-pocket expenses affect the demand for medical services. In specifying a demand function for medical services, two factors need to be recognized. First, the choice of medical provider is not affected by the amount of insurance premium (some HMOs, however, may
limit the pool of providers to choose from). Second, the providers of the service have, for most part, a considerable degree of freedom in charging a unit price for the service. From the equilibrium conditions derived above, it is clear that the specification of a demand function for medical service should exhibit two important characteristics. First, contrary to demand for most goods, the demand for health care is inelastic at high prices, is elastic for some price range and becomes inelastic at relatively low prices. The third-party-payment system is the primary reason for this unusual behavior in the demand function. Second, because the provider’s marginal revenue could be positive or negative, the revenue function cannot be strictly concave. Using the proposed model in Section I, we specify an example of an inverse demand function that meets the two requirements for health care demand, as follows:4

$$p = 2q^{-2.5} - 10q + 50$$  \hspace{1cm} (4)$$

This inverse demand function is a combination of linear and constant elasticity demand functions. The demand function and the corresponding marginal revenue function

$$MR = 3.0q^{-2.5} - 20q + 50$$

are graphed in Figure 1.
Both functions are continuous and twice differentiable. The marginal revenue function has both positive and negative slopes, so the revenue function is not strictly concave. As discussed above, the marginal revenue for patients with low out-of-pocket payments for medical services is negative (absolute value of demand elasticity less than one) and increasing, becoming positive (at price and quantity levels we will call \( p_1 \) and \( q_1 \)) for those patients whose demand is elastic, and then becomes negative again (at price and quantity levels we will call \( p_2 \) and \( q_2 \)).

Figure 2 shows the behavior of the revenue function. The revenue declines to a price \( p_1 \), where it reaches a minimum, and then increases, attaining a maximum at price \( p_2 \). It is quite clear from Figure 2 that between the range \( q_1 - q_2 \) there is an upper bound for the total expenditures patients will spend on the medical service. However, there is no upper bound for the expenditure for the range \( 0 < q < q_1 \). As a result, the expenditures for the patients who purchase medical service in this range will continue to increase. Consequently, if the natural upper bound for patients with elastic demand has not been reached, we have an explanation for rising medical expenditures.

Manning et al. (1987) finds that "The per capita expenses on free plans (no out-of-pocket costs) are 45 percent higher than those on the plans with a coinsurance rate, subject to an upper limit on out-of-pocket expenses" (p. 258). Their observation is quite consistent with our model. As shown above, the higher the patient’s share for a medical service (more out-of-pocket costs, for example, a catastrophic insurance plan) the more elastic the demand. Therefore, the consumer would buy less at a given price, resulting in lower total expenditure. Manning et al. (1987) confirm this observation. According to their study, “Cost sharing affects primarily the

![Fig. 2. Behavior of the revenue function.](image)
number of medical contacts, rather than the intensity of those contacts. In other words, the differences in expenditures across plans reflect real variation in the number of contacts rather than an increase in the intensity or charge per service” (p. 258). If we interpret intensity as a proxy for the quality of service, *ceteris paribus*, at any given price the patient is inclined to purchase less without compromising the quality when the out-of-pocket costs are high. Thus, our model and the proposed demand function provide a plausible theoretical explanation for why expenditures on health care are rising and why a catastrophic plan might reduce total health care costs.

Figure 3 shows the behavior of the price elasticity of demand (defined as a positive number) for the demand function example we considered:

\[
\eta = \frac{dq}{dp} = \frac{2q^{2.5} - 10q + 50}{5q^{2.5} - 10q}. 
\]

The elasticity of demand increases and then decreases. This behavior of elasticity is supported by both our model and the empirical estimates. Two commonly specified demand functions in empirical studies are constant elasticity and linear. However, these specifications are not appropriate for medical service since elasticity of demand cannot be assumed to be either constant or decreasing for medical services because of third-party payment. Thus a combination of these two demand functions would provide one of the many plausible specifications. For medical services, the demand is inelastic at higher prices because the third party’s share is higher; it
becomes elastic as the patient’s share increases. Therefore, it is quite possible for empirical studies to exhibit a wide range of elasticity estimates, as reported by Manning et al. (1987).

HEALTH SAVINGS ACCOUNTS

Since the second half of the 1990s, we have seen an increasing move in the health marketplace toward insurance arrangements that hold some promise for alleviating the market failure argument of our work. First, medical savings accounts (MSAs) were authorized by the Health Insurance Portability and Accountability Act of 1996, otherwise known as the Kassebaum-Kennedy health insurance reform bill. Under the bill, Congress instituted a demonstration project on medical savings accounts (MSAs), allowing up to 750,000 MSAs. MSAs were made available to those without health insurance, self-employed individuals, and small employers with 2 to 50 employees. An MSA is a savings account that can be used to pay medical expenses not covered by insurance. That type of account, also known as an Archer MSA, has become quite popular with the uninsured and small employers. But its success can hardly be judged in full, owing to the very restrictive nature of the program.

The new health savings account (HSA) provisions included in the Medicare bill passed in 2003 bring about a considerable expansion of that type of account. Since the new law became effective on January 1, 2004, about 250 million non-elderly Americans have become eligible for HSAs. Account holders must have a qualified catastrophic insurance plan “underlying” the savings account. Compared with Archer MSAs, allowable deductibles are lower: $1,000 for an individual and $2,000 for a family. The maximum deductible requirement of Archer MSAs has been replaced by maximum out-of-pocket limits of $5,000 for individuals and $10,000 for families. The limits include deductibles and coinsurance for “in-network” providers. The amounts will be adjusted annually for the cost of living. There is also no restriction on the stop-loss limits for out-of-network services. Annual contributions to the HSA are limited to 100 percent of the deductible, with a maximum of $2,600 for an individual and $5,150 for a family. Account holders age 55 and up may make “catch-up” contributions of $500 in 2004, increasing the amount by $100 each year until this “catch-up” amount reaches $1,000 in 2009. The contributions may be made by the employer, by the employee, or by an individual opening the account. Employer contributions are excluded from income and individual contributions are deductible (importantly, a taxpayer does not have to itemize deductions in order to take the contribution as a deduction—i.e., this
RISING HEALTH CARE EXPENDITURES

It has long been recognized in the literature that institutional arrangements cause distortion in consumer demands (Pauly, 1968; Feldstein, 1973). Rarely, however, has the effect on the profit-maximization problem of the health provider been pointed out. Our simple utility-maximization model provides a plausible quantitative theoretical explanation for rising medical expenditures, primarily as a result of demand distortions, and the resulting incentives for the producer (i.e., outside of the standard moral hazard
argument, as such argument has been applied to the insured, but not to the health care provider). This model indicates that the way in which costs are shared between the consumers and the insurer may be the single most important factor in capping the rapid rise in medical expenditures. Our analysis supports the recommendation proposed by Newhouse (1981): that a market-oriented strategy should structure insurance policies so that most individuals, most of the time, pay for their medical care services by including substantial deductibles in health insurance policies (e.g., a catastrophic plan or indemnity insurance).

As noted by Newhouse (1981, p. 8), insurance as presently structured tends to reduce the amount of price competition in the marketplace. As a result, the health care provider, like a monopolist facing a highly inelastic demand curve, competes on the basis of quality, with nearly no regard for price. Our model shows why this can happen. The proposed demand function also explains why the wide range of elasticity estimates reported in the literature is quite possible, depending on the coinsurance rates.

As noted by Manning et al. (1987), “one does not observe insurance policies that do or do not cover new procedures, so there is no straightforward test of willingness to pay for new technology. Although virtually all policies do not cover ‘experimental’ procedures, once efficacy and ‘safety’ are demonstrated, insurance plans tend to cover all procedures” (p. 269). One can therefore argue that the third-party payment system may not subsidize all new products or procedures directly in the short run, but it encourages innovators to offer a large number of products in the long run. Of those products, those that succeed, can recover the costs for products that fail. Once a new product is approved, the provider can charge a higher price that covers the cost of not only the successful product but also the ones that failed. Because insurance covers those products that succeed, the third-party payment system encourages firms to engage in a larger number of exploratory drugs and procedures in the long run, in spite of the fact that only a few of such undertakings ultimately succeed.

Glied (2003) has pointed out that a substantial proportion of the decline in the rate of growth of health care expenditures came from massive implementation of managed care and Medicare payment reductions. She also notes an increase in the share of patients’ out-of-pocket payments in the study period. We believe these observations to be consistent with our conclusions. While managed care does not necessarily reduce the third-party share of monetary payment, it does increase non-monetary cost to the patients (the dissatisfaction of consumers with the restrictions imposed by HMOs and other forms of managed care is nearly legendary). The economic effect is the same: the consumers shoulder a higher share of the health care cost burden. Medicare payment reductions, on the other hand,
reduce the third-party share directly. Furthermore, following the 1991 recession, more employers required employee contributions to their health insurance premium costs, again increasing the out-of-pocket share of the consumers.

Our analysis seems to support the Health Savings Account provision of the Medicare Prescription Drug Improvement and Modernization Act of 2003, which was enacted in the hope that the HSAs would increase consumers’ coinsurance rate and encourage competition among health providers. It should be noted, however, that this legislation also provides a Medicare prescription drug benefit, which created exactly the opposite effect—a likely decrease in the consumers’ coinsurance rate in that part of the health care marketplace. Furthermore, the incentives of HSAs can be somewhat ineffective for low-income consumers, who cannot afford to set aside funds for medical expenses and in fact may fall under some form of public assistance, thus increasing their third-party share in their medical expenses. Finally, as of this writing, many consumers are simply not eligible for HSAs because their basic health insurance plans do not qualify them—and those plans are chosen by their employers, not themselves.

So far, the public debate has not produced any specific structure for national health insurance. It needs to be recognized that merely switching the third-party payment system from the private sector to the public domain will not cure the problem of rising expenditures—unless such a switch is accompanied by some form of rationing or another command approach. A national health policy that is cost effective without compromising the quality of the medical service must balance the cost-sharing between consumers and third-party payers. In fact, measures such as catastrophic plans and health savings accounts may be more cost-effective alternatives than a mandated national health policy.

ENDNOTES

1 We use an additively separable utility function for illustration. However, it is easy to verify that similar results can also be obtained using a Cobb-Douglas utility function. Thus the results are not restricted to additively separable utility functions only.

2 As is standard in utility-maximization models, we assume that the consumer takes price as given. Because of health insurance policies, and many other factors—e.g., incomplete information about the nature of the service, and the goals of the provider (many hospitals are not for-profit organizations)—the determination of the price of medical services involves a complex process. This paper is not concerned with how the prices are determined. In fact, there is no simple model that can incorporate all relevant factors that determine the price of a medical service. Newhouse (1981) has very elegantly articulated the complexities in price determination. He has argued that in general the insurer does not question the nature of medical services provided to a patient as long as the services fall within a general description of those covered by the patient’s policy.
3 $K$ may be considered as a proxy for the coinsurance rate but it is not the same. Generally, the coinsurance rate is a stipulated amount paid by the patient regardless of income.

4 We have specified a simplest form of demand function to illustrate the result. In fact, there are many other specifications that exhibit similar characteristics. For example, $p = 100q^{-1.1} + 0.1q^2 - 10q + 50$ also produces a revenue function that is not strictly concave.

5 Some theoretical and pedagogical implications of nondecreasing marginal revenue have been discussed by Formby, Layson, and (1982). Nahata, Ostaszewski, and Sahoo (1990), in the context of third-degree price discrimination, show that if the concavity assumption for the revenue or the profit function is relaxed, many counterintuitive results are possible when price discrimination is allowed. For medical service it is easy to segment the market by the coverage. Since the demand elasticity varies depending on the amount of out-of-pocket share, it is more realistic to assume that revenue function is not strictly concave.

REFERENCES


