

ON THE DISTRIBUTION AND DYNAMICS OF HEALTH CARE COSTS: EXTENDED VERSION

Eric French
Federal Reserve Bank of Chicago

John Bailey Jones*
SUNY-Albany

December 18, 2004

Abstract

Using data from the Health and Retirement Survey and the Assets and Health Dynamics of the Oldest Old survey, this paper presents estimates of the stochastic process that determines both the distribution and dynamics of health care costs. We find that the data generating process for log health costs is well represented as the sum of a white noise process and a highly persistent AR(1) process. We also find that the innovations to this process can be modelled with a normal distribution that has been adjusted to capture the risk of catastrophic health care costs. Simulating this model, we find that in any given year, 0.1% of households receive a health cost shock with a present value of at least \$125,000.

*This paper is an detailed background version of an almost-identically-titled paper. Comments on either paper are welcome at efrench@frbchi.org and jbjones@albany.edu. We thank John Rust for encouraging us to write this paper. We also thank Meredith Crowley, Jonathan Skinner, and seminar participants at the Conference on Social Insurance and Pensions Research in Aarhus, Denmark, the Econometric Society, the Chicago Fed, and the NBER summer institute for helpful comments. Three anonymous referees and especially editor Bent Jesper Christensen provided detailed and useful comments. Diwakar Choubey, Kate Godwin, Ken Housinger, Kirti Kamboj, Tina Lam, and Santadarshan Sadhu provided excellent research assistance. The research reported herein was funded in part by a grant from the U.S. Social Security Administration (SSA) to the Center for Retirement Research at Boston College. The views of the authors do not necessarily reflect those of the Federal Reserve System or the SSA. Recent versions of the paper can be obtained at <http://www.chicagofed.org/economists/EricFrench.cfm/>.

1 Introduction

The way in which uncertain health care costs affect saving and labor supply is a topic of much debate.¹ Moreover, health care costs, in and of themselves, are of great interest to many policymakers, including those considering changes to the United States' Medicare program. Despite this interest, there are relatively few estimates of the stochastic process for health care costs. In this paper, we provide new evidence on both the cross-sectional distribution and the intertemporal persistence of health care costs, using panel data from the Health and Retirement Survey (HRS) and the Assets and Health Dynamics of the Oldest Old (AHEAD) survey.

We find that the stochastic process for log health costs is well modelled as the sum of a white noise process and a highly persistent AR(1) process. We also find that the innovations to this process can be modelled with a normal distribution, but the variance of this innovation distribution, as well as the mean for the overall process, should be adjusted—in a way made precise below—to better capture the risk of catastrophic health cost shocks. Using this model to simulate a large number of health cost histories, we find that our preferred model generates more lifetime health cost risk than models found in earlier studies.

The only fully parameterized model of the health cost process currently available is the one of Feenberg and Skinner (1994), who find that log health costs are well represented by an ARMA(1,1) process. In this paper, we extend Feenberg and Skinner's results in several directions. First, our data are 20 years more recent than theirs. Second, Feenberg and Skinner observe medical costs only for those who deduct medical expenses on their taxes. Although they develop a sophisticated econometric procedure to address this selection problem, we avoid it completely, because we observe medical expenses for virtually all members of our sample. Third, our econometric approach allows us to use unbalanced panel data. This allows us to include people who die, who tend to incur high medical expenses.

Feenberg and Skinner assume that the cross-sectional distribution of health care costs is lognormal. Eichner et al. (1998), in contrast, model the cross section non-parametrically, and conclude that health cost shocks are “not approximated well by any analytic solution.” However, they do not test their non-parametric model against a parametric alternative. Rust

¹Recent papers include studies of saving by Hubbard et al. (1994, 1995), Palumbo (1999), and Dynan et al. (2002), and studies of retirement by Rust and Phelan (1997), Blau and Gilleskie (2001) and French and Jones (2003).

and Phelan (1997) argue that the right tail of the health cost distribution is better represented by a Pareto distribution. A Pareto distribution has a fatter right tail than a lognormal distribution, generating a higher probability of catastrophic health care costs at any variance. Because the extremely risk-averse may be particularly concerned about catastrophic shocks, the Pareto specification may be one of the reasons why Rust and Phelan find that health insurance has large effects on retirement behavior.² However, Rust and Phelan do not formally test their Pareto specification against a lognormal alternative, nor do they account for the persistence of health costs. Because the lifetime effect of health cost shocks depend as much on their duration as on their size, it is important to consider cross-sectional and times series properties jointly. While Eichner et al. do consider persistence, their sample has an extremely short time horizon, and, moreover, consists only of insured individuals from a single firm.³

We begin our analysis by examining the time series properties of health costs. To do this, we employ a commonly-used error components methodology that works well with short panels and imposes no distributional assumptions. We find that log health costs are reasonably well represented by the sum of an AR(1) component and a white noise component. The AR(1) component is quite persistent, so that health cost shocks can have a large impact on lifetime wealth. Because this sum can be rewritten as an ARMA(1,1) process, our results comport with Feenberg and Skinner’s findings.

We then consider the cross-sectional distribution. Using the likelihood ratio test developed by Vuong (1989), we conclude that in most cases, the (truncated) lognormal and Pareto specifications fit the upper tail of the distribution (i.e., the top decile) equally well. Moving on to our primary goal of matching the overall cross-sectional distribution, we conclude that the lognormal distribution is the better model. The lognormal distribution that best fits the overall cross section, however, understates the upper tail of the distribution, and thus understates the risk of catastrophic shocks. Fortunately, we find that a useful analytical approximation is the “fitted” lognormal distribution that matches the mean and the 99.5th percentile of the empirical distribution. This approximation captures two of the most salient features of the health cost distribution: the overall average cost, and the magnitude of a

²Rust and Phelan also rule out the possibility that households self-insure by saving.

³Palumbo (1999) conditions health costs on a persistent health state variable, but provides little description of the residual distribution.

“catastrophic shock”. Using these health cost distributions to estimate the welfare losses from health cost uncertainty, we find that the fitted distribution outperforms the “standard” lognormal distribution, which tends to understate the losses.

To complete our model of the stochastic process for health care costs, we need the distribution of health cost innovations. Because the *sum* of an AR(1) and a white noise process is not a Markov process, we cannot use our time series model to back out an empirical distribution of log health cost innovations. We are left to infer the innovation distribution from the cross-sectional distribution. In this respect, the lognormal performs particularly well: a Gaussian ARMA process for log health costs generates a lognormal cross-sectional distribution for health costs.

Our preferred model of the health cost process is thus the combination of our two-component time series model with our lognormal approximation of the cross section. Using this model to simulate health cost histories, we find that in any given year 0.1% of households receive a health cost shock that exceeds \$125,000 in present value. This is much more risk than is generated by Feenberg and Skinner’s estimates.

The rest of the paper is organized as follows. In Section 2, we describe the health cost data contained in the HRS and AHEAD surveys. In Section 3, we examine the correlation of health care costs across time, and in Section 4, we examine the cross-sectional distribution. In Section 5, we estimate lifetime health cost risk by simulating our preferred health cost model. We conclude in Section 6.

2 Data

We use data from the HRS and AHEAD surveys. These data contain detailed information on health care costs, health insurance, and demographics.

The HRS is a sample of individuals that were non-institutionalized and aged 51-61 in 1992. Spouses of these individuals were also interviewed, regardless of their age. The HRS includes a nationally representative core sample, and additional samples of blacks, Hispanics, and Florida residents. A total of 12,652 individuals in 7,608 households were interviewed in 1992. These individuals were again interviewed in 1994, 1996, 1998, and 2000, creating up to five separate responses for each individual.

The AHEAD is a nationally representative sample of individuals that were non-institutionalized

and aged 70 or older in 1993. Because the same researchers at the University of Michigan collect both the HRS and the AHEAD data, the two data sets have similar sample designs. A total of 8,222 individuals in 6,047 households were interviewed for the AHEAD survey in 1993. These individuals were again interviewed in 1995,⁴ 1998, and 2000, creating up to four separate responses for each individual.

In 1998 and 2000 all individuals in the HRS and AHEAD (as well as an additional sample of older individuals) were asked the same questions. In the HRS and AHEAD waves before 1998, many of the questions asked were the same across the two datasets, allowing us to merge them together. Because the health insurance and health cost data are incomplete in wave 1 of both datasets, we use waves 2 through 5 in the analyses below.

Table 1 presents means and standard deviations of variables that measure health care costs, health insurance coverage, health care utilization, and demographic features. All of the variables are measured at the household level. To annualize the data, we divide the health cost and health care utilization measures by the number of years since the individual was last interviewed—which on average is two—or by two if the individual was never previously interviewed.

Virtually all Americans aged 65 and older (as well as those eligible for government-provided disability insurance) are eligible for insurance through the government’s Medicare program. We therefore split the sample between those younger than 65 and those older than 65. Note that after age 65, nearly everyone has some form of insurance, although the fraction of individuals with employer-provided coverage falls, as many workers lose some or all of their employer-provided coverage when they leave their job and/or become eligible for Medicare. The other major form of publicly-provided health insurance is Medicaid, which is available to individuals with low income and very few assets. Those who report not having any insurance are assigned to the “none” category.⁵

The central variable of interest in this study is the total amount of health care costs paid

⁴ Throughout our analysis, we include the 1995 wave of the AHEAD with the 1994 wave of the HRS. We also tried including the 1995 wave of the AHEAD with the 1996 wave of the HRS, which had only a small effect on the results.

⁵ Many individuals have multiple forms of insurance. We assigned respondents to health insurance type according to the following hierarchy: employer-provided coverage (this category includes insurance to current or former government employees); Medicaid; privately-purchased but not employer-provided; and Medicare. We classify married households on the basis of the husband’s insurance. See French and Kamboj (2002) for a fuller description of the construction of all of these variables.

Variable	Age < 65		Age ≥ 65	
	Mean	Std. Dev.	Mean	Std. Dev.
Annual health care costs (in 1998 dollars)	2,365	(4,271)	2,805	(6,072)
Male head of household	0.64	(0.48)	0.51	(0.50)
Married	0.48	(0.50)	0.37	(0.48)
Age	58.5	(3.6)	76.9	(8.1)
No insurance (none)	0.15	(0.36)	0.01	(0.11)
Employer-provided insurance	0.61	(0.49)	0.28	(0.45)
Privately-purchased insurance	0.10	(0.31)	0.25	(0.43)
Medicaid	0.09	(0.28)	0.15	(0.36)
Medicare	0.05	(0.22)	0.31	(0.46)
Income (in 000s of 1998 dollars)	49.8	(100.5)	29.6	(45.5)
Assets (in 000s of 1998 dollars)	249.8	(756.9)	298.0	(1,113)
Annual doctor visits	6.2	(10.3)	7.2	(9.9)
Annual nursing home nights	0.7	(14.9)	8.0	(48.6)
Annual hospital nights	1.2	(4.9)	2.1	(6.8)
	N = 15,990		N = 18,903	

Table 1: SAMPLE STATISTICS

by the household. Health care costs are the sum of what the household spends on insurance premia, drug costs, and costs for hospital, nursing home care, doctor visits, dental visits, and outpatient care. For our sample, mean household health care costs for those younger than 65 are \$2,365 and mean health care costs for those aged 65 and older are \$2,805. This compares to the US per capita average of \$2,832 for households headed by a non-institutionalized individual aged 65 or older (Federal Interagency Forum on Aging-Related Statistics, 2000). Note that even though most individuals in our sample are insured, there is a great deal of variation in health care costs. The standard deviation of health care costs is \$4,271 and \$6,072 for those younger and older than 65, respectively. This is not surprising, as most health insurance plans have deductible and/or co-pay provisions. Moreover, many insurance plans (such as Medicare) do not cover prescription drugs.

One important reason why average health care costs in the HRS/AHEAD data are below the national average is that individuals in the HRS/AHEAD spend far fewer nights in a nursing home.⁶ In our sample, individuals aged 65 or older spent 8.0 nights per year in a nursing home per year, as opposed to the national average of 15.8 nights (National Center for Health Statistics, 1999). Because the HRS/AHEAD sample was initially drawn from

⁶Selden et al. (2001) find that 9% of total aggregate health costs and 13% of costs paid out-of-pocket arise from nursing home visits. Because of the skewness of nights spent in a nursing home, Palumbo (1999) argues that nursing homes are a significant source of health cost uncertainty for the elderly.

the non-institutionalized population—which excludes individuals in nursing homes—it is not surprising that its members spend relatively fewer nights in a nursing home. HRS/AHEAD members who enter a nursing home after the initial interview, however, are retained in the sample, and re-interviewed. By the later waves, many HRS/AHEAD members have entered a nursing home. In wave 5, individuals aged 65 or older spent 12.7 nights in a nursing home, which is much closer to the aggregate value.⁷

3 The Persistence of Health Care Costs

We construct the health cost process in two steps, first finding its autocorrelation structure, and then finding its innovation distribution. Feenberg and Skinner (1994) find that autocorrelation structure of logged health costs is well represented by an ARMA(1,1) process. To re-examine their findings, we evaluate several time series models with a commonly-used error components methodology.⁸ This approach works well with short panels, and it requires no distributional assumptions.

We estimate the following error components model:

$$\ln hc_{it} = X'_{it}\beta + R_{it}, \quad (1)$$

$$R_{it} = f_i + a_{it} + u_{it}, \quad (2)$$

$$a_{it} = \rho a_{it-1} + \epsilon_{it}, \quad (3)$$

$$u_{it} = \psi_{it} + \phi\psi_{it-1}, \quad (4)$$

where $X'_{it}\beta$ is the expectation of health costs conditional on the covariate vector X_{it} , and R_{it} is the residual, which can be decomposed into: f_i , a permanent person-specific component; a_{it} , an autoregressive component; and u_{it} , a moving average component. Note that t denotes a two-year period.

The estimation procedure has two stages. In the first stage, we estimate the parameter vector β in equation (1) by regressing log health costs on demographic and health insurance

⁷For nights spent in a hospital, the HRS/AHEAD matches the national statistics rather well. For individuals older than 65, average nights in a hospital are 2.1 in the HRS/AHEAD, as opposed to the national average of 2.3 (National Center for Health Statistics, 1999).

⁸See Abowd and Card (1989), Pischke (1995) and Baker (1997) for similar approaches.

variables that households can use to forecast future health costs.⁹ Table 2 presents the parameter estimates. We use the OLS estimates throughout the paper, but the GLS estimates are very similar.¹⁰

Of particular interest is the coefficient on log income of 0.179.¹¹ Households have some control over the quality of care they receive; ideally, the variation induced by this choice should be omitted from our measure of health cost risk.¹² Most of the remaining parameter estimates are of the expected sign. The one surprising finding is that those with no health insurance have lower health care costs than those with employer-provided insurance. This is particularly surprising given that these estimates exclude employer expenditures, which average over \$2,700 per employee (Employee Benefit Research Institute, 1999). As French and Kamboj (2002) show, one reason for this is that those with employer-provided insurance are more likely to obtain health care services. Households receiving Medicaid spend significantly less. Given that the government provides Medicaid for free to those with low income and assets, this is hardly surprising.

In the second stage of the estimation procedure, we estimate the covariance matrix of the residuals from the first step regression, and fit to it the model described in equations (2)–(4). For tractability, we assume that a_{it} is a stationary process ($|\rho| < 1$ and ϵ_{it} is homoskedastic) and that the components in equations (1)–(4) are mutually orthogonal.¹³ We also assume

⁹In all the analyses that follow, health care costs below \$250 (including reports of no expenditures) were recoded to \$250. For more interpretation of these low cost reports, see French and Kamboj (2002). One alternative bottom-coding scheme is the one employed by Hubbard et al. (1994), who drop zero-cost observations, but recode none of the others. Applying their rule would lead us to drop about 10% of our observations.

¹⁰The reported GLS estimates use the empirical covariance matrix reported in Table 3, although we also estimated GLS coefficients using the estimated covariance matrix implied by the model in column 2 of 4 for the weighting matrix.

¹¹Feenberg and Skinner find an elasticity of around 0.38. See Cutler and Zeckhauser (2000) for other estimates of income elasticities.

¹²Moreover, health insurance coverage is not completely exogenous. This could lead to inconsistent estimates of β and the health cost risk that individuals face. On the one hand, individuals may purchase health insurance in response to a health shock. On the other hand, negative health shocks could cause the loss of a job and the health insurance associated with that job. The latter effect seems to be more important. Of individuals aged 65 or younger with no private health insurance in the previous wave, only 5.0% of those whose health got worse purchased private health insurance, whereas 6.3% of those whose health did not change obtained private health insurance and 5.8% of those whose health improved purchased private health insurance. Of individuals aged 65 or younger with no employer-provided insurance in the previous wave, only 8.6% of those whose health got worse obtained employer provided health insurance, whereas 13.5% of those whose health did not change obtained employer provided health insurance and 14.2% of those whose health improved obtained employer-provided health insurance.

¹³Specifically, we assume $E(X_{it}R_{it}) = E(f_i a_{it}) = E(f_i u_{it}) = E(a_{it} u_{it}) = E(a_{it-1} \epsilon_{it}) = E(\psi_{it} \psi_{it-1}) = 0$.

Variable	OLS Estimates		GLS Estimates	
	Coefficient	(Robust S.E.)	Coefficient	(S.E.)
Male	-0.121	(0.020)	-0.106	(0.019)
Married	0.729	(0.020)	0.718	(0.019)
Age	0.0451	(0.009)	0.062	(0.008)
Age ²	-0.00020	(0.00006)	-0.00032	(0.00005)
Employer-provided × (age < 65)	0.227	(0.026)	0.269	(0.023)
Privately-purchased × (age < 65)	1.34	(0.035)	1.18	(0.031)
Medicaid × (age < 65)	-0.309	(0.039)	-0.281	(0.034)
None or Medicare × (age ≥ 65)	-0.019	(0.033)	-0.043	(0.029)
Employer-provided × (age ≥ 65)	0.186	(0.033)	0.201	(0.030)
Privately-purchased × (age ≥ 65)	0.972	(0.034)	0.844	(0.031)
Medicaid × (age ≥ 65)	-0.481	(0.037)	-0.438	(0.033)
Log income	0.179	(0.009)	0.155	(0.008)
Wave dummies included				
$N = 34,893$	$R^2 = 0.30, \sigma = 1.05$		$R^2 = 0.30, \sigma = 1.05$	

Table 2: LEAST SQUARES REGRESSIONS OF LOG HEALTH COSTS

that the permanent person-specific effect f_i is unchanging over time, so that:

$$\text{Var}(f_i) = \sigma_f^2; \quad \text{Var}(a_{it}) = \sigma_a^2; \quad \text{Var}(\epsilon_{it}) = \sigma_\epsilon^2. \quad (5)$$

Note that $\sigma_\epsilon^2 = \sigma_a^2(1 - \rho^2)$. We allow for heteroskedasticity, however, in the innovation to the MA(1) component:

$$\text{Var}(\psi_{it}) = \sigma_{\psi t}^2. \quad (6)$$

The variances and autocovariances implied by the model are:

$$\text{Var}(R_{it}) = (\sigma_f^2 + \sigma_a^2 + \sigma_{\psi t}^2 + \phi^2 \sigma_{\psi t-1}^2), \quad \text{for all } t, \quad (7)$$

$$\text{Cov}(R_{it}, R_{i,t+1}) = (\sigma_f^2 + \rho \sigma_a^2 + \phi \sigma_{\psi t}^2), \quad \text{for all } t, \quad (8)$$

$$\text{Cov}(R_{it}, R_{i,t+k}) = (\sigma_f^2 + \rho^k \sigma_a^2), \quad \text{for all } k > 1, t. \quad (9)$$

Table 3 shows the empirical covariance matrix. Covariances appear below the diagonal of this matrix, variances appear along the diagonal, and autocorrelations appear above it. Standard errors are in parentheses. Because the data are unbalanced, Table 3 also shows the number of observations in each cell (in brackets).¹⁴ The covariance matrix gives us 10

¹⁴Wave 3 has very few observations because AHEAD respondents were not interviewed in wave 3. See footnote 4. The number of households increases across waves for two reasons. First, fewer observations are

moment conditions to match.¹⁵

	Wave 2	Wave 3	Wave 4	Wave 5
Wave 2	1.1257 (0.0185) [7, 935]	0.3854	0.3221	0.3082
Wave 3	0.4386 (0.0220) [2, 804]	1.1504 (0.0286) [3, 938]	0.4113	0.3281
Wave 4	0.3552 (0.0171) [5, 358]	0.4585 (0.0214) [3, 037]	1.0806 (0.0168) [10, 826]	0.4139
Wave 5	0.3391 (0.0160) [5, 809]	0.3649 (0.0196) [3, 164]	0.4462 (0.0138) [8, 489]	1.0754 (0.0157) [12, 194]
Covariances lie below the diagonal, correlations above Standard errors in parentheses Sample sizes in brackets				

Table 3: EMPIRICAL COVARIANCE MATRIX

Using a minimum distance estimator, we fit several variants of the error components model to this covariance matrix. Details of the estimation procedure are in appendix B; the underlying asymptotic theory can be found in Chamberlain (1984). Table 4 shows parameter estimates and values of the overidentification test statistic.¹⁶ When the model is true, this statistic will converge to a χ^2 distribution, with degrees of freedom equal to the number of moment conditions less the number of parameters.

The first column of Table 4 shows the results for a simple stationary AR(1) model, where $\sigma_{\psi t}^2 = \sigma_f^2 = 0$. This model is overwhelmingly rejected by the data; the overidentification test statistic is 394.9, with a p-value of 0. The reason for this failure can be seen in Table 3, which shows that while there is a large decline from the variance to the first autocovariance, the decline between the first and second (and between the second and third) autocovariances is much smaller.¹⁷ An AR(1) model, which generates a geometrically declining series of

dropped in later waves because of missing information. Second, the HRS/AHEAD added new households in wave 4.

¹⁵Although combining moment conditions across waves would be more efficient under stationarity, matching wave-specific moments allows us to consider non-stationary models.

¹⁶The standard errors shown here have not been adjusted to reflect uncertainty in the estimate of β .

¹⁷The average variance is 1.11, the average first autocovariance is 0.45, and the second autocovariance is 0.34.

autocovariances, cannot replicate this progression.

Parameter	Model					
	1	2	3	4	5	6
σ_a^2	1.046 (0.010)	0.522 (0.019)	0.574 (0.453)	0.395 (0.043)	0.519 (0.020)	0.399 (0.044)
σ_ϵ^2	0.825 (0.014)	0.145 (0.047)	0.552 (0.513)	0.039 (0.115)	0.141 (0.049)	0.041 (0.114)
ρ	0.459 (0.009)	0.849 (0.018)	0.197 (0.186)	0.949 (0.043)	0.854 (0.018)	0.948 (0.042)
σ_{ut}^2		0.575 (0.019)	0.189 (0.470)	0.702 (0.059)	0.589 (0.027)	0.672 (0.070)
$\sigma_{\psi t}^2$		0.575 (0.019)	0.189 (0.470)	0.694 (0.039)	0.589 (0.027)	0.659 (.049)
ϕ				0.104 (0.031)		0.102 (0.031)
σ_f^2			0.334 (0.022)			
χ^2 -statistic	394.9	18.1	10.9	10.9	7.3	0.2
Degrees of freedom	8	7	6	6	4	2
p-value	0.000	0.012	0.091	0.091	0.120	0.909
Standard errors in parentheses σ_a^2 = variance of autoregressive component of log health costs σ_ϵ^2 = innovation variance of the autoregressive component ρ = autoregressive coefficient of log health costs σ_{ut}^2 = variance of moving average component $\sigma_{\psi t}^2$ = innovation variance of the moving average component In models 5 and 6, σ_{ut}^2 and $\sigma_{\psi t}^2$ report averages across waves ϕ = moving average coefficient of log health costs σ_f^2 = variance of permanent person-specific component						

Table 4: PARAMETER ESTIMATES OF ERROR COMPONENTS MODELS

The above reasoning suggests adding a moving average component to the AR(1) model. We begin with the simplest case, setting $\phi = 0$, so that the moving average component is white noise, and assuming that this white noise component is homoskedastic across waves: $\sigma_{ut}^2 = \sigma_{\psi t}^2 = \sigma_\psi^2$. The sum of these two processes is equivalent to the ARMA(1,1) process studied by Feenberg and Skinner.¹⁸ Estimates from this model are reported in the second column of Table 4. The overidentification test statistic is 18.1, implying a considerably better fit than the AR(1). Given that we have only 7 degrees of freedom, however, the model is still strongly rejected.

¹⁸See Hamilton (1994, p. 393) for a derivation.

A common error components model of wages (see Abowd and Card, 1989, and Baker, 1997, for example) includes a permanent person-specific effect, f_i , and allows the moving average component of wages to follow an MA(1) process instead of white noise.¹⁹ Columns 3 and 4 of Table 4 show the effects of these two changes. These models fit the data better than the AR(1) with white noise, although they are still rejected at the 10% level.²⁰

What also improves goodness of fit is allowing the variance of the white noise component, ψ_{it} , to differ across waves.²¹ Such heteroskedasticity could reflect variation in the survey questions used to generate the health cost measure, as the questions differ from wave to wave. Results from this model are shown in column 5 of Table 4.²² Given that the empirical variance of health costs changes significantly from wave to wave, allowing for heteroskedasticity significantly improves the fit; the χ^2 statistic falls to 7.3. This model is not rejected at the 10% level. The variances and covariances of the fitted model are shown in Table 5.

Model Predicted Covariances of the Residuals of log Health Costs, Waves 2-5				
	Wave 2	Wave 3	Wave 4	Wave 5
Wave 2	1.1279			
Wave 3	0.4427	1.1481		
Wave 4	0.3779	0.4427	1.0823	
Wave 5	0.3226	0.3779	0.4427	1.0730

Table 5: COVARIANCE MATRIX IMPLIED BY AR(1) PLUS HETEROSKEDASTIC WHITE NOISE

One last attempt to improve the fit of the model is to allow the moving average component of health costs to be an MA(1) with heteroskedastic innovations. Estimates are in column 6. The model does fit the data better, but introduces two additional parameters, leaving us with only two degrees of freedom. Fortunately, the parameters ρ , σ_a^2 , and σ_{ut}^2 seem reasonably stable across models 2, 4, 5 and 6. This means that even though the AR(1)-plus-white-noise model is rejected in favor of more complicated models, all of the models have similar time series implications. We return to these implications in our discussion of lifetime health cost

¹⁹Note that $\sigma_{ut}^2 = \sigma_{\psi t}^2 + \phi^2 \sigma_{\psi t-1}^2$, which means that $\sigma_u^2 = (1 + \phi^2) \sigma_{\psi}^2$ when ψ is homoskedastic.

²⁰We also tried estimating a model with an AR(1), MA(1) and a permanent person specific effect. This model fit the data no better (p-value of 10.9) than either model 3 or model 4, and the parameters were very imprecisely estimated.

²¹We did not estimate a model with heteroskedasticity in ϵ_{it} for two reasons. First, there is more variability in the variances across waves than there is in the autocovariances, suggesting that heteroskedasticity in ϵ_{it} would not likely help the fit very much. Second, classical measurement error that varies across waves would cause heteroskedasticity in ψ_{it} but not in ϵ_{it} .

²²The reported estimates and standard errors for $\sigma_{\psi t}^2$ and σ_{ut}^2 are averages across waves.

risk in Section 5.

4 Cross-Sectional Distribution

For the risk-averse, the possibility of catastrophic health care costs may be a matter of great concern. This means that in modeling the cross-sectional distribution of health care costs, special attention must be given to fitting the far right tail. Moreover, even if one prefers to estimate the distribution nonparametrically, the scarce data of the upper tail might require a parametric model. We thus proceed in two steps, considering first the upper tail, and then the entire distribution.

4.1 The Upper Tail

Previous studies have identified two statistical models for the upper tail of the health cost distribution. Feenberg and Skinner (1994) use the lognormal distribution.²³ This implies that the density function for large health costs, $f(\cdot)$, is

$$f(\ln hc | \ln hc \geq \ln hc_L) = \frac{1}{1 - \Phi([\ln hc_L - \mu]/\sigma)} \phi([\ln hc - \mu]/\sigma) \frac{1}{\sigma}, \quad (10)$$

where Φ and ϕ are the standard normal cdf and pdf, respectively; μ and σ are the mean and standard deviation, respectively, of the untruncated distribution; and hc_L is the truncation point used to define the upper tail. Rust and Phelan (1997) use the Pareto distribution, which has the density

$$\tilde{g}(hc | hc \geq hc_L) = \gamma hc_L^\gamma hc^{-(1+\gamma)}. \quad (11)$$

A change of variables shows that if hc follows a Pareto distribution, its logarithm follows an exponential distribution:

$$g(\ln hc | \ln hc \geq \ln hc_L) = \gamma e^{-\gamma[\ln hc - \ln hc_L]}. \quad (12)$$

Choosing between the two models boils down to seeing whether the right tail of the log health cost distribution is better modeled with a truncated normal or with an exponential

²³By assuming that log health costs follow a stationary Gaussian ARMA process, Feenberg and Skinner's approach implies that the upper tail of the cross-sectional distribution is truncated lognormal.

distribution.

The two models can be compared formally with the likelihood ratio test developed by Vuong (1989) and extended by Rivers and Vuong (2002). Consider a sample of log health costs of size N . Let $L_N(\hat{\mu}_N, \hat{\sigma}_N^2)$ and $L_N(\hat{\gamma}_N)$ denote the maximized sample log-likelihoods for the truncated normal and exponential models, respectively. Suppose that $\frac{1}{N}\hat{\omega}_N^2$ consistently estimates the variance of $\frac{1}{N}[L_N(\hat{\mu}_N, \hat{\sigma}_N^2) - L_N(\hat{\gamma}_N)]$, the mean log-likelihood difference. Since the two models in question are strictly non-nested, it follows from Vuong (Theorem 5.1) and Rivers and Vuong (Theorems 1 and 3) that the adjusted statistic²⁴

$$D_N \equiv N^{-1/2} \left(\frac{1}{\hat{\omega}_N} [L_N(\hat{\mu}_N, \hat{\sigma}_N^2) - L_N(\hat{\gamma}_N)] - 1 \right) \quad (13)$$

will converge in distribution to a standard normal variable if the two models are equivalent. On the other hand, if the truncated normal model better represents the data generating process for log health costs, D_N will converge to infinity, and if the exponential model is better, D_N will converge to negative infinity.

To perform the Vuong test, we assemble our health cost data into a “pseudo” cross section, and compute D_N for the N observations in the top decile. Although a household’s health care costs are correlated across waves, we can calculate the likelihood values as if the observations were independent—the Vuong test is valid even if both of the competing models are misspecified. The variance estimate $\hat{\omega}_N^2$ must be adjusted, however, to reflect any correlation; in practice, this adjustment is generally quite small.²⁵

We account for the effects of age, gender, marital status, income, wave, and health insurance type in two ways. The first approach is to repeat the linear regression shown in Table 2, compute the residuals, and add back the mean. The first line of Table 6 presents parameter estimates, log-likelihoods, and p-values of the Vuong statistic D_N for this modified cross section. Given the way D_N has been constructed, a low p-value should be taken as evidence in support of the normal specification, and a high p-value should be taken as evidence in support of the exponential specification. The p-value of 0.266 shown in the first line of Table

²⁴The -1 term is a variant of the Akaike correction factor.

²⁵Letting $m_i(\ln hc_i; \hat{\mu}_N, \hat{\sigma}_N^2, \hat{\gamma}_N) = \ln f(\ln hc_i; \hat{\mu}_N, \hat{\sigma}_N^2) - \ln g(\ln hc_i; \hat{\gamma}_N)$ denote the estimated log-likelihood difference for observation i , it follows that $\frac{1}{N}\hat{\omega}_N^2$ is the estimated variance of $\frac{1}{N} \sum_{i=1}^N m_i(\ln hc_i; \hat{\mu}_N, \hat{\sigma}_N^2, \hat{\gamma}_N)$. Our approach for finding $\hat{\omega}_N^2$ is analogous to the procedure for finding the matrix \hat{S} , which we describe in Appendix A (equation (24)).

6 suggests that the two models are roughly equivalent. The second line of Table 6 presents results for the wave 5 data alone—recall that these data potentially do a better job of capturing nursing home costs, which could skew the health cost distribution the right. Perhaps not surprisingly, the fatter-tailed exponential model better fits the wave 5 data, although the difference is not significant at standard confidence levels.

The second way in which we account for the conditioning variables is to break the data into cells by age, marital status, and health insurance type.²⁶ The bottom 24 lines of Table 6 show the results for cells with 30 or more observations. At a 5% (two-sided) significance level, the truncated normal model dominates ($p < 0.025$) in two cells, and the exponential model dominates in one cell. At a 10% significance level, the exponential dominates in one additional cell, and at 20%, the truncated normal dominates in four more cells.²⁷ But this leaves 16 cells where neither model clearly dominates. Taken as a whole, the two models fit equally well.

The cell-by-cell results also show that the p-values tend to be somewhat higher for those with employer-provided or privately-purchased insurance. This is somewhat surprising, but consistent with Rust and Phelan’s (1997) finding that insurance does not completely remove the risk of catastrophic costs.

²⁶We controlled for gender, income, and wave with a standard partial regression approach. Using OLS regression, we first found the components of gender, income and wave that were orthogonal to the other explanatory variables. We then linearly regressed log health costs on these orthogonal components, and computed the residuals. Omitting these adjustments does not change our qualitative results.

²⁷In some of these cells, the truncated normal distribution was difficult to estimate; the likelihood function was too irregular to yield the exact maximum implied by first-order conditions. A look at the underlying data suggests that the difficulties often arose because those data were in fact better approximated by the exponential model.

Subsample	$\ln hc_{0.9}$	N	$\hat{\mu}_N$	$\hat{\sigma}_N^2$	$L_N(\hat{\mu}_N, \hat{\sigma}_N^2)$	$\hat{\gamma}_N$	$L_N(\hat{\gamma}_N)$	$p(D_N)$
Entire Sample	8.41	3490	-24.13	21.70	-1938.2	1.56	-1938.7	0.2664
Wave 5	8.47	1220	-24.29	22.31	-707.8	1.52	-707.4	0.7925
Age < 65, Unmarried								
Employer-provided	8.03	446	-16.31	12.02	-112.5	2.11	-112.4	0.6306
Privately-purchased	8.67	83	-17.81	14.66	-30.8	1.87	-30.8	0.4682
No insurance	7.61	147	-29.41	29.03	-105.6	1.33	-105.6	0.4851
Medicaid	7.34	119	6.46	1.55	-79.7	1.36	-82.3	0.0933
Medicare	8.42	47	7.27	2.12	-37.3	1.21	-38.2	0.0641
Ages 65-79, Unmarried								
Employer-provided	8.10	182	-21.28	17.54	-80.6	1.75	-80.4	0.8327
Privately-purchased	8.50	142	-12.57	8.99	-15.1	2.46	-14.2	0.9983
Medicaid	7.48	122	4.54	3.34	-89.6	1.29	-90.6	0.1864
Medicare	8.03	207	-25.13	22.33	-117.7	1.54	-117.7	0.5373
Age > 79, Unmarried								
Employer-provided	8.68	107	8.92	1.08	-91.9	1.08	-98.7	0.0081
Privately-purchased	8.80	141	8.40	1.36	-104.6	1.25	-109.2	0.0531
Medicaid	7.80	115	6.95	2.20	-102.8	1.08	-105.7	0.0549
Medicare	8.38	173	7.50	2.06	-146.4	1.14	-150.5	0.0192
Age < 65, Married								
Employer-provided	8.77	537	0.19	4.56	-142.6	2.08	-143.1	0.3150
Privately-purchased	9.65	84	8.63	0.46	18.7	3.37	18.0	0.2439
No insurance	8.42	93	4.89	2.41	-34.1	1.88	-34.5	0.3723
Medicare	8.90	32	2.87	3.59	-10.6	1.95	-10.6	0.3580
Ages 65-79, Married								
Employer-provided	8.83	201	-15.73	12.27	-53.2	2.09	-53.0	0.8810
Privately-purchased	9.21	146	-9.34	7.01	2.6	2.77	2.9	0.8654
Medicaid	8.55	34	5.45	2.53	-16.7	1.66	-16.8	0.3111
Medicare	8.73	163	-19.10	15.79	-64.1	1.84	-63.9	0.9566
Age > 79, Married								
Employer-provided	9.02	41	-14.10	10.85	-8.2	2.23	-8.1	0.7952
Privately-purchased	9.39	47	-11.66	9.11	-5.9	2.40	-5.9	0.6462
Medicare	8.82	47	4.35	4.60	-35.0	1.28	-35.3	0.3882
$\ln hc_{0.9}$ = the 90th percentile of log health costs N = the number of observations in the top decile $\hat{\mu}_N, \hat{\sigma}_N^2, \hat{\gamma}_N$ = estimated likelihood parameters defined in the text $L_N(\hat{\mu}_N, \hat{\sigma}_N^2), L_N(\hat{\gamma}_N)$ = estimated log-likelihoods $p(D_N)$ = p-value of the Vuong test statistic D_N								

Table 6: PARAMETER ESTIMATES AND LOG-LIKELIHOOD VALUES FOR THE TOP DECILE

4.2 The Entire Cross Section

Although the Pareto and lognormal models fit the upper tail of the empirical health cost distribution equally well, the overall cross section provides more discrimination. Figure 1 shows the cross-sectional distribution for the entire sample. Once the effects of the conditioning variables have been removed, the empirical distribution is fairly close to lognormal.

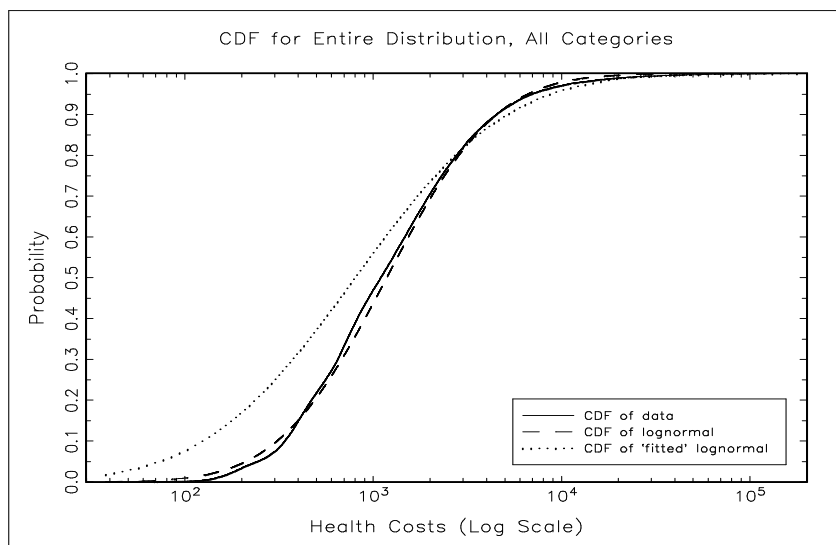


Figure 1: DISTRIBUTION OF HEALTH CARE COSTS

Even in those subsamples where the Pareto distribution best fits the cross section, there are two additional problems. First, the Pareto models that are estimated on entire cross sections generally do much worse at fitting the upper tail than the lognormal models estimated on entire cross sections. Second, in almost all cases, the estimates of γ that are found with the entire cross section are less than 1, which implies that expected health care costs, $\frac{\gamma}{\gamma-1}hc_L$, are unbounded. This is completely at odds with the estimates from the upper tail shown in Table 6.

Finally, even if we could somehow extend the Pareto models of the top decile to the entire health cost distribution, we would have difficulty combining them with the time series

models estimated in Section 3.²⁸ While a stationary ARMA process with normal innovations (that is common across households) will generate a normally-distributed cross section, to our knowledge there is no closed-form innovation distribution for log health costs that generates an exponential cross section.²⁹ These concerns lead us to abandon the Pareto as a model of the overall cross section.

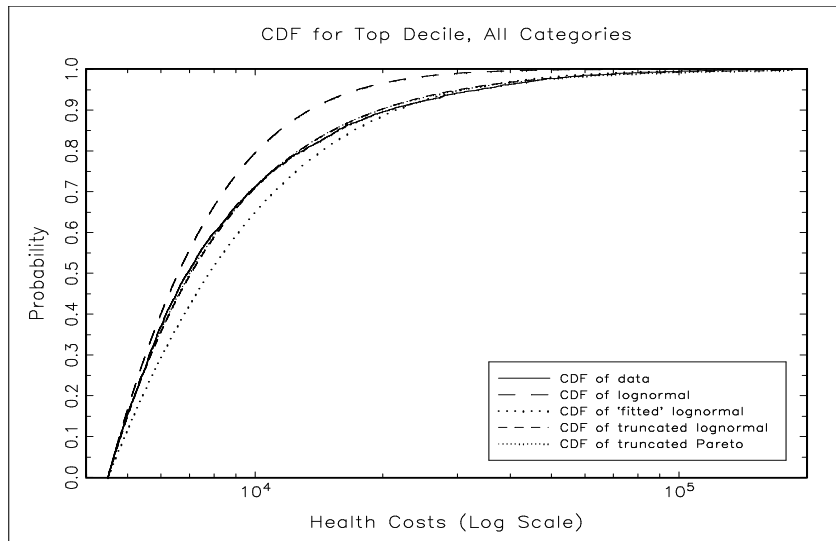


Figure 2: CONDITIONAL DISTRIBUTION OF THE TOP DECILE OF HEALTH CARE COSTS

Unfortunately, the lognormal distribution that best fits the overall cross section does not fit the upper tail very well, even if it outperforms the Pareto. Figure 2 shows the conditional distribution for the top decile of the health cost data shown in Figure 1. While the Pareto and truncated lognormal models that were estimated from the top decile fit closely, the upper tail implied by the “standard” lognormal model is too thin. This conclusion can be reinforced

²⁸Although Rust and Phelan (1997) fit a Pareto distribution to the upper tail of their data, they use a discrete approximation to model the rest of the distribution. Because they assume health care costs are independent across time, this bifurcated model does not hinder their structural estimation.

²⁹A comparison of characteristic functions shows that the exponential distribution does not belong to the class of *stable* distributions, which are preserved under unweighted summation. It can also be shown for many cases that ARMA processes with stable innovations have stable unconditional distributions. (See Billingsley, 1979; Brockwell and Davis, 1991; and McCulloch, 1996.) All of this suggests that there is no closed-form innovation distribution that yields an exponential cross section. We have not considered non-normal stable distributions in this paper, because, as is well known (see, e.g., Tsionas, 1999), in many applications they are difficult to use.

by comparing the log-likelihoods shown in the first column of Table 7 with the log-likelihoods shown in Table 6: in most cases, the standard lognormal fits the top decile much less closely.

This consideration leads us to an alternative estimate. We find the mean and variance of the lognormal distribution that matches both the mean and 99.5th percentile of health costs. In particular, we pick values $\tilde{\mu}$ and $\tilde{\sigma}^2$ such that

$$e^{\tilde{\mu} + \tilde{\sigma}^2/2} = \widehat{E}(hc), \quad (14)$$

$$\Phi\left(\frac{\ln \widehat{hc}_{0.995} - \tilde{\mu}}{\tilde{\sigma}}\right) = 0.995, \quad (15)$$

where $\widehat{E}(hc)$ and $\widehat{hc}_{0.995}$ are the mean and the 99.5th percentile of the empirical cross section.³⁰ Table 8 shows the parameters of this “fitted” distribution, along with the parameters of the standard lognormal distribution, and standard errors for both sets of parameters. (Appendix A describes how the standard errors for the fitted parameters are calculated.) Given that they rely on the far upper tail of the data distribution, where the data are more sparse, it is not surprising that the parameter estimates for the fitted lognormal model are less precise.

³⁰Defining $z_{0.995} = \Phi^{-1}(0.995)$ to be the 99.5th percentile of the standard normal distribution, we can solve equations (14) and (15) to find

$$\begin{aligned} \tilde{\sigma} &= z_{0.995} - \sqrt{z_{0.995}^2 - 2[\ln \widehat{hc}_{0.995} - \ln(\widehat{E}(hc))]}, \\ \tilde{\mu} &= \ln \widehat{hc}_{0.995} - \tilde{\sigma} z_{0.995}. \end{aligned}$$

We find $\ln \widehat{hc}_{0.995}$ with a GAUSS function that interpolates between data points to get the exact percentile.

Subsample	LLH: Top Decile		LLH: Full Sample		$\hat{E}(hc)$		$hc_{0.995}$ (in 000s)	
	Std	Ftd	Std	Ftd	Std	Ftd	Std	Ftd
Entire Sample	-2224.1	-1997.0	-51125	-55385	2036	2300	17.5	33.8
Wave 5	-836.3	-737.4	-17722	-19788	2250	2558	18.9	40.5
Age < 65, Unmarried								
Employer-provided	-121.3	-120.6	-6261	-6314	1283	1376	10.0	12.7
Privately-purchased	-32.8	-32.1	-1150	-1162	3379	3276	25.8	28.7
No insurance	-156.2	-107.5	-2039	-2389	751	1044	5.8	18.3
Medicaid	-141.4	-80.2	-1474	-1757	536	711	3.3	10.2
Medicare	-41.8	-37.6	-770	-791	1868	2316	21.6	37.7
Ages 65-79, Unmarried								
Employer-provided	-92.4	-86.2	-2599	-2737	1540	1685	12.6	21.0
Privately-purchased	-28.7	-27.2	-1711	-1722	2707	2777	15.8	18.4
Medicaid	-148.0	-89.8	-1603	-2021	622	886	4.3	16.6
Medicare	-139.2	-122.8	-3062	-3432	1191	1463	10.5	25.1
Age > 79, Unmarried								
Employer-provided	-106.7	-98.6	-1779	-1791	2659	3142	32.3	45.4
Privately-purchased	-130.7	-108.0	-1966	-2087	3797	4030	29.5	48.7
Medicaid	-149.4	-104.0	-1715	-1930	848	1294	7.7	24.5
Medicare	-175.9	-147.3	-2803	-2985	1645	2266	18.4	42.4
Age < 65, Married								
Employer-provided	-148.7	-147.5	-8086	-8088	3075	3020	28.2	26.8
Privately-purchased	10.4	18.5	-1068	-1231	9154	8399	58.9	32.9
No insurance	-34.4	-34.5	-1438	-1441	1698	1866	16.9	20.6
Medicare	-11.0	-10.8	-477	-478	4121	3814	39.8	33.1
Ages 65-79, Married								
Employer-provided	-59.0	-59.1	-2983	-2986	3447	3373	30.7	31.1
Privately-purchased	-2.7	-2.9	-1706	-1712	5787	5630	32.0	33.1
Medicaid	-16.8	-16.8	-547	-550	1871	2170	21.8	29.4
Medicare	-70.3	-70.8	-2479	-2535	3046	3126	28.9	38.3
Age > 79, Married								
Employer-provided	-10.5	-9.5	-646	-669	3794	3618	41.3	28.9
Privately-purchased	-7.0	-6.9	-596	-598	6937	6572	45.4	39.4
Medicare	-37.4	-36.6	-759	-760	3654	3988	41.4	47.2
"Std" refers to standard lognormal estimates, "Ftd" to fitted estimates								
Log-likelihoods ("LLH") calculated with log health costs								

Table 7: SUMMARY STATISTICS GENERATED BY DIFFERENT LOGNORMAL DISTRIBUTIONS

Subsample	Standard Lognormal				“Fitted” Lognormal			
	$\hat{\mu}$		$\hat{\sigma}^2$		$\tilde{\mu}$		$\tilde{\sigma}^2$	
Entire Sample	7.07	(0.0056)	1.10	(0.0083)	6.69	(0.032)	2.11	(0.072)
Wave 5	7.18	(0.0094)	1.07	(0.0137)	6.69	(0.069)	2.32	(0.157)
Age < 65, Unmarried								
Employer-provided	6.67	(0.0148)	0.97	(0.0206)	6.62	(0.040)	1.21	(0.074)
Privately-purchased	7.65	(0.0338)	0.95	(0.0465)	7.53	(0.068)	1.12	(0.067)
No insurance	6.15	(0.0255)	0.95	(0.0352)	5.64	(0.193)	2.63	(0.413)
Medicaid	5.93	(0.0245)	0.71	(0.0292)	5.54	(0.131)	2.05	(0.187)
Medicare	6.74	(0.0582)	1.58	(0.1034)	6.55	(0.228)	2.40	(0.266)
Ages 65-79, Unmarried								
Employer-provided	6.83	(0.0238)	1.03	(0.0342)	6.57	(0.070)	1.73	(0.137)
Privately-purchased	7.57	(0.0217)	0.66	(0.0250)	7.54	(0.058)	0.79	(0.034)
Medicaid	6.02	(0.0261)	0.82	(0.0335)	5.35	(0.241)	2.87	(0.302)
Medicare	6.51	(0.0235)	1.14	(0.0354)	6.00	(0.110)	2.57	(0.151)
Age > 79, Unmarried								
Employer-provided	7.05	(0.0397)	1.67	(0.0727)	7.02	(0.159)	2.07	(0.176)
Privately-purchased	7.76	(0.0263)	0.97	(0.0366)	7.47	(0.075)	1.67	(0.088)
Medicaid	6.15	(0.0322)	1.18	(0.0495)	5.70	(0.366)	2.92	(0.779)
Medicare	6.65	(0.0297)	1.52	(0.0516)	6.29	(0.268)	2.87	(0.586)
Age < 65, Married								
Employer-provided	7.44	(0.0149)	1.19	(0.0230)	7.44	(0.093)	1.14	(0.206)
Privately-purchased	8.74	(0.0301)	0.76	(0.0370)	8.86	(0.039)	0.36	(0.053)
No insurance	6.78	(0.0378)	1.32	(0.0614)	6.78	(0.141)	1.49	(0.300)
Medicare	7.69	(0.0640)	1.27	(0.1020)	7.69	(0.177)	1.11	(0.396)
Ages 65-79, Married								
Employer-provided	7.57	(0.0239)	1.15	(0.0363)	7.52	(0.051)	1.20	(0.086)
Privately-purchased	8.36	(0.0206)	0.61	(0.0228)	8.30	(0.033)	0.67	(0.030)
Medicaid	6.73	(0.0695)	1.60	(0.1242)	6.73	(0.966)	1.91	(2.069)
Medicare	7.40	(0.0276)	1.24	(0.0435)	7.20	(0.085)	1.70	(0.073)
Age > 79, Married								
Employer-provided	7.51	(0.0605)	1.47	(0.1037)	7.69	(0.120)	1.00	(0.121)
Privately-purchased	8.46	(0.0409)	0.77	(0.0509)	8.45	(0.157)	0.69	(0.334)
Medicare	7.43	(0.0577)	1.54	(0.1013)	7.48	(32.423)	1.62	(65.033)
Standard errors in parentheses								

Table 8: PARAMETERS FOR LOGNORMAL DISTRIBUTIONS

Figure 2 shows that in addition to replicating average health costs, this “fitted” specification fits the far upper tails of the data distribution fairly well. Figure 1 shows that one weakness of this specification is that it can provide a poor fit of the distribution’s lower tail. In practical terms, this is a relatively minor cost; at the lower tail of the distribution, large differences in logged health costs lead to relatively small changes in health costs themselves. Table 7 compares the standard and fitted lognormal distributions in more detail. The first two columns of Table 7 show that in most cases, the fitted distribution better fits the upper decile. The last four columns compare the two models’ predictions of mean health care costs and the 99.5th percentile. While the fitted model matches these two statistics by construction, the standard lognormal model often misses by a large margin.³¹ For the full sample, the standard lognormal implies a 99.5th percentile that is half of what is seen in the data.³²

It is also useful to compare the parameter estimates in Table 8 to the lognormal estimates for the top decile shown in Table 6. The lognormal parameters for the top decile are quite different, and are unlikely to fit the overall distribution very well. For example, the lognormal parameters shown on the first line of Table 6 imply that mean health care costs are less than one cent. All of these factors suggest that in terms of simultaneously matching both the overall distribution and its upper tail, the fitted lognormal provides the best approximation.

4.3 Welfare Effects of Health Cost Uncertainty

A good model of the health cost distribution should be able to accurately measure the welfare losses associated with health cost uncertainty. These measures depend on the expectation

$$E(V(A - e^{\ln hc})) = \int V(A - e^{\ln hc}) f(\ln hc) d \ln hc, \quad (16)$$

where $V(\cdot)$ is a value function, A is assets and $f(\cdot)$ is the pdf of log health costs. It follows that a useful test of our fitted lognormal model is to see whether the integral calculated with this distribution has the same welfare implications as the sample mean $\frac{1}{N} \sum_{i=1}^N V(A - e^{\ln hc_i})$.

³¹The mean implied by the fitted distribution, \$2,300, is below the raw data mean of \$2,600, because the fitted distribution is estimated with data that have been purged of income, wave and demographic effects. Because this filtering reduced the variance of log health costs without changing their mean, health costs themselves are lower.

³²When the data are bottom-coded with Hubbard et al.’s rules (see footnote 9), the standard lognormal model fits the upper tail much better, and, moreover, is very close to the alternative model we develop here. Our alternative model, however, is estimated mostly from the upper tail, and does not rely on bottom-coding decisions.

To conduct this experiment, we specialize the value function as

$$V(A - hc) = \frac{1}{1 - \zeta} [\max\{A - hc, A_{min}\}]^{1-\zeta}, \quad \zeta \geq 0, \zeta \neq 1, \quad (17)$$

which combines a CRRA function with an asset floor given by A_{min} ; this is similar to, if more stylized than, value functions used in Hubbard et al. (1994, 1995), Palumbo (1999), and French and Jones (2002), where social insurance limits the effect of excessive health care costs. In the figures shown below, the asset floor is fixed at \$25,000 and ζ is set to 2 or 5.³³ To measure the welfare effects of health cost uncertainty, we compute the equivalent differential (EQD), the decrease in assets that would leave a consumer facing no uncertainty no better off than a consumer facing uncertainty:

$$V(A_{PHC} - EQD) \equiv E(V(\max\{A - hc, A_{min}\})), \quad (18)$$

where $A_{PHC} = \frac{1}{N} \sum_{i=1}^N \max\{A - hc_i, A_{min}\}$ denotes the average post-health-cost asset balance.

One advantage of analytical distributions is that they allow the use of quadrature, where the integral is computed by evaluating the value function at a small number of well-chosen points (nodes) and taking a weighted average. For the case at hand, a promising approach is Gauss-Hermite (GH) quadrature, where the nodes and weights are picked so that when $f(\cdot)$ is the standard normal pdf and $V(\cdot)$ is a low-order polynomial in $\ln hc$ the approximation is exact.³⁴

Figure 3 plots the equivalent differential as a function of pre-health-cost assets, A . Figure 3 shows functions calculated with three different estimates of $E(V(A - e^{\ln hc}))$: the sample average; the integral found by combining the standard lognormal model with 10-node GH quadrature; and the integral found by combining our fitted lognormal model with 7-node GH quadrature. Although they have different numbers of nodes, the two quadrature techniques take averages over the same interval of health costs, and are thus comparable.³⁵ All of these

³³The correct value of A_{min} is not obvious in such a stylized setting. We experimented with different values of A_{min} and found similar qualitative results.

³⁴Judd (1998) provides a nice review of Gauss-Hermite quadrature and numerical integration in general.

³⁵GH quadrature nodes are points on the support of the standard normal distribution, with larger collections of nodes having a wider range. As discussed below, we picked the number of nodes so that the highest quadrature node matched the upper bound of the empirical distribution. The standard lognormal model

estimates utilize the entire sample, the latter two using parameters from the first line of Table 8.

For both values of ζ , the equivalent differential found by combining the fitted lognormal distribution and 7-node GH quadrature, EQD_F , closely matches the equivalent differential generated by the data integral, EQD_D . However, the equivalent differential of the standard lognormal, EQD_S , matches the data integral poorly. For most asset levels, the equivalent differential implied by the standard lognormal is less than $\frac{1}{10}$ as large as is implied by either the fitted lognormal or the observed distribution. The welfare losses implied by the standard lognormal are small. Never is the equivalent differential over \$1,000, and it is usually less than \$10. In many cases EQD_S is negative and falls off the graph. The mismatch occurs both because the standard lognormal distribution does not match mean health care costs, and because it fits the upper tail very poorly. To disentangle these two effects, Figure 4 shows equivalent differentials calculated with the “shifted” lognormal distribution, where $\hat{\sigma}^2$ is left unchanged, but $\hat{\mu}$ is adjusted to let the distribution match mean health costs. While matching mean health costs greatly improves the fit, the inability to match the upper tail leaves the shifted distribution an inferior alternative. In short, the fitted lognormal model provides a better numerical approximation.

It turns out that the choice of 7 nodes is essential to the fitted lognormal’s good performance. The reason for this is that the largest node in this quadrature scheme generates a health cost roughly equivalent to the largest observed health expenditure, hc_{max} . When households are highly risk-averse (e.g. $\zeta = 5$), EQD_D is driven by hc_{max} . While hc_{max} is quite large (roughly \$200,000), it is nonetheless finite, and the risk effects of it generates eventually shrink as pre-health-cost assets, A , continue to grow.³⁶ By matching hc_{max} , EQD_F is able to match EQD_D .

requires more of these nodes because it utilizes a smaller estimate of σ^2 , so that $\frac{1}{\sigma}[\ln hc_{max} - \hat{\mu}]$, where $\ln hc_{max}$ is the largest observed health cost, is larger.

³⁶Conversely, when A is small, many health cost realizations will be offset by the asset floor A_{min} . This generates the result that EQD_D initially increases in A ; people with more assets have more to lose.

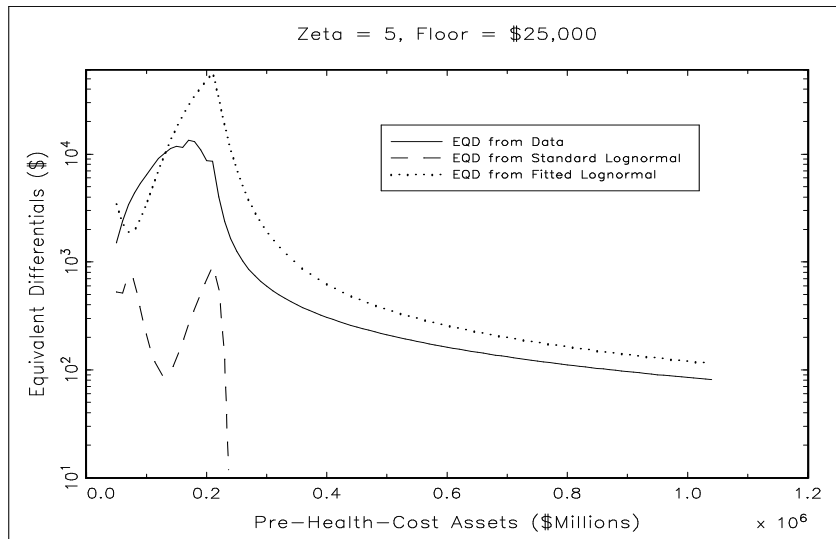
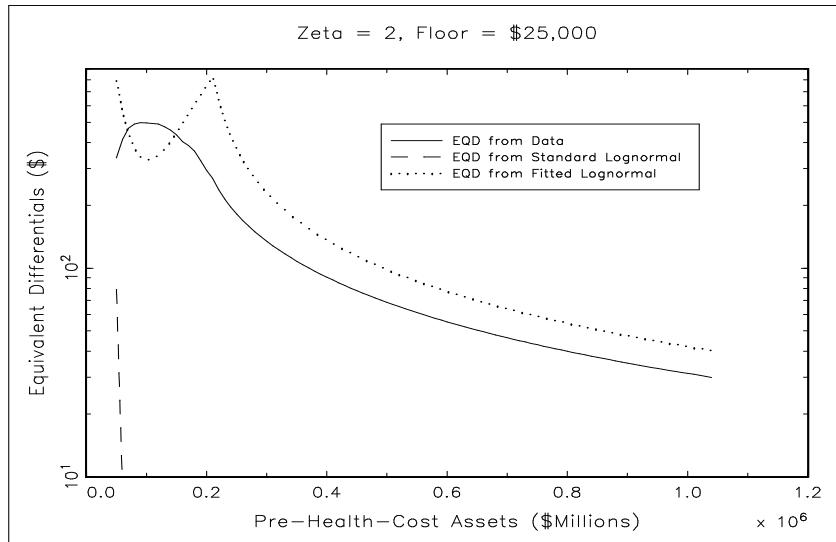


Figure 3: EQUIVALENT DIFFERENTIALS FROM DIFFERENT INTEGRATION METHODS

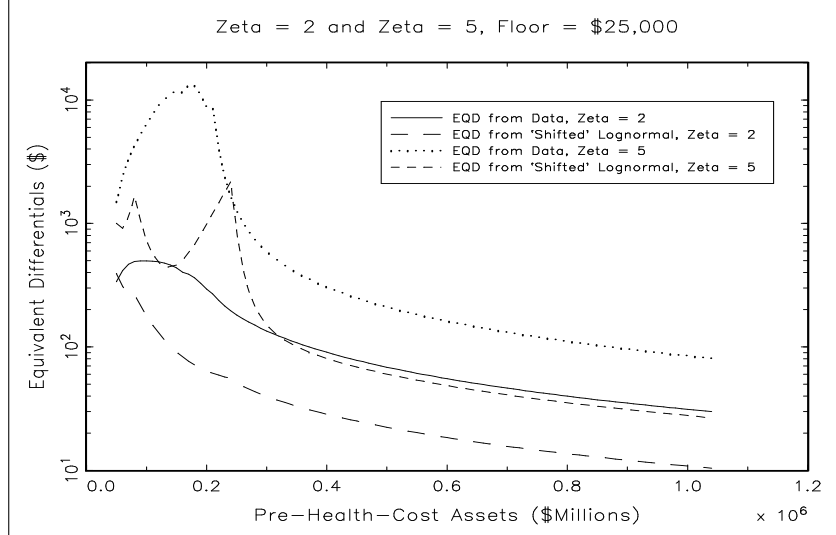


Figure 4: EQUIVALENT DIFFERENTIALS FROM THE “SHIFTED” LOGNORMAL MODEL

It can be shown that if one uses more than 7 nodes, thus increasing the upper bound on health costs in the numerical approximation, EQD_F will exceed EQD_D at higher asset levels, particularly when $\zeta = 5$. This suggests that a finite sample with a finite largest element may greatly understate the degree of health cost risk that wealthy agents actually face, unless one can use a parametric model to extrapolate. On the other hand, if one believes that the largest cost in our 35,000-observation sample accurately estimates the upper bound of health cost distribution, the numerical methods one uses should be tailored to this extreme order statistic. In essence, the fitted lognormal approximation requires estimates of not two but three parameters: $\tilde{\mu}$, $\tilde{\sigma}^2$, and the number of nodes.³⁷

³⁷Another case of strong sensitivity to extreme order statistics is the estimation of search and matching models (see, e.g., Christensen and Kiefer, 1997), where theory predicts a bounded support.

5 Lifetime Health Cost Risk

5.1 The Annual Stochastic Process

The stochastic process for log health costs can be found by combining the time series model estimated in Section 3 with a model for the distribution of health cost innovations. Unfortunately, the error components model that we estimate does not allow us to back out an empirical distribution of log health cost innovations, because the sum of an AR(1) and an MA process is not a Markov process. On the other hand, the lognormal model that we find in Section 4 implies that the innovations are normal: if log health costs are normally distributed in the cross-section and follow a stationary ARMA process, then the innovations to that process must be normally distributed.³⁸ Therefore, our preferred model of the health cost process combines the time series model with the fitted lognormal approximation of the cross section.

For the exercises below, we use the homoskedastic AR(1)-plus-white noise time series model. Although the heteroskedastic model provides a better fit, much of this heteroskedasticity reflects wave-specific differences in the wording of questions, rather than underlying differences in health cost risk. The homoskedastic model is also more parsimonious, and is more easily compared to other studies.

An important limitation of this model is that the health cost data which it fits consist of two-year averages. In most studies of household decision-making, the relevant time period is one year. We therefore fit an annual model of log health costs to the data, using the Method of Simulated Moments. Specifically, for a given set of parameter values we simulate a panel of annual health costs, where annual health costs are the exponentiated sum of an AR(1) component, a white noise component, and a mean shifter. Taking two-year averages of the simulated annual data yields data of the same form as the HRS/AHEAD data. We then find the mean and 99.5th percentile of health costs, and the first three autocorrelations of the log health cost residuals, of the simulated data, and compare them to the same statistics from the HRS/AHEAD data. The parameters that minimize a GMM criterion function are our

³⁸If households share a common stationary and ergodic health cost process, the unconditional health cost distribution for an individual household must equal the cross-sectional distribution across the population. A rigorous discussion of the necessary conditions for a stationary distribution can be found in Stokey and Lucas (1989). (Although the discussion there is couched in terms of Markov processes, one can derive stationary distributions for each component of our model, and then consider the sum.)

estimated parameters. Appendix C describes the details of our approach. We refer to this model as the “fitted lognormal” model in the discussion below, as it is analagous to the fitted lognormal model derived in Section 4. We also compute a one-year analog to the standard lognormal model, where we match mean log health costs, and the variance and first three autocovariances of the log health cost residuals. Comparing the two models gives some sense of how explicitly matching extreme health shocks affects the measurement of health cost risk.

Table 9 presents parameters for the annual health cost process estimated from the entire data set. While the persistent component of the standard lognormal model is effectively the persistent component of the two-year model with ρ rescaled to a one-year frequency, the annual transitory component has twice the variance of its two-year counterpart, as it gets averaged out in two-year data. As with the two-year data, the fitted estimates contain a much higher level of variance than the standard estimates. The last column of Table 9 shows overidentification test statistics, along with p-values. Although the one-year models do not aggregate exactly into the two-year models we derived earlier, the fits they generate resemble the ones shown in Table 4.³⁹

Table 9 also includes estimates from Feenberg and Skinner (1994) and Hubbard et al. (1994), both of which are measured at one-year frequencies. Feenberg and Skinner analyze total household medical expenditures for tax filers who deduct medical expenditures, whereas Hubbard et al. analyze total household medical expenses collected from survey data. We are aware of no other study that presents estimates of these parameters.⁴⁰

Feenberg and Skinner’s estimates of both σ_a^2 and σ_u^2 are much smaller than our estimates. There are several potential reasons for this. First, their data are from 1968-1973, when medical spending was lower and potentially less volatile. Second, they use a balanced panel in their analysis whereas we use an unbalanced panel. Given that a major reason for attrition in our analysis is death, and those who die likely have higher medical expenses, Feenberg and

³⁹The overidentification statistics for the fitted and standard lognormal models should be compared to each other and to the Table 4 statistics with some care, as they are calculated with different moment conditions.

⁴⁰We decompose Feenberg and Skinner’s ARMA(1,1) process into AR(1) and white noise components by utilizing the discussion in Hamilton (1994, p. 393). Neither Feenberg and Skinner or Hubbard et al. calculate the analog to μ . It is worth noting that average health care costs (in 1998 dollars, adjusted with the BEA’s consumption price index) are roughly \$2,600 in our study (\$2,650 after bottom-coding), \$2,980 in Feenberg and Skinner, and \$3,030 in Hubbard et al. These differences reflect, among other things, differences in demographics across the samples. In the simulations below, we set μ so that the latter two models generate the same mean health costs as our fitted model.

	σ_a^2	σ_u^2	ρ	μ	χ_{10}^2 -statistic
HRS/AHEAD: Standard Lognormal	0.524 (0.0195)	1.039 (0.0281)	0.922 (0.0100)	6.852 (0.0613)	18.54 [0.0466]
HRS/AHEAD: Fitted Lognormal	0.909 (0.0485)	1.819 (0.0746)	0.925 (0.0034)	6.366 (0.0709)	24.05 [0.0075]
Feenberg and Skinner	0.269	0.100	0.896	N.A.	N.A.
Hubbard et al.	0.930	0.220	0.901	N.A.	N.A.
Standard errors in parentheses					
p-values in brackets					

Table 9: PARAMETER ESTIMATES FOR ONE-YEAR HEALTH COST PROCESSES

Skinner likely underestimate the variance of health care costs.⁴¹ Third, their sample consists only of individuals whose health care costs are high enough to be itemized on their income tax returns. The adjustments they make for truncation might not recover all of the underlying variance. Alternatively, this tax data could be less noisy than standard survey data. Note that their estimate of σ_u^2 is much smaller than ours, in both absolute and relative terms, and if measurement error is transitory, datasets with less measurement error will have lower values of σ_u^2 .

Hubbard et al. (1994, 1995) use data from the 1977 National Health Care Expenditures Survey and the 1977 National Nursing Home Survey to estimate the cross sectional variance of health care costs. Their estimated cross-sectional variance, $\sigma_a^2 + \sigma_u^2$, is smaller than our standard lognormal estimate, to which it most closely compares. We are not sure what causes this discrepancy, although their data are 20 years older than our data, and the variance of health care costs has potentially grown over time. In addition, Hubbard et al. do not explicitly match extreme health cost events, although their bottom-coding decisions largely attenuate this problem.⁴² Because Hubbard et al. allocate the total variance between σ_a^2 and σ_u^2 partly on the basis of Feenberg and Skinner’s estimates, they attribute much more of the cross-sectional variance to the autoregressive component, a_{it} , than we do.⁴³

⁴¹The HRS/AHEAD survey utilizes follow-up interviews of the deceased’s survivors. When restricting our sample to those who have non-missing health care costs in all waves, the variance of health costs drops by 10% and the 99.5th percentile drops by 6%.

⁴²When we use Hubbard et al.’s bottom-coding rule (see footnote 9), the cross-sectional variance of the standard lognormal model increases from 1.56 to 2.35.

⁴³Background calculations graciously provided by Jonathan Skinner reveal that the allocation is not an exact proportional rescaling.

5.2 Simulation Estimates of Lifetime Health Cost Risk

Given both the distribution and dynamics of health care costs, we can estimate the lifetime health cost risk that households face. Using the stochastic processes described in Table 9, we simulate 30-year health cost sequences for 1 million households. Each household begins at age 64 with a draw of a_{i64} from the invariant distribution and then realizes a 30-year sequence of innovations, $\{\epsilon_{it}, u_{it}\}_{t=65}^{94}$. Adding μ to these sequences of shocks and exponentiating yields a health cost history.⁴⁴ To measure lifetime health care costs, we discount this sequence back to age 65, using an annual interest rate of 3% and age- (but not health-) specific mortality adjustments described in French (2003). Holding all other variables fixed, we then recompute the sequence with one or both of the age-65 innovations, $(\epsilon_{i65}, u_{i65})$, set to zero. The difference between the two discounted sequences gives the lifetime effects of the age-65 innovations.

Table 10 shows the effects of the age-65 innovations on age-65 and lifetime health care costs, for the four models shown in Table 9. The first column shows results for the standard lognormal process. When the AR(1) and white noise innovations are considered together, the lifetime cost variation induced by the age-65 shocks has a standard deviation of \$6,570. This is considerably larger than the standard deviation of the age-65 variation, \$3,630, indicating that persistence in health care costs is important. Moreover, the variation induced by the AR(1) innovation, ϵ_{it} , has a lifetime standard deviation of \$5,580 and an age-65 standard deviation of \$1,190. Although transitory shocks generate most of the cross-sectional and short-term variance, it is the persistent shocks (reflecting chronic conditions) that generate most of the lifetime health cost risk. Turning to catastrophic shocks, we find that under our standard lognormal model 1% of the population will receive an age-65 shock to lifetime health costs of at least \$23,900, and 0.1% will receive a shock of at least \$54,700. Given that that the standard lognormal model undershoots the upper tail, a better measure of catastrophic risk appears in column 2, which shows results for our preferred model, the fitted lognormal. The fitted model implies that the top 1% will receive a shock to lifetime expenses of at least \$43,500, and the top 0.1% will receive a shock of at least \$124,700.

The amount of health cost risk implied by our estimates is considerably higher than that found by Feenberg and Skinner (1994). Redoing the simulations with Feenberg and

⁴⁴To restore the age effects that have been removed from the stochastic processes in Table 9, we let μ vary by age, using the coefficients given in Table 2. We have not attempted to account for other differences within or across individuals.

	HRS/AHEAD: Standard	HRS/AHEAD: Fitted	Feenberg- Skinner	Hubbard et al.
Standard Deviation of Age-65 Health Care Costs (in \$000s)				
Due to ϵ_{i65}	1.19	2.84	0.61	1.59
Due to $\epsilon_{i65} + u_{i65}$	3.63	7.99	1.01	2.29
Standard Deviation of Lifetime Health Care Costs (in \$000s)				
Due to ϵ_{i65}	5.58	10.44	3.74	8.70
Due to $\epsilon_{i65} + u_{i65}$	6.57	12.87	3.82	8.86
Additional Lifetime Health Care Costs Due to $\epsilon_{i65} + u_{i65}$ (in \$000s)				
99th percentile	23.9	43.5	11.8	31.7
99.9th percentile	54.7	124.7	19.9	71.1
Additional Lifetime Costs Given a \$1 Increase in Age-65 Costs				
Median ratio	\$1.55	\$1.61	\$3.01	\$3.82

Table 10: EFFECTS OF AGE-65 SHOCKS ON LIFETIME HEALTH CARE COSTS

Skinner’s parameter values, we find that the lifetime cost effects have a standard deviation of \$3,820, of which \$3,740 is attributable to the AR(1) innovation. When Hubbard et al.’s (1994) parameter values are used, the standard deviations rise to \$8,860 and \$8,700, which fall between the standard deviations implied by our two versions of the lognormal model. Because Hubbard et al. attribute most of their model’s variance to its autoregressive component, their model generates more lifetime health cost risk than the standard lognormal model, even though it has a smaller cross-sectional variance.

Lastly, we compute the additional lifetime health care costs associated with an additional \$1 in health care costs at age 65. For each simulated household we divide the change in lifetime costs generated by $\epsilon_{i65} + u_{i65}$ by the change in age-65 costs caused by the same two shocks. Taking the median of this ratio, we find that a \$1 shock to current health care costs leads to somewhere between \$0.55 and \$0.61 of *future* health care costs.⁴⁵ Using Feenberg and Skinner’s parameter values and our methodology, we find that a \$1 health cost shock today leads to \$2.01 of future health costs.⁴⁶ Using Hubbard et al.’s parameter values, the corresponding figure is \$2.82. Given that our estimates attribute a much smaller fraction of the variance to the autoregressive component, it is not surprising that health cost shocks

⁴⁵We report the median of this ratio across households because the combination of a positive (negative) AR(1) innovation and a negative (positive) white noise innovation can lead to a large positive—or negative!—ratio. The median will be more robust to these sorts of outliers.

⁴⁶Feenberg and Skinner report that a \$1 shock to current health care costs leads to \$2.65 in future health care costs. The discrepancy seems to come from the fact that they do not account for mortality in their simulations. When using Feenberg and Skinner’s estimates and not accounting for mortality risk, we find that a \$1 shock to current health costs leads to a \$2.62 increase in future health costs.

have less persistent effects in our model.

6 Conclusion

Using data from the Health and Retirement Survey (HRS) and Assets and Health Dynamics of the Oldest Old (AHEAD), this paper presents estimates of the stochastic process that determines the distribution and dynamics of health care costs. Similarly to Feenberg and Skinner (1994), we find that the data generating process for log health costs is well represented as the sum of an AR(1) and a white noise process, which can be rewritten as an ARMA(1,1). Contrary to Feenberg and Skinner, however, we find that most of the cross-sectional variation in health care costs comes from transitory variation; the effects of a health cost shock are much less persistent in our model.

We find that the innovations to the log health cost process can be modelled with a normal distribution, but the variance of this innovation distribution, as well as the mean for the overall process, should be adjusted so that the model matches the mean and the 99.5th percentile of the empirical health cost distribution. This fitted lognormal distribution matches the upper tail—the catastrophic portion—of the health cost distribution much better than the standard lognormal model. Simulating this fitted distribution reveals significant catastrophic health cost risk: in any given year 0.1% of households suffer a shock that costs at least \$125,000 over their lifetimes. The risk implied by our model is considerably more than that implied by previous estimates.

Before concluding, we note four important caveats to our analysis. First, there are non-trivial measurement problems with our data. If some of the transitory variation in health care costs is merely measurement error, we are overstating the variability of health care costs. Alternatively, because the initial sample excluded those who were in nursing homes, we may be understating health costs from this source, leading us to underestimate health cost variability.

The remaining three problems are more conceptual, and they all suggest that our estimates overstate the amount of health cost uncertainty that households face. The second problem is that the quantity of health care services consumed is to some extent a choice. This means that households can reduce their health care costs by reducing medical services they consume. Third, low income, low wealth households have access to Medicaid, making

health care services very inexpensive. While we have conditioned our estimates on several factors, including income and health insurance type, we might not have completely removed these effects. Lastly, those with high health care costs often die shortly after their health cost shock. Because they die so soon, people who suffer from massive health cost shocks face little risk of being financially destitute (Pauly, 1990).

Appendix A: Distribution of the Fitted Lognormal Parameters

Our approach is to modify the standard GMM framework to incorporate quantiles; useful background references include Buchinsky (1998) and Powell (1994). Suppose we have a sample of size N . Let $\lambda_0 = (\mu_0, \sigma_0^2)$ denote the population parameter vector for the fitted lognormal distribution and let $\tilde{\lambda} = (\tilde{\mu}, \tilde{\sigma}^2)$ denote the sample estimates. To find the standard errors of $\tilde{\lambda}$, it is useful to rewrite equations (14) and (15) as:

$$\widehat{E}(h(hc; \tilde{\lambda})) = 0, \quad (19)$$

$$h(hc; \tilde{\lambda}) \equiv \begin{pmatrix} hc - \exp(\tilde{\mu} + \tilde{\sigma}^2/2) \\ 1\{hc \leq \exp(\tilde{\sigma}z_{0.995} + \tilde{\mu})\} - 0.995 \end{pmatrix}, \quad (20)$$

where $z_{0.995} = 2.5758$ is the 99.5th percentile of the standard normal distribution, and $1\{\mathcal{A}\}$ is the 0 – 1 indicator function that returns 1 when event \mathcal{A} occurs.

Noting that the elements of λ are exactly identified, it follows from standard arguments (e.g., Newey and McFadden, 1994) that

$$\sqrt{N}(\tilde{\lambda} - \lambda_0) \rightsquigarrow_D \mathcal{N}(0, \Sigma), \quad (21)$$

where

$$\Sigma = [D'S^{-1}D]^{-1}, \quad (22)$$

$$D = \frac{\partial E(h(hc; \lambda_0))}{\partial \lambda}, \quad (23)$$

and S is a weighted sum of variances and autocovariances of $h(hc; \lambda_0)$.

Suppose the data consist of a panel, where each household appears up to $J + 1$ times,

arranged into a pseudo cross section. S can be then estimated as

$$\hat{S} = \widehat{AC}_0 + \widehat{AC}_1 + \widehat{AC}'_1 + \dots + \widehat{AC}_J + \widehat{AC}'_J, \quad (24)$$

where \widehat{AC}_j is an estimate of the autocovariance of $h(hc; \tilde{\lambda})$ with an observation j steps away in the pseudo cross section. If the data are sorted by household and wave, for $j > 0$ \widehat{AC}_j will be a weighted average of the j th autocovariance for observations from the same household and the j th autocovariance of uncorrelated observations from different households. In large samples, this should be approximately the same as the j th autocovariance for observations from the same household multiplied by the probability that observations j steps apart come from the same household. It is straightforward to show that the latter quantity is the proper input for calculating the variance of a sample mean. If observations for a household are not necessarily consecutive, \widehat{AC}_j will be a more complicated average, but again will be the proper estimate of $N \times$ the variance of the sample mean of $h(hc; \lambda)$.

The first row of D is given by

$$\begin{bmatrix} D_{11} & D_{12} \end{bmatrix} = -\exp(\mu_0 + \sigma_0^2/2) \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}. \quad (25)$$

To get the second row, rewrite the second element of $E(h(hc; \lambda))$ as

$$F(\exp(\sqrt{\sigma_0^2} z_{0.995} + \mu_0)) - 0.995 = 0, \quad (26)$$

where $F(\cdot)$ is the true distribution function of hc . While $F(\cdot)$ must have a continuous density, which we denote by $f(\cdot)$, it need not follow the lognormal distribution. It immediately follows that

$$\begin{bmatrix} D_{21} & D_{22} \end{bmatrix} = f(\exp(\sqrt{\sigma_0^2} z_{0.995} + \mu_0)) \exp(\sqrt{\sigma_0^2} z_{0.995} + \mu_0) \begin{bmatrix} 1 & z_{0.995} \frac{1}{2} \sigma_0^{-1} \end{bmatrix}. \quad (27)$$

Simplifying, we get

$$D = \begin{bmatrix} -E(hc) & -\frac{1}{2}E(hc) \\ f(hc_{0.995})hc_{0.995} & f(hc_{0.995})hc_{0.995} z_{0.995} \frac{1}{2} \sigma_0^{-1} \end{bmatrix}. \quad (28)$$

To estimate this matrix, we simply replace population quantities with their sample analogs.

Unless we assume that $F(\cdot)$ is in fact a lognormal distribution, the density $f(hc_{0.995})$ will have to be estimated nonparametrically, using a kernel density estimator. We use a kernel estimator for GAUSS written by Ruud Koning.

Appendix B: Distribution of Error Components Estimates

This appendix describes the estimation procedure for the error components model discussed in Section 3. The procedure is the same as the one in Abowd and Card (1989) or Pischke (1995), except that it allows the data to be unbalanced.

Recall that we are interested in fitting a model to the covariance matrix of health cost residuals, R_{it} , shown in Table 3. Defining T as the number of years of data, we have $L = T(T+1)/2 = 10$ moment conditions, which are the unique elements of the covariance matrix in Table 3.⁴⁷ Define $\theta = (\sigma_f^2, \sigma_a^2, \sigma_{\psi t}^2, \rho, \phi)$ as the parameter vector,⁴⁸ and $m_{it,t+k}(\theta)$ as the contribution of household i to the moment condition that defines the covariance of medical expenses in years t and $t+k$. This moment condition depends upon the medical expense residuals R_{it} and $R_{i,t+k}$, and the parameter vector θ .

To incorporate unbalanced data, we simply omit missing observations from the appropriate moment conditions. Let $1\{R_{it} \neq \text{missing}\}$ be the indicator function that returns 1 when R_{it} is observed for household i . Given the model in equations (2)–(4) and the restrictions in equations (5) and (6), household i 's contributions to the moment conditions are:

$$m_{it,t}(\theta) = [R_{it}^2 - (\sigma_f^2 + \sigma_a^2 + \sigma_{\psi t}^2 + \phi^2 \sigma_{\psi t-1}^2)] \times 1\{R_{it} \neq \text{missing}\}, \quad t \in \{1, 2, 3, 4\}, \quad (29)$$

$$m_{it,t+1}(\theta) = [R_{it}R_{i,t+1} - (\sigma_f^2 + \rho\sigma_a^2 + \phi\sigma_{\psi t}^2)] \times 1\{R_{it}, R_{i,t+1} \neq \text{missing}\}, \quad t \in \{1, 2, 3\}, \quad (30)$$

$$m_{it,t+2}(\theta) = [R_{it}R_{i,t+2} - (\sigma_f^2 + \rho^2\sigma_a^2)] \times 1\{R_{it}, R_{i,t+2} \neq \text{missing}\}, \quad t \in \{1, 2\}, \quad (31)$$

$$m_{it,t+3}(\theta) = [R_{it}R_{i,t+3} - (\sigma_f^2 + \rho^3\sigma_a^2)] \times 1\{R_{it}, R_{i,t+2} \neq \text{missing}\}, \quad t = 1. \quad (32)$$

Let N be the number of households observed in any year. The sample moment condition

⁴⁷Note that we do not group all four variances together, all three first autocovariances together, or the two second autocovariances together. If ψ_{it} is homoskedastic, this leads to inefficient estimates of the model parameters. However, if ψ_{it} is not homoskedastic, grouping all the autocovariances together will result in a misspecified model. Moreover, matching wave-specific moments gives us overidentification restrictions to test the stationarity hypothesis.

⁴⁸Recall that σ_{ut}^2 and σ_ϵ^2 can be written as functions of other parameters in θ .

corresponding to equation (29) is

$$m_{Nt,t}(\theta) = \frac{1}{N} \sum_{i=1}^N [R_{it}^2 - (\sigma_f^2 + \sigma_a^2 + \sigma_{\psi t}^2 + \phi^2 \sigma_{\psi t-1}^2)] \times 1\{R_{it} \neq \text{missing}\}. \quad (33)$$

Letting $N_{t,t+k}$ denote the number of households observed in both years t and $t+k$, we can rewrite equation (33) as

$$m_{Nt,t}(\theta) = \frac{N_{t,t}}{N} \left(\frac{1}{N_{t,t}} \sum_{i:R_{it} \neq \text{missing}} R_{it}^2 - (\sigma_f^2 - \sigma_a^2 + \sigma_{\psi t}^2 + \phi^2 \sigma_{\psi t-1}^2) \right). \quad (34)$$

It is straightforward to derive $m_{Nt,t+k}(\theta)$ for $k > 0$.

Assume that the share $\frac{N_{t,t+k}}{N}$ remains constant as N grows. Denoting $m_N(\theta)$ as the $L \times 1$ vector of all sample moment conditions, we minimize

$$Nm_N(\theta)'W_N m_N(\theta), \quad (35)$$

where W_N is a weighting matrix. Denoting by $\hat{\theta}$ the estimated vector of coefficients and by θ_0 the true vector, the estimator has a sampling distribution given by

$$\sqrt{N}(\hat{\theta} - \theta_0) \rightsquigarrow_D \mathcal{N}(0, V), \quad (36)$$

$$V = (D'WD)^{-1}D'W\Phi WD(D'WD)^{-1}, \quad (37)$$

$$D = \frac{\partial m(\theta_0)}{\partial \theta}, \quad (38)$$

where Φ is the fourth moment matrix of the data. We estimate Φ using its sample analogs. For example, for a moment condition corresponding to a variance,

$$\hat{\Phi}_{(t,t),(t,t)} = \frac{1}{N} \sum_{i:R_{it} \neq \text{missing}} \left(R_{it}^2 - \hat{E}(R_{it}^2) \right)^2, \quad (39)$$

where $\hat{E}(R_{it}^2)$ is calculated as $\frac{1}{N_{t,t}} \sum_{i:R_{it} \neq \text{missing}} R_{it}^2$. We estimate D using $\hat{D} = \frac{\partial m(\hat{\theta})}{\partial \theta}$. Assuming that the model is properly specified, Newey (1985) shows that

$$Nm(\hat{\theta})'Q^{-1}m(\hat{\theta}) \quad (40)$$

is distributed $\chi^2_{L-\text{rank}(D)}$, where in this case $\text{rank}(D)$ equals the number of parameters, Q^{-1} is the generalized inverse of Q and

$$Q = P\Phi P', \quad (41)$$

$$P = I - D(D'WD)^{-1}D'W. \quad (42)$$

Again, to estimate Q , we replace the objects in equation (41) with their sample analogs.

We use two different weighting schemes in our analysis. First, we use an “optimal” weighting matrix, i.e. $W = \Phi^{-1}$. Under optimal weighting, $V = (D'\Phi^{-1}D)^{-1}$ and $Q^{-1} = \Phi^{-1}$. Although the optimal weighting matrix is asymptotically efficient, it can be severely biased in small samples, with the bias most severe when the tails of the empirical distribution are fat (see Altonji and Segal, 1996, for details). Second, we use a “diagonal” weighting matrix, as suggested by Pischke (1995). The diagonal weighting scheme uses the inverse of the matrix that is the same as Φ along the diagonal and has zeros off the diagonal of the matrix. Although not asymptotically efficient, it likely has better small sample properties. In practice, however, the choice of weighting matrix has only a small effect on the results, and we show only the estimates found while using the optimal weighting matrix.

Appendix C: Distribution of One-year Parameters

The one-year model of health costs utilizes the parameter vector $\xi = (\sigma_{a1}^2, \sigma_{u1}^2, \rho_1, \mu_1)$. By simulating a large number of time series at a one-year frequency and aggregating them into two-year data, we can find the summary statistics for two-year data implied by any set of one-year parameters. We therefore estimate ξ by finding the parameter values that come closest to replicating the summary statistics found in the two-year data sample.

We generate the simulated two-year data as follows. For individual i , we simulate a sequence of one-year residuals,

$$R_{1is} = a_{1is} + u_{1is}, \quad s = 1, 2, \dots, 8, \quad (43)$$

where: $u_{1is} \sim \mathcal{N}(0, \sigma_{1u}^2)$ is the simulated one-year white noise component; $a_{1i1} \sim \mathcal{N}(0, \sigma_{1a}^2)$ and $a_{1is} = \rho a_{1is-1} + \epsilon_{1is}$, $s = 2, 3, \dots, 8$, are the simulated one-year AR(1) components; and $\epsilon_{1is} \sim \mathcal{N}(0, \sigma_{1\epsilon}^2)$ is the simulated time- $s+1$ innovation to the AR(1) process. Taking averages,

we compute two-year health costs as

$$hc_{i1} = \frac{1}{2}(\exp(R_{1i1} + \mu_1) + \exp(R_{1i2} + \mu_1)), \quad (44)$$

$$hc_{i2} = \frac{1}{2}(\exp(R_{1i3} + \mu_1) + \exp(R_{1i4} + \mu_1)), \quad (45)$$

and so on. Repeating this sequence many times—in practice, we simulate 1 million histories—yields a panel of simulated two-year data.

With the simulated data in hand it is straightforward to calculate the summary statistics implied by the underlying one-year parameters. One can then combine these summary statistics with the data sample to form moment conditions. We use two groups of moment conditions. The first group converts the autocovariance conditions described in Appendix B into autocorrelation conditions. For example, the moment variable $m_{it,t+1}(\theta)$ defined in equation (30) becomes

$$m_{it,t+1}^1(\xi) = \left[\frac{R_{it}R_{i,t+1}}{\sigma_{Rt} \sigma_{Rt+1}} - CR_{t,t+1}(\xi), \right] \times 1\{R_{it}, R_{i,t+1} \neq \text{missing}\}, \quad t \in \{2, 3\}, \quad (46)$$

where $CR_{t,t+1}(\xi)$ is the two-year residual autocorrelation implied by the one-year model, calculated from simulated data, and σ_{Rt} is the standard deviation of R_{it} , calculated from the data sample. Equations (31) and (32) are modified in an analogous fashion. (Equation (29) is dropped.)

The second group of moment conditions consists of the conditions described in Appendix A, converted for convenience to fit a panel structure. In particular, each of the two conditions in equation (20) is converted into 4 conditions, one for each wave of the two-year data.

The asymptotic distribution of $\hat{\xi}$ is similar to the distribution of $\hat{\theta}$ that was discussed in Appendix B, and thus we do not provide a detailed derivation. The only significant difference is that moment conditions for the mean and 99.5th percentile are added. When computing the gradient matrix D we add:

$$\frac{N_t}{N} \left[\begin{array}{c} -\frac{\partial E_{hc,t}^*(\xi)}{\partial \xi} \\ f(hc_{0.995,t}) \frac{\partial hc_{0.995,t}^*(\xi)}{\partial \xi} \end{array} \right], \quad (47)$$

where $E_{hc,t}^*(\xi)$ and $hc_{0.995,t}^*(\xi)$ denote the model-predicted mean and 99.5th percentile, respectively, for wave t , and N_t denotes the number of non-missing observations. Equation (47)

must be repeated for each wave. One final, lesser difference is that to account for simulation error, the estimated variances must be multiplied by $1 + \frac{N}{N_S}$, where N_S is the number of simulations utilized.

References

- [1] **Abowd, J., and David Card** (1989), “On the Covariance Structure of Earnings and Hours Changes”, *Econometrica*, 57(2), 411-445.
- [2] **Altonji, J., and L. Segal** (1996), “Small Sample Bias in GMM Estimation of Covariance Structures”, *Journal of Business and Economic Statistics*, 14(3), 353-366.
- [3] **Baker, M.** (1997), “Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings”, *Journal of Labor Economics*, 15(2), 338-375.
- [4] **Billingsley, P.** (1979), *Probability and Measure*, John Wiley and Sons.
- [5] **Blau, D. and D. Gilleskie** (2000), “A Dynamic Structural Model of Health Insurance and Retirement”, manuscript.
- [6] **Brockwell, P., and R. Davis** (1991), *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag.
- [7] **Buchinsky, M.,** (1998), “Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research,” *Journal of Human Resources*, 33, 88-126.
- [8] **Christensen, B., and N. Kiefer** (1997), “Inference in Non-linear Panel Models with Partially Missing Observations”, *Journal of Econometrics*, 79, 201-219.
- [9] **Cutler, D. and R. Zeckhauser** (2000), “The Anatomy of Health Insurance”, in A. Culyer and J. Newhouse, eds., *Handbook of Health Economics*, 563-643.
- [10] **Dynan, K., J. Skinner, and S. Zeldes** (2002), “The Importance of Bequests and Life-Cycle Saving in Capital Accumulation: A New Answer”, *American Economic Review*, 92(2), 274-278.
- [11] **Eichner, M., M. McClellan, and D. Wise** (1998), “Insurance or Self-Insurance? Variation, Persistence, and Individual Health Accounts”, in D. Wise, ed., *Inquiries in the Economics of Aging*, 19-45.
- [12] **Employee Benefit Research Institute** (1999), *Health Benefits Databook*, EBRI-ERF.
- [13] **Feenberg, D., and J. Skinner** (1994), “The Risk and Duration of Catastrophic Health Care Expenditures”, *The Review of Economics and Statistics*, 76(4), 633-647.
- [14] **Federal Interagency Forum on Aging-Related Statistics** (2000), *Older Americans 2000: Key Indicators of Well-Being*, U.S. Government Printing Office, available at <http://www.agingstats.gov/chartbook2000/default.htm>.
- [15] **French, E.** (2003), “The Effects of Health, Wealth and Wages on Labor Supply and Retirement Behavior”, manuscript.

- [16] **French, E., and J. Jones** (2003), “The Effects of Health Insurance and Self-Insurance on Retirement Behavior”, manuscript.
- [17] **French, E., and K. Kamboj** , “The Effect of Health Insurance on Health Costs and Health Care Utilization of the Elderly”, *Economic Perspectives* (forthcoming).
- [18] **Hamilton, J.** (1994), *Time Series Analysis*, Princeton University Press.
- [19] **Hubbard, R., J. Skinner, and S. Zeldes** (1994), “The Importance of Precautionary Motives in Explaining Individual and Aggregate Saving”, *Carnegie-Rochester Series on Public Policy*, 59-125.
- [20] **Hubbard, R., J. Skinner, and S. Zeldes** (1995), “Precautionary Saving and Social Insurance”, *Journal of Political Economy*, 103(2), 360-399.
- [21] **Judd, K.** (1998), *Numerical Methods in Economics*, MIT Press.
- [22] **Kimball, M.** (1994), “The Importance of Precautionary Motives in Explaining Individual and Aggregate Saving: A Comment”, *Carnegie-Rochester Series on Public Policy*, 127-132.
- [23] **McCulloch, J.** (1996), “Financial Applications of Stable Distributions”, in G. Maddala and C. Rao (eds.), *Handbook of Statistics*, 14, Elsevier, 393-425.
- [24] **National Center for Health Statistics** (1999), *Health and Aging Chartbook*, Hyattsville, MD.
- [25] **Newey, W.** (1985), “Generalized Method of Moments Specification Testing”, *Journal of Econometrics*, 29(3), 229-256.
- [26] **Newey, W. and D. McFadden** (1994), “Large Sample Estimation and Hypothesis Testing”, in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, Elsevier.
- [27] **Palumbo, M.** (1999), “Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle”, *Review of Economic Studies*, 66(2), 395-421.
- [28] **Pauly, M.** (1990), “The Rational Non-Purchase of Long Term-Care Insurance”, *Journal of Political Economy*, 98(1), 153-168.
- [29] **Pischke, J-S** (1995), “Measurement Error and Earnings Dynamics: Some Estimates From the PSID Validation Study”, *Journal of Business & Economics Statistics*, 13(3), 305-314.
- [30] **Powell, J.,** (1994), “Estimation of Semiparametric Models”, in R. Engle and D. McFadden (eds.) *Handbook of Econometrics*, Vol. 4., Elsevier.
- [31] **Rivers, D. and Q. Vuong** (1997), “Model Selection Tests for Nonlinear Dynamic Models”, *Econometrics Journal*, 5, 1-39.
- [32] **Rust, J. and C. Phelan** (1997), “How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets”, *Econometrica*, 65(4), 781-831.

- [33] **Selden, T., K. Levit, J. Cohen, S. Zuvekas, J. Moeller, D. McKusick, and R. Arnett** (2001), “Reconciling Medical Expenditure Estimates from the MEPS and NHA, 1996”, *Health Care Financing Review*, 23(1), 161-178.
- [34] **Stokey, N., and R. Lucas** (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press.
- [35] **Tsionas, E.** (1999), “Monte Carlo Inference in Econometric Models with Symmetric Stable Disturbances”, *Journal of Econometrics*, 88, 365-401.
- [36] **Vuong, Q.** (1989), “Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses”, *Econometrica*, 57(2), 307-333.