

# HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

---

THE UNIVERSITY *of* York

WP 12/05

## Empirical evidence for decreasing returns to scale in a health capital model

Titus J. Galama  
Patrick Hulle  
Erik Meijer  
Sarah Outcault

April 2012

# Empirical evidence for decreasing returns to scale in a health capital model\*

Titus J. Galama<sup>a</sup>, Patrick Hulle<sup>b</sup>, Erik Meijer<sup>a,†</sup>, and Sarah Outcault<sup>a</sup>

<sup>a</sup>*RAND Corporation*

<sup>b</sup>*Tilburg University and VU University Amsterdam*

February 28, 2012

## Abstract

We estimate a health investment equation, derived from a health capital model that is an extension of the well-known Grossman model. Of particular interest is whether the health production function has constant returns to scale, as in the standard Grossman model, or decreasing returns to scale, as in the Ehrlich-Chuma model and extensions thereof. The model with decreasing returns to scale has a number of theoretically and empirically desirable characteristics that the constant returns model does not have. Although our empirical equation does not point-identify the decreasing returns to scale curvature parameter, it does allow us to test for constant versus decreasing returns to scale. The results are suggestive of decreasing returns and in line with prior estimates from the literature. But when we attempt to control for the endogeneity of health by using instrumental variables, the results become inconclusive. This brings into question the robustness of prior estimates in this literature.

*Keywords:* health investment; lifecycle model; Grossman model; optimal control

*JEL:* I12; J24; D91

---

\*We thank Jeff Tanner for assisting with data preparation and Isaac Ehrlich, Brian Ferguson, Michael Hurd, Arie Kapteyn, Tobias Klein, Peter Kooreman, Susann Rohwedder, Eddy van Doorslaer, and seminar participants at the RAND Corporation, Tilburg University, and the 20th European Workshop on Econometrics and Health Economics for stimulating discussions and constructive comments from which we have greatly benefited. We thank the US National Institute on Aging for funding under research grants R01 AG037398 and R01 AG030824.

†Corresponding author. RAND Corporation, 1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138, meijer@rand.org

# 1 Introduction

The canonical model of the demand for health and health investment (e.g., medical care) arises from Grossman (1972a, 1972b, 2000). In Grossman's human capital framework individuals invest in health (e.g., invest time and consume medical goods and services) for the consumption benefits (health provides utility) as well as production benefits (healthy individuals have higher earnings) that good health provides. The model provides a conceptual framework for interpretation of the demand for health and medical care in relation to an individual's resource constraints, preferences, and consumption needs over the life cycle.

While Grossman's model has great theoretical and intuitive appeal and has led to a rich body of literature and many useful insights in health economics, it also has several limitations. For example, (i) in empirical work it is generally found that health and the demand for medical care are negatively related, whereas Grossman's model appears to predict a positive relationship (Wagstaff, 1986a; Zweifel & Breyer, 1997; Galama & Kapteyn, 2011); (ii) empirically, health declines faster for individuals with lower socio-economic status, and the model does not predict this (Case & Deaton, 2005); and (iii) even though the model assumes forward looking rational agents, the model solutions do not depend on past or future values of exogenous variables or initial health and wealth (e.g., Usher, 1976); for example, structural and reduced form equations for health depend only on present time conditions, such as the current wage rate and current prices, see (42), (45), and (47) in Grossman (2000). The lack of history in the model's solution implies that, whatever the initial health state, health immediately jumps to its equilibrium path.

A particularly devastating criticism appears to be the claim of Ehrlich and Chuma (1990) and Galama (2011) that the model does not have a unique solution, and that in particular the equilibrium path typically used as the model's solution is in fact not a correct solution. This would invalidate many theoretical and empirical analyses based on the model. However, this claim has been disputed by Ried (1998) and Grossman (1998, 2000). Grossman (2000) also provides a review and rebuttal of some of the other limitations mentioned. It is not the aim of this paper to revisit this theoretical debate and we refer the interested reader to these authors.

Despite the limitations, theoretical extensions and competing economic models are still relatively few. Promising adaptations of the model are the models of Ehrlich and Chuma (1990) and Galama (2011), who have extended the Grossman model to include a health production process that is characterized by decreasing returns to scale (DRTS), whereas the standard model assumes a linear health production function with constant returns to scale (CRTS). Substantively, DRTS is appealing, because great improvements in health can be made with low levels of health investment (e.g., improving sanitation), whereas at high levels of health investment, very expensive treatments often provide only a relatively small improvement in health (e.g., Perloth, Goldman, & Garber, 2010), although this may still be economically valuable (e.g., Goldman et al., 2010). It is generally understood that health production is subject to the law of diminishing returns (e.g., Wagstaff, 1986b). A model incorporating decreasing returns may thus provide a more realistic representation of real-world health production processes. For decreasing returns, the marginal cost of investment is found to be an increasing function of investment: the marginal cost of investment is higher for higher levels of investment, because the higher the level of investment, the smaller the health gain. As a result, healthy and high socio-economic status individuals face different effective health costs.

In addition to these direct substantive advantages, introducing decreasing returns also removes the limitations of the Grossman model mentioned above (Galama, 2011): The model with DRTS predicts (i) a negative correlation between health investment and health; (ii) that the wealthy and educated live longer and experience slower declines in health; and (iii) that current health status is a function of the initial level of health and the histories of prior health investments made. Furthermore, the DRTS model predicts the empirical stylized fact that health investment rapidly increases near the end of life and that length of life is finite. Finally, because the Hamiltonian is strictly concave and the first-order condition for health investment is monotonically increasing in health investment, it is uncontroversial that the model solution for DRTS is finite and unique.

Empirical tests of the health production literature have thus far been based on the equilibrium equation derived under the assumption of a linear health production process (e.g., Grossman, 1972b; Wagstaff, 1986a). In this paper we test the predictions of a theory of health capital with decreasing returns to scale in health production. To this end, we employ an equation for health investment that was derived by Galama (2011). We estimate this equation using the Panel Study of Income Dynamics (PSID) and contrast our findings with those of a relatively small existing empirical literature.

Our contribution is as follows. First, to the best of our knowledge, no prior attempts have been made to empirically confront the predictions of a theory of health capital with decreasing returns to scale. Second, we carefully account for the endogenous nature of health in the demand for medical care. Only a few papers in the empirical literature have estimated direct relationships between medical care and health, and none of the papers that have tested the predictions of health capital theory have attempted to address the inherent endogeneity of health.

We proceed as follows. In section 2 we give a concise overview of prior empirical work. In section 3, we describe the structural model on which our estimates are based. Section 4 describes our empirical strategy. Section 5 describes the data we use and some measurement issues. Section 6 presents the estimation results and section 7 concludes. The appendix contains more detailed information about the variables used and model specifications.

## **2 Prior empirical tests of health capital theory**

In this section, we describe previous empirical results for the Grossman model and its extensions. Reduced form models have been estimated by Grossman (1972a), Van de Ven and Van der Gaag (1982), Wagstaff (1986a, 1993), Leu and Doppmann (1986), Leu and Gerfin (1992), Van Doorslaer (1987), Erbsland, Ried, and Ulrich (1995), Nocera and Zweifel (1998), Gerdtham, Johannesson, Lundberg, and Isacson (1999), and Gerdtham and Johannesson (1999). These papers use a large variety of methodologies and data from diverse cultural and institutional environments. Despite this, the studies are broadly in agreement with one another and in line with the predictions of the Grossman model. Health is found to increase with income (wages, lifetime earnings) and education, and decrease with age, the price of medical goods and services, and with environmental factors such as physically and mentally demanding work environments, manual labor, and psychological stress factors. In addition, these studies find that health is better among individuals

who participate in sports and have healthy eating and sleeping habits, and is lower for individuals who are overweight and who smoke. Singles have worse health than married individuals and females have worse health than males. Furthermore, moderate alcohol consumption is found to have either a positive or a negligible association with health.

In contrast to the relative abundance of estimates of reduced form equations, few studies have estimated structural equations.<sup>1</sup> The only ones we are aware of are Van de Ven and Van der Gaag (1982), Wagstaff (1986a), Erbsland et al. (1995), and Nocera and Zweifel (1998). As is the case with the relation we estimate, the structural equations for health investment contain health as an explanatory variable, whereas the reduced form equations do not, because the endogenous variable health is substituted out of the equation. The structural relations also provide a prediction for the sign of the relationship between health investment and health.

While the reduced form estimates are generally in agreement with the predictions of the Grossman model, this is not true for estimates of structural equations. Van de Ven and Van der Gaag (1982), Wagstaff (1986a), and Erbsland et al. (1995) find statistically significant negative relations between health investment and health. This contrasts sharply with the theory, which predicts that this relationship is positive under the assumption of a linear health production process as is widely utilized in the health production literature; see, for example, (13) in Wagstaff (1986a). A partial exception is Nocera and Zweifel (1998), who report a negative but insignificant relation between health investment and health for their sample I, but a positive and statistically significant relation for their sample II. However, the significance of the latter is likely inflated by the use of deterministically interpolated data in sample II. Thus, taken together, the evidence from these studies seems to be in the direction of a negative relation, as predicted by the DRTS model, as opposed to the prediction of a positive relation in the equilibrium equation of the CRTS model.

An important limitation of all of these studies is that they do not account for the endogeneity of health. Health is an endogenous variable, because health and health investment are joint outcomes of the lifetime utility maximization problem. Therefore, in addition to estimating regression models that take health as exogenous, we will estimate equations that account for the endogeneity of health through instrumental variables methods.

### **3 Structural health capital model**

Our approach is based on the theoretical model presented in Galama (2011), which is an extension of the Grossman model in that it allows for a decreasing returns to scale health production process. Here we highlight the main features of the model. For consistency and comparability with prior work we adopt the usual assumptions made in this literature, mostly those made by Grossman (1972a, 1972b, 2000); Cropper (1977) and Wagstaff (1986a). Our review of the literature shows that the assumptions made by these authors are commonly made in the Grossman literature. Whenever we deviate from the literature, we do so in order

---

<sup>1</sup>The health capital literature uses the terminology *structural equation* (or relation) for an equation that contains some endogenous explanatory variables. While we follow this terminology, it must be noted that these structural equations are often derived equations and not the original model equations that define the structural model. Furthermore, the coefficients in these structural equations are generally functions of the structural parameters, but not structural parameters themselves.

to relax assumptions rather than impose additional restrictions. We discuss the implications of the important assumptions in some detail. Many assumptions in the literature are made for convenience, generally to obtain linear expressions for empirical testing. Thus, the major assumption that the extant literature and this work rely on is that such linear expressions are reasonable approximations of the model. The assumptions we adopt from the literature in this section mainly serve to contrast predictions between our theoretical results and those of the larger literature. However, in our empirical specification, we do not impose the sign restrictions that are predicted by the theoretical model.

Individuals are assumed to derive utility from consumption  $C_t$  and health  $H_t$ . The *period utility* or *instantaneous utility* is denoted by  $U(C_t, H_t)$ . Individuals live for  $T$  periods, starting with period 0. Individuals maximize their lifetime utility  $V_0$ , which is additively separable, and discount future utility at a per-period rate of  $\beta$ . Thus, lifetime utility takes the form

$$V_0 = \sum_{t=0}^{T-1} \left( \frac{1}{1+\beta} \right)^t U(C_t, H_t).$$

We use the isoelastic utility function

$$U(C_t, H_t) = \frac{1}{1-\rho} (C_t^\zeta H_t^{1-\zeta})^{1-\rho},$$

where  $\zeta$  is the relative “share” of consumption versus health in the utility function ( $0 \leq \zeta \leq 1$ ). This utility function is also known as the *power utility* function and as the *constant relative risk aversion* (CRRA) utility function, where  $\rho > 0$  is the *coefficient of relative risk aversion*. However, because our model is deterministic, the individual does not face any risk. The parameter  $\rho$  thus reflects preference for utility smoothing across time, rather than risk aversion. The isoelastic utility function is arguably the most widely used utility function in intertemporal utility maximization problems (e.g., Hansen & Singleton, 1982; Attanasio & Weber, 1989; Rust & Phelan, 1997; Keane, 2011). It is increasing and strictly concave, and the parameter  $\rho$  allows a wide range of curvature, which translates into flexibility regarding the amount of utility smoothing across time. See Wakker (2008) for an extensive discussion of its properties and Chiappori and Paiella (2011) for empirical support for this functional form.

When utility depends on multiple components (typically consumption and leisure in labor supply models, consumption and health in our case), the period utility can be specified as additive or multiplicative in these components (see, e.g., MaCurdy, 1981, pp. 1064–1065). Our multiplicative specification, which most closely follows French (2005), is able to account for the observation that the marginal utility of consumption declines as health deteriorates (e.g., Finkelstein, Luttmer, & Notowidigdo, 2008). This would rule out the strongly separable (additive) functional form for the utility function typically employed in the literature (e.g., Wagstaff, 1986a), where the marginal utility of consumption is independent of health. However, it turns out that this is not essential for the empirical equation we estimate, in which the marginal utility of consumption does not enter explicitly.

Initial health  $H_0$  is given, and health evolves according to the dynamic transition equation

$$H_{t+1} - H_t = I_t^\alpha - d_t H_t, \tag{1}$$

where  $d_t$  is an age-specific deterioration rate (biological aging rate) and  $I_t$  is investment in health. The parameter  $\alpha$  reflects the returns to scale. The literature, starting with Grossman (1972a, 1972b), typically assumes that  $\alpha = 1$ , which, combined with the investment function (2) below, amounts to constant returns to scale. By introducing the parameter  $\alpha$ , we relax this restriction, allowing for decreasing returns to scale ( $0 < \alpha < 1$ ). Note, however, that our solution does not apply to  $\alpha = 1$ , which illustrates the special nature of constant returns to scale.

We adopt the next several assumptions from the Grossman literature to allow for comparability of results and to arrive at relations that can eventually be log-linearized. Investment in health  $I_t$  is assumed to be produced by combining time inputs  $\tau_{I,t}$  with goods and services purchased in the market  $m_t$  (e.g., medical care), according to a Cobb-Douglas production function

$$I_t = \mu_{I,t} m_t^{1-k_I} \tau_{I,t}^{k_I}, \quad (2)$$

where  $\mu_{I,t}$  is an efficiency factor and  $1 - k_I$  and  $k_I$  are the elasticities of investment in health with respect to goods and services purchased in the market and with respect to time inputs, respectively. Effectively this assumption allows one to log-linearize investment  $I_t$  into its components, goods and services  $m_t$  and time inputs  $\tau_{I,t}$ .

We assume that the more educated are more efficient consumers and producers of health investment. This is based on the interpretation of education as a productivity factor in time inputs and in identifying and seeking effective care (Grossman, 1972a, 2000). We adopt the usual relation

$$\mu_{I,t} = \mu_I^* \exp(\rho_I E), \quad (3)$$

where  $E$  is the level of education (e.g., years of schooling) and  $\rho_I$  is the parameter indicating the efficiency effect of education. Note that we make the standard assumption that education is given and constant across time, and thus in this specification,  $\mu_{I,t}$  does not depend on  $t$ . This assumption is most applicable to individuals who have completed their schooling (adults).

We assume the biological aging rate to be of the form

$$d_t = \exp(\lambda_0 + \lambda_1 t + \lambda_2 EC_t), \quad (4)$$

where  $EC_t$  is a vector of environmental variables (working and living conditions, hazardous environment, etc.) that affect the biological aging rate (Cropper, 1977; Wagstaff, 1986a).

For consumption, we specify a similar production function as for health investment. Thus, consumption  $C_t$  is assumed to be produced by combining time inputs  $\tau_{C,t}$  (leisure) with goods and services purchased in the market  $X_t$ , according to a Cobb-Douglas production function

$$C_t = \mu_{C,t} X_t^{1-k_C} \tau_{C,t}^{k_C}, \quad (5)$$

where  $\mu_{C,t}$  is an efficiency factor and  $1 - k_C$  and  $k_C$  are the elasticities of consumption with respect to goods and services purchased in the market and with respect to time inputs, respectively. For the efficiency factor, we use an analogous specification as for the health investment efficiency factor

$$\mu_{C,t} = \mu_C^* \exp(\rho_C E), \quad (6)$$

with  $\rho_C$  capturing the effect of education on consumption efficiency.

In addition to the *consumption benefit* of health through its inclusion in the utility function, health also has a *production benefit* through its reduction of sick time. Sick time is assumed to be a power law in health

$$s_t = \tau_{\text{tot}} \left( \frac{H_t}{H_{\min}} \right)^{-\gamma}, \quad (7)$$

where  $\tau_{\text{tot}}$  is the total time budget (the total number of hours in a period). It is assumed that the parameter  $\gamma$  is positive, so that sick time decreases with health. This choice of functional form for sick time has the desirable properties  $\lim_{H_t \rightarrow \infty} s_t = 0$  and  $\lim_{H_t \downarrow H_{\min}} s_t = \tau_{\text{tot}}$ , that is, with infinite health, the individual is never sick, and near the minimum possible level of health, the individual is always sick. The specifications (2)–(7) are employed in the literature in order to arrive at expressions that can eventually be log-linearized, which we will also do in section 3.1.

In the Grossman model, sick time governs the relation between income and health. Individuals work  $\tau_{w,t}$  hours, which is limited by the time budget constraint

$$\tau_{w,t} + \tau_{I,t} + \tau_{C,t} + s_t = \tau_{\text{tot}}.$$

Thus, increased sick time due to worse health must be offset by reduced work time, time spent investing in health, or time devoted to consumption (leisure). Income in period  $t$  equals the hourly wage rate  $w_t$  times the number of work hours  $\tau_{w,t}$

$$Y_t = w_t \tau_{w,t}.$$

We assume wage rates are exogenously given and do not depend on work hours. However, they may depend on other exogenous variables, such as education. Note that the current version of our model focuses on workers and it limits their income to earnings.

The final component of the structural model is assets. Initial assets  $A_0$  are given, and assets evolve according to the dynamic transition equation

$$A_{t+1} - A_t = \delta A_t + Y_t - p_{X,t} X_t - p_{m,t} m_t, \quad (8)$$

where  $\delta$  is the rate of return on capital and  $p_{X,t}$  and  $p_{m,t}$  are the prices of consumption goods and services, and medical goods and services, respectively. Individuals do not face a borrowing constraint in our model, but they are subject to the lifetime budget constraint  $A_T = A^*$ , where  $A^*$  is a given constant.

Individuals die when health reaches the minimum health level  $H_t = H_{\min}$ . Length of life  $T$  is endogenous and is determined by maximizing lifetime utility with respect to  $T$ . For the purpose of this paper, endogenous length of life  $T$  affects the shadow price of initial wealth  $q_0^A$ , which acts as an individual-specific constant.

### 3.1 Structural relation between medical care and health

In this paper, we study the relationship between the demand for medical goods and services  $m_t$  and health  $H_t$ . Galama (2011) derived a structural equation for this relationship for the model presented above

$$b_{1t} m_t^{1-\alpha} - (1-\alpha) m_t^{1-\alpha} \tilde{m}_t = b_{2t} H_t^{-\rho\phi} + b_{3t} H_t^{-(1+\gamma)}, \quad (9)$$



with the notation  $\tilde{x}_t \equiv 1 - x_{t-1}/x_t$  for a variable  $x_t$ , and the following functions

$$\begin{aligned} b_{1t} &\equiv \exp(\lambda_0 + \lambda_1 t + \lambda_2 EC_t) + \delta - (1 - \alpha k_I) \tilde{p}_{m,t} - \alpha k_I \tilde{w}_t; \\ b_{2t} &\equiv b_2^* (q_0^A)^{-\phi} e^{[(\phi-1)\rho_C + \alpha\rho_I]E} p_{m,t}^{-(1-\alpha k_I)} w_t^{-[k_C(\phi-1) + \alpha k_I]} p_{X,t}^{-(1-k_C)(\phi-1)} \left( \frac{1 + \beta}{1 + \delta} \right)^{-\phi t}; \\ b_{3t} &\equiv b_3^* e^{\alpha\rho_I E} p_{m,t}^{-(1-\alpha k_I)} w_t^{1-\alpha k_I}; \end{aligned}$$

with the following constants

$$\begin{aligned} b_2^* &\equiv (1 - \zeta) \zeta^{(1-\rho)(\rho\phi-1)/\rho} \alpha k_I^{\alpha k_I} (1 - k_I)^{1-\alpha k_I} (\mu_I^*)^\alpha \left[ \mu_C^* k_C^{k_C} (1 - k_C)^{1-k_C} \right]^{\phi-1}; \\ b_3^* &\equiv \gamma \tau_{\text{tot}} H_{\text{min}}^\gamma \alpha k_I^{\alpha k_I} (1 - k_I)^{1-\alpha k_I} (\mu_I^*)^\alpha, \\ \phi &\equiv \frac{1}{1 - \zeta + \rho\zeta}. \end{aligned}$$

As mentioned above, the constant  $q_0^A$  is the shadow price of initial wealth. It is part of the *adjoint function* (analogous to a Lagrange multiplier; see, e.g., Varaiya, 1998, pp. 78–80) of the assets transition equation and it emerges in the optimization process.

## 4 From theoretical to empirical model

Ideally, we would like to estimate the structural relation (9) between health investment and health directly. However, the equation is highly nonlinear in both variables and parameters, which implies that it is difficult to establish whether the structural parameters are identified from this equation. Given the results for our linearized equation below, we suspect that they may not be, but we have been unable to prove this. Unlike for the linearized equation below, it is not clear whether and how it would be possible to obtain useful partial results if the parameters are not identified. Furthermore, assuming identification, estimating the parameters from this equation is computationally difficult, because it may lead to convergence problems and local optima. Another drawback of estimating the nonlinear equation is that it is less straightforward to compare our results with empirical results found in the literature, which are based on linearized equations. Therefore, we leave estimating the nonlinear equation for future research and follow the literature by deriving linearized approximations to the equation of interest and estimating these resulting linear equations.

### 4.1 Pure investment and pure consumption

Analytical solutions for the Grossman model are usually based on two sub-models: the *pure investment* model and the *pure consumption* model. In the pure investment model, health does not provide utility and hence  $\zeta = 1$  and  $b_{2t} = 0$ , whereas in the pure consumption model, health does not provide a production benefit and  $b_{3t} = 0$ .

We proceed by assuming that  $(1 - \alpha)\tilde{m}_t$  is negligible compared to  $b_{1t}$ . This is the case if changes in wages and prices are small and changes in health investment are smaller than the biological aging rate plus the rate

of return to capital. Lifecycle models predict considerable consumption smoothing, and this carries over to health investment, so this is likely consistent with the predictions of the structural model. It is, however, difficult to assess the validity of this assumption empirically, because the biological aging rate depends on the scale of health, which is not well-defined. Some tentative computations suggest that it is likely satisfied for out-of-pocket (OOP) and total medical expenditures, but maybe not for hospital nights. Under this assumption, the structural relation simplifies and we can obtain an approximately linear structural relation for the demand for health investment goods and services  $m_t$  in the pure investment and pure consumption models. This also facilitates comparisons with the results in the literature.

For the pure investment model,  $\zeta = 1$  and  $b_{2t} = 0$ , and with  $(1 - \alpha)\tilde{m}_t$  negligible compared to  $b_{1t}$ , (9) reduces to

$$\begin{aligned} \log m_t \approx & \frac{\log b_3^*}{1 - \alpha} - \frac{1 + \gamma}{1 - \alpha} \log H_t + \frac{\alpha \rho_I}{1 - \alpha} E - \frac{1 - \alpha k_I}{1 - \alpha} \log p_{m,t} + \frac{1 - \alpha k_I}{1 - \alpha} \log w_t \\ & - \frac{\lambda_0}{1 - \alpha} - \frac{\lambda_1}{1 - \alpha} t - \frac{\lambda_2}{1 - \alpha} \text{EC}_t + R_t \end{aligned} \quad (10)$$

where

$$R_t = -\frac{1}{1 - \alpha} \log \left( 1 + \frac{\delta - (1 - \alpha k_I) \tilde{p}_{m,t} - \alpha k_I \tilde{w}_t}{\exp(\lambda_0 + \lambda_1 t + \lambda_2 \text{EC}_t)} \right).$$

Analogously, for the pure consumption model,  $b_{3t} = 0$  and with  $(1 - \alpha)\tilde{m}_t$  negligible compared to  $b_{1t}$ , (9) reduces to

$$\begin{aligned} \log m_t \approx & \frac{\log b_2^*}{1 - \alpha} - \frac{\rho \phi}{1 - \alpha} \log H_t - \frac{\phi \log q_0^A}{1 - \alpha} + \frac{(\phi - 1) \rho_C + \alpha \rho_I}{1 - \alpha} E - \frac{1 - \alpha k_I}{1 - \alpha} \log p_{m,t} \\ & - \frac{k_C(\phi - 1) + \alpha k_I}{1 - \alpha} \log w_t - \frac{(1 - k_C)(\phi - 1)}{1 - \alpha} \log p_{X,t} \\ & - \frac{\lambda_0}{1 - \alpha} - \frac{\lambda_1}{1 - \alpha} t - \frac{\lambda_2}{1 - \alpha} \text{EC}_t - \left[ \frac{\phi}{1 - \alpha} \log \left( \frac{1 + \beta}{1 + \delta} \right) \right] t + R_t. \end{aligned} \quad (11)$$

It is customary to view  $\lambda_0$  as an error term (see, e.g., Wagstaff, 1986a, and Grossman, 1972a, 1972b, 2000). Furthermore, it is often assumed that  $R_t$  is either small or constant (Grossman, 1972a, 2000). This is the case if the rate of return on capital  $\delta$  and changes in the wage rate  $w_t$  and the price  $p_{m,t}$  are much smaller than the health deterioration rate  $d_t$  or if the rate of return to capital  $\delta$  and changes in the wage rate  $w_t$  and the price  $p_{m,t}$  follow the same pattern as  $d_t$ , so that their ratio is approximately constant. Alternatively, it has been assumed that  $R_t$  is approximately a linear function of age  $t$  (e.g., Wagstaff, 1986a). Neither of these assumptions about  $R_t$  is likely to hold across the lifecycle. Health deterioration is assumed to accelerate across the lifecycle (e.g., Grossman, 2000) and our functional form is consistent with this. Asset returns do not show a consistent time trend so are best thought of as constant, and similarly for price changes. Wage tends to increase at younger ages, then stay relatively constant, with a slight drop off before retirement. Thus,  $R_t$  is not constant or small across the entire lifecycle: at young ages it is approximately constant but nonzero, whereas at older ages it is closer to zero but not constant. Over short time spans, it is approximately

linear or even constant, so in a panel data analysis with a few closely spaced waves the linearity assumption should work fine. In cross sections or longer panels, it is better to either include the term  $R_t$  in the model and estimate it as a nonlinear regression model, or approximate it by including a flexible function of age. We do the latter and include a quadratic in age below.

## 4.2 Model predictions

In the previous section, we derived two approximations, (10) and (11) of the structural relation between health investment and health, depending on different simplifying assumptions. Both equations are of the form

$$\log m_t = \theta_1 + \theta_2 \log H_t + \theta_3 E + \theta_4 \log w_t + \theta_5 \log p_{m,t} + \theta_6 \log p_{X,t} + \theta_7 t + \theta_8 EC_t + u + R_t, \quad (12)$$

where  $u$  is an individual-specific error term and  $R_t$  is a nonlinear residual term. As mentioned above, we approximate  $R_t$  by a quadratic in age, which subsumes the linear age term already in the model.

The coefficients in (12) are functions of the structural parameters, and the exact form of these functions depends on whether we assume the pure investment or the pure consumption model. For example, in the pure investment model,  $\theta_6 = 0$ , whereas in the pure consumption model, it is generally nonzero. We cannot fully recover the structural parameters from our econometric model. However, the structural model does provide predictions about the signs of most of the coefficients in the econometric model. In particular, according to both the pure investment and pure consumption models, the sign of the coefficient of log health is negative if there are decreasing returns to scale: individuals in better health invest less in health, *ceteris paribus*. This contrasts with the literature that assumes constant returns to scale, which predicts a positive coefficient of log health. Note that we cannot simply set  $\alpha = 1$  in our equation to arrive at this constant returns to scale result.

In the pure investment model, higher education leads to a higher demand for medical goods and services, whereas in the pure consumption model, the sign is ambiguous and depends on the relative efficiency gains from education for consumption and health investment. This contrasts with the usual prediction of the Grossman model (e.g., Feldstein, 1993, p. 78; Zweifel, Breyer, & Kifmann, 2009, pp. 83–84) that education unambiguously reduces this demand. Similarly, the sign of the coefficient of log wage is unambiguously positive in the pure investment model, but because generally  $\phi - 1 < 0$  (assuming a coefficient of relative risk aversion larger than 1, as is typically found in the literature), the coefficient of log wage is not guaranteed to be positive in the pure consumption model. The coefficients of age and environmental factors can be either positive or negative, because these partially pick up the effect of the nonlinear residual term  $R_t$ . In the pure investment model (10), if we disregard  $R_t$ , the coefficient of age is negative: at older ages, health investment is lower (*ceteris paribus*, in particular, controlling for health). However,  $\partial R_t / \partial t$  may be positive, so after approximating  $R_t$  with a quadratic in age, the model does not provide a clear prediction for the sign of the coefficient of age. This is also true for the pure consumption model, where the equation has an additional term in age and its sign depends on whether the discount rate  $\beta$  is larger than the asset return rate  $\delta$ .

## 5 Data

Our analysis is based on the Panel Study of Income Dynamics, a longitudinal survey of a representative sample of individuals in the U.S. and the households in which they reside. The PSID started in 1968 and was conducted annually up to 1997, after which it has been conducted biennially.

The PSID covers nearly the entire life cycle of individuals; from childhood through old age. In addition, the PSID collects rich data on income, wealth, demographic characteristics, labor force participation, nature of work, education level, consumption behavior (e.g., alcohol, smoking), medical expenditures, and detailed information on health (e.g., self-reported health, childhood health, height, weight, activities of daily living).

The most detailed information is collected for the household heads. For married couples, the PSID typically assigns the head status to the male. As a result female heads of household in the survey are not representative of the U.S. female population (i.e., typically not married/partnered, in a same-sex couple, or married to an incapacitated male). Therefore, we limit our analyses to male heads. Furthermore, detailed health data are available for all ages only since the 2003 wave. Therefore, the estimates presented below are based on the pooled 2003, 2005, and 2007 waves. The PSID interviewed 7,822 families in 2003, 8,002 in 2005, and 8,289 in 2007. Of all the households interviewed, 5,483 are headed by males in 2003, 5,594 in 2005 and 5,761 in 2007.

It follows from (1) and (8) that stock variables such as health are measured at the beginning of the period. However, in the data, health variables refer to the situation at the time of the interview, whereas flow variables such as medical expenditures refer to the past. Hence, to make the data consistent with the concepts in the theoretical model, we need to use medical expenditures and other flow variables measured in 2005 and 2007 in combination with stock variables measured a wave earlier (2003 and 2005, respectively). This implies that we effectively have two waves of data for the regression analyses reported below.

Our analyses are unweighted.

### 5.1 Measurement of health investment

We consider three different dependent variables. The first is out-of-pocket medical costs, which is the part of total medical expenditures that is paid for by the respondent, and this is the total of three different categories: (a) nursing home and hospital bills; (b) doctor, outpatient surgery, and dental bills; and (c) prescriptions, in-home medical care, special facilities, and other goods and services. The insurance premium is excluded from the out-of-pocket costs because this in itself does not indicate any health investment. The second dependent variable is total medical expenditures. This includes out-of-pocket medical costs and what has been covered by the insurer. Our third dependent variable is nights spent in a hospital.

Total medical expenditures conceptually comes closest to our theoretical variable, but respondents often have little information about what the insurer pays. Out-of-pocket costs may be a good proxy for true total medical expenditures. Especially if the individual makes active decisions about health care utilization based on out-of-pocket costs, this may be an even better reflection of (the goods and services part of) true health investment as conceptualized in the model. A drawback of these two expenditure-based measures is that they are only measured at the household level in the PSID, whereas our model is based on the individual.

Section 5.1.1 below describes the imputation method we employ to deal with this issue.

The number of hospital nights is unlikely to be an accurate indicator of our theoretical concept. It may be a measure of bad health rather than investment in health. Also, it does not capture a potentially large part of health investment that does not involve spending a night in a hospital. However, the measure is not meaningless. The model predicts that in response to bad health, individuals will invest more in health, and in principle, anything that improves health is health investment. Arguably individuals spend nights in a hospital because that is better for their health than not doing so. The measure has a few additional advantages: (1) it is likely to be measured better (less noisy) than medical expenditures; (2) it is measured at the individual level in the PSID and thus does not require imputation; (3) it can serve as a test for the sensitivity of the regression results to different operationalizations of the dependent variable; and (4) it is a measure that is commonly used measure of health investment (e.g., Van de Ven & Van der Gaag, 1982; Wagstaff, 1986a, 1993; Erbsland et al., 1995), thus allowing us to compare our results with prior estimates in the literature.

### **5.1.1 Imputation of individual medical expenditures**

As noted above, the PSID records total medical expenditures and medical out-of-pocket costs at the *household* level instead of at the individual level. However, the model is for *individuals*. Therefore, we use data from the Medical Expenditure Panel Survey (MEPS) to impute individual-level total expenditures and out-of-pocket costs for individuals in the PSID. The MEPS is a series of short panels (up to two years per household) that are nationally representative samples of the U.S. population. In addition to, among others, demographics, employment, income, and health, it contains detailed high quality information about health insurance, medical care utilization, and medical expenditures at the individual level. Part of this information is obtained from medical providers or employers. MEPS collected data on 32,681 individuals in 2003, 32,320 in 2005 and 29,370 in 2007. We use the following imputation procedure to obtain individual-level medical expenditure data for (male) heads in the PSID:

1. Define cells based on age, sex, and individual health insurance status (note that individuals within the household may have different health insurance status).
2. Based on these cells, randomly match a MEPS individual to each PSID household member (hot deck imputation).
3. Use the MEPS individual's total medical expenditures and out-of-pocket costs as imputations for the individual medical expenditures and out-of-pocket costs in the PSID.
4. Rescale so that individual total medical expenditures and out-of-pocket costs sum to the observed household total medical expenditures and out-of-pocket costs in the PSID. Thus, we effectively only impute *allocation percentages*.

We repeat this procedure five times. The econometric models, discussed below, are estimated using one imputed variable at a time. Since we have five imputed variables for total medical expenditures and out-of-pocket medical expenditures, we obtain five sets of regression results, which are then combined

according to the “Rubin rules” (Rubin, 1987). This takes imputation uncertainty into account in computing the standard errors and thus better reflects the sampling variability of the estimators than single imputation. This results in more random noise in the dependent variable than if the individual amounts were reported, but it does not affect the unbiasedness of the estimators and only leads to larger standard errors.

It is difficult to check the quality of the imputations, because we only impute allocation factors. We computed the within-household sums of the intermediate imputations in step 3 and compared their distributions with the distributions of the reported household-level variables. There is some variation between the five imputation samples, but taking the average of the five samples, the means of the imputed amounts are slightly (2–4%) higher than the reported ones in 2005 and a bit lower (7–12%) in 2007. For total medical expenditures, the average of the standard deviations was lower in both years (17% in 2005 and 33% in 2007), whereas for out-of-pocket costs, it was 12% higher in 2005 and 6% lower in 2007. Note, however, that after rescaling in step 4 these differences become zero by construction.

Imputation was not necessary in two instances: for single households, as the household coincides with the individual, and for individuals in households that report zero household medical expenditures (or out-of-pocket costs, respectively). There is no imputation uncertainty for these households. For 13% of the sample for individual total medical expenditures and for 15% of the sample for individual OOP, imputation was not required. (Note that the samples differ because of differences in missing values in family medical expenditures and out-of-pocket costs.)

For a small number of households—a maximum of 3% across waves for both medical expenditures and out-of-pocket costs—the imputed values in step 3 above were zero for all household members, while the reported family medical expenditures or out-of-pocket costs in the PSID were positive. In these cases, the procedure above does not lead to a valid allocation of the family expenditures or out-of-pocket costs, and we randomly assigned the reported family value (with equal probabilities) to one of the household members.

## 5.2 Measurement of health

Next to health investment, *health* is the most important variable in our study, and thus measuring health as accurately as possible is crucial to our endeavor. The health variable most often used in empirical studies is self-reported general health (SRH), which is administered in virtually all surveys. It is a categorical variable, in which respondents assess their health using five categories: (1) excellent; (2) very good; (3) good; (4) fair; (5) poor. While this variable is a fairly good measure of general health, as judged by its correlations with other health variables, such as chronic conditions, as well as by its predictive power for mortality and later-life outcomes (e.g., Idler & Benyamini, 1997), it is too coarse for our purposes, because the health transition equation in the theoretical model requires that health is a continuous and positively valued variable.

Therefore, we follow Börsch-Supan, McFadden, and Reinhold (1996), Soldo, Mitchell, Tfray, and McCabe (2007), and Meijer, Kapteyn, and Andreyeva (2011) and estimate a health measurement model, with true health as a latent variable, and subsequently compute a health index as the best possible estimate of individual health, given the parameter estimates of the health measurement model. We use self-reported health and difficulties with activities of daily living (ADLs) as indicators of health. We use

several health-related behaviors and a set of socio-demographic covariates as explanatory variables. The former include body mass index (BMI), drinking, and smoking. The socio-demographic covariates are age, household size, race, and education. The appendix gives more details about the variables used, the specification and estimation of the health measurement model, and the definition of the health index.

The thus constructed health index is a continuous variable. However, because it is based on linear models with normal error terms, its support is the whole real line, with an arbitrary mean and variance normalization. The theoretical health variable in the health capital model is positive. Therefore, our health measure for person  $n$  in wave  $t$  in the substantive regressions is  $\hat{H}_{nt} = \exp(A + B \cdot \text{HI}_{nt})$ , or, equivalently,  $\log \hat{H}_{nt} = A + B \cdot \text{HI}_{nt}$ , where  $\text{HI}_{nt}$  is the health index and  $A$  and  $B$  are constants, which are subsumed in the intercept and slope coefficients, respectively, in the regression equation, so that in practice we simply use the health index for  $\log \hat{H}_{nt}$  in the model.

### 5.3 Endogeneity of health

In the econometric models we estimate, health is an endogenous variable, because according to our theoretical model, health and health investment are joint outcomes of the lifetime utility maximization problem. Therefore, in addition to ordinary regressions we also perform instrumental variables analyses. We use as instruments self-reported childhood (age 0 to 16) health, which uses the same five-point scale as adult self-reported health (i.e., Excellent, Very Good, Good, Fair, Poor), and a binary indicator for whether or not one's parents smoked during childhood.

The PSID contains more information concerning childhood health, but the two measures used can be considered strong instruments according to the first stage  $F$ -statistic and the remaining variables do not add predictive power. A large literature has established a strong correlation between early childhood conditions and later-life health, see for example Van den Berg et al. (2006, 2010). Self-reported childhood health was asked retrospectively in 2003 and 2007. We take the average if it was reported in both years. The more detailed information was collected only in the 2007 wave.

Concerns regarding whether respondents can accurately recall events that happened during childhood are understandable; answers may also be biased towards current health status. Analyses presented in Smith (2009), however, suggest that these concerns are too pessimistic and that this type of information can be usefully exploited. Another concern is that health during childhood may affect health investments through additional channels beyond health (and other variables in the model, particularly education). This could occur, for example, through higher health investments during childhood and habit formation or path dependence thereafter, possibly violating the exclusion restriction. We believe that health and education are the main channels through which childhood health is related to medical investments and therefore, we believe our IV results are valid. Moreover, even if we cannot control for all potential pathways, exploring differences between IV results and standard OLS results is still useful for an assessment of the potential severity of the endogeneity problem.

## **5.4 Exogenous covariates**

Exogenous explanatory variables in the regression models include age, education, education of the household head's wife, race, an indicator for marital status (treating cohabitators as married), household size, urbanicity, the wage rate, and a status dummy for private health insurance (either provided by an employer or purchased directly) to proxy the price variables. The appendix provides more details about the precise definitions of these variables.

## **5.5 Health measurement sample**

If age, race, or education was missing for a respondent in one of the waves, we checked whether this information could be derived from other waves. After this "logical imputation", observations for which age or race were still missing were excluded from the sample. There were 8 such cases in 2003, 6 in 2005, and 7 in 2007. The resulting sample sizes for the health measurement model were 5,475 in 2003, 5,588 in 2005, and 5,754 in 2007. Note that the sample sizes vary from year to year. Heads may appear in some years and not in others due to changes in household composition (household splits or marriages), non-response, or exclusion from the sample due to missing data, as described above.

Summary statistics of the health measurement sample are presented in Table 1, columns 2–4.

## **5.6 Regression sample**

While the health measurement model was estimated on all male household heads that had nonmissing race and age, the regression analysis was conducted on a more restricted sample. The regression sample selection is based on four additional criteria resulting in a sample of male household heads who are younger than 65, have health insurance (public or private), are working, and have an hourly wage of at least \$4. Male heads with an hourly wage less than \$4 are dropped because these appear to be extreme cases. We also limit our sample to male heads who are working since our structural model does not contain retirement (or unemployment). Individuals face very different prices for medical goods and services depending on whether or not they have health insurance. In the absence of detailed data about how those prices actually differ, we have chosen to restrict our analysis to the insured only. Our regression sample is further limited to heads under 65 years of age since Medicare eligibility creates a very different set of circumstances for those 65 and over. Limiting the sample to those that meet the selection criteria for the regression analysis excludes roughly 2,000 individuals from each year's sample. Table 2 gives a breakdown of the sample selection. Summary statistics of the resulting regression sample are presented in Table 1, columns 5–7.

# **6 Results**

In this section, we describe the estimation results. As before, we start with the health measurement model and then describe the regression models.



Table 1: Characteristics of the health measurement sample and the regression sample

|                                     | Health measurement |       |       | Regression <sup>a</sup> |       |
|-------------------------------------|--------------------|-------|-------|-------------------------|-------|
|                                     | 2003               | 2005  | 2007  | 2003                    | 2005  |
| Age, mean                           | 44.3               | 44.6  | 44.7  | 41.1                    | 41.2  |
| <i>Race:</i>                        |                    |       |       |                         |       |
| White (%)                           | 66.3               | 65.6  | 65.3  | 71.3                    | 72.2  |
| Black (%)                           | 24.6               | 25.4  | 25.8  | 20.9                    | 20.5  |
| Other (%)                           | 9.0                | 9.0   | 8.9   | 7.7                     | 7.3   |
| Married (%)                         | 78.4               | 79.4  | 78.9  | 82.0                    | 82.8  |
| Household size, mean                | 2.9                | 2.9   | 2.9   | 3.1                     | 3.0   |
| Years of education, mean            | 13.0               | 13.0  | 13.0  | 13.6                    | 13.6  |
| Less than high school (%)           | 18.0               | 18.1  | 17.9  | 9.7                     | 9.8   |
| High school or some college (%)     | 56.9               | 56.6  | 56.7  | 58.3                    | 58.8  |
| College degree or more (%)          | 25.1               | 25.3  | 25.5  | 32.0                    | 31.4  |
| BMI, mean                           | 27.7               | 27.9  | 28.0  | 27.8                    | 28.0  |
| Underweight (%)                     | 0.5                | 0.6   | 0.6   | 0.4                     | 0.5   |
| Normal weight (%)                   | 28.0               | 27.2  | 26.3  | 26.4                    | 25.6  |
| Overweight (%)                      | 46.1               | 46.1  | 45.2  | 47.7                    | 47.3  |
| Obese (%)                           | 17.8               | 18.3  | 19.3  | 17.9                    | 18.8  |
| Severely obese (%)                  | 7.5                | 7.8   | 8.7   | 7.6                     | 7.7   |
| Current cigarette consumption, mean | 3.8                | 3.5   | 3.3   | 3.3                     | 3.0   |
| Past cigarette consumption, mean    | 5.1                | 4.8   | 4.6   | 4.0                     | 3.5   |
| Drink alcohol (%)                   | 67.3               | 67.6  | 67.1  | 71.8                    | 71.7  |
| <i>ADLs:</i>                        |                    |       |       |                         |       |
| Difficulty dressing (%)             | 1.7                | 2.4   | 2.1   | 0.3                     | 0.4   |
| Difficulty bathing (%)              | 1.9                | 2.0   | 2.1   | 0.3                     | 0.2   |
| Difficulty eating (%)               | 0.6                | 0.7   | 0.6   | 0.2                     | 0.1   |
| Difficulty getting out of chair (%) | 3.0                | 3.6   | 3.2   | 1.0                     | 0.7   |
| Difficulty walking (%)              | 5.7                | 6.8   | 6.3   | 2.0                     | 1.9   |
| Difficulty getting out of house (%) | 1.4                | 1.9   | 1.9   | 0.2                     | 0.3   |
| Difficulty using toilet (%)         | 1.0                | 0.9   | 1.1   | 0.2                     | 0.2   |
| Self-reported health, mean          | 2.3                | 2.4   | 2.3   | 2.1                     | 2.1   |
| Sample size                         | 5,475              | 5,588 | 5,754 | 2,368                   | 2,493 |

<sup>a</sup> Intermediate sample; see Table 2

Table 2: Sample selection for regression analysis

|   | 2003  | 2005  |
|---|-------|-------|
| Male household heads                      | 5,483 | 5,594 |
| Health measurement sample                 | 5,475 | 5,588 |
| Not in next wave (for $m_t$ )             | 600   | 540   |
| Age $\geq 65$                             | 506   | 508   |
| Uninsured                                 | 710   | 734   |
| Non-working                               | 451   | 427   |
| Hourly wage $< \$4$                       | 16    | 17    |
| Missing selection variables               | 561   | 594   |
| Missing covariates                        | 263   | 275   |
| Intermediate sample size                  | 2,368 | 2,493 |
| Missing family total medical expenditures | 975   | 1,065 |
| Missing family out-of-pocket med.exp.     | 155   | 207   |
| Missing hospital nights                   | 8     | 12    |
| Sample size total medical expenditures    | 1,393 | 1,428 |
| Sample size out-of-pocket med.exp.        | 2,213 | 2,286 |
| Sample size hospital nights               | 2,360 | 2,481 |

*Note:* Categories in lines 3–9 are mutually exclusive and assigned with higher rows taking precedence.

## 6.1 Health measurement model

We estimated the health measurement model using the stacked data of all respondents in all three waves that satisfied the criteria described in section 5. The pseudo- $R^2$  of this model was 0.199, which is slightly lower than the corresponding pseudo- $R^2$  values in Meijer et al. (2011), which utilized a similar model. We suspect that this can be explained by the smaller number of health indicators in the PSID than in the European SHARE data used by Meijer et al. Another potential explanation may be that we estimate a model for all (adult) ages whereas the model of Meijer et al. only included individuals of 50 years and older. The ADLs have less variation for younger individuals and thus they are less informative in our sample.<sup>2</sup>

Estimation results are given in Table 3. The factor loadings all have the expected signs: difficulty with an ADL and a higher value on the self-reported health scale indicate worse health (note that higher values in the latent health variable and health index denote better health). They are highly statistically significant. The regression parameters also have the expected signs: health is worse at higher ages, is better for the more educated and for whites, and worse for more obese individuals and smokers. The positive effect of (moderate) drinking is also in line with the literature.

The resulting reliability of the health index is 0.66. This is a bit lower than the typical rule of thumb of 0.70 for an acceptable psychological test, and also lower than the reliabilities of the health indexes in Meijer et al. (2011). Again, this is due to the much smaller number of health indicators that is available in the PSID. This is unfortunate as it means that the estimates in the regression models will be subject to more sampling variability (larger standard errors). Fortunately, however, this does not lead to biases. Unlike the classical measurement error case (e.g., Meijer & Wansbeek, 2000), the construction of the health index as the conditional expectation means that the following linear regressions satisfy the *Berkson model* (e.g., Wansbeek & Meijer, 2000, pp. 29–30). For this model, the ordinary least squares estimator is consistent, despite the less than perfect measurement of the regressor. For the nonlinear models presented below, the Berkson model leads to an attenuation by a constant scale factor (Wansbeek & Meijer, 2000, p. 330), but since we are focusing on the signs of the coefficients, not their size, the results below remain valid.

## 6.2 Regression models

The outcome variables considered in this paper are nonnegative, have a high frequency of zeros and have a right-skewed distribution. Furthermore, the number of nights spent in a hospital is an integer-valued variable. Common strategies to deal with such data structures are to estimate a two-part model or a bivariate sample selection model, and for nonnegative integer-valued variables to estimate count data models.

For total medical expenditures and out-of-pocket expenditures, we estimated both Heckman-type bivariate sample selection models and two-part models. In both cases, the participation equation used a probit specification and the outcome equation was specified as a linear regression equation with the logarithm of the outcome variable as the dependent variable. The outcome equation is only estimated for observations with positive outcome values. In our case, the results of these two approaches were very similar, and we only present the results from the bivariate selection models. We have estimated the standard

---

<sup>2</sup>We thank an anonymous reviewer for pointing this out.

Table 3: Estimation results from the health measurement model

| Factor analysis submodel                   |          |                 | Regression submodel                         |          |                 |
|--|----------|-----------------|---|----------|-----------------|
| Parameter                                  | Estimate | <i>t</i> -value | Parameter                                   | Estimate | <i>t</i> -value |
| <i>Intercepts</i> ( $\tau_i$ )             |          |                 | <i>Regression coefficients</i> ( $\gamma$ ) |          |                 |
| Bathing                                    | 0.004    | 0.1             | $a = (\text{Age} - 50)/10$                  | -0.494   | -10.3           |
| Dressing (reference)                       | 0        | n.a.            | $a^2$                                       | -0.004   | -0.4            |
| Eating                                     | 1.309    | 19.8            | $a^3$                                       | -0.020   | -4.2            |
| Getting out of chair                       | -0.296   | -3.6            | Secondary education                         | 0.777    | 9.0             |
| Walking                                    | -1.158   | -10.0           | Tertiary education                          | 1.688    | 11.4            |
| Getting out of the house                   | 0.300    | 3.7             | Household size                              | 0.092    | 4.4             |
| Using toilet                               | 0.821    | 11.4            | Underweight                                 | -1.367   | -4.5            |
| Self-reported health                       | 0        | n.a.            | Overweight                                  | -0.020   | -0.3            |
|  |          |                 | Obese                                       | -0.625   | -7.6            |
|  |          |                 | Severely obese                              | -1.673   | -12.3           |
| <i>Factor loadings</i> ( $\lambda_i$ )     |          |                 | Cigarettes now                              | -0.472   | -11.2           |
| Bathing                                    | -1.026   | -14.5           | Cigarettes past                             | -0.161   | -6.8            |
| Dressing (reference)                       | -1       | n.a.            | Drink alcohol                               | 0.447    | 6.7             |
| Eating                                     | -0.401   | -9.7            | Race: white                                 | 0.537    | 4.3             |
| Getting out of chair                       | -0.808   | -13.0           | Race: black                                 | -0.104   | -0.9            |
| Walking                                    | -0.877   | -11.5           | Missing education                           | 1.214    | 7.8             |
| Getting out of the house                   | -0.866   | -11.6           | Missing BMI                                 | -1.202   | -5.4            |
| Using toilet                               | -0.794   | -11.3           | Missing cigs now                            | -1.055   | -2.9            |
| Self-reported health                       | -0.449   | -13.2           | Missing cigs past                           | 0.154    | 1.0             |
|  |          |                 | Constant                                    | 4.057    | 15.9            |
| <i>Thresholds for self-reported health</i> |          |                 | Residual s.d. ( $\sqrt{\psi}$ )             | 2.207    | 13.2            |
| $\alpha_1$                                 | -0.259   | -4.9            |   |          |                 |
| $\alpha_2$                                 | 0.844    | 14.5            |   |          |                 |
| $\alpha_3$                                 | 2.197    | 31.0            |   |          |                 |
| $\alpha_4$                                 | 3.589    | 41.2            |   |          |                 |

Note. n.a. = not applicable; sample size is 16,817.

Table 4: Regression models for out-of-pocket costs

| Variable                   | Bivariate sample selection |                 |                          |                 | IV bivariate sample selection <sup>a</sup> |                 |                          |                 |
|----------------------------|----------------------------|-----------------|--------------------------|-----------------|--|-----------------|--------------------------|-----------------|
|                            | Participation eq.          |                 | Outcome eq.              |                 | Participation eq.                          |                 | Outcome eq.              |                 |
|                            | Coeff                      | <i>t</i> -value | Coeff                    | <i>t</i> -value | Coeff                                      | <i>t</i> -value | Coeff                    | <i>t</i> -value |
| Health index               | 0.00                       | 0.04            | -0.09                    | -2.38           | -0.04                                      | -0.50           | 0.05                     | 0.32            |
| Private insurance          | 0.61                       | 5.27            | 0.08                     | 0.28            | 0.61                                       | 5.27            | 0.08                     | 0.29            |
| Log wage                   | 0.02                       | 0.42            | 0.14                     | 1.51            | 0.04                                       | 0.74            | 0.08                     | 0.91            |
| $a = (\text{Age} - 50)/10$ | 0.27                       | 6.35            | -0.10                    | -1.45           | 0.26                                       | 4.22            | -0.04                    | -0.38           |
| $a^2$                      | 0.01                       | 0.55            | -0.02                    | -0.64           | 0.02                                       | 0.70            | -0.03                    | -0.86           |
| Secondary education        | 0.11                       | 1.21            | -0.04                    | -0.21           | 0.14                                       | 1.23            | -0.13                    | -0.54           |
| Tertiary education         | 0.09                       | 0.78            | 0.15                     | 0.67            | 0.14                                       | 0.85            | -0.04                    | -0.12           |
| Wife's education           | 0.00                       | 0.16            | 0.00                     | -0.11           | 0.00                                       | 0.30            | -0.01                    | -0.36           |
| Black                      | -0.08                      | -1.26           | -0.32                    | -2.53           | -0.10                                      | -1.30           | -0.24                    | -1.79           |
| Other race                 | -0.08                      | -0.80           | -0.28                    | -1.42           | -0.08                                      | -0.82           | -0.28                    | -1.39           |
| Married                    | -0.02                      | -0.12           | 0.10                     | 0.29            | -0.05                                      | -0.23           | 0.19                     | 0.49            |
| Suburban                   | -0.02                      | -0.43           | 0.16                     | 1.61            | -0.02                                      | -0.43           | 0.15                     | 1.59            |
| Rural                      | 0.00                       | -0.01           | 0.32                     | 1.85            | -0.01                                      | -0.07           | 0.33                     | 1.97            |
| Household size             | -0.02                      | -0.85           | -0.04                    | -0.81           | -0.02                                      | -0.80           | -0.04                    | -0.92           |
| Constant                   | 0.15                       | 0.73            | 6.25                     | 16.95           | 0.29                                       | 0.78            | 5.80                     | 9.85            |
| <i>N</i>                   | 4,499                      |                 | 3,224–3,247 <sup>b</sup> |                 | 4,499                                      |                 | 3,224–3,247 <sup>b</sup> |                 |

<sup>a</sup> The first stage *F* statistic for the IV regression is 88.9.

<sup>b</sup> The number of positive observations varies between imputed data sets.

versions of these models, as well as versions that account for the endogeneity of health by using instrumental variables (IV) versions.

For the number of nights spent in a hospital we have estimated several count data models: Poisson, negative binomial, hurdle, and an IV version of the Poisson model that takes the endogeneity of health into account. The hurdle model is a two-part model, with the first equation being a probit for a positive amount versus a zero amount, and the second equation being a zero-truncated negative binomial model, that is, a negative binomial model conditional on the amount being positive.

A more detailed discussion of the estimated models can be found in the appendix and, for example, in Cameron and Trivedi (2005).

Table 4 presents the results of the regression models for out-of-pocket medical costs. Columns 2–5 show the results for the standard bivariate sample selection model, whereas columns 6–9 show results for the IV bivariate sample selection model.

The first observation from this table is that, despite the relatively large sample sizes, very few coefficients

are statistically significantly different from zero. This is partly due to the imputation uncertainty, but we also interpret this as a sign that out-of-pocket medical expenditures contain a relatively large random component. For the time-varying covariates, another part of the explanation may be the large lag between observing the covariates and observing the dependent variable (2 years). We will return to the timing issue later in this section.

Examining the coefficients of the health index, we do not find an effect of health on having positive (vs. zero) out-of-pocket expenses in the non-IV model, but a statistically significant negative effect on the dollar amount for those who have positive out-of-pocket expenses, which replicates the general finding in the literature. To put the magnitude of the coefficients in perspective, note that the standard deviation of the health index is about 1.9 in the full sample which includes individuals 65 and over, and about 1.5 in the regression sample. So a one standard deviation improvement in health is associated with a 13–17% reduction in out-of-pocket costs. Because the DRTS model predicts a negative sign and the CRTS model a positive sign, this is suggestive evidence in favor of the DRTS model. However, when we move to the IV estimates, we see that the coefficients of the health index are not statistically significantly different from zero and are of opposite sign for the participation and outcome equations.

Regarding the other coefficients in the model, in both the standard model and the IV model, having private insurance is positively related to positive medical out-of-pocket costs, but among those with positive out-of-pocket costs, the privately insured have no higher costs than the publicly insured.

The results for age point to a positive relationship between age and out-of-pocket expenditures. As discussed above, within the pure investment model, this implies an important role of (and variation in) the residual term  $R_t$ . Within the pure consumption model, there is more ambiguity as to what the source of this positive coefficient may be. Besides the role of  $R_t$ , other explanations of the positive sign may be that we do not measure health precisely enough and thus the age coefficients pick up part of the health effect, or that there are other confounding factors that are correlated with age, such as cohort effects, although the latter would presumably point in the opposite direction.

The results for total medical expenditures are given in Table 5. They are largely similar to those for medical out-of-pocket costs. One exception is that the private insurance coefficients are now much smaller and not statistically significant. Another exception is that the age coefficients are now positive in the outcome equation as well, and statistically significant (at the 10% level in the standard model and at the 1% level in the IV model). The health index coefficient is again negative and highly significant in the outcome equation of the standard model, but vanishes in the IV model.

In addition to the results presented here we have also performed a number of sensitivity analyses. We used self-reported health instead of the health index, household income instead of wages of the household head, and we included year dummies as covariates. Results are similar to those reported here and are omitted. We have also experimented with exploiting the panel nature of the data, estimating fixed effects logit models for the participation equation and fixed effects linear regression models for the outcome equations. For these models, almost all coefficients became insignificant. Our preliminary assessment of these results is that this is due to measurement error. It is well known (e.g., Wansbeek & Meijer, 2000, p. 140) that measurement error problems are greatly exacerbated in fixed effects estimators for panel data, and our possibilities for estimating panel data models are highly limited with only two effective waves of data.

Table 5: Regression models for total medical expenditures

| Variable                   | Bivariate sample selection |                 |                          |                 | IV bivariate sample selection <sup>a</sup> |                 |                          |                 |
|----------------------------|----------------------------|-----------------|--------------------------|-----------------|--|-----------------|--------------------------|-----------------|
|                            | Participation eq.          |                 | Outcome eq.              |                 | Participation eq.                          |                 | Outcome eq.              |                 |
|                            | Coeff                      | <i>t</i> -value | Coeff                    | <i>t</i> -value | Coeff                                      | <i>t</i> -value | Coeff                    | <i>t</i> -value |
| Health index               | -0.03                      | -0.74           | -0.14                    | -3.68           | -0.06                                      | -0.46           | 0.01                     | 0.08            |
| Private insurance          | 0.24                       | 1.23            | 0.01                     | 0.02            | 0.27                                       | 1.62            | 0.11                     | 0.41            |
| Log wage                   | 0.06                       | 0.66            | 0.13                     | 1.43            | 0.07                                       | 0.69            | 0.07                     | 0.55            |
| $a = (\text{Age} - 50)/10$ | 0.23                       | 3.64            | 0.21                     | 1.94            | 0.23                                       | 2.73            | 0.34                     | 2.90            |
| $a^2$                      | 0.00                       | -0.13           | 0.03                     | 0.60            | 0.00                                       | 0.11            | 0.01                     | 0.17            |
| Secondary education        | 0.12                       | 1.02            | -0.04                    | -0.19           | 0.15                                       | 1.18            | -0.10                    | -0.44           |
| Tertiary education         | 0.20                       | 1.29            | 0.06                     | 0.27            | 0.26                                       | 1.35            | -0.08                    | -0.26           |
| Wife's education           | 0.00                       | 0.03            | 0.01                     | 0.25            | 0.00                                       | 0.02            | 0.00                     | -0.12           |
| Black                      | -0.05                      | -0.40           | -0.42                    | -2.83           | -0.07                                      | -0.60           | -0.38                    | -2.16           |
| Other race                 | 0.03                       | 0.21            | -0.50                    | -2.47           | 0.01                                       | 0.06            | -0.51                    | -2.50           |
| Married                    | -0.05                      | -0.19           | -0.26                    | -0.54           | -0.06                                      | -0.19           | -0.16                    | -0.33           |
| Suburban                   | -0.01                      | -0.14           | -0.01                    | -0.06           | -0.01                                      | -0.07           | -0.02                    | -0.19           |
| Rural                      | 0.00                       | 0.02            | -0.02                    | -0.16           | 0.01                                       | 0.12            | -0.01                    | -0.09           |
| Household size             | -0.01                      | -0.19           | -0.10                    | -1.89           | -0.01                                      | -0.16           | -0.11                    | -2.25           |
| Constant                   | 0.75                       | 2.03            | 7.99                     | 12.59           | 0.83                                       | 1.58            | 7.25                     | 11.40           |
| <i>N</i>                   | 2,821                      |                 | 2,202–2,270 <sup>b</sup> |                 | 2,821                                      |                 | 2,202–2,270 <sup>b</sup> |                 |

<sup>a</sup> The first stage *F* statistic for the IV regression is 56.3.

<sup>b</sup> The number of positive observations varies between imputed data sets.

Table 6: Regression models for nights spent in a hospital

| Variable                   | Hurdle model |                 |                   |                 |        |                 |                                  |                 |                         |                 |
|----------------------------|--------------|-----------------|-------------------|-----------------|--------|-----------------|----------------------------------|-----------------|-------------------------|-----------------|
|                            | Poisson      |                 | Negative binomial |                 | Probit |                 | Zero-Truncated Negative Binomial |                 | IV Poisson <sup>a</sup> |                 |
|                            | Coeff        | <i>t</i> -value | Coeff             | <i>t</i> -value | Coeff  | <i>t</i> -value | Coeff                            | <i>t</i> -value | Coeff                   | <i>t</i> -value |
| Health index               | -0.44        | -4.03           | -0.34             | -4.72           | -0.15  | -6.28           | -0.15                            | -1.94           | -0.16                   | -0.43           |
| Private insurance          | -1.29        | -3.55           | -0.97             | -2.79           | -0.38  | -3.10           | -0.60                            | -1.79           | -1.28                   | -3.34           |
| Log wage                   | -0.08        | -0.48           | 0.17              | 0.72            | -0.06  | -0.86           | 0.26                             | 1.04            | -0.19                   | -0.88           |
| $a = (\text{Age} - 50)/10$ | 0.25         | 1.82            | 0.45              | 2.93            | 0.20   | 3.72            | -0.08                            | -0.44           | 0.36                    | 1.93            |
| $a^2$                      | -0.02        | -0.25           | -0.02             | -0.23           | 0.03   | 1.14            | -0.12                            | -1.30           | -0.04                   | -0.50           |
| Secondary education        | 0.29         | 0.93            | -0.40             | -0.88           | 0.11   | 1.01            | -0.15                            | -0.43           | 0.13                    | 0.35            |
| Tertiary education         | 0.69         | 1.25            | -0.52             | -1.09           | 0.14   | 1.01            | -0.11                            | -0.28           | 0.37                    | 0.66            |
| Wife's education           | -0.08        | -1.89           | -0.12             | -2.01           | -0.02  | -1.02           | -0.09                            | -1.12           | -0.10                   | -1.93           |
| Black                      | 0.23         | 0.97            | 0.11              | 0.47            | 0.07   | 0.92            | 0.02                             | 0.09            | 0.36                    | 1.11            |
| Other race                 | -0.33        | -0.69           | 0.44              | 0.87            | 0.09   | 0.70            | -0.09                            | -0.16           | -0.33                   | -0.66           |
| Married                    | 0.02         | 0.03            | 0.51              | 0.54            | -0.02  | -0.09           | 0.43                             | 0.38            | 0.20                    | 0.28            |
| Suburban                   | -0.25        | -1.03           | -0.03             | -0.12           | 0.01   | 0.16            | -0.18                            | -0.63           | -0.26                   | -1.08           |
| Rural                      | 0.23         | 0.62            | 0.52              | 1.69            | 0.11   | 1.11            | 0.18                             | 0.54            | 0.24                    | 0.65            |
| Household size             | 0.13         | 1.25            | 0.11              | 1.35            | 0.04   | 1.42            | 0.05                             | 0.70            | 0.12                    | 1.13            |
| Constant                   | 2.73         | 4.01            | 2.11              | 2.72            | -0.20  | -0.76           | 1.78                             | 1.86            | 1.85                    | 1.21            |
| <i>N</i>                   | 4,841        |                 | 4,841             |                 | 4,841  |                 | 282                              |                 | 4,841                   |                 |

<sup>a</sup> The first stage *F* statistic for the IV regression is 87.5.

Table 6 presents the results of the regressions for hospital nights. Not surprisingly, this shows a consistently negative effect of health. This replicates the findings in the literature. However with the IV estimator, it is not significant. Having private insurance is negatively related to spending nights in a hospital, in contrast with the positive sign we saw for out-of-pocket and total medical expenditures. Age is generally positively related to hospital nights, as it is with medical expenditures.

We compared our empirical results with those from the literature. This turned out to be difficult, because the empirical specifications generally do not correspond closely to the theoretical ones or because variables are used that are not closely related to those in the PSID. Examples of differences in specification are the inclusion of an additional health investment variable as an explanatory variable (e.g., doctor visits in the equation for hospital nights; Wagstaff, 1986a), or the use of multiple health measures as explanatory variables (e.g., both “permanent health” and number of sick days; Van de Ven & Van der Gaag, 1982). An example of a variable that is used and not closely related to those available in the PSID is the number of visits to the family doctor (e.g., Van de Ven & Van der Gaag, 1982; Wagstaff, 1986a; Erbsland et al., 1995).



With these caveats in mind, the closest comparisons appear to be models with hospital nights (or weeks) as the dependent variable, which is used by Van de Ven and Van der Gaag (1982), Wagstaff (1986a), and Erbsland et al. (1995). The first two use it in levels and the third uses the logarithm of 1 + the number of hospital nights, which should be very similar. Neither of these attempt to deal with the endogeneity of health, and thus their results should be most comparable to our non-IV results, in particular the (single equation) Poisson and negative binomial models. They obtain negative and highly statistically significant coefficients for health, with  $t$ -statistics between 3 and 19. Our  $t$ -statistics are between 4 and 6 and our results are thus, in terms of sign and statistical significance, in line with those in the literature.

Nocera and Zweifel (1998) use the logarithm of total medical expenditures, as reported by the insurance company, as their dependent variable, while accounting for the zeros through a tobit model but, again, not accounting for the endogeneity of health. Thus, their results should be most comparable to our non-IV sample selection model results for total medical expenditures, although arguably our respondent-reported measure is considerably more noisy than their measure. We find a statistically significant negative coefficient for the health index ( $t = -3.7$ ). They find a negative but insignificant effect in their sample I, which (apart from the administrative data) is based on a single cross-sectional survey that reports on the past four years and is translated into four observations per respondent. Their sample II uses answers from two surveys (1981 and 1993) with partially the same respondents, who report about the past four years, and for the intervening years, the data are linearly interpolated, thus creating a data set with 12 “observations” per respondent. In this sample, the coefficient of health is positive and significant at the 5% level ( $t = 2.26$ ). This  $t$ -statistic is likely inflated, because the covariates for a third of the observations are imputed in a way that does not take the variation into account, and standard errors do not appear to be adjusted for imputation uncertainty. Thus, unlike the results for hospital stays, which show a clear and strong relation with health, these, as well as our, results seem to indicate that medical expenditures are rather weakly related to health.

In our models, we used health investment from the next wave, because it is reported retrospectively and thus precedes the current health status, whereas in the model, health investment depends on health and thus should follow current health status. In the four papers mentioned, such corrections have apparently not been made. When we repeat our analyses for out-of-pocket and total medical expenditures with contemporaneously measured health investment as the dependent variable, the coefficients for health are more systematically negative, although still not significant for the IV models, despite the larger sample sizes (because we have three waves instead of two). One might argue that despite the reverse ordering (and thus the potential for reverse causality), contemporaneously reported medical expenses better approximate the desired measures, because health investment (esp. in the form of treatment) likely responds quickly to changes in health, and the contemporaneous report may be slanted towards the present and thus be closer to “current” investment than the report two years later.<sup>3</sup>

---

<sup>3</sup>We thank Arie Kapteyn for pointing this out.

## 7 Discussion

This paper includes the first empirical tests of health capital theory with decreasing returns to scale. We start from the structural model developed by Galama (2011), which is an extension of the model of Grossman (1972a, 1972b) incorporating a decreasing returns to scale health production function as in Ehrlich and Chuma (1990). This model has a number of desirable characteristics that Grossman's constant returns to scale model does not have. In particular, there are substantive and empirical reasons to believe returns are decreasing. The model with decreasing returns to scale is able to reproduce stylized facts that run counter to the predictions of the constant returns model, and it has some theoretical properties that are more realistic.

The model leads to a derived structural relation between the demand for medical goods and services and health and other explanatory variables. After a few simplifying assumptions, a linear model equation is obtained that can be empirically estimated. We are the first to estimate an equation from a health capital model with decreasing returns to scale, and the first who take the endogeneity of health into account.

We obtain a statistically significant negative coefficient of health when the number of nights spent in a hospital is the dependent variable and when we do not take the endogeneity of health into account. Similarly, in the models for out-of-pocket medical expenditures or total medical expenditures, the dollar amounts are negatively related to health and statistically significant, although we do not find an effect for the participation equations (i.e., whether expenditures are positive or zero). This closely resembles the methodology and the findings in the literature and appears to support decreasing returns to scale, which predicts a negative coefficient, whereas constant returns to scale is associated with a positive coefficient.

However, when we attempt to control for the endogeneity of health by using instrumental variables methods, using childhood health and parental smoking during childhood as instruments, the coefficients become statistically insignificant and not consistently negative. Since we are the first to take the endogeneity of health into account and since our results when we do not account for endogeneity are in line with those in the literature, this brings into question the robustness of prior estimates of the relation between health investment and health from the literature.

Our mixed findings suggest there is room for further research. Various technical issues need to be addressed in more detail, such as identification of sample selection equations by imposing sound exclusion restrictions, and measurement error in health investment. Also, the approximations through linearization we have employed in this paper, although comparable to those in the literature, may not be accurate enough, and thus techniques such as GMM or nonlinear least squares may need to be employed to estimate more accurate nonlinear equations. Our current model equation does not allow us to derive the structural parameters as a function of the reduced form parameters, but additional equations can be derived from the theoretical model to estimate the structural parameters.

As in most of the literature on health capital models, our results only apply to working males. Understanding health investment in this population is of great interest, because health disparities are largely formed by age 50 (Smith, 2004; Case & Deaton, 2005). On the other hand, health costs are much larger later in life and thus it would be of interest to add the period after retirement, as well as the (endogenous) decision of when to retire. For individuals 65 years and older, the availability of Medicare generally lowers out-of-pocket medical costs and premiums, encouraging health investment. In the other direction, after

retirement, the production benefit of health—the effect of health on income—disappears. The steeply increasing medical costs may be the result of quickly decreasing returns to scale, of steeply accelerating biological aging rates, or of aggressive costly attempts to extend life. Models for females need to be much more involved than models for males, because the (dynamic) labor force participation decisions cannot be ignored at any age, and correspondingly, marriage and fertility need to be included in the model (Keane, 2011).

## Appendix

### A Details of variables used in the analyses

#### A.1 Health measurement model

Self-reported health and difficulties with activities of daily living (ADLs) are used as indicators of health (dependent variables). The following ADLs are included: bathing, dressing, eating, getting out of a chair, walking, getting out of the house, and using the toilet.

Information on the following health-related behaviors (covariates) is included: BMI, drinking, and smoking. BMI is calculated as weight (in kg) divided by squared height (in meters). We used the Center for Disease Control's categorization of BMI to create the following set of BMI dummy variables: underweight ( $\text{BMI} < 18.5$ ), normal weight ( $18.5 \leq \text{BMI} < 25$ ), overweight ( $25 \leq \text{BMI} < 30$ ), obese ( $30 \leq \text{BMI} < 35$ ), and severely obese ( $\text{BMI} \geq 35$ ). The information on drinking included is an indicator for whether the respondent drinks alcohol. For smoking we included the number of cigarettes smoked per day for current smokers, and for former smokers the number of cigarettes they used to smoke per day before quitting. We scale the numbers of cigarettes smoked daily by dividing by 10.

Socio-demographic covariates included are age, education, race, an indicator for marital status (treating cohabitators as married), and household size. Specifically, age is included in the following form:  $(\text{age} - 50)/10$ ,  $((\text{age} - 50)/10)^2$ , and  $((\text{age} - 50)/10)^3$ . Based on the reported years of education (YoE), individuals are stratified into three different classes: primary education ( $\text{YoE} < 12$ ), secondary education ( $12 \leq \text{YoE} < 16$ ), and tertiary education ( $\text{YoE} \geq 16$ ). We generated three categories for race: white, black, and other.

#### A.2 Regression models

Age is included as  $(\text{age} - 50)/10$  and  $((\text{age} - 50)/10)^2$ . Other socio-demographic explanatory variables included are education, race, an indicator for marital status, and household size. See section A.1 for their definitions. Additionally, we included education of the household head's wife (measured in years), urbanicity, the wage rate, and a status dummy for private health insurance (either provided by an employer or purchased directly) to proxy the price variables.

The three dummy variables indicating the degree of urbanicity are derived from the Beale's code of urbanicity which we have aggregated as follows:

“Urban”: Central or fringe counties of metropolitan areas of 1 million population or more, and counties in metropolitan areas of 250,000 to 1 million population;

“Suburban”: Counties in metropolitan areas of less than 250,000, urban populations of 20,000 or more, and urban populations of less than 20,000 that are adjacent to metropolitan areas; and

“Rural”: Urban populations of less than 20,000 that are not adjacent to metropolitan areas and rural populations.

PSID respondents are asked whether they are salaried (and paid by week, two weeks, month, or year) or paid by the hour. To create a measure of (hourly) wage, we assume that, for those who are salaried, they work full-time (i.e., 40 hours per week).

Feldstein (1993) discusses three measures for the price of medical goods and services: the medical care price index (MCPI), the Scitovsky index, and health insurance premiums. It is, however, doubtful whether these accurately reflect the relevant prices for our purposes. Because of these considerations and because we only have a limited number of time periods, we proxy for the price variables by including a private health insurance (including employer-provided, as opposed to public insurance such as Medicare and Medicaid) status dummy that captures price differences for medical goods and services as faced by the individual. We also performed sensitivity analyses including year dummies as additional proxies for prices, which did not affect the results.

## **B Details of the health measurement model and construction of the health index**

The specification and estimation of our health measurement model and the construction of our health index closely follow Meijer et al. (2011). We use self-reported health and difficulties with activities of daily living (ADLs) as indicators of health. Unlike some other surveys, such as the Health and Retirement Study (HRS) or the Survey of Health, Ageing and Retirement in Europe (SHARE), the PSID does not measure grip strength or other objective health variables, so we are not able to include such continuous measures in the model, which would have allowed us to scale health to make it comparable with the European data used by Meijer et al. We use several health risk factors (BMI, past and present smoking, drinking alcohol), education, age, race, and household size as covariates.

The model specification is as follows. Let  $y_{nt,i}$  be the  $i$ -th health indicator for individual  $n$  at time  $t$ ,  $\eta_{nt}$  be latent true health of individual  $n$  at time  $t$ . Then, analogous to binary probit models, we assume that

$$\Pr(y_{nt,i} = 1 \mid \eta_{nt}) = \Phi(\tau_i + \lambda_i \eta_{nt}), \quad (13)$$

where  $\Phi(\cdot)$  is the standard normal cumulative distribution function. The parameters  $\tau_i$  are intercepts and the parameters  $\lambda_i$  are *factor loadings*. Similarly, for self-reported health, we assume an ordinal probit structure

$$\Pr(y_{nt,i} = j \mid \eta_{nt}) = \Phi(\alpha_j - \lambda_i \eta_{nt}) - \Phi(\alpha_{j-1} - \lambda_i \eta_{nt}), \quad (14)$$

where  $\alpha_j$ ,  $j = 1, \dots, 4$ , are threshold parameters, with  $\alpha_0 = -\infty$  and  $\alpha_5 = +\infty$ . Because location and scale of  $\eta_{nt}$  are arbitrary, we must fix one intercept parameter and one factor loading. We set the intercept of the dressing ADL to zero and its factor loading to 1.

The relation between  $\eta_{nt}$  and the covariates  $x_{nt}$  is given by the reduced form equation

$$(\eta_{nt} | x_{nt}) \sim \mathcal{N}(\gamma' x_{nt}, \psi), \quad (15)$$

We assume conditional independence (given  $\eta_{nt}$ ) of (13)–(15), although we allow for within-individual correlation across time by using standard errors clustered at the individual level.

Conditional on  $\eta_{nt}$ , the likelihood of observation  $(n, t)$  is the product of the relevant probabilities from (13) and (14) and the normal density following from (15). The likelihood of the observed data of observation  $(n, t)$  is then obtained by integrating  $\eta_{nt}$  out of the conditional likelihood. We estimate the model with maximum simulated likelihood (MSL), using 100 Halton draws of  $\eta_{nt}$ , which gives higher accuracy than a corresponding number of pseudo-random draws (Train, 2003).

We construct the health index as the conditional expectation of the latent health variable, given all the observed variables:

$$\text{HI}_{nt} = \text{E}(\eta_{nt} | y_{nt}, x_{nt}; \hat{\theta}),$$

where  $\hat{\theta}$  is the vector collecting the estimates of all parameters ( $\{\tau_i, \lambda_i\}, \{\alpha_j\}, \gamma, \psi$ ) of the health measurement model.

Missing data in the health indicators are accommodated using the common selection on observables (or missing at random, MAR) assumption. This amounts to setting the loglikelihood contribution of such a variable to zero. Missing covariates are not as easily accommodated, and we use the same practical solution as Meijer et al. (2011): plug in an arbitrary value, so that the covariate becomes nonmissing, and add a dummy for missingness to compensate for its average effect. For this arbitrary value, we use the median, which makes interpreting the coefficients of the missingness dummies easier, but the other coefficients are unaffected by this value.

See Meijer et al. (2011) for more details about the methodology behind the health measurement model and detailed formulas for construction of the health index.

## C Details on the regression models

For total medical expenditures and out-of-pocket expenditures, we estimated both Heckman-type bivariate sample selection models and two-part models. In both cases, the participation equation used a probit specification and the outcome equation was specified as a linear regression equation with the logarithm of the outcome variable as the dependent variable. The outcome equation is only estimated for observations with positive outcome values. The sample selection models explicitly estimate the correlation between the participation equation and the expenditure equation, but rely on strong assumptions about correctness of the functional form specification, homoskedasticity, and normality. The two-part models require weaker assumptions in this respect, but it is less clear to what extent they can accommodate correlation between the

two equations. See Duan, Manning, Morris, and Newhouse (1983, 1984) and Hay and Olsen (1984) for an early discussion of some of these issues, and Jones (2000) for an overview of the subsequent debate.

To account for the endogeneity of health, we have additionally estimated an instrumental variables version of the sample selection model. This estimates a system of three equations. The first two equations are the same as in the bivariate sample selection model

$$\begin{aligned} y_1^* &= x\beta_1 + w'\gamma_1 + \varepsilon_1, & y_1 &= 1(y_1^* > 0); \\ y_2^* &= x\beta_2 + w'\gamma_2 + \varepsilon_2, & y_2 &= y_2^* \text{ if } y_1 = 1, \text{ and unobserved otherwise,} \end{aligned}$$

where  $y_1$  indicates participation (positive OOP or total medical expenditures),  $y_2$  is the outcome variable (logarithm of OOP or total medical expenditures),  $x$  is the endogenous explanatory variable (health), and  $w$  is the vector of exogenous covariates. The third equation of the IV model is the first-stage equation

$$x = z'\pi + w'\gamma_3 + \varepsilon_3,$$

where  $z$  is the vector of instrumental variables (childhood health and parental smoking during childhood). The model is completed by assuming that the error terms  $\varepsilon_1$ ,  $\varepsilon_2$ , and  $\varepsilon_3$  are jointly normally distributed. Allowing nonzero correlations between these error terms explicitly introduces correlations between the endogenous explanatory variable  $x$  and the error terms  $\varepsilon_1$  and  $\varepsilon_2$ , which solves the endogeneity issue. We estimated this model with the `cmp` command in Stata (Roodman, 2009).

For the number of nights spent in a hospital we have estimated several count data models. The simplest count data model is Poisson regression, which assumes that, conditional on the explanatory variables  $x$ , the dependent variable follows a Poisson distribution with both mean and variance equal to  $\exp(x'\beta)$ . However, equality of conditional mean and variance is often violated. Therefore, we have also estimated a negative binomial regression model, which is an extension of the Poisson model that allows for greater dispersion through an additional variance parameter. A typical characteristic of many count variables is a spike at zero that is not well reproduced by these count data regression models. Nights spent in hospital shows this quite extremely, with only 282 observations with a positive amount, out of a total sample of 4,841. To accommodate this, we also estimated a hurdle model. This is a two-part model, with the first equation being a probit for a positive amount versus a zero amount, and the second equation being a zero-truncated negative binomial model, that is, a negative binomial model conditional on the amount being positive. A more detailed discussion of the estimated models can be found in Cameron and Trivedi (2005).

Finally, we estimate a count data model that takes into account the endogeneity of health. This involves specifying models for both the dependent variable and the endogenous regressor. We assume a Poisson model for the dependent variable and a linear reduced-form equation for the endogenous variable. The model is estimated using a two-step procedure as explained in Cameron and Trivedi (2009).

## References

- Attanazio, O. P., & Weber, G. (1989). Intertemporal substitution, risk aversion and the Euler equation for consumption. *Economic Journal*, 99(Suppl.), 59–73.
- Börsch-Supan, A., McFadden, D. L., & Reinhold, S. (1996). Living arrangements: Health and wealth effects. In D. A. Wise (Ed.), *Advances in the economics of aging* (pp. 193–218). Chicago: University of Chicago Press. (with a comment by S. F. Venti)
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. New York, NY: Cambridge University Press.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata*. College Station, TX: Stata Press.
- Case, A., & Deaton, A. (2005). Broken down by work and sex: How our health declines. In D. A. Wise (Ed.), *Analyses in the economics of aging* (pp. 185–212). Chicago, IL: University of Chicago Press.
- Chiappori, P.-A., & Paiella, M. (2011). Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association*, 9, 1021–1052.
- Cropper, M. L. (1977). Health, investment in health, and occupational choice. *Journal of Political Economy*, 85, 1273–1294.
- Duan, N., Manning, W. G., Jr., Morris, C. N., & Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, 1, 115–126.
- Duan, N., Manning, W. G., Jr., Morris, C. N., & Newhouse, J. P. (1984). Choosing between the sample-selection model and the multi-part model. *Journal of Business & Economic Statistics*, 2, 283–289.
- Ehrlich, I., & Chuma, H. (1990). A model of the demand for longevity and the value of life extension. *Journal of Political Economy*, 98, 761–782.
- Erbsland, M., Ried, W., & Ulrich, V. (1995). Health, health care, and the environment. Econometric evidence from German micro data. *Health Economics*, 4, 169–182. (Reprinted in A. M. Jones & O. O'Donnell, 2002, *Econometric analysis of health data*, Wiley, New York, pp. 25–36.)
- Feldstein, P. J. (1993). *Health care economics* (4th ed.). Albany, NY: Delmar.
- Finkelstein, A., Luttmer, E. F. P., & Notowidigdo, M. J. (2008). *What good is wealth without health? The effect of health on the marginal utility of consumption* (Working Paper No. 14089). Cambridge, MA: National Bureau of Economic Research.
- French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *Review of Economic Studies*, 72, 395–427.
- Galama, T. J. (2011). *A contribution to health capital theory* (Working Paper No. WR-831). Santa Monica, CA: RAND Corporation.
- Galama, T. J., & Kapteyn, A. (2011). Grossman's missing health threshold. *Journal of Health Economics*, 30, 1044–1056.
- Gerdtham, U.-G., & Johannesson, M. (1999). New estimates of the demand for health: Results based on a categorical health measure and Swedish micro data. *Social Science & Medicine*, 49, 1325–1332.
- Gerdtham, U.-G., Johannesson, M., Lundberg, L., & Isacson, D. (1999). The demand for health: Results from new measures of health capital. *European Journal of Political Economy*, 15, 501–521.

- Goldman, D. P., Jena, A. B., Lakdawalla, D. N., Malin, J. L., Malkin, J. D., & Sun, E. (2010). The value of specialty oncology drugs. *Health Services Research, 45*, 115–132.
- Grossman, M. (1972a). *The demand for health: A theoretical and empirical investigation* (Occasional Paper No. 119). New York, NY: National Bureau of Economic Research.
- Grossman, M. (1972b). On the concept of health capital and the demand for health. *Journal of Political Economy, 80*, 223–255.
- Grossman, M. (1998). On optimal length of life. *Journal of Health Economics, 17*, 499–509.
- Grossman, M. (2000). The human capital model. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics* (Vol. 1, pp. 347–408). Amsterdam, Netherlands: North-Holland.
- Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica, 50*, 1269–1286.
- Hay, J. W., & Olsen, R. J. (1984). Let them eat cake: A note on comparing alternative models of the demand for medical care. *Journal of Business & Economic Statistics, 2*, 279–282.
- Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior, 38*, 21–37.
- Jones, A. M. (2000). Health econometrics. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics* (Vol. 1, pp. 265–344). Amsterdam, Netherlands: North-Holland.
- Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature, 49*, 961–1075.
- Leu, R. E., & Doppmann, R. J. (1986). Gesundheitszustandsmessung und Nachfrage nach Gesundheitsleistungen. In E. Wille (Ed.), *Informations- und Planungsprobleme in öffentlichen Aufgabenbereichen* (pp. 1–90). Frankfurt am Main, Germany: Lang.
- Leu, R. E., & Gerfin, M. (1992). Die Nachfrage nach Gesundheit: Ein empirischer Test des Grossman-Modells. In P. Oberender (Ed.), *Steuerungsprobleme im Gesundheitswesen* (pp. 61–78). Baden-Baden, Germany: Nomos.
- MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy, 89*, 1059–1085.
- Meijer, E., Kapteyn, A., & Andreyeva, T. (2011). Internationally comparable health indices. *Health Economics, 20*, 600–619.
- Meijer, E., & Wansbeek, T. (2000). Measurement error in a single regressor. *Economics Letters, 69*, 277–284.
- Nocera, S., & Zweifel, P. (1998). The demand for health: An empirical test of the Grossman model using panel data. In P. Zweifel (Ed.), *Health, the medical profession, and regulation* (pp. 35–49). Boston, MA: Kluwer.
- Perlroth, D. J., Goldman, D. P., & Garber, A. M. (2010). The potential impact of comparative effectiveness research on U.S. health care expenditures. *Demography, 47*, S173–S190.
- Ried, W. (1998). Comparative dynamic analysis of the full Grossman model. *Journal of Health Economics, 17*, 383–425.
- Roodman, D. (2009). *Estimating fully observed recursive mixed-process models with cmp* (Working Paper No. 168). Washington, DC: Center for Global Development.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.



- Rust, J., & Phelan, C. (1997). How Social Security and Medicare affect retirement behavior in a world of incomplete markets. *Econometrica*, *65*, 781–831.
- Smith, J. P. (2004). Unraveling the SES–health connection. *Population and Development Review*, *30*(Suppl.), 108–132.
- Smith, J. P. (2009). Reconstructing childhood health histories. *Demography*, *46*, 387–403.
- Soldo, B. J., Mitchell, O. S., Tfraily, R., & McCabe, J. F. (2007). Cross-cohort differences in health on the verge of retirement. In B. Madrian, O. S. Mitchell, & B. J. Soldo (Eds.), *Redefining retirement: How will boomers fare?* (pp. 138–158). Oxford, UK: Oxford University Press.
- Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge, UK: Cambridge University Press.
- Usher, D. (1976). Comments on “The correlation between health and schooling”. In N. E. Terleckyj (Ed.), *Household production and consumption* (pp. 212–220). New York, NY: National Bureau of Economic Research. Retrieved July 4, 2011, from <http://www.nber.org/chapters/c3962.pdf>
- Van de Ven, W. P. M. M., & Van der Gaag, J. (1982). Health as an unobservable: A MIMIC-model of demand for health care. *Journal of Health Economics*, *1*, 157–183.
- Van den Berg, G. J., Deeg, D. J. H., Lindeboom, M., & Portrait, F. (2010). The role of early-life conditions in the cognitive decline due to adverse events later in life. *Economic Journal*, *120*, F411–F428.
- Van den Berg, G. J., Lindeboom, M., & Portrait, F. (2006). Economic conditions early in life and individual mortality. *American Economic Review*, *96*, 290–302.
- Van Doorslaer, E. K. A. (1987). *Health, knowledge and the demand for medical care: An econometric analysis*. Assen, Netherlands: Van Gorcum.
- Varaiya, P. P. (1998). *Lecture notes on optimization*. University of California, Berkeley. Retrieved May 31, 2011, from <http://paleale.eecs.berkeley.edu/~varaiya/papers.ps.dir/N00.pdf> (Previously published in 1972 as *Notes on optimization*, Van Nostrand Reinhold, New York.)
- Wagstaff, A. (1986a). The demand for health: Some new empirical evidence. *Journal of Health Economics*, *5*, 195–233.
- Wagstaff, A. (1986b). The demand for health: Theory and applications. *Journal of Epidemiology and Community Health*, *40*, 1–11.
- Wagstaff, A. (1993). The demand for health: An empirical reformulation of the Grossman model. *Health Economics*, *2*, 189–198.
- Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, *17*, 1329–1344.
- Wansbeek, T., & Meijer, E. (2000). *Measurement error and latent variables in econometrics*. Amsterdam, The Netherlands: North-Holland.
- Zweifel, P., & Breyer, F. (1997). *Health economics*. New York, NY: Oxford University Press.
- Zweifel, P., Breyer, F., & Kifmann, M. (2009). *Health economics* (2nd ed.). Berlin, Germany: Springer.