Testing the Mechanisms of Structural Models: The Case of the Mickey Mantle Effect

By HANMING FANG, MICHAEL KEANE, AHMED KHWAJA, MARTIN SALM AND DAN SILVERMAN[*]

A well-known method of validating econometric models (structural or otherwise) is to examine their performance in out-of-sample prediction. That is, given a change in the policy environment, do the key endogenous variables of the model move in ways that are in some sense "reasonably close" to the model's forecasts? Unfortunately however, as noted by Michael P. Keane and Kenneth I. Wolpin (forthcoming), the examination of models' predictive validity is not especially common in the micro-econometrics literature.

A common feature of structural econometric models is that latent variables, not observed and therefore not fit by the model, are key determinants of agents' behavior. For example, in Keane and Wolpin (2001), parental transfers to college-aged youth are key drivers of college attendance decisions; policy changes affect attendance partly through their effect on transfers. Yet their model is not actually fit to parental transfers, as these are unobserved in their data. Instead they are a latent variable (inferred from income, changes in assets, etc.). Thus, while Keane and Wolpin (2001) can check if policies affect college attendance as their model predicts, they cannot directly test the policy-to-transfers-to-behavior mechanism embedded in the model.

Here, we suggest one approach to validating structural econometric models is to seek evidence that such latent behavioral mechanisms embedded in a model are in fact operative. Of course, this requires collecting data on the relevant latent variable(s).[1] We present an example of this idea motivated by Ahmed Khwaja (2001)'s structural model of health over the life cycle. In the model, agents make sequential decisions on health insurance, health investments (preventive care and healthy/unhealthy behaviors) and medical treatment. Estimation uses the Health and Retirement Study (HRS). Simulation of the model generates the surprising result that provision of free health insurance would <u>not</u> cause people to engage in more risky behaviors like drinking and smoking, or less in healthy behaviors like exercise. This runs counter to the usual "moral hazard" story in static insurance models, where insurance induces more risky behavior.

The point is that, in a dynamic model, better insurance may increase life expectancy, as it allows one to afford more preventive care and better treatment in the event of illness. Increased life expectancy, in turn, enhances one's incentive to invest in health (i.e., in any dynamic model, a longer planning horizon – in this case, life-span – increases returns to investment).[2] This dynamic effect counteracts the static moral hazard effect of insurance on investment in health.

We call the mechanism where greater life expectancy increases investment in health the "Mickey Mantle effect," after the great Yankee slugger of the 50s and 60s. Mantle (1931-95) was a phenomenal natural talent. But a string of injuries, combined with heavy drinking and serious disregard for his health, ended his career prematurely. After years of alcoholism leading ultimately to liver failure, Mantle finally died of cancer at the age of 63. He explained his reckless behavior by noting he never expected to live past his early 40s, as most males in his father's line died young due to Hodgkin's disease.[3] He was surprised to live into his 60s, and observed: "If I knew I was going to live this long, I'd have taken better care of myself."[4]

Khwaja's model predicts that if we could observe the latent variable life expectancy, we would see the Mantle effect – a *ceteris paribus* positive effect of life expectancy on investment in health. If this effect were not present, it would cast serious doubt on the validity of the model, while if it is present we would gain additional confidence in the model.

The HRS collects data on subjective life expectancy. Our goal is to seek evidence of the Mantle effect using these data, by estimating health investment decision rules that include life expectancy.[5] Of course, life expectancy may be endogenous in the decision rule for investment in health, for two reasons. First, reverse causality: investment in health increases life expectancy. Second, omitted variables: A healthier person, *ceteris paribus*, expects to live longer and so has a greater return to investments that increase quality of life in old age. Thus, inadequate controls for health may create spurious correlation between life expectancy and health investment. Moreover, survey questions presumably measure subjective life expectancy with considerable error.

To clarify, consider the following simple schematic representation of Khwaja's model:

(A) Investment in health = f(lagged health, price of investment in health, income, taste for health, <u>life expectancy</u>)

(B) Life expectancy = g(lagged health, price of health care, <u>investment in health</u>, environmental risk factors, *genetic/hereditary factors*)

(C) Current health = h(lagged health, <u>investment in health</u>, environmental risk factors, *genetic/hereditary factors*, exogenous shocks to health)

(D) Insurance coverage = I(lagged health, insurance plan options, income, risk aversion, taste for health, tastes for insurance plan options)

Our interest is in estimating (A). Assume the error term in (A) arises because (i) the "taste for health" and (ii) some part of lagged health are unobserved. According to (B), life expectancy is affected by investment in health. Thus, life expectancy is endogenous in (A), as a person with (i) a high unobserved taste for health and/or (ii) higher than observed lagged health will tend to have both a high rate of investment in health and high life expectancy (creating spurious correlation).

A valid instrument for life expectancy in (A) is a variable that affects investment in health only through its affect on life expectancy (and not through any other channel). In the system (A)-(D), one's genetic/hereditary health endowment plays this role. Thus, motivated by the Mantle story, we instrument for subjective life expectancy using parents' age at death (or current ages if still alive), which serve as a proxy for the health endowment.[6, 7] Of course, the assumption that genetic/hereditary factors do not enter (A) directly is a strong one, but we think it is not unreasonable, given adequate controls for health and life expectancy.[8, 9]

I. Description of the HRS Data

We use the first six waves (1992-2002) of the HRS, which began as a panel study of the 1931 through 1941 U.S. birth cohorts (see www.hrsonline.isr.umich.edu). Participants in the first wave ranged from 51 to 61 years old, and were re-interviewed every two years. Spouses received an identical interview and could be of any age. In 1998 new cohorts born between 1942 and 1947 were added. We restrict our sample to all persons age 51 to 65 at the interview date who had complete information.[10] Our dependent variables are binary indicators for whether a respondent currently smokes, drinks heavily (average of 3 or more drinks per day), or is obese (Body Mass Index > 30). Our analysis sample contains 44,238 observations (which is reduced to 43,963 for the smoking regression, as this variable is missing for 275 cases).

Our measure of subjective life expectancy is a respondent's assessment of the percent chance he/she will live to age 75 or more. Prior studies show that such longevity probabilities are reasonable predictors of actual longevity.[11] Control variables include the respondents' age, gender, race, ethnicity, and marital status, household income, net household wealth, education, indicators for whether the respondents' father and mother had high school degrees and a measure of relative risk aversion (see Robert Barsky, F.T. Juster, and Miles Kimball, 1997).[12]

We control for health using a detailed set of health indicators. These include self-assessed health (i.e., excellent, very good or better, good or better, or fair or better), as well as a large number of objective measures, such as: whether the respondent had a recent overnight stay in a hospital, the number of limitations in activities of daily living (ADLs),[13] and binary indicators for whether the respondent was ever diagnosed with hypertension, diabetes, cancer, lung disease, heart problems, stroke, mental disease, or arthritis/rheumatism. We also include a number of variables measuring changes in health status since the last interview, and binary variables set to one if any of the health measures is missing.

Our instruments include the age at death of respondents' parents (or their current age if still alive), as well as age^2, age^3 and binary indicators of whether the father or mother died at an age that fell in the range of <65, 66 to 70, 71 to 75, 76 to 80, 81 to 85, or 86+.

Table 1 contains a complete list of variables used in the analysis, along with means and standard deviations. 58% of respondents are female. The average age is 58 and 65% expect to live to age 75+. 78% report being in good or better health, while 49% report very good or better. The average age at death of respondents' mothers is 74.4 while that of fathers was 70.7.[14]

II. Empirical Results

Given the large number of health measures, interpreting their coefficients in regressions for life expectancy or health investment is difficult. Thus, we conduct a factor analysis described in the **Appendix**. We kept the first 4 factors, which explain the bulk of the covariance amongst the health indicators. Factor 1 is by far the most important. It is a poor health factor with large negative loadings on self-reported health and large positive loadings on the physician diagnosed conditions. Factors 2 through 4 are all positive health factors whose interpretation is subtler.[15]

Table 2 reports the 1st stage results from two-stage least squares (2SLS). The dependent variable is expected probability of living to age 75+. The first column reports results using the 4 health factors, while the second column includes all the separate health indicators from Table 1.

Clearly, health Factor 1 is a far more important determinant of life expectancy than the other factors. A one standard deviation increase in (poor) health Factor 1 reduces expected probability of living to age 75+ by (-11.058)(.862)= 9.5 percentage points. Women's subjective probability of living to 75+ is about 3.5 points greater than men, *ceteris paribus*, while that for Blacks is almost 7 points greater than for whites. An additional 4 years of education raises this probability more than 2 points. Interestingly, marriage and assets are not significant and income, while significant, has a very small effect. The point estimates imply that roughly a 300 thousand dollar increase in annual income is needed to raise the subjective probability by just 1 point.

Parents' ages at death have large and significant effects in the expected direction. For instance, having a father whose age at death was 65 or less reduces the subjective probability of living to 75+ by 6.6 percentage points, *ceteris paribus*. The F-test for the joint significance of the parental age at death variables is 23.19 in column (1) and 24.06 in column (2).[16]

Table 3 reports OLS regressions of the health investment measures (smoking, heavy drinking, high-BMI) on the subjective probability of living to age 75+, along with controls for socio-demographics and health. The results provide modest support for the Mantle effect. For instance, in the smoking regressions in columns (1)-(2), subjective probability of living to age 75+ has *t*-statistics in the 9 to 10 range. The point estimates imply a 10-percentage point increase in this subjective probability reduces probability of smoking by about 1 percentage point.

Point estimates for heavy drinking are highly significant, but an order of magnitude smaller. However, the percent of respondents who report heavy drinking is also an order of magnitude smaller than that who report smoking (see Table 1), so in percentage terms the effect on behavior is similar. For high-BMI our results are not significant, statistically or quantitatively.

Table 4 reports our main IV results using parents' age at death as instruments for life expectancy. Here, the results are mixed. Those for smoking in columns (1)-(2) seem to provide strong support for the Mantle effect. Subjective life expectancy is highly significant, and the point estimates imply that OLS greatly understates the strength of the effect. Specifically, they imply that, *ceteris paribus*, a 10-percentage point increase in subjective probability of living to age 75+ reduces the probability of smoking by about 2.3 to 2.7 percentage points. As the percent of respondents who smoke is 22.6% (see Table 1) these represent decreases of 10 to 12%.

In contrast, for heavy drinking we obtain point estimates that are insignificant and of the wrong sign. The evidence for high-BMI is mixed. The point estimates are quantitatively large (at least half as great as for smoking) and of the right sign, but only marginally significant at best.

III. Conclusion

We have argued that testing the latent mechanisms of structural models, independent of full-blown structural estimation, can be a valuable model validation tool. This perspective has the benefit that it can potentially rationalize much of the descriptive or IV-based empirical work being done in economics as contributing to the structural research program. As an example of this idea, we attempt to find evidence for the "Mantle effect" that plays a key role in Khwaja's (2001) structural model of investment in health. We find clear evidence for the effect with respect to smoking, but mixed evidence with respect to heavy drinking and high-BMI.

REFERENCES

Barsky, Robert B., F.T. Juster, and Miles Kimball. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics*, 112 (2): 537-579.

Bloom, David E., David Canning, Michael Moore, and Younghwan Song. 2006. "The Effect of Subjective Survival Probabilities on Retirement and Wealth in the United States." National Bureau of Economic Research Working Paper No. 12688.

Hurd, Michael D., and Kathleen McGarry. 1995. "Evaluation of the Subjective Probabilities of Survival." *Journal of Human Resources,* 30 (Supplement): S268–92.

Hurd, Michael D., and Kathleen McGarry. 2002. "The Predictive Validity of Subjective Probabilities of Survival.'' *Economic Journal,* 112 (October): 966–85.

Keane, Michael P., and Kenneth I. Wolpin. Forthcoming. **"**Exploring the Usefulness of a Non-Random Holdout Sample for Model Validation: Welfare Effects on Female Behavior." *International Economic Review*.

Keane, Michael P., and Kenneth I. Wolpin. 2001. **"**The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review*, 42(4): 1051-1103.

Khwaja, Ahmed. 2001. "Health Insurance, Habits and Health Outcomes: A Dynamic Stochastic Model of Investment in Health." Ph.D. dissertation, University of Minnesota, 2001.

Khwaja, Ahmed, Dan Silverman, and Frank A. Sloan. 2006. "Are Smokers Misinformed?" mimeo, Duke University.

Smith, V. Kerry, Donald H. Taylor, Jr., and Frank A. Sloan. 2001. "Longevity Expectations and Death: Can People Predict Their Own Demise?" *American Economic Review,* 91(4): 1126-1134.

Table 1: Summary Statistics

Variable	Mean	Std. Dev.
Dependent Variables		
Smoking	0.226	0.418
Heavy Drinking	0.034	0.183
High BMI	0.269	0.443
Explanatory Variables		
Subjective probability of living to age 75+	65.182	29.063
Female	0.579	0.493
Black	0.146	0.354
Hispanic	0.075	0.264
Age	58.159	3.927
Years of education	12.466	2.981
Married	0.731	0.443
Household income (in 1,000 USD)	58.034	95.760
Household wealth (in 1,000 USD)	295.961	882.010
Risk tolerance	0.242	0.143
Father has high school degree	0.441	0.496
Mother has high school degree	0.458	0.498
Health Indicators		
Self reported health excellent	0.182	0.386
Self reported health very good or better	0.492	0.499
Self reported health good or better	0.782	0.412
Self reported health fair or better	0.932	0.250
Change in self reported health (1=better, 5=worse)	2.950	0.852
Number of ADL limitations	0.151	0.581
Change in number of ADL limitations	0.027	0.473
Hospital stay in last two years	0.171	0.376
Hypertension diagnosis ever	0.396	0.489
Diabetes diagnosis ever	0.115	0.320
Cancer diagnosis ever	0.070	0.255
Lung disease diagnosis ever	0.069	0.255
Heart problems diagnosis ever	0.136	0.342
Stroke diagnosis ever	0.032	0.176
Mental diseases diagnosis ever	0.112	0.316
Arthritis diagnosis ever	0.445	0.497
Hypertension first diagnosis last 2 years	0.032	0.176
Diabetes first diagnosis last 2 years	0.016	0.125
Cancer first diagnosis last 2 years	0.012	0.109
Lung disease first diagnosis last 2 years	0.010	0.099

Heart problem first diagnosis last 2 years	0.019	0.136
Stroke first diagnosis last 2 years	0.006	0.077
Mental disease first diagnosis last 2 years	0.015	0.122
Arthritis first diagnosis last 2 years	0.044	0.206
Hypertension worse	0.009	0.098
Diabetes worse	0.007	0.085
Cancer worse	0.001	0.037
Lung disease worse	0.007	0.088
Heart problems worse	0.007	0.085
Mental disease worse	0.005	0.076
Stroke remaining problems worse	0.009	0.097
Health status variable missing	0.003	0.057
Change in health variable missing	0.037	0.189
Health Factors§		
Health factor 1	-0.050	0.861
Health factor 2	-0.041	0.732
Health factor 3	0.037	0.720
Health factor 4	-0.005	0.699
Instrumental Variables		
Age of mother+	74.37	13.43
Age of father+	70.67	14.035
Mother age at death ≤ 65	0.200	0.400
Mother age at death 66 to 70	0.070	0.256
Mother age at death 71 to 75	0.096	0.295
Mother age at death 76 to 80	0.100	0.300
Mother age at death 81 to 85	0.098	0.297
Mother age at death > 85	0.079	0.271
Father age at death ≤ 65	0.312	0.463
Father age at death 66 to 70	0.112	0.315
Father age at death 71 to 75	0.137	0.344
Father age at death 76 to 80	0.121	0.327
Father age at death 81 to 85	0.101	0.301
Father age at death > 85	0.082	0.274
Number of observations++	43,963	

§ The health factors are constructed from the health indicators via factor analysis (see Appendix).

+ Age is either current age or age at death. 36% of mothers and 13% of fathers are still alive.

++ Sample statistics refer to the sample used in the smoking regression.

Table 2: 1st Stage IV Regression Results: Predicting Life Expectancy§

Dependent Variable: Subjective probability of living to age 75+

Independent Variable	(1) Health factors	(2)+ Health variables
Female	3.747***	3.497***
	(0.395)	(0.395)
Black	6.678***	6.865***
	(0.647)	(0.644)
Hispanic	-3.904***	-3.523***
	(0.934)	(0.932)
Age	0.392***	0.365***
	(0.050)	(0.051)
Years of education	0.652***	0.558***
	(0.075)	(0.075)
Married	0.688	0.643
	(0.443)	(0.437)
Health factor 1	-11.058***	
	(0.242)	
Health factor 2	0.693***	
	(0.228)	
Health factor 3	2.720***	
	(0.233)	
Health factor 4	0.662***	
	(0.258)	
Household income (in 1,000 USD)	0.003***	0.003**
	(0.001)	(0.001)
Household wealth (in 1,000 USD)	0.0001	0.0001
	(0.0001)	(0.0001)
Risk tolerance	-0.394	-0.462
	(1.311)	(1.296)
Father has high school degree	0.625	0.622
	(0.448)	(0.442)
Mother has high school degree	1.033**	0.992**
	(0.454)	(0.449)
Mother age at death ≤ 65++	-7.288***	-7.487***
	(1.404)	(1.393)
Mother age at death 66 to 70	-4.856***	-4.986***
	(1.049)	(1.037)
Mother age at death 71 to 75	-3.505***	-3.556***
	(0.803)	(0.794)

Mother age at death 76 to 80	-3.112***	-3.116***
	(0.706)	(0.699)
Mother age at death 81 to 85	0.103	0.138
	(0.631)	(0.622)
Mother age at death > 85	-0.824	-0.840
	(0.746)	(0.734)
Father age at death ≤ 65[++]	-6.623***	-6.628***
	(1.532)	(1.509)
Father age at death 66 to 70	-3.933***	-3.919***
	(1.124)	(1.111)
Father age at death 71 to 75	-3.391***	-3.363***
	(0.902)	(0.892)
Father age at death 76 to 80	-1.349**	-1.310*
	(0.771)	(0.763)
Father age at death 81 to 85	0.025	0.013
	(0.695)	(0.686)
Father age at death > 85	0.905	0.835
	(0.792)	(0.786)
Age of mother[+++]	-0.609	-0.497
	(0.542)	(0.541)
Age of mother ^ 2	0.005	0.003
	(0.009)	(0.009)
Age of mother ^ 3	0.00001	>-0.00001
	(0.00004)	(0.00004)
Age of father[+++]	-0.339	-0.311
	(0.622)	(0.614)
Age of father ^ 2	0.001	0.0007
	(0.010)	(0.010)
Age of father ^ 3	0.00001	0.00001
	(0.00005)	(0.00005)
R-squared	0.177	0.187
Partial R-squared (identifying instruments)	0.021	0.022
F-test for excluded instruments	23.19	24.06

Note: * denotes significance at the 10% level; ** denotes 5%; *** denotes 1%. Huber-White standard errors are in brackets, clustered at respondent level. Wave and birth region fixed effects are included but not shown.

[§] The table reports 1st stage results for the sample used in the smoking regression, which has 43,963 observations. 1st stage results for heavy drinking and high BMI are very similar, as the sample size is only increased to 44,238.

[+] All the health indicators in Table 1 are included in the regression in column 2, but the coefficients are not shown.

[++] The omitted categories for the father and mother age at death dummies are mother still alive and father still alive.

[+++] Age in the age polynomials is either current age or age at death.

Table 3: Ordinary Least Squares Regression Results

	Smoking		Heavy drinking		High BMI	
	(1)	(2) +	(3)	(4) +	(5)	(6) +
	Health factors	Health variables	Health factors	Health variables	Health factors	Health variables
Subj. Prob. of living	-0.0010***	-0.0009***	-0.0001***	-0.0001***	-0.0001	0.00006
to age 75+	(0.0001)	(0.0001)	(0.00004)	(0.00004)	(0.0001)	(0.0001)
Female	-0.046***	-0.049***	-0.055***	-0.057***	0.017**	0.012
	(0.007)	(0.007)	(0.003)	(0.003)	(0.007)	(0.007)
Black	-0.030***	-0.013	-0.005*	-0.005*	0.107***	0.076**
	(0.011)	(0.011)	(0.003)	(0.003)	(0.012)	(0.012)
Hispanic	-0.98***	-0.091***	-0.012**	-0.012**	0.043***	0.035**
	(0.016)	(0.015)	(0.005)	(0.005)	(0.016)	(0.016)
Age	-0.008***	-0.008***	-0.0004	-0.0005	-0.007***	-0.008***
	(0.0008)	(0.0008)	(0.0003)	(0.0003)	(0.0008)	(0.0008)
Years of education	-0.016***	-0.014***	-0.001***	-0.001***	-0.002	-0.002*
	(0.001)	(0.001)	(0.0005)	(0.0005)	(0.001)	(0.001)
Married	-0.110***	-0.107***	-0.017***	-0.016***	0.013	0.011
	(0.008)	(0.008)	(0.003)	(0.003)	(0.008)	(0.008)
Health factor 1	0.015***		-0.002**		0.082***	
	(0.004)		(0.001)		(0.004)	
Health factor 2	-0.011***		-0.002**		-0.012***	
	(0.003)		(0.001)		(0.003)	
Health factor 3	-0.009**		0.001		-0.040***	
	(0.003)		(0.001)		(0.004)	
Health factor 4	0.025***		0.003***		0.061**	
	(0.004)		(0.001)		(0.004)	
Household income	-0.00009***	-0.00008***	<0.00001	<0.00001	-0.00004	-0.00003
(in 1,000 USD)	(0.00002)	(0.00002)	(0.00001)	(0.00001)	(0.00003)	(0.00003)
Household wealth	-0.00001***	-0.00001***	<0.00001	<0.00001	<0.00001	<0.00001
(in 1,000 USD)	(<0.00001)	(<0.00001)	(<0.00001)	(<0.00001)	(<0.00001)	(<0.00001)
Risk tolerance	0.068***	0.066***	0.019**	0.020**	-0.042*	-0.033
	(0.024)	(0.023)	(0.009)	(0.0009)	(0.024)	(0.023)
Father has high	0.003	0.006	0.0009	0.001	-0.018**	-0.019**
school degree	(0.008)	(0.008)	(0.003)	(0.003)	(0.008)	(0.008)
Mother has high	0.0002	-0.0002	0.003	0.003	-0.0006	0.001
school degree	(0.008)	(0.008)	(0.003)	(0.003)	(0.008)	(0.008)
R – squared	0.05	0.07	0.03	0.03	0.06	0.09
Observations	43,963	43,963	44,238	44,238	44,238	44,238

Note: * denotes significant at the 10% level; ** denotes 5%; *** denotes 1%. Huber-White standard errors are in brackets, clustered at respondent level. Wave and birth region dummies are included but coefficients are not shown.

+ All the health indicators listed in Table 1 are included, but their coefficients are not shown.

Table 4: IV Regression Results: 2nd Stage

	Smoking		Heavy drinking		High BMI	
	(1)	(2) +	(3)	(4) +	(5)	(6) +
	Health factors	Health variables	Health factors	Health variables	Health factors	Health variables
Subj. Prob. of living to age 75+	-0.0023** (0.0009)	-0.0027*** (0.0008)	0.0004 (0.0003)	0.0003 (0.0003)	-0.0018** (0.0009)	-0.0012 (0.0009)
Female	-0.041*** (0.008)	-0.043*** (0.008)	-0.058*** (0.003)	-0.059*** (0.003)	0.025*** (0.008)	0.016** (0.008)
Black	-0.022* (0.013)	-0.001 (0.013)	-0.009*** (0.003)	-0.009** (0.003)	0.120*** (0.013)	0.085*** (0.013)
Hispanic	-0.102*** (0.016)	-0.097*** (0.016)	-0.010* (0.005)	-0.010* (0.005)	0.035** (0.017)	0.031* (0.016)
Age	-0.008*** (0.0009)	-0.007*** (0.0009)	-0.0007** (0.0003)	-0.0007** (0.0003)	-0.006*** (0.0009)	-0.007*** (0.0009)
Years of education	-0.015*** (0.001)	-0.013*** (0.001)	-0.002*** (0.0005)	-0.002*** (0.0005)	-0.0061 (0.001)	-0.001 (0.001)
Married	-0.109*** (0.008)	-0.106*** (0.008)	-0.017*** (0.003)	-0.017*** (0.003)	0.014* (0.008)	0.012 (0.008)
Health factor 1	-0.001 (0.011)		0.004 (0.004)		0.059*** (0.011)	
Health factor 2	-0.010*** (0.003)		-0.003** (0.001)		-0.010*** (0.004)	
Health factor 3	-0.005 (0.004)		0.0003 (0.001)		-0.035*** (0.004)	
Health factor 4	-0.026*** (0.004)		0.003** (0.001)		-0.060*** (0.004)	
Household income (in 1,000 USD)	-0.00008*** (0.00002)	-0.00008*** (0.00002)	<0.00001 (<0.00001)	<0.00001 (0.00001)	-0.00003 (0.00003)	-0.00003 (0.00003)
Household wealth (in 1,000 USD)	-0.00001*** (<0.00001)	-0.00001*** (<0.00001)	<0.00001 (<0.00001)	<0.00001 (<0.00001)	<0.00001 (<0.00001)	<0.00001 (<0.00001)
Risk tolerance	0.068*** (0.024)	0.065*** (0.024)	0.019** (0.009)	0.020** (0.009)	-0.042* (0.024)	-0.033 (0.023)
Father has high school degree	0.004 (0.008)	0.007 (0.008)	0.0005 (0.003)	0.0007 (0.003)	-0.017* (0.008)	-0.018** (0.008)
Mother has high school degree	0.001 (0.008)	0.001 (0.008)	0.002 (0.003)	0.002 (0.003)	0.002 (0.008)	0.002 (0.008)
R-squared	0.05	0.06	0.02	0.02	0.05	0.08
Hansen J – Statistic (P – value)	13.542 (0.699)	14.300 (0.645)	24.135 (0.115)	23.690 (0.128)	20.038 (0.272)	20.328 (0.257)

Note: * denotes significant at the 10% level; ** denotes 5%; *** denotes 1%. Huber-White standard errors are in brackets, clustered at respondent level. Wave and birth region dummies are included but coefficients are not shown.
+ All the health indicators listed in Table 1 are included, but their coefficients are not shown.

ENDNOTES:

---

[*] Fang: Department of Economics, Yale University, New Haven, CT 06520-8264; and SHUFE (email: hanming.fang@yale.edu); Keane: ARC Federation Fellow, University of Technology Sydney and Research Fellow, Arizona State University (email: michael.keane@uts.edu.au); Khwaja: Fuqua School of Business, Duke University, 1 Towerview Drive, Durham, NC 27708 (email: ahmed.khwaja@duke.edu); Salm: University of Mannheim, 68131 Mannheim, Germany (email: salm@mea.uni-mannheim.de); Silverman: Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109-1220 (email: dansilv@umich.edu). This work was supported by Australian Research Council grant FF0561843 to Michael Keane.

[1] Keane and Wolpin (2001) did examine historical data on transfers from other sources, and they argue that qualitatively it follows the patterns predicted by their model. But the additional data was inadequate to directly test the policy-to-transfers-to-schooling behavioral mechanism.

[2] More intuitively, if one expects to live longer, it creates an incentive to invest in health to enhance quality of life in old age. On the other hand, the expectation of greater longevity can also reduce the marginal value of additional years of life, a mechanism that would reduce investment in health. In Khwaja's model the investment increasing effects of a greater expected life span dominate.

[3] Mantle's father died of Hodgkin's disease at age 39, while his two uncles died at 32 and 41. Of Mantle's four sons, one died of Hodgkin's disease a year before Mantle (at age 36), while the other died of cancer in 2000. Mantle's two surviving sons are active in the Mickey Mantle Foundation, which promotes organ donations.

[4] This could be *ex post* rationalization, but Mantle made related statements in his youth – e.g., as a rookie he told the player representative: "You don't have to talk to me about pensions. I won't be around long enough to collect one."

[5] In principle, Khwaja might have used these data in estimation, but modeling expectations entails great difficulties.

[6] According to the framework (A)-(D), measures of environmental risk factors are also potential instruments. Living in a risky environment may reduce life expectancy, but conditional on life expectancy it should not affect investment in health directly. Of course, this assumes that risky environment is not endogenous in the sense that people with low tastes for health will also choose to live in a risky environment.

[7] After completing this work we became aware of a recent paper by David E. Bloom et al. (2006) that uses similar instruments to estimate effects of life expectancy on saving and the timing of retirement.

[8] Our key identifying assumption is that investment is conditionally mean independent of the genetic health endowment, given the controls in (A) – i.e., life expectancy, measured health, income, prices. But a family history of congenital disease might affect investment in health through other channels (e.g., having fewer financial resources in youth if parents were ill). However, it can be plausibly argued that this problem is resolved by conditioning on current health status in (A), as this would control for effects of family background on prior investments in health.

[9] Insurance coverage may tend to be correlated with unobserved tastes for health. Thus, there may be a selection bias whereby people with greater taste for health also have more comprehensive insurance (and hence, a lower cost of investment in health). In that case, consistent estimation of (A) would require us to deal with this selection problem. That in turn, would require estimating (A) jointly with the choice model for insurance coverage in (D).

[10] In the HRS wave 1-6, there were 56,567 observations in the 51-65 age range. Of these, 7,564 were dropped because of missing information on longevity expectations, 3,087 because of missing information on age and death of parents, 1,557 because they had not answered the question about risk aversion. In addition 26 observations had missing information on education, 48 on birth region, and 38 had non-responses to questions on worsening of health conditions, giving an analysis sample of 44,238. There are 275 missing observations for the smoking question.

[11] See Michael D. Hurd and Kathleen McGarry, 1995, 2002; V. Kerry Smith, Donald H. Taylor, Jr., and Frank A. Sloan, 2001 and Ahmed Khwaja, Dan Silverman and Frank A. Sloan, 2006.

[12] Note that the variables "price of investments in health" in (A) and "price of health care" in (B) depend on prices of alcohol and tobacco, proximity to and cost of healthy food, proximity to athletic facilities, etc., as well as insurance coverage. We do not measure these variables directly, but instead proxy for them using time and region dummies (determinants of prices) and variables like education, income and risk tolerance (which drive insurance coverage).

[13] ADLs are whether the individual is able to independently walk, dress, bathe, eat, get into bed, and use the toilet.

[14] These figures include current age for parents who are still alive (36% of mothers and 13% of fathers).

[15] Factors 2 and 3 load positively on self-reported health but also load positively on stroke indicators. But while 2 loads negatively on the *change* in self-reported health, 3 loads positively. Factor 4 loads positively on self-reported health and negatively on hypertension and diabetes indicators. But it also loads positively on ADL limitations.

[16] While the six parental age polynomial coefficients are not individually significant, their joint F-test is 10.07 in column (1) and 9.90 in column (2).