Medical Expenditure Puzzle*

Xiaoshu Han
Department of Economics,
University of Texas at Austin

Incomplete

(February, 2006)

Abstract

What does your medical expenditure do to your health? Researchers often get significant negative sign on the relative coefficient in the reduced form health production regression. The puzzling result motivates this simple dynamic quantitative general equilibrium model to study the relationships between health status, medical expenditure and employment. The structural parameters are estimated by an indirect inference procedure. This paper finds that the simulated coefficient of medical expenditure in the health equation is negative even though in the health evolution equation of the structural model, medical expenditure only impacts the health in the positive way.

*I am grateful to my thesis supervisor Russell Cooper for his support and guidance. I would like to thank Stephen Bronars, Dean Corbae, Stephen Donald, Burhanettin Kuruscu, Gerald Oettinger and John Rust for their valuable comments and suggestions. I also thank all the participants in the macroeconomics writing course at University of Texas at Austin for helpful comments. All the errors remain with author.

1 Introduction

As we all know, medical expenditure in U.S. now is almost 15 percent of its total GDP.[1] Hence, in order to effectively control the skyrocketing cost and at the same time, improve individuals' qualities of life, it is very important for researchers and policy makers to understand the role of medical expenditure in individuals' lives. What does your medical expenditure do to your health? People would respond very naturally that of course it will do good to health or there is just no point of spending any money on medical care. However, researchers often get significant negative sign on the relative coefficient in the reduced form health production regression. Before we can explain the medical expenditure puzzle, let us do some literature review.

Medical expenditure has been suggested to be viewed as one form of human capital decades ago (Mushkin 1962, pp. 129-49; Becker 1964, pp. 33-36; Fuchs 1966, pp. 90-91).[2] Grossman (1972) is the first to construct a life cycle model of the demand for health capital itself. In his model, health is viewed as a capital stock which yields an output of "healthy days". Individuals may invest in health by combining time (e.g., for doctor's visits) with purchased inputs(e.g., medical services). The incentive for investing in health is that by increasing the health stock the individual increases the amount of time available for earning income or for producing consumption goods. This approach has enabled him to derive some propositions about the pattern of medical expenditures over an individual's lifetime and to describe the behavior of health capital over the life cycle. Cropper (1977) adds in a few new features into the previous life cycle model such as the randomness of illness and the disutility associated with illness and derives some propositions about the pattern of medical expenditures over an individual's lifetime. However, neither of there papers has a specific

[1]Source: CMS, Office of the Actuary, National Health Statistics Group.
[2]Quoted from Grossman 1972.

utility function. Their papers concentrate only on some analytical works.

Grossman (1972b) then runs a reduced form regression on health production function and finds that the coefficient on medical expenditure is significant negative which contradicts the expectation from his previous paper. He explains this is a result of the correlation between the medical expenditure and the error term (health depreciation rate). The correlation causes a downward bias and hence, the coefficient could be negative. In order to correct the wrong sign on medical expenditure, a lot of labor and health economists have devoted enormous amount of efforts to find the right data set and right instruments to a certain group of individuals. Rosenzweig and Schultz (1983, 1988, 1991) and Grossman and Joyce(1990) consider the effects of some health inputs such as mother's prenatal care and smoking behavior on baby's health. They use baby's weight as a measure of baby's health and use income, education and price of medical care service as instruments for medical care. They get the consistent estimates though the validity of baby's weight as an instrument for baby's health has been questioned.

Instead of spending all the efforts on searching for the good instruments and trying to correct the sign as the most researchers do, this paper studies individuals' choices of medical expenditure and how their medical expenditures affect their health and in turn their job opportunities and qualities of life. This paper reviews the impacts and causalities between health status, medical expenditure and employment in a simple dynamic quantitative general equilibrium model. The structural parameters in the model are estimated by matching the model's implications with individual observations from the Medical Expenditure Panel Survey (MEPS) as part of a minimum distance estimation routine.

With the estimated structural parameters, a panel data set is simulated. The paper finds that the simulated coefficient of medical expenditure in the health equation is negative even though in the health evolution equation of the structural model, medical expenditure only impacts the health in the positive way. This study shows that measurement without theory, without taking into consideration of people's preference, behavior, market structure and government policy, is just not that informative (Rosenzweig and Wolpin 2000).

The paper is arranged as follows: Section 2 outlines the theoretical model. Section 3 describes the data and model parameterization strategy. Section 4 presents the estimation results. Section 5 concludes the paper and discusses the future work.

2 The Theoretical Model

2.1 The Economic Environment

This paper considers a dynamic quantitative model with heterogeneous agents. The economy is populated by a large number of individuals who are ex ante heterogeneous with respect to their health status. The agents in this economy are infinite-lived and maximize the following expected value of their discounted utility:

$$E\sum_{t=0}^{\infty}\beta^t U(c_t, h_t) \tag{1}$$

where c_t is consumption, h_t is health stock, β is the discount factor and $0 < \beta < 1$, $U(\cdot, \cdot)$ is the momentary utility function.

In this simple model, agents can't save. Their budget constraint is given by

$$c_t = y_t^d - m_t \tag{2}$$

where c_t is the consumption in the current period, y_t^d is the disposable income in the current period and m_t is the medical care expenditure in the current period (which includes all expenditures that may affect an agent's health status, such as time and money spent on health clubs, appropriate nutrition, medical insurance, and other expenses related to health care.)

Agents' health stocks evolve according to the following equation:

$$h_{t+1} = \phi_t(1-\delta)h_t + am_t^b \tag{3}$$

where h_t is the health stock at the beginning of the current period, δ is the health depreciation rate, ϕ_t is the health shock at the beginning of the current period, m_t is the medical care expenditure in the current period, a and b are parameters and h_{t+1} is the health stock at the beginning of the next period.

In each period of their lives, agents face a stochastic employment opportunity. Let s denote the employment state of an individual. If $s = e$, the agent is employed and if $s = u$, the agent is unemployed. Conditional on agents' employment status last period and health stock at the beginning of this period, the employment probabilities in this period are denoted by $\pi(e'|e, h')$, $\pi(u'|e, h')$, $\pi(u'|u, h')$ and $\pi(e'|u, h')$. They are estimated by Nadaraya Watson Nonparametric Regression from the MEPS data.

Figure 1 shows the employment probabilities this period conditional on whether the individual is employed (top) or unemployed (bottom) last period and their health status. The vertical axis is the estimated employment probabilities and the horizontal axis is the actual health stock divided by 100. The lowest health level is 11.73 and the highest level is 67.24 in the MEPS data (see Table 2).

The probabilities are continuous in h'. From the graphs, we can see that everything else being equal, the healthier agent has a higher chance of getting the job and the agent who had a job last period has a better chance of being employed this period.

If the agent is employed, his disposable income is $y - \tau$, where y is his income and τ is the income tax he has to pay. If the agent is unemployed, his disposable income is just his unemployment insurance from the government θy, where θ is the replacement ratio of unemployment insurance.

Agents face an i.i.d. stochastic health shock in each period. Let $\phi \in \Phi = \{\phi_g, \phi_b\}$[3] denote the health shock state.

Let η denotes the employment status of the agent. If the agent is employed, $\eta = 1$,

[3] ϕ_g is good shock, i.e. new health technology being invented and ϕ_b is bad shock, i.e. car accident or broken leg. $\phi_g > 1$ and $\phi_b \in [0, 1)$.

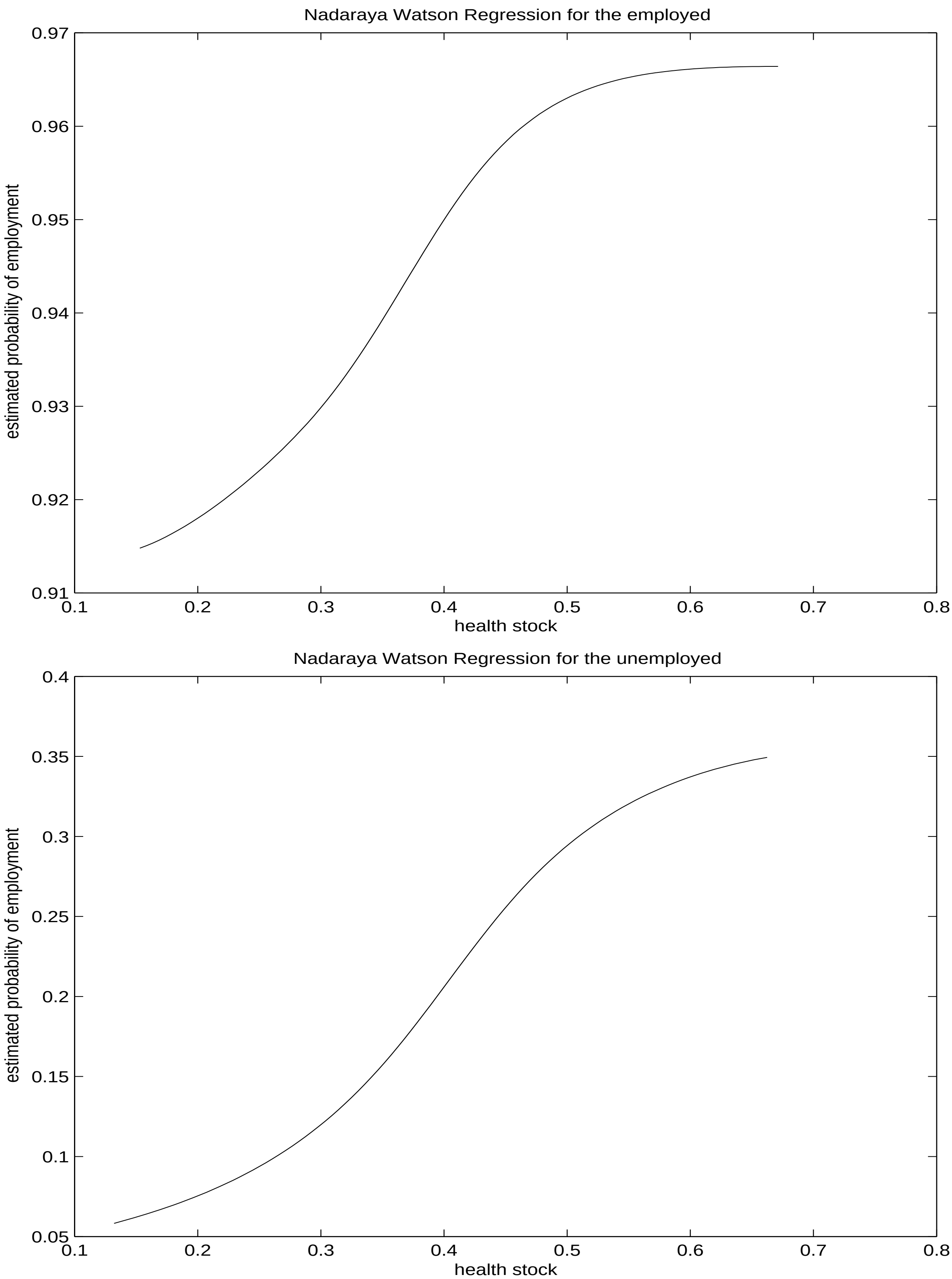

Figure 1: The employment transition probabilities for employed and unemployed individuals

otherwise $\eta = 1$.

The timing of the model is the following: The agent enters period t with a health stock of h_t, then his employment probability can be calculated and he gets a health shock ϕ_t. His health depreciates at a rate of δ and he makes his investment on medical care m_t to maintain his health. Then he enters period t+1 with a health stock of h_{t+1} which evolves according to equation (3).

The maximization problem can be written as a dynamic programming problem. Note that the state variable are health stock h, employment opportunity s and health shock ϕ. The dynamic programming problem is:

$$\begin{aligned} V(\phi, e, h) &= \max_m u[y(1-\tau) - m, h] + \beta \sum_{\phi'} \chi(\phi'|\phi) \sum_{s'} \pi(s'|e, h') V(\phi', s', h') \\ V(\phi, u, h) &= \max_m u[\theta y - m, h] + \beta \sum_{\phi'} \chi(\phi'|\phi) \sum_{s'} \pi(s'|u, h') V(\phi', s', h') \end{aligned}$$

subject to

$$m \geq 0 \tag{4}$$

Definition: The stationary equilibrium for this economy is the set of decisions rules $c(\phi, s, h)$, $m(\phi, s, h)$, a time-invariant measure $\lambda(\phi, s, h)$ of individuals at state (ϕ, s, h) and a tax rate τ such that:

1. Given the tax rate τ, individuals solve the maximization problem in (4).

2. The goods market clears:

$$\sum \lambda(\phi, s, h) c(\phi, s, h) + \sum \lambda(\phi, s, h) m(\phi, s, h) = \sum \lambda(\phi, s, h) \eta(\phi, s, h) y(\phi, s, h) \tag{5}$$

3. Government finances UI benefits by taxing income. So, the total amount of UI benefits should be equal to the taxes paid by the employed individuals. The government budget constraint is satisfied:

$$\sum \lambda(\phi, s, h) \eta(\phi, s, h) \tau y(\phi, s, h) = \sum \lambda(\phi, s, h)[1 - \eta(\phi, s, h)] \theta y(\phi, s, h) \tag{6}$$

4. The invariant measure $\lambda(\phi, s, h)$ solves the following equations:

$$\begin{aligned}
\lambda'(\phi_g, e, h') &= \sum_\Omega \lambda(\phi, s, h)\pi(e|s, h)\chi(\phi_g|\phi) \\
\lambda'(\phi_b, e, h') &= \sum_\Omega \lambda(\phi, s, h)\pi(e|s, h)\chi(\phi_b|\phi) \\
\lambda'(\phi_g, u, h') &= \sum_\Omega \lambda(\phi, s, h)\pi(u|s, h)\chi(\phi_g|\phi) \\
\lambda'(\phi_b, u, h') &= \sum_\Omega \lambda(\phi, s, h)\pi(u|s, h)\chi(\phi_b|\phi) \qquad (7)
\end{aligned}$$

where $\Omega(\phi, s, h) = \{(\phi, s, h) : h' = h'(\phi, s, h)\}$.

2.2 Properties of Utility Function, Value Function and Decision Rules

For utility function, this paper assumes $u(\cdot, \cdot)$ is twice differentiable and continuous, $u_c > 0$, $u_h > 0$, $u_{cc} < 0$, $u_{hh} < 0$, $u(0, \cdot) = \infty$, $u(\cdot, 0) = \infty$, $u(\infty, \cdot) = 0$ and $u(\cdot, \infty) = 0$. Based on these properties of the utility function, certain properties for value function and decision rules could be analyzed. A two period version of this model is used to do the analysis in order to simplify the problem.

The value function in period 1 is defined by

$$V(\phi_1, s_1, h_1) = \max_{m_1} u(c_1(\phi_1, s_1, h_1), h_1)$$

subject to

$$c_1(\phi_1, s_1, h_1) = s_1 y - m_1(\phi_1, s_1, h_1) \qquad (8)$$

Proposition 1 In a two period version of this model, it is optimal for the agents to invest nothing on their medical care in period 1. V_1 is continuous, increasing and concave in both s_1 and h_1. ϕ_1 doesn't affect V_1.

Proof:

1. Take the first order condition of V_1 with respect to m_1. This gives us $-u_c \leq 0$, $m_1 \geq 0$ and $-u_c m_1 = 0$ which implies $m_1^* = 0$.

2. $\frac{\partial V_1}{\partial s_1} = u_c y > 0$ and $\frac{\partial^2 V_1}{\partial s_1^2} = u_{cc} y^2 < 0$.

3. $\frac{\partial V_1}{\partial h_1} = u_h > 0$ and $\frac{\partial^2 V_1}{\partial h_1^2} = u_{hh} < 0$.

4. ϕ_1 affects V_1 through m_1. $m_1^* = 0$ implies ϕ_1 doesn't affect V_1.

The value function in period 0 is defined by

$$V(\phi_0, s_0, h_0) = \max_{m_0} u(c_0(\phi_0, s_0, h_0), h_0) + \beta \sum_{\phi_1} \chi(\phi_1|\phi_0) \sum_{s_1} \pi(s_1|s_0, h_1) V(\phi_1, s_1, h_1)$$

subject to

$$c_0(\phi_0, s_0, h_0) = s_0 y - m_0(\phi_0, s_0, h_0)$$
$$h_1 = (1-\delta) h_0 \phi_0 + a m_0(\phi_0, s_0, h_0)^b \tag{9}$$

Proposition 2 In period 0, the medical care investment is not monotone in either health stock or employment state. Its value could be increasing or decreasing depending on the starting health stock and certain parameter values. However, V_0 is continuous and increasing in s_0, ϕ_0 and h_0.

Proof: The proof can be found in Appendix 2.

3 Parameterizations

3.1 Approach

The structural model parameters are estimated using the method of indirect inference. For arbitrary values of the vector of parameters Θ, the dynamic programming problem is solved and policy functions are generated. Using these policy functions, the decision rule is simulated, given arbitrary initial conditions, to create a simulated version of the data to match. One then chooses a descriptive statistical model that provides a rich description of the patterns of covariation in the data. Such a descriptive model can be estimated on both the simulated data from the structural model, and on the actual observed data. This then gives us two sets of coefficients to match, $\Psi^s(\Theta)$ and Ψ^d.

The estimate $\hat{\Theta}$ is pinned down by minimizing the weighted distance between the actual and simulated coefficients from the descriptive models. Formally, it solves

$$\pounds(\Theta) = \min_{\Theta} [\Psi^d - \Psi^s(\Theta)]' W [\Psi^d - \Psi^s(\Theta)] \tag{10}$$

where W is a weighting matrix. The method of indirect inference will generate a consistent estimate of θ. The weighting matrix,W, is based on the variances of the coefficients estimated from the MEPS.[4] Assuming that the covariance between coefficients is zero, the weighting matrix is constructed as the inverse of a matrix in which the variances of the coefficients are on the diagonal and all off-diagonal elements are zero.

Since the $\Psi^s(\Theta)$ function is not analytically tractable, the minimization is performed using numerical techniques. A simulated annealing algorithm is used to perform the optimization in order to obtain the global minimum in parameter space no matter what the starting values are.

3.2 Data and Descriptive Model

3.2.1 Data

The data used for this paper come from the Household Component of the Medical Expenditure Panel Survey. The MEPS HC is a nationally representative survey of the U.S. civilian noninstitutionalized population, collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment. The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a -year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for two calendar years are collected from each household. This series of data collection rounds is launched

[4] The respective variances of the five MEPS coefficients displayed in Section 3.2.2 are (0.00016865, 0.0022495, 0.010501, 0.000015119, 0.00016078).

each subsequent year on a new sample of households to provide overlapping panels of survey data and, when combined with other ongoing panels, will provide continuous and current estimates of health care expenditures. MEPS HC panel 6 covers two years' data (2001 and 2002) for 21,959 individuals. Since this paper's main goal is to explore the relationships for the working age individuals, the sample of individuals whose ages are either below 18 or over 65 is dropped. Students are out of sample too. After cleaning the sample with the missing information, 8896 data points are left.

3.2.2 Descriptive Model

The descriptive model consists two linear equations which are extensively estimated in the literature of health economics.[5] These equations are health equation and medical expenditure equation. They take the following forms:

$$\begin{aligned} h_{t+1} &= \alpha_1 m_t + \alpha_2 h_t + \alpha_3 h_t m_t + X_1 \alpha_4 + \varepsilon_1 \\ m_t &= \gamma_1 h_t + \gamma_2 s_t + X_2 \gamma_3 + \varepsilon_3 \end{aligned} \tag{11}$$

where h is individual's health, s is individual's employment status and m is individual's out-of-pocket medical expenditure.[6] X_1 and X_2 are control variables in these equations. Since these control variables are not modeled in the structural model, it is assumed that the agents in the structural model are homogenous in them. The MEPS coefficients Ψ^d, which are going be matched by the simulated coefficients $\Psi^s(\Theta)$, are $\{\alpha_1\ \alpha_2\ \alpha_3\ \gamma_1\ \gamma_2\}$. Please refer Table 1 and Table 2 for the definitions and the summary statistics of all the variables in these equations.

In the health equation, medical expenditure is expected to have a positive effect on health but the coefficient by OLS regression is usually negative. Researchers try to reverse the sign

[5] Currie and Madrian (1999) has a detailed review of them. Stratmann (1999) estimates the effect of doctor visits on work day loss using the types of health insurance as instruments for doctor visits.

[6] Here, I use out-of-pocket medical expenditure to match the structural model medical expenditure since in the structural model, there are no any forms of medical subsidies.

by using different kinds of instruments. Last period health is expected to have a positive effect on this period health.

In the medical expenditure equation, conditional on the same employed status, healthier individuals are expected to spend less on their medical bill and therefore, the coefficients is expected to be negative. Conditional on the same health level, the employed individuals are expected to spend more on their medical care and therefore the coefficient is expected to be positive but in most studies, the coefficient is significantly negative.

The coefficients on the control variables in the two equations are exogenous to the respective equation. They all have their expected signs.

These two equations are estimated separately by OLS regression. The results are reported in Table 3. According to the OLS results,

$$\Psi^d = \{-.031616\ 0.545791\ 0.045417\ -1.48747\ -.060408\}.$$ [7]

[7] In the actual computation, in order for the two sets of coefficients match each other, the variables must be level free. As a result, we use the share of medical expenditure to income instead of the medical expenditure itself in the computation regression. Please refer to Appendix 2 for the regression results. The MEPS coefficients used in the compution are $\Psi^d = \{-.29389\ 0.556889\ 0.392657\ -.11865\ -.052437\}$. In order to use the medical expenditure share, the individuals without any income or with the medical expenditure share greater than 1 are out of sample. This leaves us with 8168 data points. The expenditure regression and the share regression uses the same sample.

Table 1. Variable definitions:

Dependent variables	
$health_t$	physical component summary index at the end of period t, 0 is the lowest heath level, 100 is the highest
$medical_t$	total amount of out-of-pocket medical expenditure
Independent Variables appearing in all the equations	
age	Individual's actual age
sex	1 if male, 0 otherwise
race	1 if white, 0 otherwise
mar	1 if married, 0 otherwise
edu	1 if with a college degree, 0 otherwise
Additional independent variables appearing in the health equation	
smoke	1 if smoke, 0 otherwise
phyact	1 if currently spends half hour or more on moderate to vigorous physical activities at least three times a week, 0 otherwise
$health_{t-1}$	physical component summary index at the end of period t-1, 0 is the lowest heath level, 100 is the highest
Additional independent variables appearing in the medical expenditure equation	
$employed_t$	employment status at period t: 1 if in labor force, 0 otherwise
inscov	whether the individual has health insurance coverage: 1 if yes, 0 otherwise
spousein	spouse' actual income: if no spouse then spousein=0

Table 2. Descriptive statistics of the sample:

Variables in the Model	Sample Mean	Standard dev.	Minimum	Maximum
Dependent variables				
$health_t$	50.07	9.39	11.73	67.24
$medical_t$	507	1130	0	37128
Independent Variables appearing in all the equations				
age	42.27	11.95	19	65
sex	0.46	0.50	0	1
race	0.81	0.39	0	1
mar	0.63	0.48	0	1
edu	0.23	0.42	0	1
Additional independent variables appearing in the health equation				
smoke	0.23	0.42	0	1
phyact	0.55	0.50	0	1
$health_{t-1}$	50.11	9.41	13.21	67.13
Additional independent variables appearing in the medical expenditure equation				
$employed_t$	0.76	0.43	0	1
inscov	0.84	0.37	0	1
spousein	20372	28647	0	280777

Table 3. Estimation results of OLS estimation:

Variables in The Equations	Coefficients of The Variables	Standard Error
Health equation		
$cons$	0.253715	0.006395
$medical_t/1000$	-.031616	0.002966
$age/100$	-.084297	0.006780
sex	0.004634	0.001514
$race$	0.003534*	0.001933
mar	0.004828	0.001599
edu	0.012747	0.001802
$smoke$	-.007739	0.0018
$phyact$	0.010026	0.001521
$health_{t-1}/100$	0.545791	0.009914
$health_{t-1}/100 medical_t/1000$	0.045417	0.006072
R^2	0.462259	

*the coefficient is not significant at 5 percent significant level.

Table 3. Estimation results of OLS estimation(Continued):

Variables in The Equations	Coefficients of The Variables	Standard Error
Medical expenditure equation		
$cons$	0.558884	0.08524
$employed_t$	-.060408	0.029215
$health_t/100$	-1.48747	0.123018
$age/100$	1.5514	0.093923
sex	-.167855	0.021297
$race$	0.170828	0.026855
mar	-.096417	0.02625
edu	0.168123	0.025532
$inscov$	-.007437*	0.029837
$spousein$	0.001263	0.000471
R^2	0.092041	
$No\ of\ observations$	8168**	

**See footnote 8.

3.3 Parameters in the Structural Model

- The utility function used in the computation has the following form:

$$U(c,h) = \frac{(c^{1-\sigma}h^{\sigma})^{1-\rho} - 1}{1-\rho} \tag{12}$$

  $0 \leq \sigma < 1$ and $\rho > 0$ guarantee that $u_c > 0$, $u_h > 0$, $u_{cc} < 0$ and $u_{hh} < 0$ but the sign of u_{ch} depends on whether ρ is greater or less than 1.

- The time period in the model is one year. The health stock is set between 0 and 1 to match the data. The output is normalize to 1. β is set to 0.95. δ is the health depreciation rate in the health evolution equation. δ is set to 0.01 based on the fact that human beings can easily live up to 100 years if without any accident or illness.

- This paper takes policy parameters θ and τ as exogenous. θ is set to 0.25 which the U.I. replacement ratio in the United States and τ is set to 0.15 which is the average tax rate in the United States.[8]

- The parameters need to be estimated by Indirect Inference are $\Theta = \{\sigma\ \rho\ a\ b\ z\}$. σ and ρ are preference parameters. a and b are parameters in the health stock evolution equation. z is the health shock parameter. Even though the result of indirect inference doesn't depend on the initial values of the parameters, this paper still search for the most reasonable values possible for the starting values to reduce the computation time.

  - σ is set to 0.15 arbitrarily and degree of risk aversion ρ is set to 2.5 following Mehra and Prescott(1985).
  - a is set to 0.5 from the coefficient on medical expenditure in health equation of the descriptive model and b is set to 0.99 arbitrarily.
  - In this paper, $\phi_g = 1 + z$, $\phi_g = 1 - z$ and z is set to 0.2 arbitrarily.

[8] In the actual computation, I use only partial equilibrium instead of the general equilibrium in the model.

4 Results

4.1 Estimation Results

The estimation procedure described in the previous section gives us the following results of the structural parameters and the simulated coefficients:

Table 6a. Structural Parameter Estimates:

	σ	ρ	a	b	$z*$
Initial Value	0.15	2.5	0.5	1	0.2
Estimated Value	0.1416	3.5346	0.789	0.7944	0.201

*$\sigma \in [0,1]$, $\rho > 0$, $a \in [0,1]$, $b > 0$, $z \in [0,1]$.

Table 6b. Data and Simulated Coefficients:

	α_1	α_2	α_3	γ_1	γ_2	$\pounds(\Theta)$
MEPS	-.29389	0.556889	0.392657	-.11865	-.052437	
Simulated	-.30829	0.53033	0.33786	-.24121	-.07161	122

From Table 6b, we can see that the simulated coefficients match the MEPS coefficients quite well. Interestingly, the simulated α_1, the coefficient of medical expenditure in the health equation, is negative even though in the health evolution equation of the structural model, medical expenditure only impacts the health in the positive way because parameters a and b are both positive. This is probably caused by the endogeneity problem in the reduced form health production function regression. Something in the error term might be correlated with the medical expenditure in the reduced form health production function such as health depreciation rate or individuals' initial health level. The fact reflects that reduced form regression results sometimes give us the wrong information without knowing the structural model behind it because of some correlation issues.

From a statistical perspective, the model is rejected since the reported value of $\pounds(\Theta)$ is still high compare to the cut off value. However, in this setting, this reflects the fact that the coefficients are calculated from a very large panel data set, implying very small standard deviations of the coefficients (and a very large W). Given how precisely the micro coefficients are estimated from the actual data, virtually any model would be formally rejected with even very modest deviations of the simulated coefficients from the actual coefficients. As we have emphasized above, the fit of the model in the last line of Table 6b is actually quite good in terms of matching the data coefficients on both a qualitative and quantitative basis.

5 Conclusion

This paper is motivated by the "medical expenditure puzzle". Instead of trying to correct the wrong sign in the reduced form health production function, this paper studies individuals' choices of medical expenditure and how their medical expenditures affect their health and in turn their job opportunities and qualities of life in a dynamic general equilibrium model. Through this study, we can see clearly how the medical expenditure impacts individuals' health and what factors determine how much the individuals are going to spend on their medical care.

The structural parameters are estimated by the method of Indirect Inference which minimizes the distance function of the MEPS coefficients and the simulated coefficients. The simulated coefficient of medical expenditure in the health equation is negative even though in the health evolution equation of the structural model, medical expenditure only impacts the health in the positive way. The fact reflects that reduced form regression results sometimes give us the wrong information without knowing the structural model behind it.

This paper concentrates on understanding the role of medical expenditure in individuals' lives. There are not many government policies included in the model other than the tax and unemployment insurance. But with the estimated structural parameters, it will be easy to accommodate the health insurance and social insurance policies into the model. This model provides room for government policies.

References

[1] Abdulkadiroglu, A., B. Kuruscu and A. Sahin, "Unemployment Insurance and the Role of Self-Insurance," *Review of Economic Dynamics*, Vol. 5, July 2002, 681-703.

[2] Adda, J. and R. Cooper, *Dynamic Economics: Quantitative Methods and Applications*, The MIT Press, Cambridge, Massachusetts, 2003.

[3] Cai, L. and Kalb, G (2004). "Health Status and Labour Force Participation: Evidence from the HILDA Data." Melbourne Institute Working Paper No. 4/04.

[4] Cropper, M. L., "Health, Investment in Health, and Occupational Choice," *The Journal of Political Economy*, Vol.85, No. 6 (Dec., 1977), 1273-1294.

[5] Currie, J. and Madrian, B. C. (1999). Health, Health Insurance and the Labor Market. *Handbook of Labor Economics.* Ashenfelter, O. and Card, D., Elsevier Science B.V. 3: 3310-3415.

[6] Erbsland, M., W. Ried and V. Ulrich, "Health, Health Care, and the Environment. Econometric Evidence from German Micro Data." *Health Economics*, Vol. 4, 1995, 169-182.

[7] Gan, L. and G. Gong, "Motality Risk and Educational Attainment of Black and White Men." *NBER working paper* No. 10381.

[8] Gourieroux, C. and A. Monfort, *Simulation-Based Econometric Methods*, Oxford University Press Inc., New York, 2002.

[9] Grossman, M., "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, Vol.80, No.2, 1972, 223-255.

[10] Grossman, M. and T. J. Joyce, "Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City," *Journal of Political Economy*, Vol. 98, No.5, Part 1, Oct. 1990, 983-1007.

[11] Grossman, M., *The Demand for Health: A Theoretical and Empirical Investigation*, Columbia University Press for the National Bureau of Economics Research, New York, 1972.

[12] Grossman, M.. The Human Capital Model of The Demand for Health (1999), *NBER working parper* No. 7078.

[13] Hansen, Gary D., and Imrohoroglu, Ayse. "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard." *Journal of Political Economy*, Vol.100, 1992, 118-142.

[14] Keane, M. and A. A. Smith, Jr., "Generalized Indirect Inference for Discrete Choice Models." *manuscript* (Yale University).

[15] Kydland, Finn E., and Prescott, Edward C., "Time to Build and Aggregate Fluctuations." *Econometrica* 50 (1982), 1345-1370.

[16] Rosenzweig, M. R. and K. L. Wolpin (Dec., 2000). "Natural "Natural Experiments" in Economics." *Journal of Economic Literature*, Vol. 38, No. 4, 827-874.

[17] Rosenzweig, M. R. and T. P. Schultz, "Estimating a Household Production: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight," *Journal of Political Economy*, Vol. 91, No.5, Oct. 1983, 723-746.

[18] Stern, S. (1989). "Measuring the Effect of Disability on Labour Force Participation." *The Journal of Human Resources* 24(3): 361-395.

[19] Stratmann, T. (1999). "What Do Medical Services Buy? Effects of Doctor Visits on Work Day Loss." *Eastern Economic Journal* 25(1).

[20] Wagstaff, A., "The Demand for Health: Some New Empirical Evidence", *Journal of Health Economics*, Vol. 5, 1986, 195-233.

[21] Wagstaff, A., "The Demand for Health: An Empirical Reformulation of the Grossman Model", *Health Economics*, Vol. 2, 1993, 189-198.

Appendix 1

The value function in period 0:

$$
\begin{aligned}
V_0(\phi_0, s_0, h_0) &= \max_{m_0} u(s_0 y - m_0, h_0) + \beta \sum_{\phi_1} \chi(\phi_1|\phi_0) \sum_{s_1} \pi(s_1|s_0, h_1) u(s_1 y, (1-\delta) h_0 \phi_0 + a m_0^b) \\
&= \max_{m_0} u(s_0 y - m_0, h_0) + \beta \sum_{\phi_1} \chi(\phi_1|\phi_0) \{\pi(e|s_0, h_1) u(y(1-\tau), (1-\delta) h_0 \phi_0 + a m_0^b) \\
&\quad + \pi(u|s_0, h_1) u(y\theta, (1-\delta) h_0 \phi_0 + a m_0^b)\}
\end{aligned}
$$

1. Take the first order condition of V_0 with respect m_0, and let $x = \alpha_1 h_1 + \alpha_2 s_0 + \alpha_3$. This gives us the implicit function of the endogenous variables h_0, m_0 and s_0.

$$
\begin{aligned}
F &= -u_c(s_0 y - m_0, h_0) + \beta \sum_{\phi_1} \chi(\phi_1|\phi_0) \{ \frac{\alpha_1 e^x}{(1+e^x)^2} a b m_0^{b-1} u(y(1-\tau), h_1) + \\
&\quad \pi(e|s_0, h_1) u_h(y(1-\tau), h_1) a b m_0^{b-1} - \frac{\alpha_1 e^x}{(1+e^x)^2} a b m_0^{b-1} u(y\theta, h_1) + \\
&\quad \pi(u|s_0, h_1) u_h(y\theta, h_1) a b m_0^{b-1}\} \\
&= -u_c(s_0 y - m_0, h_0) + \beta \sum_{\phi_1} \chi(\phi_1|\phi_0) \{ \frac{\alpha_1 e^x}{(1+e^x)^2} [u(y(1-\tau), h_1) - u(y\theta, h_1)] \\
&\quad + \pi(e|s_0, h_1) u_h(y(1-\tau), h_1) + \pi(u|s_0, h_1) u_h(y\theta, h_1)\} a b m_0^{b-1} = 0
\end{aligned}
$$

Take the first order condition of F with respect to h_0, m_0 and s_0 respectively first and then the signs of $\frac{\partial m_0}{\partial h_0}$ and $\frac{\partial m_0}{\partial s_0}$ can be determined:

$$
\begin{aligned}
\frac{\partial F}{\partial h_0} &= -u_{ch}(s_0 y - m_0, h_0) + \beta \sum_{\phi_1} \chi(\phi_1|\phi_0) \{ \frac{\alpha_1^2 e^x (e^{2x} - 1)(1-\delta)\phi_0}{(1+e^x)^4} [u(y(1-\tau), h_1) \\
&\quad - u(y\theta, h_1)] + \frac{\alpha_1 e^x}{(1+e^x)^2} [u_h(y(1-\tau), h_1) - u_h(y\theta, h_1)](1-\delta)\phi_0 + \\
&\quad \frac{\alpha_1 e^x (1-\delta)\phi_0}{(1+e^x)^2} [u_h(y(1-\tau), h_1) - u_h(y\theta, h_1)] + \\
&\quad \pi(e|s_0, h_1) u_{hh}(y(1-\tau), h_1)(1-\delta)\phi_0 + \pi(u|s_0, h_1) u_{hh}(y\theta, h_1)(1-\delta)\phi_0\} a b m_0^{b-1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial F}{\partial m_0} &= u_{cc}(s_0 y - m_0, h_0) + \beta \sum_{\phi_1} \chi(\phi_1|\phi_0) [\{ \frac{\alpha_1 e^x}{(1+e^x)^2} [u(y(1-\tau), h_1) - u(y\theta, h_1)] \\
&\quad + \pi(e|s_0, h_1) u_h(y(1-\tau), h_1) + \pi(u|s_0, h_1) u_h(y\theta, h_1)\} a b m_0^{b-2}(b-1) + \\
&\quad \{ \frac{\alpha_1^2 e^x (e^{2x} - 1) a b m_0^{b-1}}{(1+e^x)^4} [u(y(1-\tau), h_1) - u(y\theta, h_1)] + \frac{\alpha_1 e^x}{(1+e^x)^2} [u_h(y(1-\tau), h_1) \\
&\quad - u_h(y\theta, h_1)] a b m_0^{b-1} + \frac{\alpha_1 e^x a b m_0^{b-1}}{(1+e^x)^2} [u_h(y(1-\tau), h_1) - u_h(y\theta, h_1)] + \\
&\quad \pi(e|s_0, h_1) u_{hh}(y(1-\tau), h_1) a b m_0^{b-1} + \pi(u|s_0, h_1) u_{hh}(y\theta, h_1) a b m_0^{b-1}\} a b m_0^{b-1}]
\end{aligned}
$$

$$\frac{\partial F}{\partial s_0} = -u_{cc}(s_0 y - m_0, h_0)y + \beta\sum_{\phi_1}\chi(\phi_1|\phi_0)\{\frac{\alpha_1\alpha_2 e^x(e^{2x}-1)}{(1+e^x)^4}[u(y(1-\tau),h_1) - u(y\theta,h_1)] + \frac{\alpha_2 e^x}{(1+e^x)^2}[u_h(y(1-\tau),h_1) - u_h(y\theta,h_1)]\}abm_0^{b-1}$$

$$\frac{\partial m_0}{\partial h_0} = -\frac{\partial F/\partial h_0}{\partial F/\partial m_0}???0$$

$$\frac{\partial m_0}{\partial s_0} = -\frac{\partial F/\partial s_0}{\partial F/\partial m_0}???0$$

Since the signs of the first order condition of F with respect to h_0, m_0 and s_0 respectively are all indeterminate depending on the parameter values, the signs of $\frac{\partial m_0}{\partial h_0}$ and $\frac{\partial m_0}{\partial s_0}$ are ambiguous too.

2.

$$\begin{aligned}\frac{\partial V_0}{\partial s_0} &= u_c(y - \frac{\partial m_0^*}{\partial s_0}) + \beta\sum_{\phi_1}\chi(\phi_1|\phi_0)\{\frac{e^x}{(1+e^x)^2}(\alpha_1 abm_0^{b-1}\frac{\partial m_0^*}{\partial s_0} + \alpha_2)[u(y(1-\tau),h_1) - \\ &\quad u(y\theta,h_1)] + \pi(e|s_0,h_1)u_h(y(1-\tau),h_1)abm_0^{b-1}\frac{\partial m_0^*}{\partial s_0} + \pi(u|s_0,h_1)u_h(y\theta,h_1)abm_0^{b-1}\frac{\partial m_0^*}{\partial s_0}\} \\ &= u_c y + \beta\sum_{\phi_1}\chi(\phi_1|\phi_0)\{\frac{\alpha_2 e^x}{(1+e^x)^2}[u(y(1-\tau),h_1) - u(y\theta,h_1)]\} > 0\end{aligned}$$

3.

$$\begin{aligned}\frac{\partial V_0}{\partial \phi_0} &= -u_c\frac{\partial m_0^*}{\partial \phi_0} + \beta\sum_{\phi_1}\chi(\phi_1|\phi_0)\{\frac{\alpha_1 e^x}{(1+e^x)^2}abm_0^{b-1}\frac{\partial m_0^*}{\partial s_0}[u(y(1-\tau),h_1) - u(y\theta,h_1)] \\ &\quad +[\pi(e|s_0,h_1)u_h(y(1-\tau),h_1) + \pi(u|s_0,h_1)u_h(y\theta,h_1)](abm_0^{b-1}\frac{\partial m_0^*}{\partial s_0} + (1-\delta)h_0)\} \\ &= \beta\sum_{\phi_1}\chi(\phi_1|\phi_0)[\pi(e|s_0,h_1)u_h(y(1-\tau),h_1) + \pi(u|s_0,h_1)u_h(y\theta,h_1)](1-\delta)h_0 > 0\end{aligned}$$

4.

$$\begin{aligned}\frac{\partial V_0}{\partial h_0} &= -u_c\frac{\partial m_0^*}{\partial h_0} + u_h + \beta\sum_{\phi_1}\chi(\phi_1|\phi_0)\{\frac{\alpha_1 e^x}{(1+e^x)^2}((1-\delta)\phi_0 + abm_0^{b-1}\frac{\partial m_0^*}{\partial h_0}) \\ &\quad [u(y(1-\tau),h_1) - u(y\theta,h_1)] + [\pi(e|s_0,h_1)u_h(y(1-\tau),h_1) + \\ &\quad \pi(u|s_0,h_1)u_h(y\theta,h_1)](abm_0^{b-1}\frac{\partial m_0^*}{\partial h_0} + (1-\delta)h_0)\} \\ &= u_h + \beta\sum_{\phi_1}\chi(\phi_1|\phi_0)\{\frac{\alpha_1 e^x}{(1+e^x)^2}[u(y(1-\tau),h_1) - u(y\theta,h_1)] + \\ &\quad \pi(e|s_0,h_1)u_h(y(1-\tau),h_1) + \pi(u|s_0,h_1)u_h(y\theta,h_1)\}(1-\delta)\phi_0 > 0\end{aligned}$$

Appendix 2

Table 4. Estimation results of OLS estimation (medical expenditure share to income as medical expenditure):

Variables in The Equations	Coefficients of The Variables	Standard Error
Health equation		
$cons$	0.250286	0.006272
$medical_t/income_t$	-.29389	0.035197
$age/100$	-.090106	0.006713
sex	0.003976	0.001525
$race$	0.002375*	0.001931
mar	0.005312	0.001601
edu	0.010824	0.001799
$smoke$	-.006402	0.001802
$phyact$	0.010071	0.001523
$health_{t-1}/100$	0.556889	0.009607
$(health_{t-1}/100)(medical_t/income_t)$	0.392657	0.078247
R^2	0.46066	

*the coefficient is not significant at 5 percent significant level.

Table 4. Estimation results of OLS estimation (medical expenditure share to income as medical expenditure)(continued):

Variables in The Equations	Coefficients of The Variables	Standard Error
Medical expenditure equation		
$cons$	0.120045	0.006965
$employed_t$	-.052437	0.002387
$health_t/100$	-.11865	0.010052
$age/100$	0.057322	0.007674
sex	-.015325	0.00174
$race$	0.006225	0.002194
mar	-.003984*	0.002145
edu	-.002006*	0.002086
$inscov$	-.010903	0.002438
$spousein$	0.000158	0.000038
R^2	0.136309	
$No\ of\ observations$	8168**	

**See footnote 8.