Is There "Too Much" Inequality in Health Spending Across Income Groups?*

Laurence Ales
Carnegie Mellon University
ales@cmu.edu

Roozbeh Hosseini
Arizona State University
rhossein@asu.edu

Larry E. Jones
University of Minnesota
lej@umn.edu

March 14, 2012

Abstract

In this paper we study the efficient allocation of health resources across individuals. We focus on the relation between health resources and income (taken as a proxy for productivity). In particular we determine the efficient level of the health care social safety net for the indigent. We assume that individuals have different life cycle profiles of productivity. Health care increases survival probability. We adopt the classical approach of welfare economics by considering how a central planner with an egalitarian (ex-ante) perspective would allocate resources. We show that, under the efficient allocation, health care spending increases with labor productivity, but only during the working years. Post retirement, everyone would get the same health care. Quantitatively, we find that the amount of inequality across the income distribution in the data is larger that what would be justified solely on the basis of production efficiency, but not drastically so. As a rough summary, in U.S. data top to bottom spending ratios are about 1.5 for most of the life cycle. Efficiency implies a decline from about 2 (at age 25) to 1 at retirement. We find larger inefficiencies in the lower part of the income distribution and in post retirement ages.

*We would like to thank George-Levi Gayle, Martin Gaynor, Fatih Guvenen, Berthold Herrendorf, Chad Jones, Ellen McGrattan, Serdar Ozkan, Juan Pablo Nicolini, Chris Phelan, Jim Poterba, Jose-Victor Rios-Rull, Seth Richards-Shubik, Todd Schoellman, Sam Schulhofer-Wohl, Erick Sager, Fallaw Sowell, Gustavo Ventura, Motohiro Yogo and seminar participants at the SED meetings in Montreal, Midwest Macro Meetings and the Federal Reserve Bank of Minneapolis for comments and many useful conversations. We thank Erick Sager for his excellent research assistance. Financial support from NSF grant SES-0962432 is gratefully acknowledged.

1 Introduction

A key policy debate in the U.S. in the last few decades has been over the redesign of the health care component of the social safety net. In the U.S., inequality in health expenditures is large.[1] There is also a small, but significant, positive correlation between income and health expenditures, leading to increased inequality over the course of the working life. In this paper, we adopt a normative approach to address a related question: how would a social planner seeking to maximize an ex-ante egalitarian social welfare function allocate health expenditures as a function of income?

To answer this questions we follow the recent approach of Hall and Jones (2007) and assume that the sole impact of health care is in terms of increased survival. We assume that individuals have different life cycle profiles of productivity. We model these profiles as deterministic conditional on an initial productivity level and assume that they follow the standard hump-shaped pattern. We adopt the classical approach of welfare economics and consider how a central planner with an egalitarian (ex-ante) perspective would allocate resources between leisure and work, consumption and health care.

We characterize the full information planner's optimum (the efficient allocation) and find two main theoretical results. The first is that the planner would devote more resources towards providing health care for individuals with higher productivities. The idea is simple; the planner allocates higher levels of health care to the highly productive since the opportunity cost of an hour of work from those types is higher to society. Thus, there is a pure (productive) efficiency reason for having different levels of health care spending across the productivity distribution. The second result is that, under the efficient allocation, retired workers receive the same amount of health care independent of their productivity. Since all workers contribute the same output at retirement (namely each worker contributes zero), there is no efficiency gain from having any difference in health care expenditures between them. In this sense, the outcome of this model is similar to the uniform provision of health care post retirement as in Medicare.

Our goal is to compare the efficient allocation with what we observe in the data. Our data source is the Medical Expenditure Panel Survey (MEPS) administered by the U.S. department of Health & Human Services. This dataset collects information on total health

[1] In terms of variance of logarithm, inequality in health expenditure is roughly twice that of income. Inequality for individuals with positive health expenditure is smaller. It is roughly the same as inequality in income.

expenditures at an individual level (these expenditures include those directly by individuals, insurance companies and government programs). In order to calibrate the model, we back out the age specific distribution of output from the data. We go on to estimate age specific parameters for the survival production function using spending levels from MEPS and age specific survival rates from the CDC. Preference parameters are chosen by matching average hours worked and the value of a statistical life (similar to Hall and Jones (2007)).

Quantitatively the model delivers two key findings. First, we find that the efficient health care safety net is fairly generous – under the efficient allocation, the share of healthcare spending for the bottom quartile of the income distribution is about 18% of total healthcare spending for ages 25 to 35 and it increases gradually up to 25% at retirement. As expected, resources at retirement are equally divided among income quartiles. Second, spending is higher on more productive 'types,' but, the efficient level of inequality across different productivity types is fairly small — the ratio of spending on top to bottom income quartiles during the working life ranges between 2 (at age 20) decreasing to 1.0 (at age 65).

We next compare the implications of the efficient allocation to what we observe in U.S. survey data. We begin by looking at levels. We find (consistent with Hall and Jones (2007) at the aggregate level) that the efficient level of overall spending is actually quite close to that seen in the data for ages up to around 65. The one exception to this general rule is that spending on younger ages is currently higher than the (ex ante) efficient level. We next turn to differences across individuals. We find that the amount of inequality across the income distribution that is seen in the data is larger than what would be justified solely on the basis of productive efficiency, but not drastically so. As a rough summary, in the data top to bottom spending (by quartiles) ratios is about 1.5 over the life cycle (efficiency would imply a slow decline from about 2 to 1 at retirement). The distribution of health spending across the income distribution in the data and the efficient outcome are also similar, but this becomes less true for higher ages. For example, spending on health care for the poor (the bottom 25%) is about 22% at all ages while model efficient quantities imply spending share of 18% of the total (at ages 25 to 35) and rise to 25% by retirement. In this sense, the distribution of health care spending across incomes is close to efficient.

This last finding depends to some extent on how income is measured in the data – total income vs. wage income. When only wage income is used to rank individuals in the data, the allocation in the data is more unequal than in the efficient allocation. As an example, the efficient allocation calls for the bottom 25% of the distribution at ages 55 to 65 to receive

24% of all spending while in the data, the bottom 25% of the population by wage income only receives 17% of all spending, i.e., considerably less than efficient.

We also compare data quantities to what we call the 'Laissez Faire' allocation, in which each individual splits his own income between consumption and health care over his life cycle. This allocation gives rise to drastically higher levels of inequality than what is seen in the data. In this sense, the current allocation goes a long way in providing the efficient amount of social insurance (relative to each man for himself), but there are still differences between the (ex ante) efficient outcome and what we observe in data, particularly for higher ages at the lower end of the income distribution for people who primarily rely on wages as their source of income.

The model considered delivers stark implications for the efficient level of inequality in health expenditures following retirement. A natural criticism is that society might use inequality in health (and hence in life expectancy) as a form of incentive in the presence of moral hazard problems over the course of the working life. To address this concern we consider an extension of our baseline model adding private information about productivity types. Truthful revelation of types is now induced with compensation in the form of consumption, leisure and health expenditures. We find that our key results still hold under this formulation.

Quantitatively, we can use the model to highlight the effects of one of the most elusive parameters. This parameter can be interpreted as the fixed flow of utility that agents derive when alive. Higher values of flow utility progressively shift priority of the planner from increasing production levels to keeping the highest number of individuals alive. In the presence of convex costs, this implies a reduction of inequality of health expenditures. Our benchmark results are based on values of the flow utility estimated in Hall and Jones (2007). In this paper the authors pin down the flow utility by targeting a value of a statistical life (VSL) of three million dollars. In our model as we move towards higher estimates of (VSL) inequality drastically decreases.[2]

Related papers

The paper closest to ours is Hall and Jones (2007). As in their paper the principle motivation for health care is to reduce mortality risk. Our paper can be seen as an extension of Hall and Jones (2007) to the cross-section. Hall and Jones (2007) focus on a representative agent for-

[2]Currently, the department of transportation (see, http://ostpxweb.dot.gov/policy/reports/080205.htm adopts a VSL of 5.8 million dollars.

mulation to derive normative implications for the overall spending on health care over time comparing this with US aggregate data. We introduce heterogeneity, pursue the implications of the model across the income distribution and compare this with household-level data.

Ozkan (2011) also looks at an environment similar to ours. The focus of the paper however, is in developing a positive description of household health care decisions. Health care expenditure patterns in his model are driven by wealth heterogeneity and moral-hazard-like market inefficiencies. Additional positive papers that study how health care affects lifetime decisions are Yogo (2009) focusing on the elderly portfolio composition, and De Nardi, French, and Jones (2010) looking at the effect of health expenditures on life-cycle asset accumulation.

The question of the efficient allocation of health care has also been addressed by Arrow (1973) and Daniels (1985). The former discuss how the approach of Rawls (1971) extended to health care would be problematic: society would be dealing with a "bottomless pit" in which an ever increasing amount of resources are diverted to the most ill individual. The latter moves towards a characterization of a more desirable allocation of health-care. This is done by proposing an ad-hoc remedy to the Rawlsian-like approach: minimal levels of health expenditure.

Our paper is also related to a large literature that documents inequality in health outcomes and health status and it's relation to income. Refer to Deaton (1999), Deaton and Paxson (1998, 1999), Wagstaff (2002), Skinner and Zhou (2004).

This paper is organized as follows: in section 2, we construct and characterize a normative model of health expenditures. In section 3 we calibrate the model and assess its quantitative implications. In section 4 we review the data and compare the health spending allocations in the data with the efficient allocations that our model generates. In section 5 we perform some robustness calculations on our results. Finally, we conclude in section 6. In the appendix we review additional information on our datasets and include proofs of theoretical results.

2 A Dynamic Model of Health and Mortality

We begin by studying–at a qualitative level–the relation between income and health expenditures implied by an ex ante efficient allocation with full information. In the model considered the sole role played by health care spending is to decrease mortality risk. In the sections 3 and 4 below, we asses the quantitative magnitude of health inequality in the U.S. relative to this benchmark.

2.1 Environment

The environment we study is similar to Hall and Jones (2007). We depart from their setup in two basic ways. First, we assume that individuals are heterogeneous in their labor productivities. Second, we allow labor supply to be endogenous. We also limit our attention to study the problem of one cohort.

The economy is populated by a continuum of finitely lived workers. Workers are heterogeneous with respect to their life-cycle productivity profile. Productivity types will be labeled according to $\theta \in \Theta = \{\theta_1, ..., \theta_I\}$ of relative size $\pi(\theta_i)$. These 'types' are drawn at date zero when workers are born and they are permanent– they determine a conditionally non-stochastic life cycle profile of labor productivities. Each workers faces the probability of death at the end of each age a we assume that every worker cannot survive for more than $A + 1$ periods. Following Hall and Jones (2007), we assume that an agent's survival rate depends on health spending which we allow to be type and age specific and we denote by $h_a(\theta)$. We assume $h_a(\theta)$ is the total health spending in terms of consumption good.[3] The (conditional) survival rate of an individual who has health spending h is given by $P_a(h)$, an age dependent strictly increasing and strictly concave function. Let $N_a(\theta)$ denote the unconditional survival probability to age a of an individual worker of type θ. By the law of large numbers, the fraction of type θ workers of age a who survive to age $a + 1$ can be defined recursively by:

$$N_{a+1}(\theta) = P_a(h_a(\theta))N_a(\theta), \quad \forall\, \theta \in \Theta, \quad 0 < a \leq A, \tag{1}$$

$$N_0(\theta) = 1, \quad \forall\, \theta \in \Theta. \tag{2}$$

The economy lasts for A periods at which time, all workers die. Workers of type θ who are alive at age a receive $c_a(\theta)$ units of consumption and work for $l_a(\theta)$ hours.

[3]Therefore, we assume without loss of generality that relative price of health spending in terms of consumption good is one.

Definition 1 *An allocation-for workers who are alive- is the collection of maps*

$$\begin{aligned} c_a &: \Theta \to \mathbb{R}_+, \quad 0 \le a \le A \\ l_a &: \Theta \to [0,1], \quad 0 \le a \le A \\ h_a &: \Theta \to \mathbb{R}_+, \quad 0 \le a \le A \\ N_a &: \Theta \to [0,1], \quad 0 \le a \le A \end{aligned}$$

Given an allocation $\{c_a(\theta), l_a(\theta), h_a(\theta), N_a(\theta)\}_{a=0}^{A}$, preferences of worker of type θ are

$$\sum_{a=0}^{A} \beta^a N_a(\theta) u\left(c_a(\theta), l_a(\theta)\right).$$

Without loss of generality we assume Θ is ordered such that $\theta_{i+1} > \theta_i$ for all $i = 1, \ldots, I$ and that productivity is increasing in θ up to a fixed retirement age, a_{ret}.[4] Thus,

$$\begin{aligned} w_a(\theta') &> w_a(\theta) \quad \text{for all } a < a_{ret} \text{ and all } \theta', \theta \text{such that } \theta' > \theta \\ w_a(\theta) &= 0 \qquad\quad \text{for all } a \ge a_{ret}. \end{aligned}$$

A worker of type θ at age a who supplies l hours of work produces $w_a(\theta)l$ units of output. There is a storage technology with fixed rate of return $R = 1/\beta$.[5]

Definition 2 *An allocation* $\{c_a(\theta), l_a(\theta), h_a(\theta), N_a(\theta)\}_{\theta,a}$ *is feasible if*

$$\sum_{\theta \in \Theta} \pi(\theta) \sum_{a=0}^{A} \frac{1}{R^a} N_a(\theta) \left[c_a(\theta) + h_a(\theta) - w_a(\theta) l_a(\theta)\right] \le 0 \tag{3}$$

and (1) *and* (2) *hold.*

2.2 The Full information Ex Ante Efficient Allocation

In this section we derive the properties of ex-ante efficient allocation. To do this we study the problem of a social planner who maximizes the ex ante welfare of individuals subject to a feasibility constraint. In particular, we study how the provision of health care varies across

[4] Here, for simplicity, we assume that the age of retirement, a_{ret}, is fixed and the same for all types. In the quantitative work below, we assume that $a_{ret} = 65$. An interesting extension would be to add endogenous, type specific retirement ages to the model. See Shourideh and Troshkin (2011) for a version of this with private information, but no health spending.

[5] The assumption that $R\beta = 1$, can be relaxed. It simplifies the arguments in this section since, in this case, consumption is constant across age.

productivity types. The planning problem is

$$\max_{\{c_a(\theta),l_a(\theta),h_a(\theta),N_a(\theta)\}_{\theta,a}} \sum_{\theta\in\Theta} \pi(\theta) \sum_{a=0}^{A} \beta^a N_a(\theta) u\left(c_a(\theta), l_a(\theta)\right) \tag{4}$$

subject to

$$\sum_{\theta\in\Theta} \pi(\theta) \sum_{a=0}^{A} \frac{1}{R^a} N_a(\theta) \left[c_a(\theta) + h_a(\theta) - w_a(\theta) l_a(\theta)\right] \leq 0 \tag{5}$$

$$\begin{aligned} N_{a+1}(\theta) &= P_a(h_a(\theta))N_a(\theta), \quad \forall\, \theta \in \Theta, \quad 0 < a \leq A, \\ N_0(\theta) &= 1, \quad \forall\, \theta \in \Theta. \end{aligned}$$

Let $\{c^*_a(\theta), l^*_a(\theta), h^*_a(\theta), N^*_a(\theta)\}_{\theta,a}$ be the solution to the planner's problem. In order to be able to derive analytical results on properties of the efficient allocation we make the following assumptions. First we assume that the per period utility function is always positive.

Assumption 1 $u : \mathbb{R}^2_+ \times [0, 1] \to \mathbb{R}_+$

Second we assume that the per period utility is additively separable in consumption and leisure.

Assumption 2 *We assume the following on u:*

$$u(c, l) = u(c) + v(1 - l)$$

where $u', -u'', v', -v'' > 0$.

Finally

Assumption 3 $P_a(h)$*, is strictly increasing and strictly concave and satisfies the following:*

$$\lim_{h\to 0} P'_a(h) = \infty, \quad \forall\; 0 \leq a \leq A.$$

Assumptions 1, 2 and 3 have strong implications on the efficient allocation as the following lemma states:

Lemma 1 *The efficient allocation features:*

1. $h^*_A(\theta) = 0$ *for all* $\theta \in \Theta$.

2. $c_a^*(\theta) = c^*$ *for all* $\theta \in \Theta$*, and all* $0 \leq a \leq A$*.*

3. $l_a^*(\theta) = 0$ *for all* $\theta \in \Theta$*, and all* $a \geq a_{ret}$*.*

4. $l_a^*(\theta') > l_a^*(\theta)$ *for all* $\theta, \theta' \in \Theta$ *with* $\theta' > \theta$ *and all* $0 \leq a < a_{ret}$*.*

5. $h_a^*(\theta) > 0$ *for all* $\theta \in \Theta$*, and all* $0 \leq a \leq A - 1$

Proof. Statements *1.* and *2.* follow directly from Assumption 2 and the assumption that no agent can survive to age $A+1$. Statement *3.* follows from the assumption that $w_a(\theta) = 0$ for all $a_{ret} \leq a \leq A$ and for all $\theta \in \Theta$. From first order condition for consumption and leisure we have $w_a(\theta)u'(c^*) = v'(1 - l_a^*(\theta))$ for all θ and all $a < a_{ret}$. Which implies *4.*. Finally *5.* follows from Assumption 3. ■

From Lemma 1 we have that

Corollary 1 *For all* $0 \leq a \leq A$ *and for all* $\theta \in \Theta$*,* $N_a(\theta) > 0$*.*

In the first proposition, we show that after retirement all workers must have the same health expenditure. The idea here is simple. From the previous lemma we have that after retirement all workers do not contribute to production and enjoy the same amount of leisure and consumption. Therefore, the utility of each living person is the same at every age following retirement. In addition, the planner values each individual equally, so that there is no gain to the planner from having a higher number of a certain type survive to higher ages. On the other hand, the cost of producing health is convex and its return in terms of increasing survival rate are concave. "This implies that it is (ex ante) efficient for everyone to receive the same health care and survive to higher ages at the same rate." Of course, the exact nature of this result depends on our assumptions that all types of workers retire at exactly the same age. More generally, this would hold for all workers who are retired, but spending would be different at a given age between retired, and not yet retired, workers.

Proposition 1 *Suppose per period utility function satisfies the assumptions 1 and 2. Let* $h_a^*(\theta)$ *be the solution to the planer's problem for age a and type* θ*. Then*

$$h_a^*(\theta) = h_a^*(\theta')$$

for all θ *and* θ' *in* Θ*, and all* $A \geq a \geq a_{ret} - 1$*.*

Proof. In Appendix A.1. ■

We now look at health expenditures before retirement. Our next proposition establishes that more productive workers must have higher health levels at all pre-retirement ages. To show this result, it is convenient to introduce a recursive formulation for the planner's problem. The state variables will be given by: the current age of the cohort a and the fraction of type θ workers who survive to age a: $N_a(\theta)$. Finally we also include in the state vector the total (net) saving in the economy up to age a: k_a. Define the function $V_a(N_a(\theta_1), N_a(\theta_2), ..., N_a(\theta_I); k_a)$ as the solution of the following problem

$$\begin{aligned} V_a(N_a(\theta_1), N_a(\theta_2), ..., N_a(\theta_I); k_a) = \max \sum_{\theta \in \Theta} \pi(\theta) \sum_{a'=a}^{A} \beta^{a'-a} N_{a'}(\theta) \left[u(c_{a'}(\theta)) + v(1 - l_{a'}(\theta)) \right] \\ s.t \qquad \sum_{\theta \in \Theta} \pi(\theta) \sum_{a=0}^{A} \frac{1}{R^a} N_a(\theta) \left[c_a(\theta) + h_a(\theta) - w_a(\theta) l_a(\theta) \right] \leq k_a \\ N_{a+1}(\theta) = P_a(h_a(\theta)) N_a(\theta) \quad \forall \, \theta \in \Theta, \forall \, a \leq A. \end{aligned}$$

It can be shown that the solution to the planner's problem can be characterized as the solution of the following Bellman's equation:

$$\begin{aligned} & V_a(N_a(\theta_1), N_a(\theta_2), ..., N_a(\theta_I); k_a) = \\ & \max \sum_{\theta \in \Theta} \pi(\theta) N_a(\theta) \left[u(c_a(\theta)) + v(1 - l_a(\theta)) \right] + \beta V_{a+1}(N_{a+1}(\theta_1), N_{a+1}(\theta_2), ..., N_{a+1}(\theta_I); k_{a+1}) \\ & s.t. \quad \sum_{\theta \in \Theta} \pi(\theta) N_a(\theta) \left[c_a(\theta) + h_a(\theta) \right] + k_{a+1} \leq R k_a + \sum_{\theta \in \Theta} \pi(\theta) N_a(\theta) w_a(\theta) l_a(\theta) \\ & N_{a+1}(\theta) = P_a(h_a(\theta)) N_a(\theta) \quad \forall \, \theta \in \Theta, \forall \, a \leq A; \qquad k_0 = 0. \end{aligned}$$

During pre-retirement ages, higher productivity workers provide more labor effort and produce more output. As we will show in the next proposition, the planner always prefers to have fewer low productivity workers and higher number of high productivity workers. The reason is that, when there are more high productivity workers, the planner can afford to have everyone work less and enjoy more leisure. The planner manipulates the type-composition of the population through providing different amounts of health care (and hence, survival probabilities) for the various types. Given that the higher number of more productive workers is desired, it is clear that the planner wants them to be healthier so that they survive to higher ages with higher probability. The following proposition formalizes this argument.

Proposition 2 *Suppose per period utility function satisfies assumptions 1 and 2. Then*

1. *For all ages* $a < a_{ret}$ *and for all* $\varepsilon > 0$, $(N_a(\theta_1), N_a(\theta_2), ..., N_a(\theta_I))$,

$$V_a\left(N_a(\theta_1), ..., , N_a(\theta_i) - \varepsilon, ..., N_a(\theta_{i'}) + \frac{\pi(\theta_i)}{\pi(\theta_{i'})}\varepsilon, \ldots, N_a(\theta_I); k_a\right) >$$
$$V_a\left(N_a(\theta_1), ..., , N_a(\theta_i), ..., N_a(\theta_{i'}), \ldots, N_a(\theta_I); k_a\right) \tag{6}$$

for all $\theta_{i'} > \theta_i$.

2. *For all ages* $a < a_{ret} - 1$ *, and all* $\theta_{i'} > \theta_i$, $h_a(\theta_{i'}) > h_a(\theta_i)$.

Proof. In Appendix A.2 ■

So far we have established that the amount of inequality in health spending is not zero in the ex ante efficient allocation. Society must treat more productive workers differently than less productive workers while they are active. We also showed that all workers must receive the same health during their retirement years.[6]

2.3 The Laissez Faire Allocation

The full information, ex ante efficient allocation discussed in the previous section features an extreme level of insurance: each agent receives full insurance against the realization of productivity type θ. In this section, we introduce an alternative benchmark allocation. This benchmark instead, will feature no redistribution across productivity types; each individual must choose his own path for leisure, consumption and health expenditures from his own present value budget constraint. We will call this the "Laissez Faire" allocation. We assume that each type has access to perfect annuity markets.[7] The maximization problem faced by

[6] Previous versions of this paper considered the case where health status appears as a state variable as in Grossman (1972). In this case, the pre-retirement behavior of the efficient allocation is similar to the present: more productive workers receive higher health spending than less productive workers. Upon retirement, the effect is reversed. Less productive workers initially receive a higher level of health spending. This is because at retirement less productive workers are endowed with a lower health stock. In the long run every agent receives the same amount of health expenditures. Details are available upon request.

[7] This implies that that consumption claims paying contingent on survival are priced at their actuarially fair value on a type by type basis.

an individual of type θ is given by:

$$\max_{c_a, l_a, h_a, N_a} \sum_{a=0}^{A} \beta^a N_a u(c_a, l_a) \tag{7}$$

$$s.t. \qquad \sum_{a=0}^{A} \frac{1}{R^a} N_a \left[c_a + h_a - w_a(\theta) l_a\right] \leq 0$$

$$N_{a+1} = P_a(h_a) N_a \; \forall \; 0 < a < A, \quad N_0(\theta) = 1.$$

As with the ex ante efficient allocation it can be shown that $h_a(\cdot)$ increasing in θ.

3 Simulation

So far we have shown that the efficient allocation generates inequality in health expenditures and that health expenditures are positively related to individual's productivity. As we will see below, this is also true in the data. In this section, we determine the size of this dependence by calibrating and simulating the model presented in the previous section. In Section 4, we compare these simulation results to data summaries.

3.1 Calibration

In the simulations a period is set to one year. Mortality is determined endogenously, however we impose that individuals live at most $A = 100$ years. Individuals are assumed to start working at age 20 and retire at age 65. Before computing the allocation, we need to calibrate the following:

1. the survival production functions;

2. the life cycle profiles of wages for each θ;

3. preference parameters.

Our main data source will be the Medical Expenditure Panel Survey – MEPS. This is a rotating panel containing individual level data on hours worked, income, demographic variables, etc., along with extensive information on medical expenditures. MEPS is a unique dataset in that it also provides detailed medical expenditure data. For each individual, expenditures are given by source of spending – out of pocket, private insurer, government program, etc. We will use this information to compute estimates of the survival production function. Refer to section 4.1 for additional information on the MEPS dataset.

3.1.1 Survival Production Functions

Let h_a denote spending level on an individual of age a. Following Hall and Jones (2007), our survival production function is given by:

$$P_a(h_a) = 1 - \frac{1}{x_a} = 1 - \frac{1}{f_a(h_a)}. \tag{8}$$

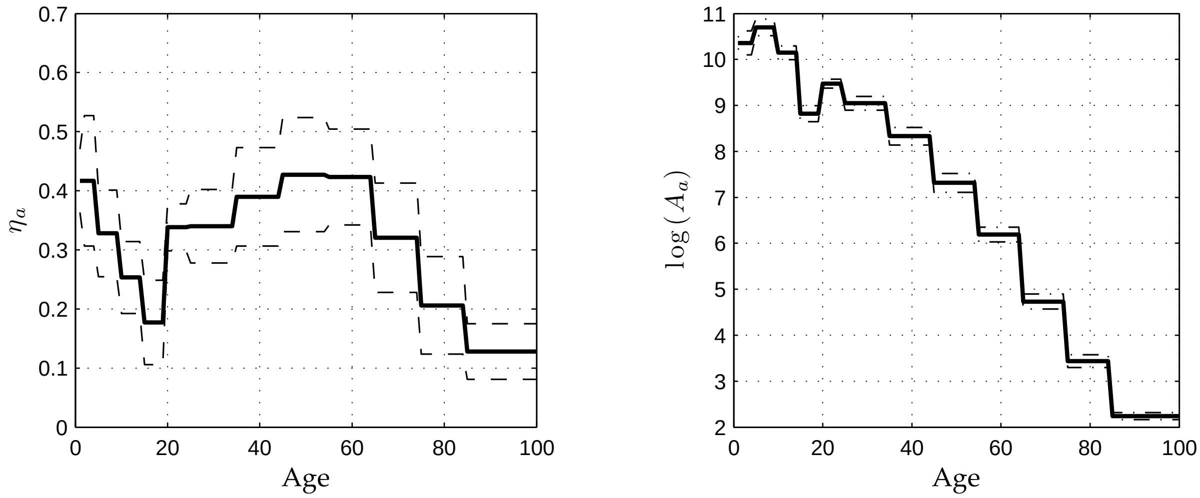
where

$$x_a = f_a(h_a) = A_a h_a^{\eta_a}. \tag{9}$$

Here, x_a is the inverse of mortality rate at age a. The estimates of Hall and Jones (2007) for both A_a and η_a are based on age specific health data (as reported by Meara, White, and Cutler (2004)) which are scaled up so as to match the aggregate values as reported by the BEA. In particular, their data includes includes nursing home expenditures as a part of health spending. Since MEPS does not include nursing home expenditures, by construction it does not match US aggregates (see Appendix C.1 for further discussion). For this reason, we estimate the parameters A_a and η_a using the MEPS data directly (instead of using Hall and Jones (2007) estimates).[8] To do this, we use variation across observable demographic characteristics in both expenditures and survival rates. In particular we define an observation (group) as a unique combination of year/age/gender/race and census region of residence. Mortality for each group and age is taken from the compressed mortality tables at the CDC.[9] We determine total mortality and subtract mortality due to homicides and suicides. Finally we drop groups that contain less than 20 individuals. This amounts to 4261 separate observations. We estimate (9) using a robust regression procedure. Parameters are estimated for age groups of 5 years up to age 20 and 10 years thereafter. Results are in Figure 1(a) and 1(b). Details are in Appendix B.

[8] The issue of how to treat nursing home expenditures is a difficult one. While some of these expenses are clearly related to health care other parts are not. Moreover, these expenditures are particularly significant for median and low income earners and post retirement individuals.

[9] This data is available at: http://wonder.cdc.gov/mortsql.html.



(a) Elasticity of survival with respect to health.

(b) Log productivity of survival production function.

FIGURE 1: Estimation of survival production function.

Our estimates of elasticity and TFP are in line with the estimates of Hall and Jones (2007). The key difference is the higher estimated elasticities for older individuals. The main reason for this is the lack of nursing home expenditure in MEPS. Hence, using only the expenditures included in MEPS to estimate the production functions makes it look like health spending is more effective at older ages– a smaller level of total spending is included but with the same overall mortality rates.

3.1.2 The Life Cycle Profiles of Wages

For internal consistency, we model individual productivity based on income data available in MEPS. For our benchmark calculations we consider wage income.[10] Since MEPS includes information on both total income and wage income we can filter out individuals that have low wages but high income.[11] To do this, we remove individuals with wage income less than 90% of total income (and more than 100%). For each age we compute the distribution of wages. We consider 99 distinct types for each of the centile threshold: $\theta \in \{\theta_1, \ldots, \theta_{99}\}$. In particular $w_a(\theta_i)$ is equal the ith percentile of wage distribution at age $a \leq 65$. We assume that $w_a = 0$ for $a > 65$.[12] As a point of comparison with Hall and Jones (2007), note that our

[10]Note that this is total wage income, not the wage rate. For a more limited part of the sample, data is available for both wage income and hours worked. This makes it possible to repeat the analysis based on wage rates. This makes some difference in the analysis because the data on wage rates has somewhat thicker right tails, particularly at higher ages. The difference is not large, however.

[11]Not doing this could cause a significant bias. For example, we could label as low productivity someone who receives the bulk of his income from non-wage sources even though his actual productivity was high.

[12]Of course, assuming that productivity is zero for people whose ages is greater than 65 is an extreme, but convenient, assumption. Since, in the data, many people are retired after age 65, we made this modeling

wage profile will be hump shaped and will also feature retirement. The mean wage profile, and the averages in the top and bottom quartile are displayed in Figure 2.

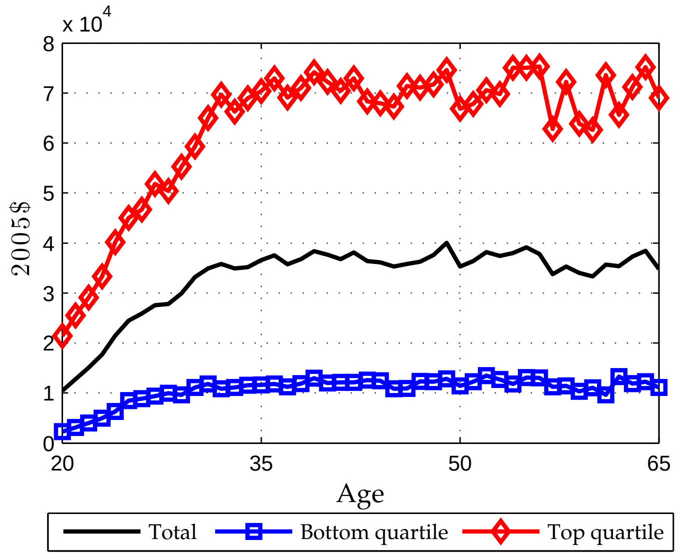


FIGURE 2: Wage profiles over age by quartile and total sample. Data from MEPS.

3.1.3 Preference Parameters

The utility function in each period is:[13]

$$U(c,l,x) = b + \frac{c^{1-\gamma}}{1-\gamma} + \psi\frac{(1-l)^{1-\epsilon}}{1-\epsilon}. \tag{10}$$

where the parameter b determines the flow value of being alive. The parameters γ and ϵ determine the curvature of the utility function with respect to consumption and leisure, respectively. Finally, ψ denotes the weight of leisure in the utility function relative to consumption.

We begin by calibrating ϵ. Using data from CEX and PSID, Heathcote, Storesletten, and Violante (2009) estimate a micro-Frisch elasticity of labor equal to 0.38. In our utility specification the Frisch elasticity (η_F) for an individual working l hours is given by

$$\eta_F = \frac{1-l}{l}\frac{1}{\epsilon} \tag{11}$$

choice partly because of data availability. Extending the analysis to include both endogenous and partial retirement decisions would be particularly useful.

[13]We have also experimented with variations on preferences in which healthcare has a direct utility effect (separable). This has little effect on the overall message of the paper. Details are available on request.

γ	ψ	ϵ	b	β	R
2.0	0.756	4.442	7.0	0.97	1.04

TABLE 1: Baseline preference parameters.

using the approach of Browning, Hansen, and Heckman (1999) we calibrate our curvature parameter to match average yearly hours in MEPS.[14] In MEPS, individuals aged 20 to 65 work an average of 39.5 hours per week. We assume that individuals, on average, work for 49 weeks per year. This implies average yearly hours equal to 1935.5. A standard value used for the number of feasible yearly hours is 5200 (normalized to 1 in our model). This implies that, on average, $l = 0.372$. Using (11) this implies that $\epsilon = 4.442$.

A key parameter of interest in this specification is b. This parameter plays a key role in determining the value of life in the model – the higher b is, the higher is the value of a life in the model. Hall and Jones (2007) choose this parameter to match the empirical value of statistical life, *VSL* hereafter, for 34-39 years olds in the year 2000 ($3 million).[15] In our benchmark calibration we estimate jointly b and ψ targeting a value of life equal to $3 million and average hours worked equal $l = 0.372$. The estimation involves minimizing the equally weighted distance between model generated values and data. We find a value of $b = 7.0$ and $\psi = 0.756$. In section 5.1 we explore the effect of changing VSL (and recalibrating ψ accordingly) on our results.[16] For the other three parameters we take β and R from the macro literature and follow Hall and Jones (2007) in setting $\gamma = 2.0$. The baseline parameters used for our simulation are given in Table 1.

3.2 Simulation Results

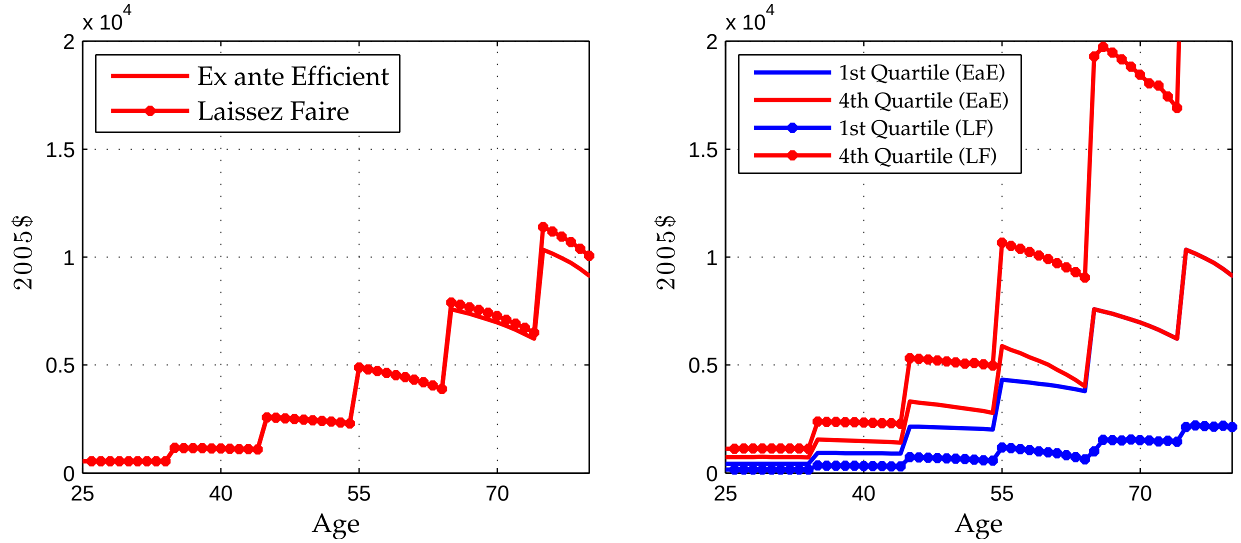
Using the parametric assumptions and the distribution of productivities described in the previous section, for every type of agent we compute both the ex ante efficient allocation (the solution to (4)) and the Laissez Faire allocation (the solution to (7)). We begin by looking at the relationship between age and average, across all types, of expenditures on

[14]Refer to Browning, Hansen, and Heckman (1999) footnote *a* in section 3.4.1.

[15]For additional discussion on the value of life refer to Hall and Jones (2007), Thaler and Rosen (1976) and Rosen (1988). Refer to Viscusi and Aldy (2003), Ashenfelter (2006) for a quantitative analysis. Technically, b is chosen so that VSL=\$3,000,000=$[B_1(\theta) + B_2(\theta)]$ where $B_1(\theta)$ is the value, in consumption terms of the discounted value of utility flows from age 37 on of a type θ agent and $B_2(\theta)$ is the value, to the planner of the excess production by a type θ. Details available from the authors on request.

[16]We choose b so that the value of statistical life at age 37 under ex ante efficient allocation is equal $3 million. The value of life under Laissez Fair allocation is higher ($3.5 million) for the chosen parameters. This is mainly due to the fact that more productive individuals have substantially higher consumption and survival rate under Laissez Fair allocation.

health. These are shown in Figure 3(a).



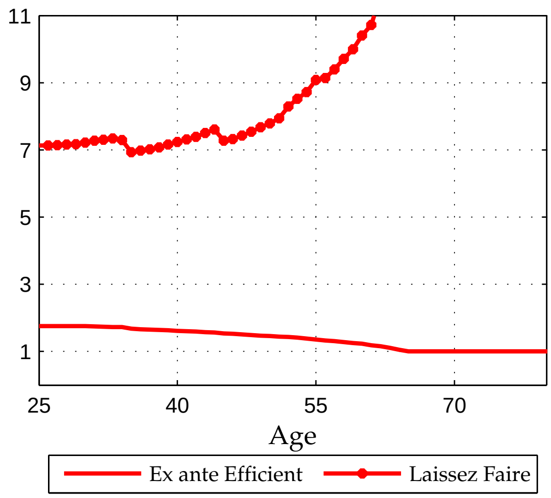
(a) Solid line is the average health spending in ex ante efficient allocation. The solid line with bullets in average health spending under laissez faire.

(b) Solid lines are average health spending under ex ante efficient allocation (EaE). Lines with bullets are the average health spending under Laissez Faire (LF).

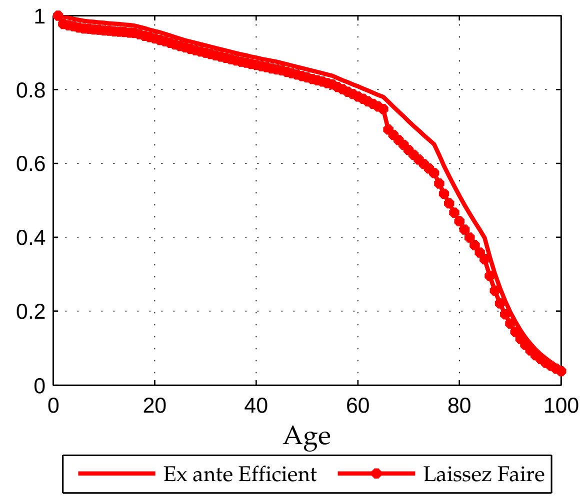
FIGURE 3: Average health spending.

As can be seen, mean expenditures increases slowly as a function of age, beginning at about \$1,000 per person per year at age 25 and increasing up to about \$10,000 per person per year at age 80. Further, the differences between the mean levels of spending for the two allocations are small. Although the mean levels of the two allocations are quite similar for all ages, there are significant differences in how these expenditures are distributed across the productivity distribution. This is shown in Figure 3(b) where mean spending in both the top and bottom quartiles of the productivity distribution are shown for each age. As can be seen in the figure, the level of inequality is small for all ages in the ex ante efficient allocation, and, as expected given the result in Proposition 1, spending is exactly equal for post retirement ages. In contrast, there is a significant amount of inequality in the laissez faire allocation and this grows, in levels, as a function of age. This can be more easily seen in Figure 4(a), where we show the ratio of mean spending for the bottom and top quartile for the two allocations. In the ex ante efficient allocation, this ratio is slightly less than 2 at $a = 25$, and slowly falls 1.0 at $a = 65$ (when retirement occurs for all types by assumption). In contrast, the corresponding ratio is about 7 at $a = 25$ and rises to 10.0 at $a = 65$ in the Laissez Faire allocation.

In both our benchmark models, expenditures on health are made to increase survival



(a) Ratio of average health spending of top income quartile relative to spending of bottom income quartile. Solid line is ex ante efficient. Line with bullets is laissez faire.

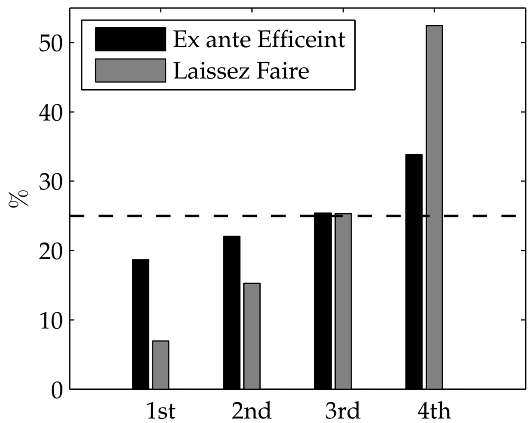


(b) Unconditional survival probability to each age. Ex ante efficient vs. Laissez faire.

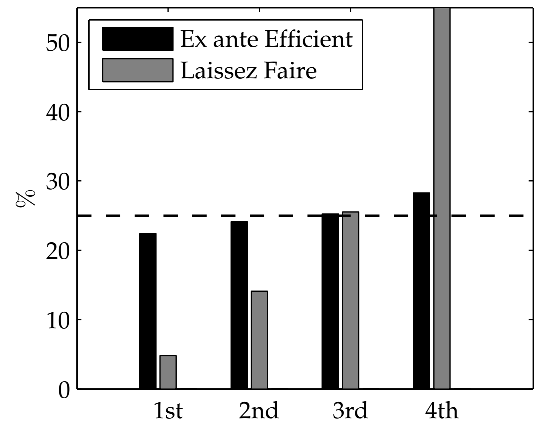
FIGURE 4: Comparison between allocations. Ex Ante Efficient and Laissez Faire.

probabilities. These (i.e., average across types of unconditional survival probabilities) are shown, as a function of age, for the two allocations in Figure 4(b). As can be seen, these are quite similar in the two allocations (on average).

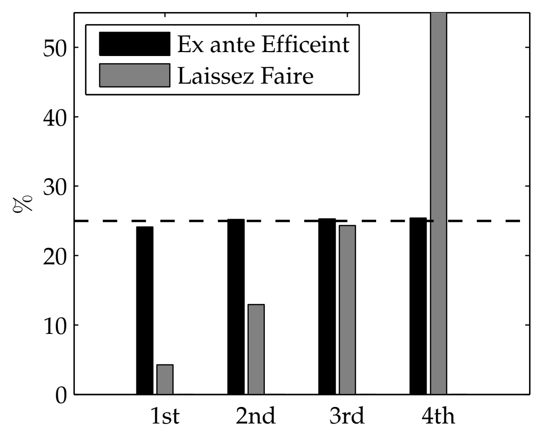
Figures 5(a), 5(b), and 5(c) break down total spending for each age across the four productivity quartiles for both the ex ante efficient and Laissez Faire allocations. E.g., the lowest productivity quartile accounts for about 18% of the health spending for individuals between $a = 25$ and $a = 35$ in the ex ante efficient allocation while about 33% of spending is on the highest types. The share going to the lowest types increases monotonically with age while that for the highest types falls. After age 65, these are, of course 25% for all of the 4 groups. Also shown is the distribution of spending across the groups in the Laissez Faire. As can be seen, these are much more unequal and this inequality is roughly independent of age (4% for the lowest quartile and 55% for the highest).



(a) Quartile of income distribution ages 25 to 35.



(b) Quartile of income distribution ages 55 to 65.



(c) Quartile of income distribution ages 70 to 80.

Figure 5: Share of health spending by income quartile for various age groups. Black bar are shares under ex ante efficient allocation. Gray bars are shares under laissez faire

4 Comparing Model and Data Quantities

In this section we compare the simulation results from the previous section to analogous quantities from the data. Of course, there is much more heterogeneity in the data than what we have allowed in our model including differences in morbidity, sickness and measurement error. Rather than complicating the model to include additional features, we try to construct data counterparts that are comparable to model quantities.

4.1 Data

As mentioned earlier, our main data source is the Medical Expenditure Panel Survey. Refer to Appendix C for additional details on this dataset.[17] For each individual in the sample, for each year, the expenditure measure taken from MEPS is the sum of actual health expenditures incurred either by the individual or provided by other sources on his behalf. MEPS provides annual data continuously from 1996 to 2008. To make observations comparable across years, all variables in the dataset are deflated to 2005 dollars using the recommended prices indices.[18] In Table 2, we summarize the number of raw observations available for each year in our data once we drop individuals with missing self reported health status (4062 observations) and missing age information (3047 observations). To limit the effects of outliers, we drop the following observations: individuals with total income exceeding the 99.9 percentile (1,971 observations); individuals with negative or zero income (135,414 observations).[19] On the remaining sample, we construct mean health spending by age and by income.[20]

Year	**Observations**	**Year**	**Observations**
1996	21,976	2003	33,854
1997	33,807	2004	34,061
1998	20,823	2005	33,574
1999	24,383	2006	33,812
2000	24,858	2007	30,668
2001	33,245	2008	32,711
2002	38,801	**Total**	**396,575**

Table 2: Observations available per year in the MEPS data.

We begin by looking at average spending on health by age. In the data expenditures are fairly high for newborns – about $4,000 per person per year – drops quickly for children aged 2-12 – about $1,000 per person per year – and then gradually rises, reaching a peak of about $9,000 per person per year after 80 years of age. In Figure 6(a) expenditures from age 25 onwards are shown.

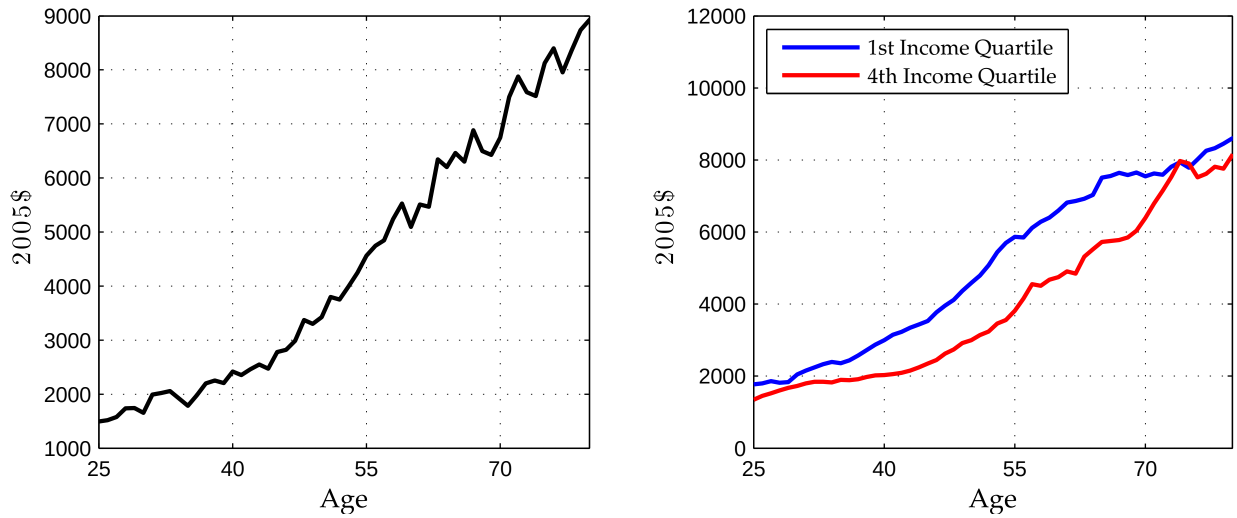
Next, in Figure 6(b), we show the breakdown of spending by age according to income.

[17] In the appendix we also look at the relation between MEPS and aggregate U.S data and the Consumer Expenditure Survey.

[18] Refer to: http://www.meps.ahrq.gov/mepswebabout_meps/Price_Index.shtml.

[19] Of these 135,414 observations, 114,661 are for individuals younger than 20 year old.

[20] In Appendix D we document that MEPS expenditure data is highly skewed. There, we also look at measures of health spending constructed looking at medians over age and income.



(a) Total sample.

(b) Top and bottom income quartile.

FIGURE 6: Health spending in MEPS data over age.

Shown is mean spending on the bottom and top quartiles of the income distribution. As can be seen, spending is higher for individuals in the lower income quartile (in blue) than in the upper income quartile (in red). This fact has been pointed out in the literature before (e.g., see Ozkan (2011)). It raises two issues. The first concerns the method of delivery of health services to lower income households. As is well known, a principal method for delivering health care to low income households is through ER visits rather than normal doctor visits. This is because ERs are required by law to treat all patients whether they are able to pay for the services or not.[21] The same is not true for normal doctor appointments/clinic visits. Thus, for very poor households, these ER bills often are not paid for by the individuals receiving the treatment. Moreover, since ER's are particularly expensive, this inflates the figure for low income households. This effect is exacerbated by the fact that since ER visits need not be paid for by poor households, they often delay treatment entirely until an ER visit is called for. This also causes costs to be higher. These features of the data is the focus of Ozkan (2011) and will not be discussed here.

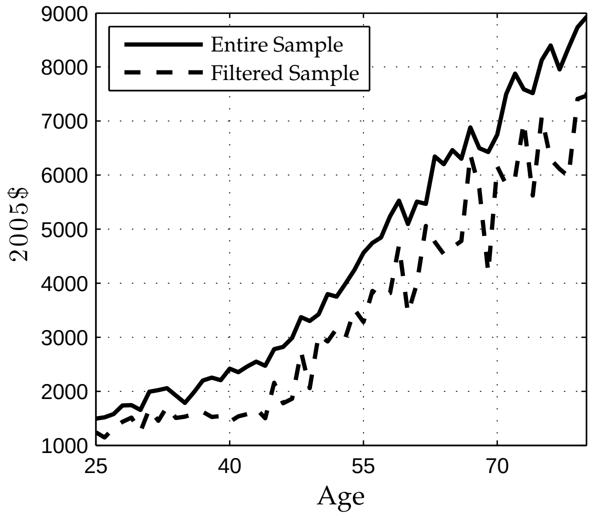
The second issue concerns a difficulty with measuring the relationship between health expenditures and income in this way: the endogeneity of income. For example, if a high income person becomes sick, their ability to work is likely compromised–their income is lower than what would be expected from their productivity (θ)–and simultaneously medical expenditures are high. This effect biases upward the estimate of spending on low income individuals

[21] See: http://www.law.cornell.edu/uscode/text/42/1395dd

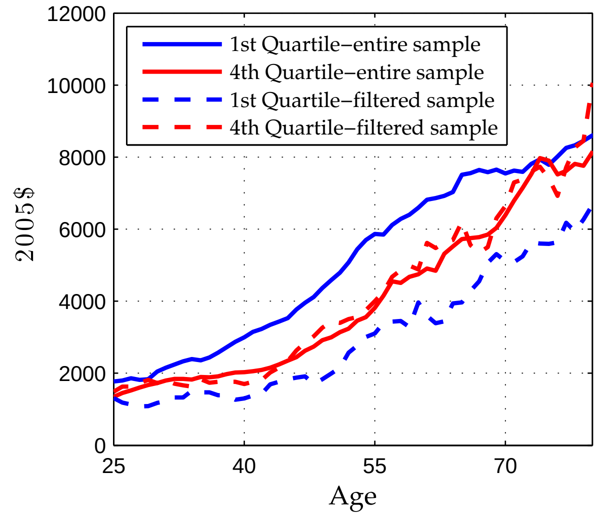
(by lumping in high θ types with them) and biases downward the estimate of spending on high income individuals (by classifying the very sick among them as low income). To try to control for this effect, we restrict the sample to individuals who enter the panel with median "health". To do this, we use the variable *self reported health status* available within MEPS. For each individual in each interview, the dataset contains detailed information on the health status of the individual. Information is collected about particular health conditions, possible impairments and overall measures of (self reported) health status. Self reported health status can take on five possible values: 5=poor, 4=fair, 3=good, 2=very good, 1=excellent.[22] In a given year an individual is interviewed multiple times. Since health status might change over the course of the year, we calculate a time weighted average of the reported values using the start and end date of each interview reference period. Each value is weighted by the fraction of time within a year for which the response of a particular interview is valid. For example, if for the first six months the individual interviewed reports a value of 1 (excellent health) and for the remaining 6 months reports 2 (very good health), the health status for the year will be 1.5, etc.. For most reported data that follows we will restrict attention to individuals whose average self reported health status is near the median of the health status for that age. This is done by using only those observations, for a given age, whose self reported health status lies between the 45th and 55th percentile of the health status distribution. Performing this last data restriction drop an additional 225,644 observations leaving us with a total of 33,546 observations. Before proceeding, we give some information on the effect of this filtering of the data.

In Figure 7(a) we display the age profile of average health expenditures. The two lines represent the means of the entire sample (solid) and the sub-sample of the population with self reported median health (dashed), respectively. We see that health expenditures increase monotonically with age for both samples. Also, individuals with median health exhibit a lower average health expenditures than the entire sample (the bulk of the health expenditures are for individuals with poor and fair health). The increasing age profile of health expenditures for individuals with median health is an indication that the self reported health status is an age relative concept rather than an absolute one. A comparison of spending in the top and bottom quartiles in both the full and restricted samples are shown in Figure 7(b). Overall, as can be seen, this selection does two things. First, it lowers overall spending (but only by 16% on average). Second, in this sub-sample, mean spending is higher for high income households as expected.

[22] The exact question asked is: "In general, would you say your health is:".



(a) Average health spending over age.



(b) Health spending on top and bottom income quartile in MEPS. Solid line is entire sample. Dashed line is filtered sample (conditioned on median self reported health status).

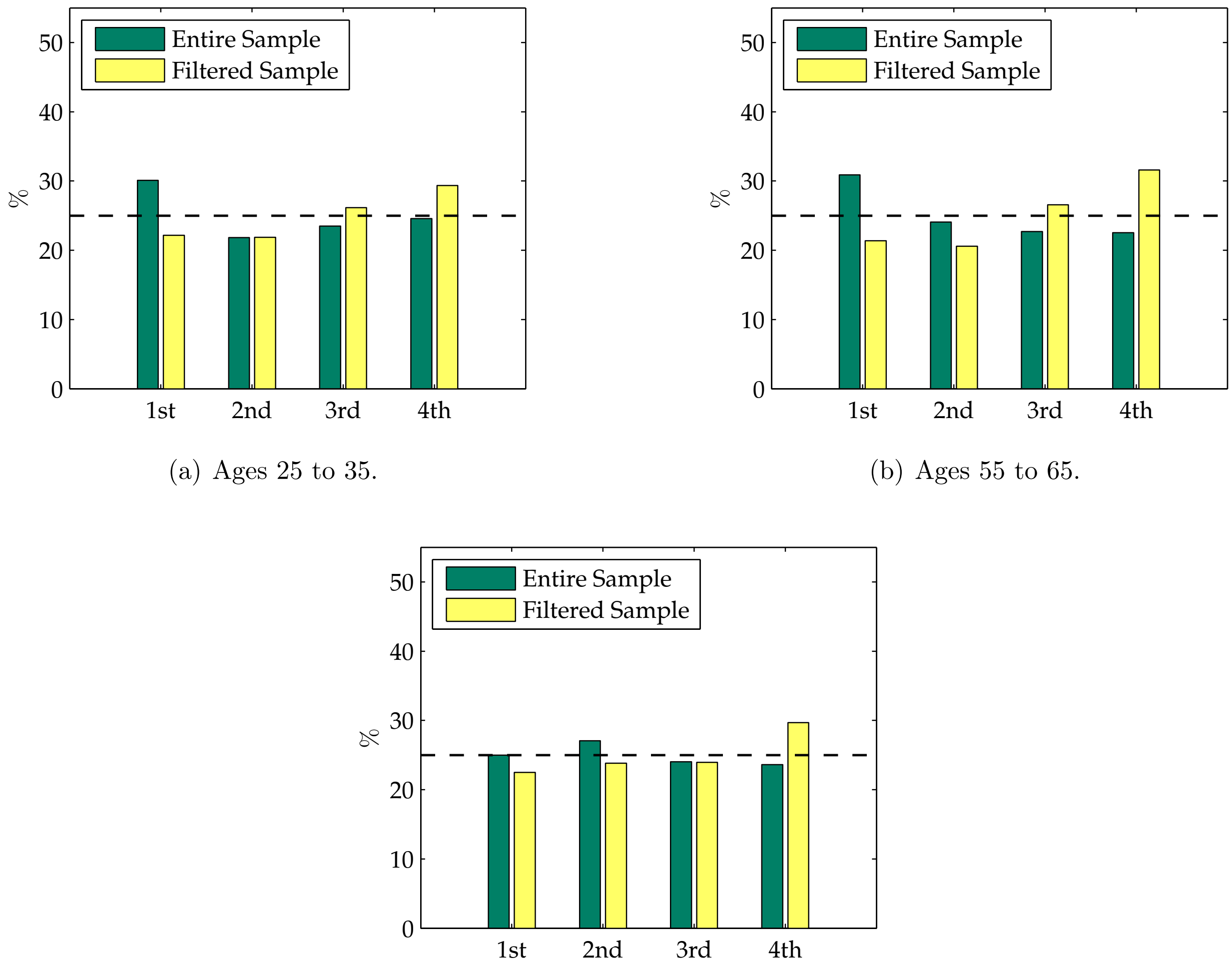
FIGURE 7: Comparison between raw and filtered MEPS data.

Figures 8(a), 8(b) and 8(c) give more detail of this comparison. Shown are the fraction of all expenditures for three different age groups that goes to each income quartile for both the full and restricted samples. The main thing to notice is that sample selection lowers the share of spending attributed to individuals in the bottom quartile of the income distribution – much of the spending on individuals in the bottom quartile overall is on individuals whose health status is below average. Assuming that these individuals have low income in part because they are in poor health, it appears that much of the bias due to income endogeneity has been removed by this selection.[23]

4.2 Model vs. Data

In this section we compare the quantities from the simulations presented in the previous section (both ex ante efficient and Laissez Faire) with data quantities for the sub-sample of MEPS with 'median health.' We find that the overall level of spending differences by income in the data is higher than what would be called for in the ex ante efficient allocation (except for lower ages). However, this difference between data and efficiency is fairly small at low

[23]To get additional direct evidence of this, we can compute the yearly dollar value of days missed from work due to health (MEPS reports days missed from work due to health). We focus on hourly payed workers. When looking at the relation of this amount with respect to health status we see that there is no statistical difference between workers that report health being excellent versus good (which is close to the median health), roughly 650$ per year. While for poor self reported health status the loss is three time as large.



(a) Ages 25 to 35.

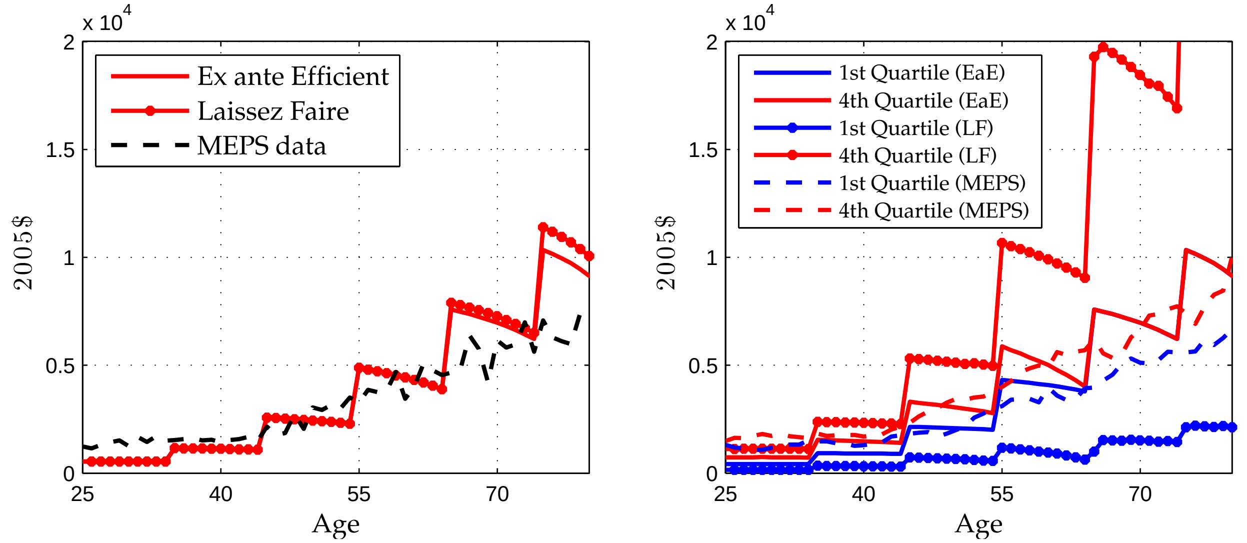
(b) Ages 55 to 65.

(c) Ages 70 to 80.

FIGURE 8: Share of health spending by income quartile for various age groups. MEPS data: total sample vs. filtered sample

ages (25-50) and grows larger as age is increased. Compared to the Laissez Faire allocation, the data exhibits much less inequality in spending. Thus, in this sense, the current allocation in the U.S. is much closer to ex ante efficient than to Laissez Faire, but there is still room for improvement. We begin by showing mean expenditures by age in Figure 9(a). As can be seen here, overall, mean health expenditures in the data track both the ex ante efficient and the Laissez Faire allocations quite well. The differences that do exist are that mean spending is higher than the (ex ante) efficient mean spending for ages 25 to 45 and lower for ages 65 to 80.

Next we look at the breakdown of health expenditure across the income distribution. Mean expenditures on the top and bottom income quartiles are shown in Figure 9(b). As

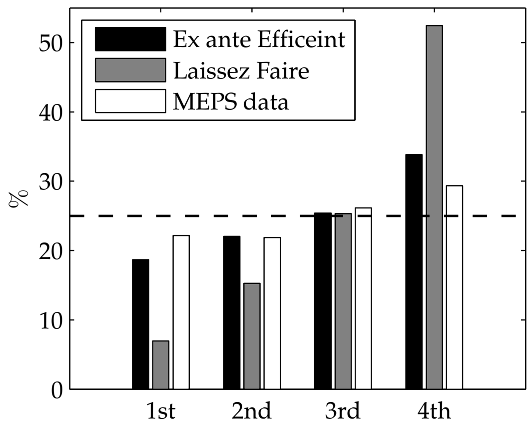


(a) Total sample.

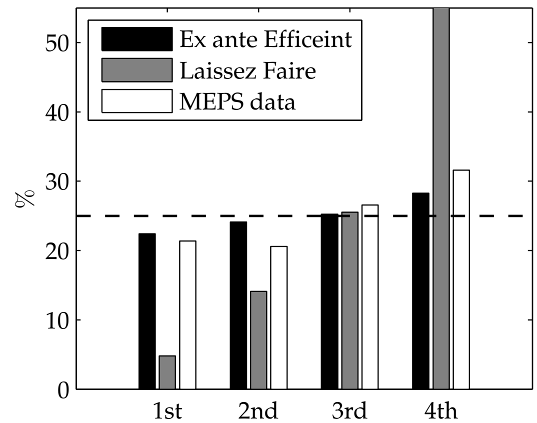
(b) Top and Bottom Quartiles of the income distribution.

FIGURE 9: Average expenditures over age. Ex Ante Efficient (EaE), Laissez Faire (LF) and MEPS data.

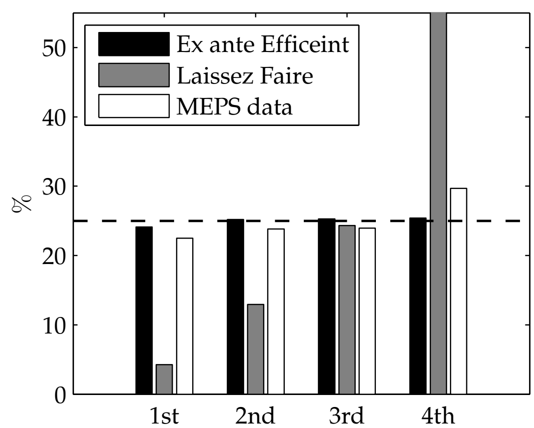
can be seen in the figure, the amount of inequality in MEPS is quite similar to that in the ex ante efficient allocation up until retirement age. After that however, there is considerably more inequality than what is called for in the efficient allocation. Notably, spending on the bottom quartile is lower than (ex ante) efficient while that on the top is about at the same level.



(a) Ages 25 to 35.



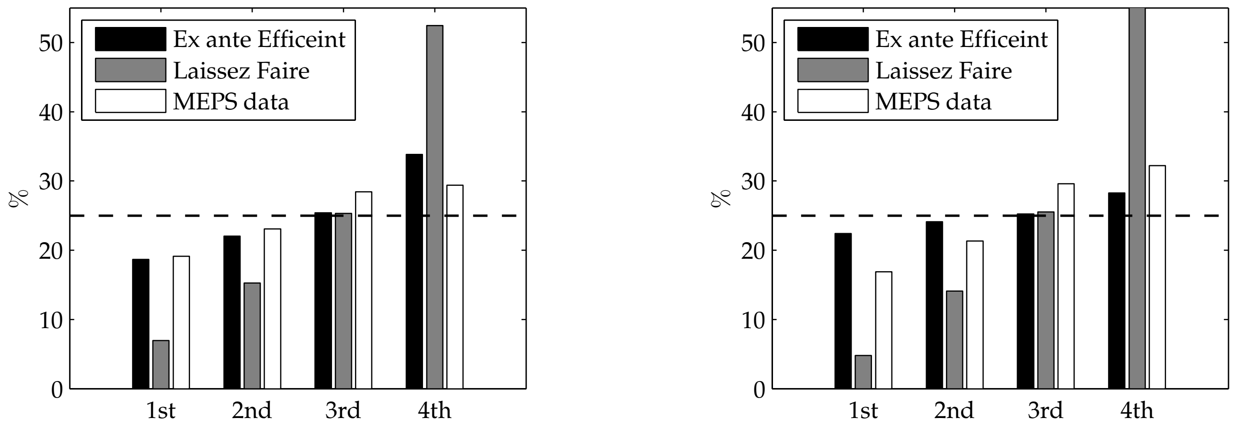
(b) Ages 55 to 65.



(c) Ages 70 to 80.

FIGURE 10: Share of health spending by income quartile for various age groups. Black bar are shares under ex ante efficient allocation. Gray bars are shares under laissez faire. White bars are from MEPS data.

The next three figures, Figure 10(a), 10(b) and 10(c), show more detail about the breakdown in spending across the income distribution for the three allocations for three different age groups. As can be seen, the distribution in the data is fairly close to that in the ex ante efficient allocation in the pre-retirement ages, but more unequal after retirement. As noted in the Introduction, this conclusion depends, at least partially, on how income is measured in the data, particularly at ages close to retirement. For example, using wage income to rank individuals in the data rather than total income, we find that the bottom 25% receives about 24% of all spending on average for ages 55 to 65, but only receives about 17% in the data. Thus, it is considerably lower than optimal when measured this way. Figures 11(a) and 11(b) show the breakdown in spending across the wage-income distribution for the three



(a) Ages 25 to 35.

(b) Ages 55 to 65.

FIGURE 11: Share of health spending by wage-income quartile for various age groups. Black bar are shares under ex ante efficient allocation. Gray bars are shares under laissez faire. White bars are from MEPS data.

allocations for two age groups.[24]

That said, it is considerably less unequal than what would hold in the Laissez Faire allocation. In this sense, it appears that the overall level of social insurance in the U.S. currently is fairly good. The one exception to this is at older ages where the inequality in health care spending is higher than what is (ex ante) efficient. This is best summarized in Figure 12(a) which shows the ratio of mean spending in the top vs. bottom income quartiles in the three allocations.

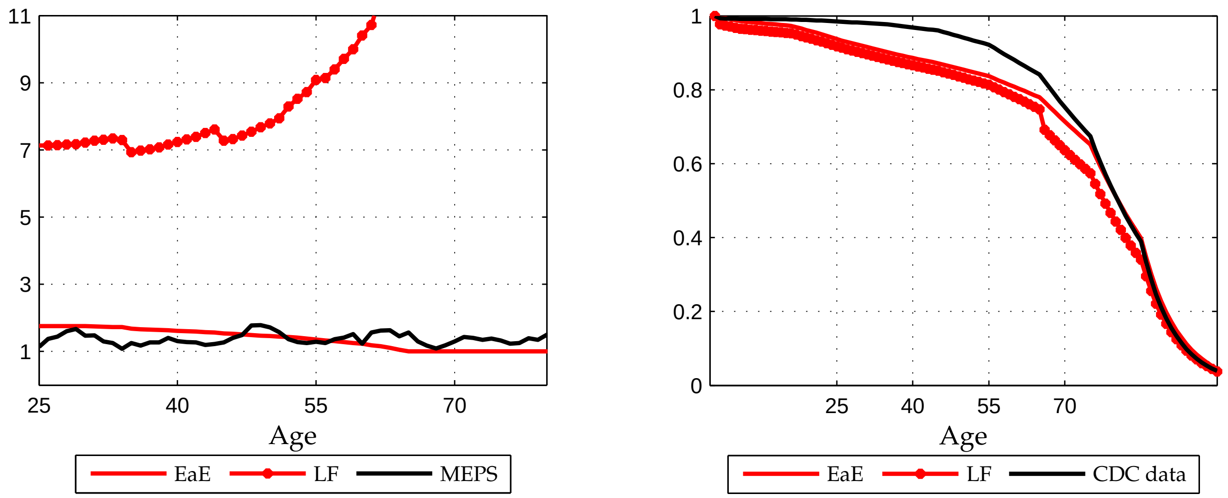
Figure 12(b) provides an additional useful overall summary. As can be seen under the ex ante efficient allocation, unconditional survival probabilities are lower (this is also true of the Laissez Faire allocation) than that seen in the data. This is due, in large part, to the lower levels of spending at young ages in the model allocations.

5 Robustness

In this section we depart from the benchmark environment and perform some robustness calculations on the results of the previous section. In particular we look at:[25]

[24] It is not clear which of these income methods is more appropriate for comparing the data to the efficient allocation. One difficulty with using wage income is that we can no longer make comparisons between data and model outcomes for ages after retirement.

[25] We have also performed other sensitivity analysis. For example, alternative values of R, alternative ways of ranking individuals in MEPS (i.e., wage rate instead of total wage income), alternative methods for



(a) Ratio of average health spending on top income quartile relative to spending at bottom income quartile.

(b) Unconditional survival probabilities to each age..

FIGURE 12: Comparison between allocations. Ex Ante Efficient (EAE), Laissez Faire (LF) and data.

1. Sensitivity to the value of life parameter in the model – VSL;

2. Assuming ability is private information.

5.1 Sensitivity to Value of Life Parameter

The flow utility from being alive is governed by the parameter b. Following Hall and Jones, this can be calibrated by looking at the micro based estimates of the Value of Statistical Life (VSL). The literature studying VSL provides a wide range of values ranging from \$0.5 million to over \$20 million. In their meta-study Viscusi and Aldy (2003) provide a reference value of \$6.7 million with a standard error of \$5.6 million (Hall and Jones use a VSL of 3 million dollars for individuals aged 35-39). In this paper, the benchmark value of $b = 7$ implies an average VSL in our environment of 3 million dollars. In Figure 13 we see that for different types, VSL ranges from \$2 to about \$8.3 million at the top of the income distribution.

We now explore alternative choice of parameter b. We target both lower and higher values for VSL – \$1.5 million and \$7.0 million. In each case, we re-calibrate the values of both b and ψ. When $VSL = 1.5$, we find $b = 3.783$ and $\psi = 0.730$. At $VSL = 7$, $b = 16.215$ and

estimating survival probability elasticities and alternative specifications of P_a in which there is a baseline survival rate (estimated from the data) independent of health expenditures. In general, these give similar results to the benchmark and hence are not reported in the main text. Details are available from the authors on request.

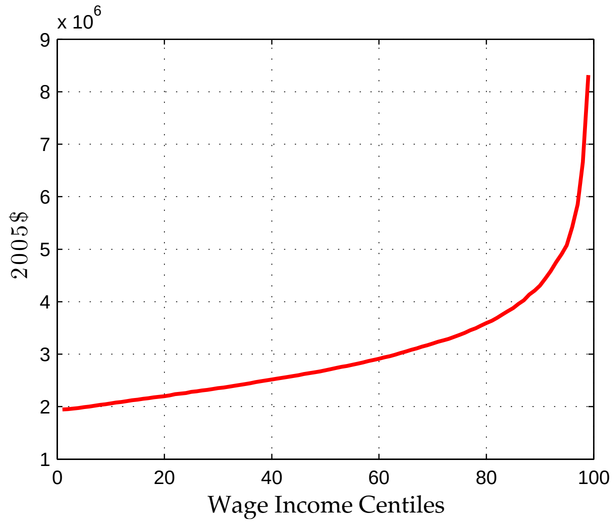


FIGURE 13: VSL over wage income at age 37.

$\psi = 0.815$. As expected as we increase VSL, the flow utility parameter b, also increases. In Figure 14 we report mean health spending by age for the three different values of b. As expected, a higher base flow value of utility increases the value, to the planner, of keeping each individual alive and this causes spending on health to increase and consumption to fall (not shown).

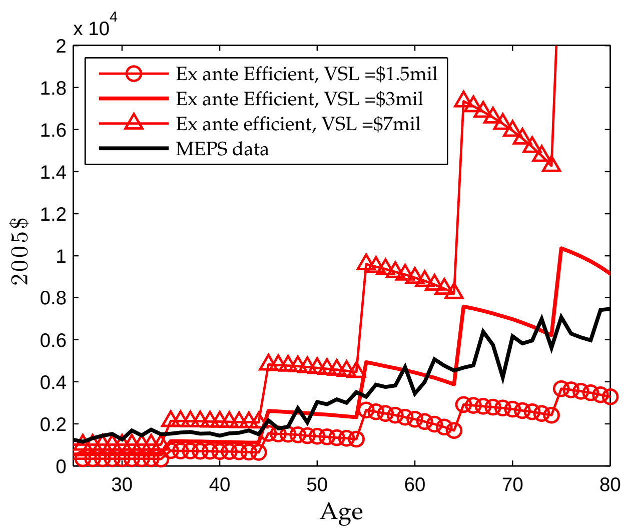
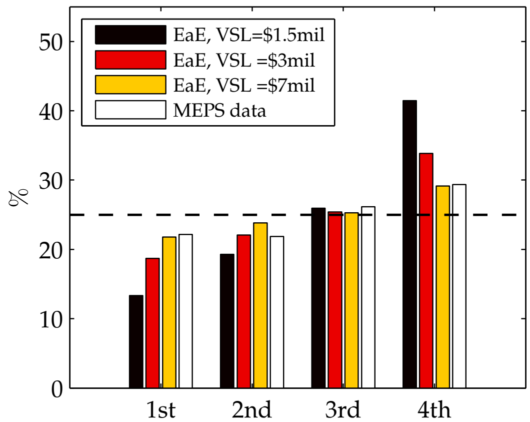
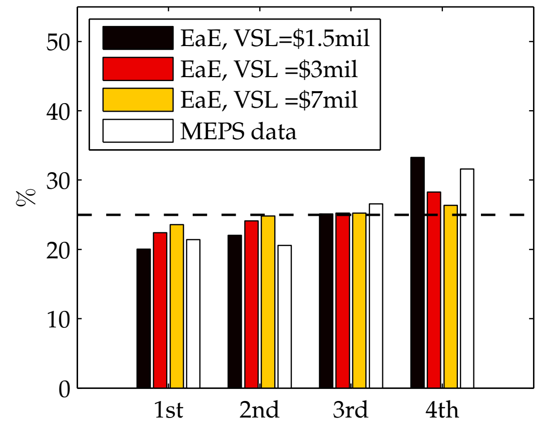


FIGURE 14: Mean Health Spending by Age: Alternative Values of VSL

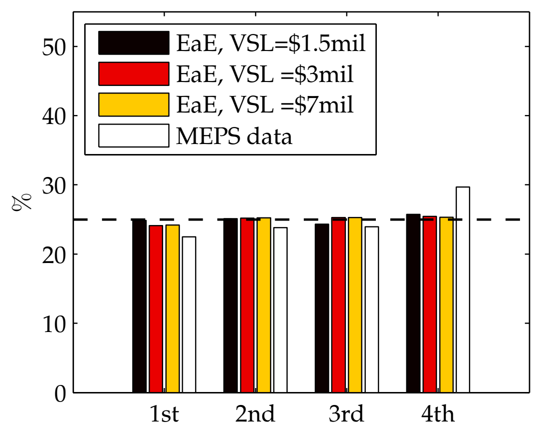
Next, we show the distribution of spending by income type for three different age groups, see Figure 15. One of the interesting feature is that not only is the level of spending sensitive to VSL, but we can see that the efficient amount of inequality is also substantially lower when VSL is larger. This is very intuitive, when b is larger, the relative value to the planner of keeping a high θ alive is smaller.



(a) Ages 25 to 35.



(b) Ages 55 to 65.



(c) Ages 70 to 80.

FIGURE 15: Share of health spending by income quartile for various age groups under the ex ante efficient allocation (EaE). Values are for alternative values of VSL.

5.2 Adding Private Information

So far, we have focused on a unique productive efficiency reason for making health care unequal – society gets more output as a whole when more productive agents survive. A possible criticism of the benchmark environment is that it misses other efficiency reasons in generating inequality in health spending. In this section we explore an incentive motive. Suppose that the productivity type of the agent is his own private information (as in Mirrlees (1971)). In such an environment, the planner can use differences in consumption, leisure and in our case health care expenditure to give agents incentives to 'reveal' their true productivity types. Note that by doing this we can expect that following retirement a positive inequality in health expenditures will still hold.

We assume that there are two productivity types $\theta_H > \theta_L$ and that these are the private information of the individual.[26] Further, as above, we assume that this productivity type is permanent. Preferences are as in (10). In this environment, the planner has to provide incentives for productive individuals to truthfully report their types. Since, as in the previous section, the realization of the type, θ, in the first period determines the entire path of labor productivity, the planner has to induce workers to reveal their types only once - in the first period. For this reason only one incentive constraint is required to ensure truthful revelation of types. The ex ante constrained efficient allocation will be given by the solution of the planner's problem as in (4) with the addition of the following incentive compatibility constraint:

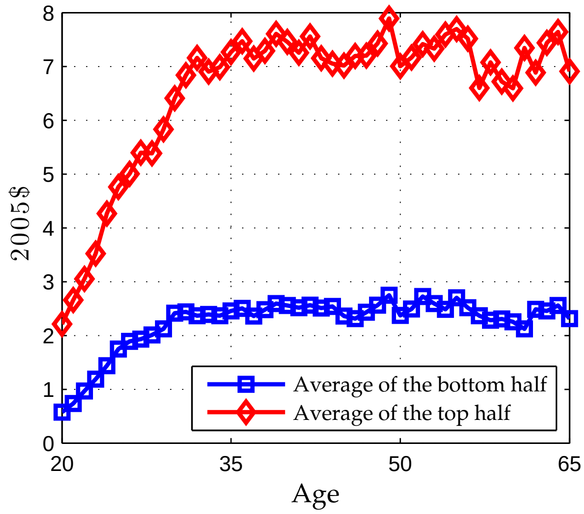
$$\sum_{a=0}^{A} \beta^a N_a(\theta_H) \left[b + \frac{c_a(\theta_H)^{1-\gamma}}{1-\gamma} + \phi \frac{(1 - l_a(\theta_H))^{1-\epsilon}}{1-\epsilon} \right] \geq \tag{12}$$
$$\sum_{a=0}^{A} \beta^a N_a(\theta_L) \left[b + \frac{c_a(\theta_L)^{1-\gamma}}{1-\gamma} + \phi \frac{(1 - l_a(\theta_L) w_a(\theta_L)/w_a(\theta_H))^{1-\epsilon}}{1-\epsilon} \right].$$

This additional constraint will provide the productive individual with higher lifetime utility, which can be provided either via consumption, leisure, or via a higher survival probability through higher health care spending or through a combination of all of the above. It can be shown that the constrained efficient allocation will feature $c_a(\theta_H) > c_a(\theta_L)$ for all ages, so that the environment will generate positive inequality in consumption.[27] Characterizing the behavior of health expenditure proves more challenging. Given this we proceed numerically.

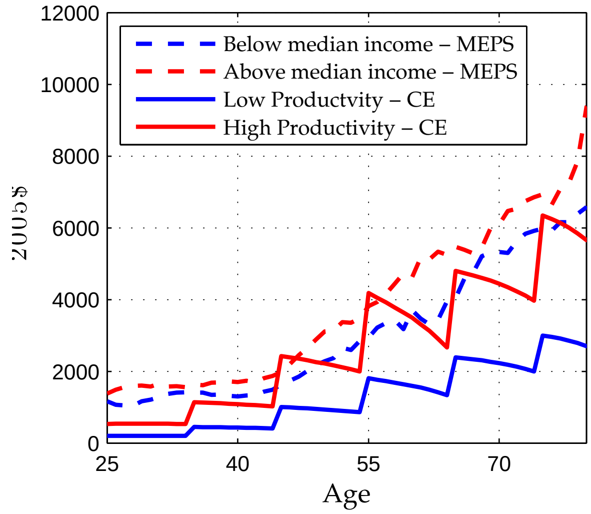
For the example that follows we use the parameters of the benchmark case (refer to Table 1). To accommodate the assumption of $I = 2$, we assume $\pi(\theta_H) = \pi(\theta_L) = 0.5$ and set $w_a(\theta_H)$ equal to the average of the top half of wages from MEPS and $w_a(\theta_L)$ equal to the average of the bottom half of wages from MEPS. Figure 16(a) shows these productivity profiles. When comparing the model with the data we use the mean health expenditure for the top and bottom halves of the wage income distribution in MEPS (as opposed to the top and bottom wage income quartiles in the previous sections). With this in mind we proceed to present the results.

[26] Assuming two types will maximize the effect of private information on health inequality.

[27] For additional considerations regarding consumption and leisure in an environment with private information also refer to Ales and Maziero (2010).



(a) Life cycle wage profiles averages above and below the median. MEPS data.

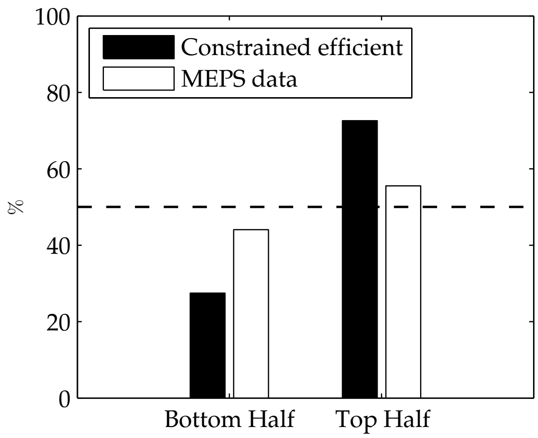


(b) Mean health expenditures averages above and below the median. Constrained efficient allocation (CE) and MEPS data

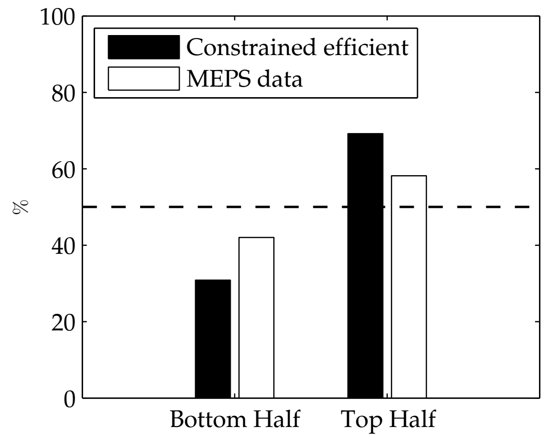
Figure 16: Adding private information.

Figure 16(b) shows mean expenditures for high and low productivity types and the corresponding quantities from MEPS. The dashed lines are the MEPS data and the solid lines are taken from the model. In terms of mean expenditures, the level of health expenditure in the data is higher than efficient. The is primarily due to an overall drop in output that occurs because of the informational friction. Hence, it is, in part, an artifact of that choice not to recalibrate the model to match output levels in this version.

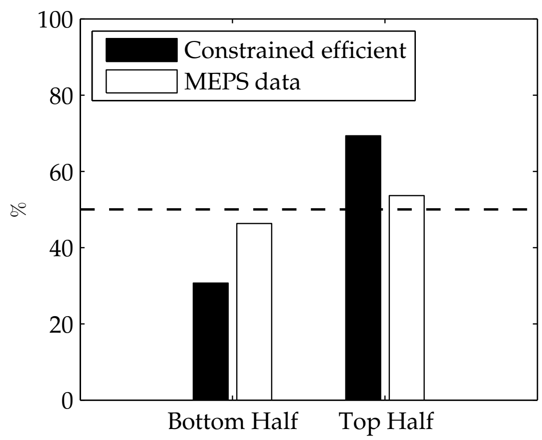
The interesting aspect that emerges in this example is that spending levels are no longer equal after retirement. As can be seen in Figures 17(a), 17(b) and 17(c) the amount of inequality in the constrained efficient allocation is roughly independent of age and is higher than what is seen in the MEPS data. There are two, related, reasons for why this is true. First the planner uses survival rates as an additional instrument, above and beyond consumption and leisure, to 'separate' the types. In addition, the planner has a further motive for high survival rates for high types post retirement – flow utility is higher for the high types is higher.



(a) Ages 25 to 35.



(b) Ages 55 to 65.



(c) Ages 70 to 80.

FIGURE 17: Share of health spending by for various age groups. Black bars are from the constrained efficient allocation. White bars are from MEPS data.

6 Conclusion

A key policy debate in the U.S. focuses on reforming the health care component of the social safety net. This paper contributes to this policy discussion by characterizing, in a normative model of health care spending, the ex-ante efficient allocation. We are interested in two dimensions of the efficient allocation: the amount of health inequality driven by income inequality and the amount of health care resources spent on the least productive segment of the population. When comparing this allocation to what we see in the data we reach two main conclusions: first, we find that the ex ante efficiency implies a fairly generous social safety net: the efficient share of healthcare spending for the bottom quartile of the income distribution is about 18% of total healthcare spending for ages 25 to 35 and it increases gradually up to 25% at retirement. Second, the amount of inequality across the income distribution that is seen in the data is larger than what would be justified solely on the basis of production efficiency, but not drastically so. This difference is higher for higher ages and is particularly pronounced for post retirement individuals.

As usual, these conclusions depend on the particular model that we have chosen to adopt to address these questions, and thus, it is of considerable interest to see how sensitive our results are to this specification. For example, we have not included health as a 'state' – spending on health today only affects survival today, not at future dates (see Yogo (2009) for a specification of the health production function with this feature). Other examples of useful extensions include, but are not limited to, explicitly including illness in the model, allowing productivity to move stochastically over the life cycle, adding a base line survival rate that is independent of spending and a more thorough of case with private information. We leave these questions for future work.

References

Albouy, V., L. Davezies, and T. Debrand (2009): "Dynamic Estimation of Health Expenditure: A new approach for simulating individual expenditure," *IRDES institut for research and information in health economics. Working Paper 20*. D.2, D.2

Ales, L., and P. Maziero (2010): "Accounting for Private Information," *Working Paper*. 27

Arrow, K. (1973): "Some ordinalist-utilitarian notes on Rawls's theory of justice," *Journal of Philosophy*, 70(9), 245–263. 1

Ashenfelter, O. (2006): "Measuring the Value of a Statistical Life: Problems and Prospects," *The Economic Journal*, 116(510), C10–C23. 15

Browning, M., L. Hansen, and J. Heckman (1999): "Micro data and general equilibrium models," *Handbook of macroeconomics*, 1, 543–633. 3.1.3, 14

Daniels, N. (1985): *Just Health Care*. Cambridge University Press. 1

De Nardi, M., E. French, and J. Jones (2010): "Why do the elderly save? the role of medical expenses," *Journal of Political Economy*, 118(1), 39–75. 1

Deaton, A. (1999): "Inequalities in Income and Inequalities in Health," *NBER Working Papers*. 1

Deaton, A., and C. Paxson (1998): "Aging and inequality in income and health," *American Economic Review*, pp. 248–253. 1

Deaton, A., and C. Paxson (1999): "Mortality, Education, Income, and Inequality Among American Cohorts," *NBER Working Paper*. 1

Duetsch, M. (2008): "Out-of-Pocket Health Care Spending Patterns of Older Americans as Measured by the Consumer Expenditure Survey," *Consumer Expenditure Survey Anthology*, pp. 46–51. 3, 31

Grossman, M. (1972): "On the concept of health capital and the demand for health," *The Journal of Political Economy*, 80(2), 223–255. 6

Hall, R., and C. Jones (2007): "The Value of Life and the Rise in Health Spending*," *The Quarterly Journal of Economics*, 122(1), 39–72. 1, 1, 2.1, 3.1.1, 3.1.1, 3.1.1, 3.1.2, 3.1.3, 15, B

Hartman, M., R. Kornfeld, and A. Catlin (2010): "Health Care Expenditures in the National Health Expenditures Accounts and in Gross Domestic Product: A Reconciliation," *BEA Working Paper*. C.1

Heathcote, J., F. Perri, and G. Violante (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 13(1), 15–51. C.2

Heathcote, J., K. Storesletten, and G. Violante (2009): "Consumption and labor supply with partial insurance: An analytical framework," *NBER Working Paper*, w15257. 3.1.3

Meara, E., C. White, and D. Cutler (2004): "Trends in medical spending by age, 1963–2000," *Health Affairs*, 23(4), 176. 3.1.1

Mirrlees, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208. 5.2

Ozkan, S. (2011): "Income Differences and Health Care Expenditures over the Life Cycle," *Mimeo, University of Pennsylvania*. 1, 4.1

Rawls, J. (1971): *A Theory of Justice*. The Belknap Press. 1

Rosen, S. (1988): "The value of changes in life expectancy," *Journal of Risk and Uncertainty*, 1(3), 285–304. 15

Selden, T., K. Levit, J. Cohen, S. Zuvekas, J. Moeller, D. McKusick, and R. Arnett (2001): "Reconciling medical expenditure estimates from the MEPS and the NHA, 1996," *Health Care Financing Review*, 23(1), 161–178. C.1

Shourideh, A., and M. Troshkin (2011): "Providing Efficient Incentives to Work: Retirement Ages and the Pension System," *University of Minnesota Working paper*. 4

Skinner, J., and W. Zhou (2004): "The Measurement and Evolution of Health Inequality: Evidence from the US Medicare Population," *NBER Working Paper*, w10842. 1

Thaler, R., and S. Rosen (1976): "The Value of Saving a Life: Evidence from the Labor Market," *NBER Chapters*, pp. 265–302. 15

Viscusi, W., and J. Aldy (2003): "The value of a statistical life: a critical review of market estimates throughout the world," *Journal of risk and uncertainty*, 27(1), 5–76. 15, 5.1

Wagstaff, A. (2002): "Inequality aversion, health inequalities and health achievement," *Journal of health economics*, 21(4), 627–641. 1

Yogo, M. (2009): "Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets," *NBER Working Paper*, w15307. 1, 6

Appendix

A Proofs of Section 2.2

A.1 Proof of Proposition 1

Proof. We first show the claim must hold for $a = A-1$. Let $h^*_{A-1}(\theta)$ and $h^*_{A-1}(\theta')$ be the solution to the planner problem for age $A-1$ and types θ and θ'. The allocation $h^*_{A-1}(\theta)$ and $h^*_{A-1}(\theta')$ must be the solution to the following maximization problem

$$\max_{\hat{h}_{A-1}(\theta),\hat{h}_{A-1}(\theta')} \pi(\theta')N^*_{A-1}(\theta')P_{A-1}(\hat{h}_{A-1}(\theta')) + \pi(\theta)N^*_{A-1}(\theta)P_{A-1}(\hat{h}_{A-1}(\theta)) \tag{13}$$

subject to

$$s.t. \quad \pi(\theta)N^*_{A-1}(\theta)\left(\hat{h}_{A-1}(\theta) - h^*_{A-1}(\theta)\right) + \pi(\theta')N^*_{A-1}(\theta')\left(\hat{h}_{A-1}(\theta') - h^*_{A-1}(\theta')\right) = 0$$

If not it can be replaced by the solution of this problem. The resulting allocation will use the same amount of resources and increases the ex ante welfare (recall that consumption is the same for all types and workers do not work after retirement). It follows immediately from strict concavity of $P_{A-1}(\cdot)$ that $h^*_{A-1}(\theta') = h^*_{A-1}(\theta)$ for θ and θ'.

Next suppose the claim is true for all ages $a+1, a+2, \ldots, A$-1. We show it must be true at age a. Note that the claim is true for all ages above a, therefore

$$\frac{N^*_{a'}(\theta)}{N^*_{a+1}(\theta)} = \frac{N^*_{a'}(\theta')}{N^*_{a+1}(\theta')} \quad \text{for all } a' > a+1$$

$h^*_a(\theta)$ and $h^*_a(\theta')$ must be the solution to the following maximization problem

$$\max_{\hat{h}_a(\theta),\hat{h}_a(\theta')} \pi(\theta')N^*_a(\theta')P_a(\hat{h}_a(\theta')) \sum_{a'=a+1}^{A} \beta^a \left(\frac{N^*_{a'}(\theta')}{N^*_{a+1}(\theta')}\right) + \pi(\theta)N^*_a(\theta)P_a(\hat{h}_a(\theta)) \sum_{a'=a+1}^{A} \beta^a \left(\frac{N^*_{a'}(\theta)}{N^*_{a+1}(\theta)}\right)$$

subject to

$$\begin{aligned}
&\pi(\theta)\frac{1}{R^a}N^*_a(\theta)\hat{h}_a(\theta) + \pi(\theta)N^*_a(\theta)P_a(\hat{h}_a(\theta)) \sum_{a'=a+1}^{A} \frac{1}{R^{a+1}}\left(\frac{N^*_{a'}(\theta)}{N^*_{a+1}(\theta)}h^*_{a+1}(\theta)\right) \\
&+\pi(\theta')\frac{1}{R^a}N^*_a(\theta')\hat{h}_a(\theta') + \pi(\theta')N^*_a(\theta')P_a(\hat{h}_a(\theta')) \sum_{a'=a+1}^{A} \frac{1}{R^{a+1}}\left(\frac{N^*_{a'}(\theta')}{N^*_{a+1}(\theta')}h^*_{a+1}(\theta')\right) = \\
&\pi(\theta)\frac{1}{R^a}N^*_a(\theta)h^*_a(\theta) + \pi(\theta)N^*_a(\theta)P_a(h^*_a(\theta)) \sum_{a'=a+1}^{A} \frac{1}{R^{a+1}}\left(\frac{N^*_{a'}(\theta)}{N^*_{a+1}(\theta)}h^*_{a+1}(\theta)\right) \\
&+\pi(\theta')\frac{1}{R^a}N^*_a(\theta')h^*_a(\theta') + \pi(\theta')N^*_a(\theta')P_a(h^*_a(\theta')) \sum_{a'=a+1}^{A} \frac{1}{R^{a+1}}\left(\frac{N^*_{a'}(\theta')}{N^*_{a+1}(\theta')}h^*_{a+1}(\theta')\right)
\end{aligned}$$

First order conditions imply

$$\frac{P_a'(h_a^*(\theta))}{1+P_a'(h_a^*(\theta))} = \frac{P_a'(h_a^*(\theta'))}{1+P_a'(h_a^*(\theta'))},$$

and therefore

$$h_a^*(\theta') = h_a^*(\theta).$$

■

A.2 Proof of Proposition 2

Proof. We break the proof in two steps. In step 1, we establish the first part of the claim for $a = a_{ret} - 1$. Then, in step 2 we show that if the claim is true in period $a+1$, it must be true in period a and the second part of the claim must also hold.

Step1

Without loss of generality and in order to make notation easier to follow we present the proof for θ_1 and θ_I. Let $(c_a^*(\theta), l_a^*(\theta), h_a^*(\theta), N_a^*(\theta))$ and k_a^* be the solution to the planner's problem. Consider the following perturbation in the distribution of types alive at age $a_{ret}-1$.

$$\begin{aligned}
\tilde{N}_{a_{ret}-1}(\theta_1) &= N^*_{a_{ret}-1}(\theta_1) - \epsilon \\
\tilde{N}_{a_{ret}-1}(\theta_I) &= N^*_{a_{ret}-1}(\theta_I) + \frac{\pi(\theta_1)}{\pi(\theta_I)}\epsilon \\
\tilde{N}_{a_{ret}-1}(\theta) &= N^*_{a_{ret}-1}(\theta) \quad \text{for all other } \theta
\end{aligned}$$

Consider an alternative allocation $(\tilde{c}_{a_{ret}-1}(\theta), \tilde{l}_{a_{ret}-1}(\theta), \tilde{h}_{a_{ret}-1}(\theta))$ such that

$$\begin{aligned}
\tilde{c}_{a_{ret}-1}(\theta) &= c^*_{a_{ret}-1}(\theta) \\
\tilde{h}_{a_{ret}-1}(\theta) &= h^*_{a_{ret}-1}(\theta) \\
\tilde{l}_{a_{ret}-1}(\theta) &= l^*_{a_{ret}-1}(\theta) \quad \theta = \theta_1, \ldots, \theta_{I-1} \\
\tilde{l}_{a_{ret}-1}(\theta_I) &= \frac{N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)w_{a_{ret}-1}(\theta_I)l^*_{a_{ret}-1}(\theta_I) + \epsilon\pi(\theta_1)w_{a_{ret}-1}(\theta_1)l^*_{a_{ret}-1}(\theta_1)}{\left(N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I) + \epsilon\pi(\theta_1)\right)w_{a_{ret}-1}(\theta_I)}
\end{aligned}$$

Couple of remark about this alternative allocation. One, note that $\tilde{l}_{a_{ret}-1}(\theta_I)$ chosen so that the total output in period $a_{ret}-1$ is unchanged. Second, we have shown in proposition 1 that $h^*_{a_{ret}-1}(\theta)$ is the same for all types. Therefore, the fact that there are more number of θ_I types under the alternative allocations does not affect the feasibility.

Note that the utility from consumption under both allocations are the same:

$$\sum_\theta \pi(\theta)\tilde{N}_{a_{ret}-1}(\theta)u(\tilde{c}_{a_{ret}-1}(\theta)) = \sum_\theta \pi(\theta)N^*_{a_{ret}-1}(\theta)u(c^*_{a_{ret}-1}(\theta)).$$

To establish the claim we need to show

$$\sum_\theta \pi(\theta)\tilde{N}_{a_{ret}-1}(\theta)v(1-\tilde{l}_{a_{ret}-1}(\theta)) > \sum_\theta \pi(\theta)N^*_{a_{ret}-1}(\theta)v(1-l^*_{a_{ret}-1}(\theta)).$$

Note that

$$\begin{aligned}
\tilde{l}_{a_{ret}-1}(\theta_I) &= \frac{N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)w_{a_{ret}-1}(\theta_I)l^*_{a_{ret}-1}(\theta_I)+\epsilon\pi(\theta_1)w_{a_{ret}-1}(\theta_1)l^*_{a_{ret}-1}(\theta_1)}{\left(N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)+\epsilon\pi(\theta_1)\right)w_{a_{ret}-1}(\theta_I)} \\
&= \frac{N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)}{N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)+\epsilon\pi(\theta_1)}l^*_{a_{ret}-1}(\theta_I) + \frac{\epsilon\pi(\theta_1)}{N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)+\epsilon\pi(\theta_1)}\frac{w_{a_{ret}-1}(\theta_1)}{w_{a_{ret}-1}(\theta_I)}l^*_{a_{ret}-1}(\theta_1) \\
&< \frac{N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)}{N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)+\epsilon\pi(\theta_1)}l^*_{a_{ret}-1}(\theta_I) + \frac{\epsilon\pi(\theta_1)}{N^*_{a_{ret}-1}(\theta_I)\pi(\theta_I)+\epsilon\pi(\theta_1)}l^*_{a_{ret}-1}(\theta_1) \equiv \hat{l}_{a_{ret}-1}(\theta_I).
\end{aligned}$$

in which we have defined $\hat{l}_{a_{ret}-1}(\theta_I)$ as the weighted average of $l^*_{a_{ret}-1}(\theta_1)$ and $l^*_{a_{ret}-1}(\theta_I)$. Then,

$$\begin{aligned}
&\sum_\theta \pi(\theta)\tilde{N}_{a_{ret}-1}(\theta)v(1-\tilde{l}_{a_{ret}-1}(\theta)) - \sum_\theta \pi(\theta)N^*_{a_{ret}-1}(\theta)v(1-l^*_{a_{ret}-1}(\theta)) = \\
&= (N^*(\theta_I)\pi(\theta_I)+\epsilon\pi(\theta_1))v(1-\tilde{l}_{a_{ret}-1}(\theta_I)) - \\
&\quad N(\theta_I)\pi(\theta_I)v(1-l^*_{a_{ret}-1}(\theta_I)) - \epsilon\pi(\theta_1)v(1-l^*_{a_{ret}-1}(\theta_I)) \\
&> (N(\theta_I)\pi(\theta_I)+\epsilon\pi(\theta_1))v(1-\hat{l}_{a_{ret}-1}(\theta_I)) - \\
&\quad N(\theta_I)\pi(\theta_I)v(1-l^*_{a_{ret}-1}(\theta_I)) - \epsilon\pi(\theta_1)v(1-l^*_{a_{ret}-1}(\theta_I)) \\
&> 0.
\end{aligned}$$

The last inequality follows from strict concavity of $v(\cdot)$.

Step 2

Suppose for all $\varepsilon > 0$, $(N_{a+1}(\theta_1), N_{a+1}(\theta_2), ..., N_{a+1}(\theta_I))$,

$$\begin{aligned}
&V_{a+1}\left(N_{a+1}(\theta_1)-\varepsilon, ..., N_{a+1}(\theta_I)+\frac{\pi(\theta_1)}{\pi(\theta_I)}\varepsilon; k_{a+1}\right) > \\
&V_{a+1}(N_{a+1}(\theta_1), ..., N_{a+1}(\theta_I); k_{a+1})
\end{aligned}$$

First we show that $h^*_a(\theta_1) < h^*_a(\theta_I)$. Let $(N^*_{a+1}(\theta_1), N^*_{a+1}(\theta_2), \ldots, N^*_{a+1}(\theta_I))$ and $(h^*_a(\theta_1), \ldots, h^*_a(\theta_I))$ be the solution given $(N_a(\theta_1), N_a(\theta_2), \ldots, N_a(\theta_I))$. Suppose $h^*_a(\theta_1) > h^*_a(\theta_I)$. Consider the following alternative allocations

$$\begin{aligned}
\tilde{c}_a(\theta) &= c^*_a(\theta) \\
\tilde{l}_a(\theta) &= l^*_a(\theta)
\end{aligned}$$

$$
\begin{aligned}
\tilde{N}_{a+1}(\theta_1) &= N^*_{a+1}(\theta_1) - \epsilon \\
\tilde{N}_{a+1}(\theta_I) &= N^*_{a+1}(\theta_I) + \frac{\pi(\theta_i)}{\pi(\theta_I)}\epsilon \\
\tilde{N}_{a+1}(\theta) &= N^*_{a+1}(\theta) \quad \text{for all other } \theta
\end{aligned}
$$

$$
\tilde{h}_a(\theta) = h^*_a(\theta) \quad \theta = \theta_2, \ldots, \theta_{I-1}.
$$

We choose $\tilde{h}_a(\theta_1)$ and $\tilde{h}_a(\theta_I)$ such that the total number of people who survive to the next period are the same

$$
\begin{aligned}
P_a(\tilde{h}_a(\theta_1)) &= \frac{\tilde{N}_{a+1}(\theta_1)}{N_a(\theta_1)} = \frac{N^*_{a+1}(\theta_1)}{N_a(\theta_1)} - \frac{\epsilon}{N_a(\theta_1)} \\
P_a(\tilde{h}_a(\theta_I)) &= \frac{\tilde{N}_{a+1}(\theta_I)}{N_a(\theta_I)} = \frac{N^*_{a+1}(\theta_I)}{N_a(\theta_I)} + \frac{\pi(\theta_1)}{\pi(\theta_I)}\frac{\epsilon}{N_a(\theta_I)}
\end{aligned}
$$

We will show next that this allocation is feasible. In particular, we will show that the cost of delivering these health status for type θ_1 and θ_I is lower than that of the original allocation.

Note that because $P_a(\cdot)$ is concave, we have

$$
\begin{aligned}
P_a(\tilde{h}_a(\theta_1)) - P_a(h^*_a(\theta_1)) &= -\frac{\epsilon}{N_a(\theta_1)} > P'_a(\tilde{h}_a(\theta_1))(\tilde{h}_a(\theta_1) - h^*_a(\theta_1)) \\
P_a(\tilde{h}_a(\theta_I)) - P_a(h^*_a(\theta_I)) &= \frac{\pi(\theta_1)}{\pi(\theta_I)}\frac{\epsilon}{N_a(\theta_I)} > P'_a(\tilde{h}_a(\theta_I))(\tilde{h}_a(\theta_I) - h^*_a(\theta_I))
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\tilde{h}_a(\theta_1) - h^*_a(\theta_1) &< -\frac{\epsilon}{N_a(\theta_1)P'_a(\tilde{h}_a(\theta_1))} \\
\tilde{h}_a(\theta_I) - h^*_a(\theta_I) &< \frac{\pi(\theta_1)}{\pi(\theta_I)}\frac{\epsilon}{N_a(\theta_I)P'_a(\tilde{h}_a(\theta_I))}.
\end{aligned}
$$

Next, we will show that $\sum_\theta \pi(\theta)N_a(\theta)\tilde{h}_a(\theta) - \sum_\theta \pi(\theta)N_a(\theta)h^*_a(\theta) < 0$. Note that

$$
\begin{aligned}
&\sum_{\theta}\pi(\theta)N_a(\theta)\tilde{h}_a(\theta)-\sum_{\theta}\pi(\theta)N_a(\theta)h_a^*(\theta)\\
=\;& N_a(\theta_1)\pi(\theta_1)\left(\tilde{h}_a(\theta_1)-h_a^*(\theta_1)\right)+N_a(\theta_I)\pi(\theta_I)\left(\tilde{h}_a(\theta_I)-h_a^*(\theta_I)\right)\\
<\;& N_a(\theta_1)\pi(\theta_1)\left(-\frac{\epsilon}{N_a(\theta_1)P_a'(\tilde{h}_a(\theta_1))}\right)+N_a(\theta_I)\pi(\theta_I)\left(\frac{\pi(\theta_1)}{\pi(\theta_I)}\frac{\epsilon}{N_a(\theta_I)P_a'(\tilde{h}_a(\theta_I))}\right)\\
=\;& \epsilon\pi(\theta_1)\left(\frac{1}{P_a'(\tilde{h}_a(\theta_I))}-\frac{1}{P_a'(\tilde{h}_a(\theta_1))}\right)\le 0
\end{aligned}
$$

Since we assumed $h_a^*(\theta_1) > h_a^*(\theta_I)$, we can always find ϵ small enough such that $\tilde{h}_a(\theta_1) > \tilde{h}_a(\theta_I)$. Then, the last inequality follows from strict concavity of $P_a(\cdot)$.

So far we have shown that if $h_a^*(\theta_1) > h_a^*(\theta_I)$, there is an alternative allocation which is feasible, uses less resources and increases the value of the planner's objective, since

$$
V_{a+1}(N_{a+1}^*(\theta_1),\ldots,N_{a+1}^*(\theta_I);k_{a+1}^*) < V_{a+1}(N_{a+1}^*(\theta_1)-\epsilon,\ldots,N_{a+1}^*(\theta_I)+\frac{\pi(\theta_1)}{\pi(\theta_I)}\epsilon;k_{a+1}^*).
$$

This is a contradiction. Therefore, at the optimal allocation we must have $h_a^*(\theta_1) \le h_a^*(\theta_I)$. Now suppose, $h_a^*(\theta_1) = h_a^*(\theta_I)$. Consider the same allocation as above with ϵ very small.

$$
\begin{aligned}
\tilde{h}_a(\theta_1) &= P_a^{-1}\left(P_a(h_a^*(\theta_1))-\frac{\epsilon}{N_a(\theta_1)}\right)\\
\tilde{h}_a(\theta_I) &= P_a^{-1}\left(P_a(h_a^*(\theta_I))+\frac{\pi(\theta_1)}{\pi(\theta_I)}\frac{\epsilon}{N_a(\theta_I)}\right)
\end{aligned}
$$

Define a function $F(\epsilon)$ as

$$
\begin{aligned}
F(\epsilon) &= N_a(\theta_1)\pi(\theta_1)\left(\tilde{h}_a(\theta_1)-h_a^*(\theta_1)\right)+N_a(\theta_I)\pi(\theta_I)\left(\tilde{h}_a(\theta_I))-h_a^*(\theta_I)\right)\\
&= N_a(\theta_1)\pi(\theta_1)\left[P_a^{-1}\left(P_a(h_a^*(\theta_1))-\frac{\epsilon}{N_a(\theta_1)}\right)-h_a^*(\theta_1)\right]+\\
&\quad N_a(\theta_I)\pi(\theta_I)\left[P_a^{-1}\left(P_a(h_a^*(\theta_I))+\frac{\pi(\theta_1)}{\pi(\theta_I)}\frac{\epsilon}{N_a(\theta_I)}\right)-h_a^*(\theta_I)\right]
\end{aligned}
$$

$F(\epsilon)$ is the extra resources that the alternative allocation uses (relative to the optimal allocation). Note that

$$
F'(0) = -\pi(\theta_1)\frac{1}{P_a'(h_a^*(\theta_1))}+\pi(\theta_I)\frac{1}{P_a'(h_a^*(\theta_I))}=0.
$$

Therefore, if $h_a^*(\theta_1) = h_a^*(\theta_I)$, the perturbation has no first order effect on cost of health.

However, we know there is a first order effect on next period value function (again, since $V_{a+1}(N^*_{a+1}(\theta_1),\ldots,N^*_{a+1}(\theta_I);k^*_{a+1}) < V_{a+1}(N^*_{a+1}(\theta_1)-\epsilon,\ldots,N^*_{a+1}(\theta_I)+\frac{\pi(\theta_1)}{\pi(\theta_I)}\epsilon;k^*_{a+1}))$. This is a contradiction. Hence at the optimum we must have $h^*_a(\theta_1) < h^*_a(\theta_I)$.

So far we have shown that if

$$\begin{aligned}&V_{a+1}\left(N_{a+1}(\theta_1)-\varepsilon,...,N_{a+1}(\theta_I)+\frac{\pi(\theta_1)}{\pi(\theta_I)}\varepsilon;k_{a+1}\right) > \\ &V_{a+1}(N_{a+1}(\theta_1),...,N_{a+1}(\theta_I);k_{a+1}),\end{aligned}$$

then $h^*_a(\theta_1) < h^*_a(\theta_I)$. To complete the proof we need to show that, this also implies

$$\begin{aligned}&V_a\left(N_a(\theta_1)-\varepsilon,...,N_a(\theta_I)+\frac{\pi(\theta_1)}{\pi(\theta_I)}\varepsilon;k_a\right) > \\ &V_a(N_a(\theta_1),...,N_a(\theta_I);k_a),\end{aligned}$$

Consider the following allocation

$$\begin{aligned}\tilde{N}_a(\theta_1) &= N_a(\theta_1)-\epsilon \\ \tilde{N}_a(\theta_I) &= N_a(\theta_I)+\frac{\pi(\theta_1)}{\pi(\theta_I)}\epsilon \\ \tilde{N}_a(\theta) &= N_a(\theta) \quad \text{for all other } \theta \\ \tilde{c}_a(\theta) &= c^*_a \quad \text{for all } \theta \\ \tilde{h}_a(\theta) &= h^*_a(\theta) \quad \theta=\theta_2,\ldots,\theta_{I-1} \\ \tilde{l}_a(\theta) &= l^*_a(\theta) \quad \theta=\theta_1,\ldots,\theta_{I-1}\end{aligned}$$

We choose $\tilde{l}_a(\theta_I)$ such that the total output in period a is unchanged, i.e.

$$\tilde{l}_a(\theta_I) = \frac{N(\theta_I)\pi(\theta_I)w_a(\theta_I)l^*_a(\theta_I)+\epsilon\pi(\theta_1)w_a(\theta_1)l^*_a(\theta_1)}{\left(N(\theta_I)\pi(\theta_I)+\epsilon\pi(\theta_1)\right)w_a(\theta_I)}.$$

Note that

$$\begin{aligned}&\tilde{N}(\theta_I)\pi(\theta_I)w_a(\theta_I)\tilde{l}_a(\theta_I)+\tilde{N}(\theta_1)\pi(\theta_1)w_a(\theta_1)\tilde{l}_a(\theta_1) = \\ &N(\theta_I)\pi(\theta_I)w_a(\theta_I)l^*_a(\theta_I)+N(\theta_1)\pi(\theta_1)w_a(\theta_1)l^*_a(\theta_1).\end{aligned}$$

Finally, we choose $\tilde{h}_a(\theta_1)$ and $\tilde{h}_a(\theta_I)$ such that the new allocation is feasible and the total number of population in period $a+1$ is unchanged. Moreover, we are going to pick $\tilde{h}_a(\theta_1)$ and $\tilde{h}_a(\theta_I)$ such that $P_a(\tilde{h}_a(\theta_1))(N_a(\theta_1)-\epsilon) < N^*_{a+1}(\theta_1)$ and $P_a(\tilde{h}_a(\theta_I))(N_a(\theta_I)+\frac{\pi(\theta_1)}{\pi(\theta_I)}\epsilon) > N^*_{a+1}(\theta_I)$. Let

$$P_a(\tilde{h}_a(\theta_1)) = \frac{N^*_{a+1}(\theta_1)}{N_a(\theta_1) - \epsilon} - \delta_1$$
$$P_a(\tilde{h}_a(\theta_I)) = \frac{N^*_{a+1}(\theta_I)}{N_a(\theta_I) + \frac{\pi(\theta_1)}{\pi(\theta_I)}\epsilon} + \delta_I$$

for some positive δ_1 and δ_I. To make sure the total number of people in period $a+1$ is unchanged we must choose δ_1 and δ_I such that

$$\pi(\theta_1)(N_a(\theta_1) - \epsilon)\delta_1 = \pi(\theta_I)\left(N_a(\theta_I) + \frac{\pi(\theta_1)}{\pi(\theta_I)}\epsilon\right)\delta_I$$

Moreover, note that $\frac{N^*_{a+1}(\theta_1)}{N_a(\theta_1)-\epsilon} > P_a(h^*_a(\theta_1))$ and $\frac{N^*_{a+1}(\theta_I)}{N_a(\theta_I)+\frac{\pi(\theta_1)}{\pi(\theta_I)}\epsilon} < P_a(h^*_a(\theta_I))$. Therefore, we can choose δ_1 and δ_I small enough so that $P_a(\tilde{h}_a(\theta_1)) > P_a(h^*_a(\theta_1))$ and $P_a(\tilde{h}_a(\theta_I)) < P_a(h^*_a(\theta_I))$. Note that, $P_a(\tilde{h}_a(\theta_1))$ and $P_a(\tilde{h}_a(\theta_I))$ have the same average as $P_a(h^*_a(\theta_1))$ and $P_a(h^*_a(\theta_I))$ and are more concentrated. Also, note that $P_a(\cdot)$ is an strictly concave and monotone function. Therefore, $P_a^{-1}(\cdot)$ is a strictly convex function. Since $P_a(\cdot)$ is strictly increasing and concave, $P_a^{-1}(\cdot)$ is a strictly convex function. This function has a lower average over $P_a(\tilde{h}_a(\theta_1))$ and $P_a(\tilde{h}_a(\theta_I))$ than $P_a(h^*_a(\theta_1))$ and $P_a(h^*_a(\theta_I))$, i.e.,

$$\tilde{N}(\theta_I)\pi(\theta_I)\tilde{h}_a(\theta_I) + \tilde{N}(\theta_1)\pi(\theta_1)\tilde{h}_a(\theta_I) < N(\theta_I)\pi(\theta_I)h^*_a(\theta_I) + N(\theta_1)\pi(\theta_1)h^*_a(\theta_1).$$

Therefore, the alternative allocation is feasible.

Note that from our assumption we know

$$V_{a+1}(N_a(\theta_1)P_a(\tilde{h}_a(\theta_1)), \ldots, N_a(\theta_I)P_a(\tilde{h}_a(\theta_I)); k^*_{a+1}) > V_{a+1}(N^*_{a+1}(\theta_1), \ldots, N^*_{a+1}(\theta_I); k^*_{a+1})$$

Using similar argument that we used for problem in age $a_{ret} - 1$ we can show that

$$\sum_\theta \pi(\theta)\tilde{N}_a(\theta)v(1 - \tilde{l}_a(\theta)) < \sum_\theta \pi(\theta)N_a(\theta)v(1 - l^*_a(\theta)).$$

Therefore the proof is complete. ■

B Estimation of Survival Production Function

The version of (9) that we estimate is:

$$x_{a,t,j} = A_a(z_t h_{a,t,j} s_{a,t})^{\eta_a}$$

Where $x_{a,t,j}$ is the inverse of non-suicide, non-homicide mortality rate for agents of age a at time t and in group j; $h_{a,t,j}$ are health expenditures z_t is aggregate productivity at time t and $s_{a,t}$ are other sources that influence mortality different from health expenditures. Our objective is to estimate A_a and η_a, for each age a. The estimation procedure follows Hall and Jones (2007) closely. One key difference are the observations used for the estimation. Hall and Jones (2007) use variation in mortality and expenditures across years using data from 1950 to 2000 (at five years interval). In our case we have cross sectional data over time only for years 1996 to 2008.[28] To introduce additional variation we partition our sample in group conditional on observable demographic characteristics. These include race, gender and us census region of residence. Taking logs of the previous equation we get

$$\log \tilde{x}_{a,t,j} = \log A_a + \eta_a \Bigg[\log z_t + \log h_{a,t,j} + \log s_{a,t} \Bigg] + \epsilon_{a,t,j}, \quad \forall a, t, j \tag{14}$$

To identify A_a and η_a we remove the effects due to z_t and $s_{a,t}$. Let $\log s_{a,t} = g_{s,a} t + \gamma_{a,t}$. We estimate $g_{s,a}$ looking at the average health expenditure growth in our time periods. Substituting in the above we have

$$\log \tilde{x}_{a,t,j} = \log A_a + \eta_a \Bigg[\log z_t + \log h_{a,t,j} + g_{s,a} t \Bigg] + \widehat{\epsilon}_{a,t,j}, \quad \forall a, t, j \tag{15}$$

Where $\widehat{\epsilon}_{a,t,j} = \epsilon_{a,t,j} + \eta_a \gamma_{a,t}$. Differencing over time (15) and taking expectations we have

$$E_t[\Delta \log \tilde{x}_{a,t,j}] = \eta_a E_t \Bigg[\Delta \log z_t + \Delta \log h_{a,t,j} + g_{s,a} \Bigg], \forall a, t \tag{16}$$

where for all a, t, j we have assumed that $E_t[\widehat{\epsilon}_{a,t+1,j} - \widehat{\epsilon}_{a,t,j}] = 0$. The key identifying assumption is that

$$(1-\mu) E_t[\Delta \log \tilde{x}_{a,t,j}] = g_{s,a} \eta_a, \qquad \forall a, t, j$$

With $\mu = 2/3$. Substituting into (16) and taking expectations over time we have for all a:

$$g_{s,a} \eta_a = (1-\mu) \eta_a E_t \Bigg[\Delta \log z_t + \Delta \log h_{a,t,j} + g_{s,a} \Bigg] \rightarrow g_{s,a} = \frac{(1-\mu)}{\mu} E_t \Bigg[\Delta \log z_t + \Delta \log h_{a,t,j} \Bigg]$$

Given the value of $g_{s,a}$ we can estimate A_a and η_a.

C The MEPS Data Set

The Medical Expenditure Panel Survey (MEPS) is a large-scale panel administered by the U.S. department of Health & Human Services since 1996. The panel is designed to be representative of the non institutionalized population in the U.S.. The data set collects information on what health services are used, how frequently are used and what are the

[28] In the estimation we also include data for 1987 taken from the National Medical Expenditure Survey 1987. This is a precursor to the MEPS discussed in this paper and shares the same structure. Additional information on NMES87 is at: http://wonder.cdc.gov/.

expenses incurred in using these services. The MEPS data is one of the main source for the National Health Expenditure Accounts (NHEA). Our main source is the household component of the data set (MEPS-HC) this component is designed with an overlapping structure. Each household is interviewed 5 times over the course of 2 years. Roughly each wave contains roughly 30.000 individuals for a total of around 12.000 families. In each interview data is collected on demographic variables, income variables, hours worked, various measures of self reported health and specific questions on health conditions. In addition detailed information on health services usage charges incurred and expenditures (including out of pocket) on these services.[29] This data is complemented with information obtained by the individual's medical providers. In particular the following payment sources are included in the MEPS:

1. Out of pocket by patient or patient's family;
2. Medicare and Medicaid;
3. Private Insurance including automobile, homeowner's, liability insurances
4. TRICARE and Veterans' administration;
5. Other federal, state and local sources;
6. Workers compensation.

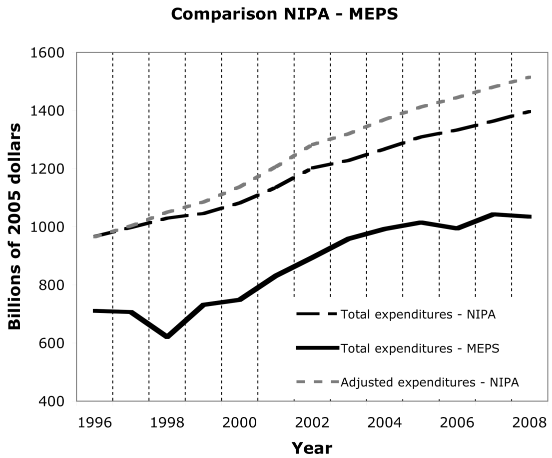
C.1 Comparison with aggregate data

We now look at the relation between the MEPS dataset and aggregate U.S. data. Our interest is to compare total level of health expenditures. For each family in MEPS we are provided with frequency weights designed to make the population comparable with the Current Population Survey. We aggregate MEPS data using these weights. Our aggregate analogue is taken from NIPA. Data on health expenditures is taken from table 2.4.5u. Each variable is deflated to 2005 dollars using it's own deflator provided in table 2.4.4u. For NIPA we report total personal consumption expenditure of health care (line 168). MEPS data is not designed to be directly comparable with aggregate measures. To alleviate discrepancy we also report an adjusted measure of health expenditures more aligned with MEPS. To do so we subtract expenditure on nursing home (line 183) and add expenditures on prescription drugs (line 120). Results are displayed in Figure 18(a). We also include, separately, information on dental expenditures (line 171) and prescription drugs expenditures for both NIPA and MEPS. These series are in plotted in Figure 18(b).

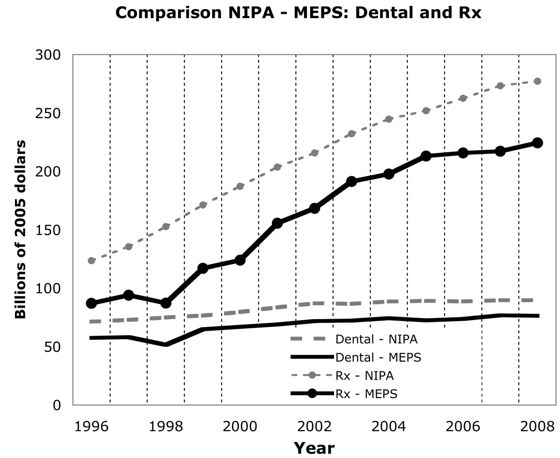
The main observation is that MEPS data underestimates the levels of aggregate data systematically. For an additional analysis of this discrepancy refer to Selden, Levit, Cohen, Zuvekas, Moeller, McKusick, and Arnett (2001) and Hartman, Kornfeld, and Catlin (2010).[30]

[29]Information on over the counter medicines is not included in the MEPS definition of expenditures.

[30] Key differences between MEPS and NHEA are: nursing homes; long term care (greater than 45 days); non-community, non-Federal Hospitals and Alternative care; government spending on health care not administered in hospitals (schools); expenditures for institutionalized population; expenditures of active duty military personnel and long term care in VA hospitals; foreign visitors to the United States.



(a) Total expenditures.



(b) Expenditures on dental and prescription drugs.

FIGURE 18: Relation between MEPS aggregated data and aggregate NIPA data.

C.2 Comparison with the the CEX

As a final comparison for the MEPS we look at another source that reports comparable health expenditure measures. The consumer expenditure survey (CEX) reports out of pocket expenditures incurred by households. Out of pocket expenditures do not include health insurance premiums. We use the CEX data set prepared by Heathcote, Perri, and Violante (2010).[31] In Table 3 we report comparisons between the MEPS and our calculation the CEX.

	MEPS (55-64)	CEX (55-64)	MEPS (65-74)	CEX (65-74)
Income before taxes	$58,707	$69,777	$45,953	$44,467
Age of reference person	59.1	58.3	69.2	69.1
Persons in reporting unit	2.15	2.26	1.85	1.7
Out of pocket	$1,949.08	$2,213.32	$2,158.4	$2,387.82
Out of pocket on drugs	$931.6	$712.73*	$1,225.9	$956.26*

TABLE 3: Comparison MEPS - CEX. Year 2005, age of reference person 55 to 64. For the MEPS we discard individual with negative reported family income, for the CEX, we discard individual that complete less than 4 interviews. Out of pocket expenditures on drugs for the CEX are taken from Duetsch (2008).

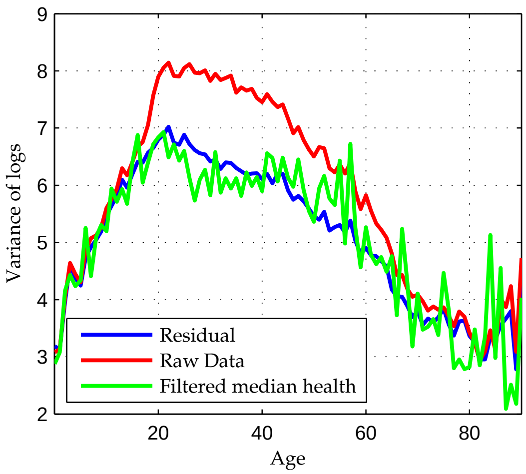
We conclude that although the MEPS is not directly comparable with aggregate US data, it is close to another widely used survey data: the CEX.

D Alternative Measures of Health Spending Inequality

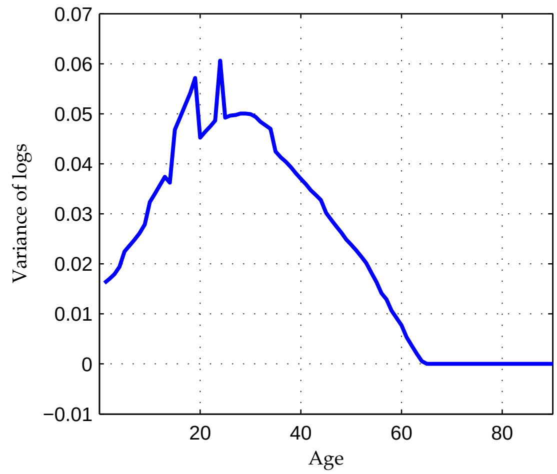
In the main body of the paper, we report comparisons between data and model quantities for spending on health. We focused on only two aspects of this, the average level of spending

[31] We use Sample A from http://www.economicdynamics.org/RED-cross-sectional-facts.htm. Also refer to Duetsch (2008).

by age, and the decomposition of this average level by income quartiles into shares. Here, we report other measures of inequality. We include information on raw levels of inequality, median spending by income by age and regression coefficients of spending on income (with other controls also included). We begin with the level of spending inequality by age as measured by variance of logs. This is shown in Figure 19(a).



(a) Total sample, individuals with reported excellent health, and controlling by observables: MEPS data.



(b) Model.

FIGURE 19: Variance of log health expenditures over ages.

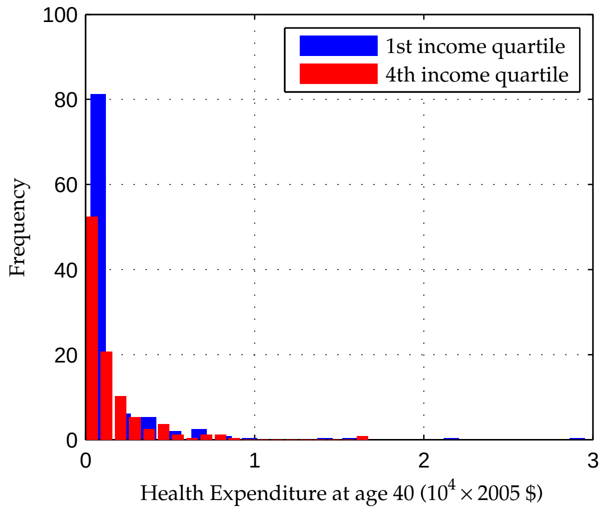
The first thing to note is that, as measured by unconditional variances, health spending is quite unequal across individuals. We now attempt to isolate the effect of income on health expenditure removing the effect due to health conditions and observed heterogeneity. We consider the following log-linear relationship between health expenditures and health status and other observables:

$$\log h_{i,a,t} = \sum_{j=1,2} \alpha_j^1 I_{i,a,t}^S(j) + \sum_{j=1,6} \alpha_j^2 I_{i,a,t}^R(j) + \sum_{j=1,4} \alpha_j^3 I_{i,a,t}^{reg}(j) + \sum_{j=1,5} \alpha_j^4 I_{i,a,t}^H(j) + \sum_{j=1,2} \alpha_j^5 I_{i,a,t}^{lim}(j) + \sum_{j=1,8} \alpha_j^6 I_{i,a,t}^{edu}(j) + \sum_{j=1,13} \alpha_j^7 I_{i,a,t}^{year}(j) + \varepsilon_{i,a,t}. \quad (17)$$

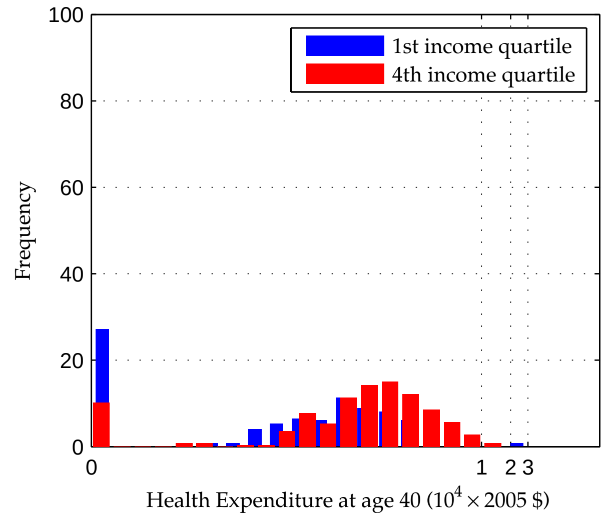
Where $h_{i,a,t}$ is total health expenditures on individual i of age a at time t. I denote a complete set of dummy variables. In particular: I^S is gender, I^R race, I^{reg} census region of residence, I^H self reported health status,[32] I^{lim} any reported limitation, I^{edu} highest education attained, I^{year} year of interview. The quantity $\varepsilon_{i,a,t}$–the residual– represents the deviation in the quantity of health expenditure for the individual from the mean of level of spending among others in the same category. In Figure in Figure 19(a), are shown: a) the variance of log expenditure conditioning only on age (red), b) the variance of log expenditure

[32] Since the annualized self reported health is a continuous variable we calculate the rounded measure of health in order to compute dummies.

restricting to individuals with median self reported health (green), c) the variance of the residual ε from (17) over age (blue).



(a) Levels of health expenditures.



(b) Log of health expenditures

FIGURE 20: Histogram of health expenditure for top wage quartile and bottom wage quartile at age 40.

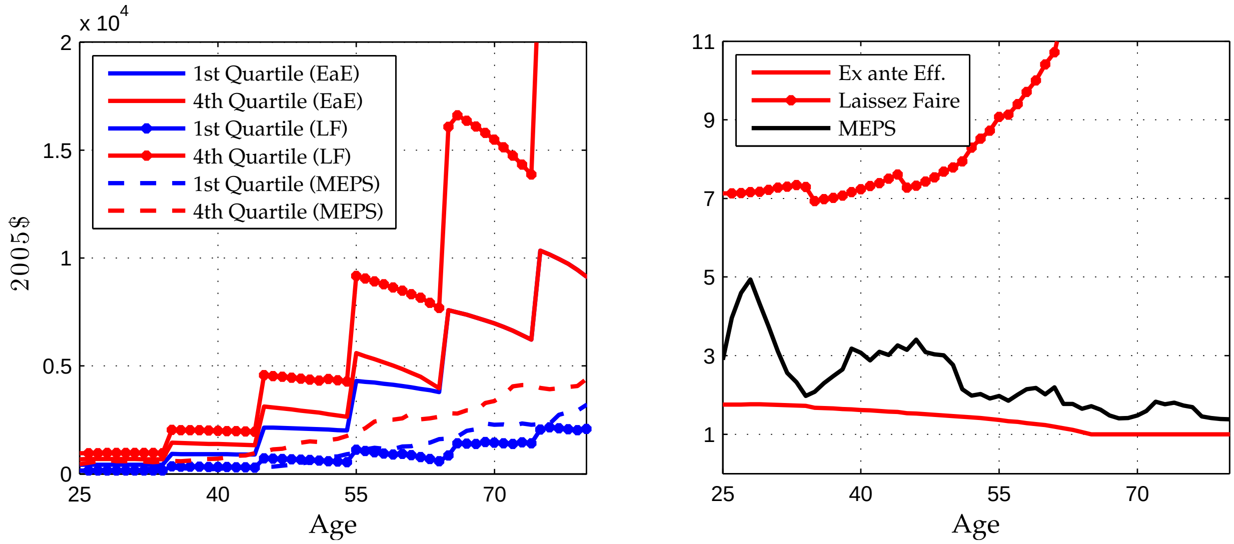
As can be seen from this figure, inequality in health expenditures is large even controlling for health status or removing observed heterogeneity. As a point of comparison, here it's around six whereas with income it is around one. Second, inequality in health expenditures rises early in life – up to about age 20 – and then slowly decreases over time through both the working life and retirement. In Figure 19(b) we plot the variance of logs from the model. The efficient allocation in the model implies diminishing inequality as we observe in the data. However, levels are dramatically different. Much of this extra inequality in the data corresponds to unobserved individual level heterogeneity.

D.1 Median health expenditures

In the main text, we restricted attention to mean health by age and by income type and found that model (ex ante efficient) and data quantities differ (particularly in older ages) but not by large amounts. This finding is due, in part, to our focus on means. To see this, we show the distribution of health expenditures at a given age. Figures 20(a) and 20(b) show the histogram of health expenditures at age 40. As can be seen, the distribution of spending at the individual level has two distinctive features. The first of these is that for a large fraction of the population spending is at, or near, $0. The second is that there are a (very) small number of observations with very high expenditures.[33]

[33]Moreover, there are more of the second type of these observations among people with low incomes. This is in keeping with what we have seen in the main body of the paper where mean spending is higher for the low income quartile than it is for the high income quartile. What these histograms show is that this is because of a small number of very sick people at the low end of the income distribution. (It is not surprising that their income is low given this.)

Given the highly skewed distribution of health expenditures in this section we focus on the median of health expenditures at a given age for a given income group. We begin showing medians for the top and bottom quartiles in both model allocations and in the data. In the model, there are no reasons for means and medians to differ. In the data, median expenditure levels are significantly lower than model quantities, particularly at higher ages. This is shown in Figure 21(a).



(a) Median spending by Income: top and bottom quartiles. ex ante efficient, Laissez Faire and MEPS data.

(b) Ratio of median health spending on top income quartile relative to spending at bottom income quartile: ex ante efficient, laissez faire and MEPS data.

FIGURE 21: Median health expenditures.

It is also true that there is more inequality in the data when measured as the ratio of medians. This is shown in Figure 21(b) where the ratio for the data is around 3 at lower ages. Thus, by this measure, the 'excess' inequality in the data (over the ex ante efficient allocation) is significantly higher.

D.2 The relation between health expenditures and income

Lt h_i be total health expenditures for individual i. Assume that

$$h_i = h(a_i, y_i, h_i^{rth}, X_i, \varepsilon_i), \tag{18}$$

where a_i is age, y_i is productivity, h_i^{rti} is self reported health status. X_i is a vector of observable characteristics and ε_i is a vector of unobservable characteristics (including measurement error). In the body of the paper the focus on averages within subgroups and controlling on health status. This procedure has the advantage of requiring little structure on h.[34] The downside of this procedure is that it forces us to drop a significant number of observations. We now proceed in the opposite way assuming a precise structure on h and maintaining all of our observations. We assume that h is either a linear or log linear function of its argument. The relation between income and health expenditure can then be easily recovered. Consider the following relation

$$h_i = \beta_0 + \alpha y_i + \beta_1 age_i + \beta_2 \frac{age_i^2}{100} + \beta_3 h_i^{rth} + \varepsilon_i \tag{19}$$

where h_i are health expenditures for individual i and y_i is either wages or income of individual i and h_i^{rth} is the annualized level of self reported health. As discussed by Albouy, Davezies, and Debrand (2009) (looking at French health expenditure data) the estimation of the above equation is challenging for two key reasons. The first reason is that a large number of individuals report zero health expenditures. The presence of zeros in the data (censoring) biases the estimate of α downwards. The second reason is that the distribution of health expenditures among those reporting positive spending is highly skewed (as reported in Figure 20(a)). Following the bulk of the literature we consider two alternative strategies to alleviate this problem. The first one is to perform a log transformation:[35]

$$\log h_i = \beta_0 + \alpha^{\log} \log y_i + \beta_1^{\log} age_i + \beta_2^{\log} \frac{age_i^2}{100} + \beta_3^{\log} h_i^{rth} + \varepsilon_i \tag{20}$$

Alternatively we can directly attack the problem of zero expenditures. Following Albouy, Davezies, and Debrand (2009), we can determine the relation between income and health expenditure in a two stage process estimating via a Tobit model. That is considering the following relation

$$h_i^{data} = \max\{0, h_i\}; \quad h_i = \alpha y_i + \beta X_i + \varepsilon_i \tag{21}$$

where h^{data} are actual health expenditures observed in the data and h is a latent variable that depends linearly on the covariates. In Table 4, we report four alternative estimates of the sensitivity of health spending to productivity: i) linear regression on levels; ii) linear regression on log-levels; iii) Tobit regression on levels; iv) Tobit regression on log-levels.[36]

[34] Using the notation in (18) we consider the following two variables

$$H^+(a) = E_i\left[h(a_i, y_i, h_i^{rth}, X_i, \varepsilon_i)\middle| y_i \geq y_a^{75}, a_i = a, h_{45}^{rth}(a) \leq h_i^{rth} \leq h_{55}^{rth}(a)\right]$$

$$H^-(a) = E_i\left[h(a_i, y_i, h_i^{rth}, X_i, \varepsilon_i)\middle| y_i \leq y_a^{25}, a_i = a, h_{45}^{rth}(a) \leq h_i^{rth} \leq h_{55}^{rth}(a)\right]$$

where y_a^{25}, y_a^{75} are respectively the 25th and 75th percentile of the wage distribution at age a; $h_{45}^{rth}(a)$, $h_{55}^{rth}(a)$ are respectively the 45th and 55th percentile of the self reported health distribution at age a. Under the assumption that $h(a_i, y_i, h_i^{rth}, X_i, \varepsilon_i) = h_1(a_i, y_i, h_i^{rth}) + h_2(X_i, \varepsilon_i)$ for some h_1 and h_2 and that the conditional expectation of h_2 is the same in each subgroup, we can compare the ratio of H^+ and H^-) with the equivalent ratio in the model.

[35] We add 1$ to each individual expenditures to avoid zeros.

[36] In this last case the censoring is at 1 rather than 0.

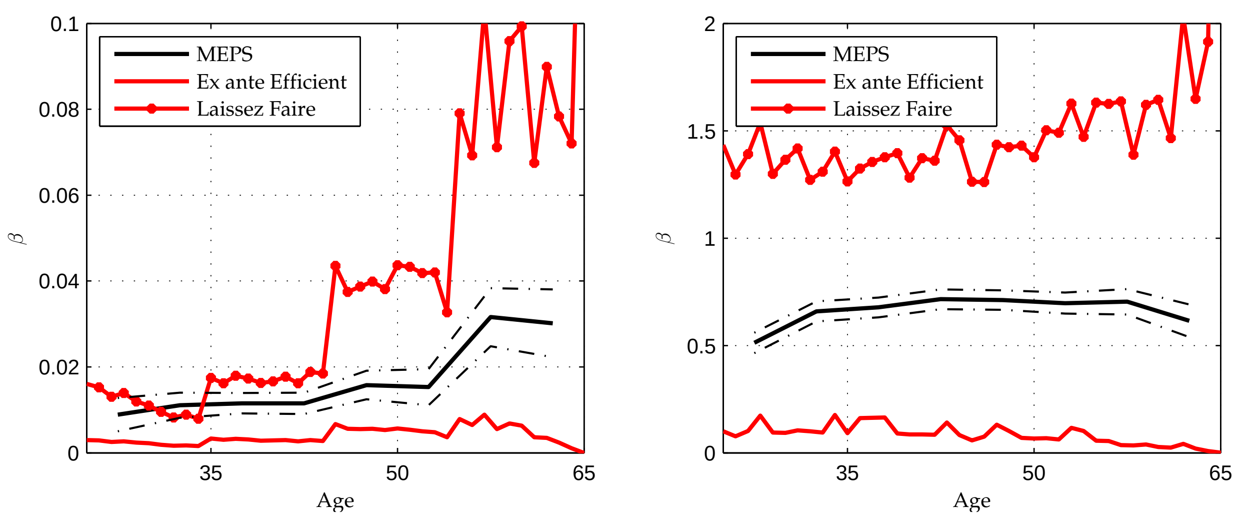
Each of these is reported for two alternative proxies for productivity: income and wages.

Estimation Type	α	Proxy for Productivity
Robust regression over Levels	0.0051 (.0001)	Income
Robust regression over Logs	0.1971 (.0055)	Income
Tobit over Levels	0.0263 (.0009)	Income
Tobit over Logs	0.2389 (.0065)	Income
Robust regression over Levels	0.0053 (.0001)	Wage
Robust regression over Logs	0.740 (.01)	Wage
Tobit over Levels	0.0314 (.0008)	Wage
Tobit over Logs	0.8383 (.0117)	Wage

TABLE 4: Estimation results from equation (19), (20) and (21).

For each of the estimates the relation between income/wage and health expenditures is positive. The main difference here is whether income or wages is used for the proxy for productivity – the estimated coefficients are uniformly and significantly larger when wages are used. Summarizing, a 1% increase in productivity as measured by wages is associated with about a .75% or .8% increase in spending on health.

Shown next are the regression coefficients, by age in both the data and in the model allocations. These are shown in both levels and logs in Figures 22(a) and 22(b) respectively.



(a) Estimation of equation (19)

(b) Estimation of equation (20)

FIGURE 22: Elasticity of health expenditures relative to income. Model vs MEPS data.

We observe that across all ages there relation between income and health expenditures is stronger in MEPS that under the ex ante efficient allocation. However this relation is smaller than what we observe within the Laissez Faire allocation.