EFACR
7/22/02 11:15AM
FRENCH/JONES

On the Distribution and Dynamics of Health Costs

Eric French
Federal Reserve Bank of Chicago

John Bailey Jones*
SUNY-Albany

May 28, 2002

Abstract

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 both the distribution and dynamics of health costs. We find that the data generating process for health costs is well represented by an ARMA(1,1). Furthermore, innovations to this process are close to lognormally distributed. In any given year, .1% of our sample receives a health cost shock that costs at least $80,000 in present value. Lastly, we discuss the accuracy of numerical solutions when integrating over health costs. Assuming lognormailty, simple approximation rules work well.

*Comments welcome at efrench@frbchi.org and jbjones@csc.albany.edu. We thank John Rust for encouraging us to write this paper. We also thank Meredith Crowley and seminar participants at the Conference on Social Insurance and Pensions Research in Aarhus, Denmark for helpful comments. Diwakar Choubey, Ken Housinger, Kirti Kamboj, 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 stochastic process that determines the distribution and dynamics of health costs is an essential input into understanding savings and retirement decisions.[1] Empirical estimates of health cost processes, however, are relatively scarce. This paper adds to the literature by simultaneously considering both the cross-sectional distribution and persistence of health costs. Using panel data from the Health and Retirement Survey (HRS) and Assets and Health Dynamics of the Oldest Old (AHEAD) surveys, we provide new evidence on both the cross-sectional distribution and the intertemporal persistence of health costs.

With the exception of Feenberg and Skinner (1994), no one has estimated a parameterized model of the time series properties of health costs.[2] Feenberg and Skinner find that health costs are well represented by an ARMA(1,1) process. They further find that the parameters of this process are such that a $1 increase in current health costs will generate an additional $2.65 of lifetime health costs. Hubbard, Skinner, and Zeldes (1994, 1995) use Feenberg and Skinner's estimates to calibrate a life cycle model of consumption and savings under uncertainty. They find that health cost uncertainty has a small effect on savings decisions, but that labor income risk has a large effect. This occurs because the dollar value of a one standard deviation shock to health costs is much smaller than a one standard deviation shock to labor income.[3]

Hubbard, Skinner, and Zeldes (1994, 1995) adopt Feenberg and Skinner's assumption that the innovations to the health cost shock process are lognormally distributed. Rust and Phelan (1997), however, 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, allowing a higher possibility of catastrophic health costs. Although

[1] Examples include studies of saving by Hubbard, Skinner, Zeldes (1994, 1995) and Palumbo (1999), and studies of retirement by Rust and Phelan (1997), Blau and Gilleskie (2001) and French and Jones (2002).

[2] Eichner et al. (1998) model health costs non-parametrically, claiming that health cost shocks are "not well approximated very well by any analytic solution." However, we find that a lognormal distribution approximates the data very well.

[3] Using estimates from Hubbard, Skinner, and Zeldes, Kimball (1994) infers that the capitalized value of a one standard deviation shock to health costs is about $\frac{1}{10}$th as large as the capitalized value of a one standard deviation shock to labor income.

a one standard deviation shock to health costs may be small vis a vis lifetime resources, a catastrophic health cost shock may be quite large. Using a Pareto distribution may be one of the reasons that Rust and Phelan find that health insurance has large effects on retirement behavior. However, Rust and Phelan do not formally test their Pareto specification against a lognormal alternative, and they do not account for the autocorrelation of health costs.

We begin our analysis by comparing the cross-sectional distribution of these two models. Using the likelihood ratio test developed by Vuong (1989), we conclude that in most cases, a lognormal specification usually fits the right tail of the distribution (i.e., the top quintile) better than a Pareto specification. Having concluded that the lognormal is an appropriate distribution for modeling catastrophic health cost shocks that potentially drive economic behavior, we then turn to our primary goal of modeling the entire cross-sectional distribution of health costs. Since catastrophic health cost shocks are potentially important for explaining behavior, we match the mean and the 99.5th percentile of the lognormal model to the empirical distribution. This approach fits the top 20% of the empirical distribution also.

Fitting the cross-sectional distribution, however, is not the only criterion for picking between distributions. We consider an additional criterion. The "best" distribution should be consistent with the time series properties of the empirical health cost process. If health costs are persistent over time, a good distribution for modeling them must capture this. Unfortunately, the Pareto distribution is an intractable distribution for modeling the dynamics of health costs. This is because there is no stochastic process for which the sum of innovations over time maps into a cross sectional Pareto distribution. Fortunately, however, the lognormal distribution has desirable properties for modeling dynamics. It is tractable because an ARMA process with lognormal innovations is distributed lognormally in the cross-section.

Estimating the time series properties of health costs, we find that health costs are reasonably well represented by an error components model that includes both 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. Assuming lognormal innovations, in any given year 0.1% of all households in our sample receive a health cost shock that exceeds

$80,000 in present value.

Because the estimates in this paper are useful inputs into dynamic models of savings and retirement behavior, which most often must be solved numerically, we also provide a practical guide to using the estimates in this paper. For example, a common way of analyzing dynamic models with uncertain health costs is to work with value functions. In these problems one must compute the expectation

$$\int V(A-hc)f(hc)dhc \tag{1}$$

where $V(.)$ is a value function, A is assets and $f(.)$ is the pdf of health costs. In most cases, this integral must be solved with numerical techniques. To evaluate the quality of these techniques, we specify a simple value function and find its expectation under different approaches. First, we integrate over the entire empirical distribution of health costs—we average across the 36,970 observations in our sample. With so many observations, this approach is very accurate. It is also slow. Second, we integrate over health costs using estimates from our lognormal model of health costs and quadrature-based techniques. Quadrature techniques evaluate integrals at a finite number of points. Although quadrature is much faster, it will be accurate only if the health cost distribution is correctly specified. The problem is compounded by curvature in $V(.)$. A quadrature method that matches, say, the expected value of health costs, may not match the expectation of $V(A-hc)$. We find, however, that combining low-order quadrature with our estimated process yields a fairly precise approximation. Our results imply that assuming a lognormal distribution for health costs is reasonable. Paradoxically, we find that in some cases *de*creasing the number of quadrature nodes improves the numerical approximation.

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, and compare it to some aggregate statistics. In section 3, we examine the cross-sectional distribution of health costs more carefully, and in section 4, we consider the correlation of health costs across time. In section 5, we see whether our distributional assumptions lead to accurate numerical integration. We conclude in section 6.

2 Data

We use the Health and Retirement Survey and Asset and Health Dynamics Among the Oldest Old (HRS/AHEAD) data. These data contain detailed information on health costs, health insurance, and demographics.

The HRS is a sample of non-institutionalized[4] individuals aged 51-61 in 1992. Spouses of these individuals were also interviewed, regardless of the spouse's age. The HRS includes both a nationally representative core sample as well as 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 non-institutionalized individuals aged 70 and older in 1993. Researchers at the University of Michigan collect these data, the same researchers who collect the HRS data. As a result, the two data sets have similar sample designs. Like the HRS, spouses of AHEAD respondents are also interviewed, regardless of age. Also like the HRS, the AHEAD includes both a nationally representative core sample as well as additional samples of blacks, Hispanics, and Florida residents. A total of 8,222 individuals in 6,047 households were interviewed in 1993. These individuals were again interviewed in 1995, 1998, and 2000, creating up to four separate responses for each individual.

In 1998 and 2000 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 the datasets 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.

[4]Institutionalized individuals include individuals in nursing homes.

In order to assess the quality of the HRS/AHEAD data, we compare the means of several key variables to aggregated per capita statistics. Table 1 presents means and standard deviations of variables that measure health insurance coverage, health costs, health care utilization, and demographic features.[5] 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.

Variable	Mean	Std. Dev
annual health costs (in 1998 dollars)	2,454	4813
male head of household	0.62	0.49
married	0.47	0.50
age	67.21	10.58
none $\times$ (age <65)	.092	.288
employer-provided $\times$ (age < 65)	0.32	0.47
private $\times$ (age < 65)	0.05	0.21
medicaid $\times$ (age < 65)	0.04	0.19
none $\times$ (age ≥ 65)	0.17	0.37
employer-provided $\times$ (age ≥ 65)	0.15	0.36
private $\times$ (age ≥ 65)	0.12	0.32
medicaid $\times$ (age ≥ 65)	0.06	0.25
annual doctor visits	6.92	9.89
annual nursing home nights	3.79	33.67
annual hospital nights	1.59	5.76
N=36,970		

Table 1: SAMPLE STATISTICS

The central variable of interest in this study is the level of health costs paid by the household. For single households, this the individual's health costs. For married households, this is the sum of husband's and wife's health costs. Health costs are the sum of insurance premia, drug costs, and costs for hospital, nursing home care, doctor visits, dental visits, and outpatient care. For our sample, mean household health costs are \$2,562 and mean health costs for those aged 65 and older are \$2,602. This compares to the US per capita average of \$2,831 for households headed by an individual aged 65 or older (United States, 2000).

One important reason why average health costs in the HRS/AHEAD data are below the

[5]For married households, insurance refers to the husband's insurance. See French and Kamboj (2002) for a fuller description of the construction of all of these variables.

national average is that individuals in the HRS/AHEAD spend far fewer nights in a nursing home. Households headed by someone aged 65 or older spent 6.9 nights in a nursing home per year in our sample versus 15.8 nights in the aggregate statistics (National Center for Health Statistics, 1999).[6] Selden 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.

3 Cross-Sectional Distribution

A critical aspect of health cost uncertainty is the possibility of catastrophic health cost shocks. This means that in modeling the cross-sectional distribution of health costs, special attention must be given to fitting the far right tail of the distribution. We thus proceed in two steps, considering first the upper tail, and then considering the entire distribution.

Previous studies have identified two statistical models for the upper tail of the health cost distribution. Feenberg and Skinner (1994) use the lognormal distribution.[7] This implies that the density function for large health costs, $f(.)$, is

$$f(\ln hc | \ln hc \geq \ln hc_L) = \frac{1}{1 - \Phi([\ln hc_L - \mu]/\sigma)} \phi([\ln hc - \mu]/\sigma), \tag{2}$$

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,

[6] Because the HRS/AHEAD sample was drawn from the non-institutionalized population—which excludes individuals in nursing homes—it is not surprising that the number of nights in a nursing home is lower in the HRS/AHEAD sample than for the national average. Nevertheless, many HRS/AHEAD household members do enter a nursing home after they were initially interviewed.

[7] While Feenberg and Skinner fit a lognormal ARMA process to their data, their approach automatically implies that the upper tail of the cross-sectional distribution is truncated lognormal.

which has the density

$$g(hc|hc \geq hc_L) = \gamma hc_L^{\gamma} hc^{-(1+\gamma)}. \tag{3}$$

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

$$g(\ln hc|\ln hc \geq \ln hc_L) = \gamma e^{-\gamma[\ln hc - \ln hc_L]}. \tag{4}$$

Choosing between the two models boils down to seeing whether logged health costs are better modeled as following a truncated normal or an exponential distribution.

The two models can be compared formally with the likelihood ratio test developed by Vuong (1989). Consider an i.i.d. sample of size N. Let $L_N(\widehat{\mu}, \widehat{\sigma^2})$ and $L_N(\widehat{\gamma})$ denote the sample log-likelihoods for the truncated lognormal and Pareto models, respectively, and define $\omega_N^2 = \frac{1}{N}\sum_{n=1}^{N}[\ln f(\ln hc_n; \widehat{\mu}, \widehat{\sigma^2}) - \ln g(\ln hc_n; \widehat{\gamma})]^2$. Since the two models at issue are strictly non-nested, it follows from Vuong's Theorem 5.1 that the statistic

$$D_N \equiv N^{-1/2}[L_N(\widehat{\mu}, \widehat{\sigma^2}) - L_N(\widehat{\gamma})]/\omega_N \tag{5}$$

will converge in distribution to a standard normal variable if the two models are equivalent. On the other hand, if the lognormal model better represents the data generating process, D_N will converge to infinity, and if the Pareto model is better, D_N will converge to negative infinity.

Table 2 presents parameter estimates, log-likelihoods, and the p-values of D_N for all the subsamples. Given the way the Vuong test statistic D_N has been constructed, a low p-value should be taken as evidence in support of the lognormal specification, and a high p-value should be taken as evidence in support of the Pareto specification. With one exception, the data are truncated to exclude health costs below \$4,000 for unmarried individuals and \$6,000 for married couples.[8] For most of the truncated subsamples shown in Table 2, the truncated lognormal model provides the better fit. The Pareto model does fit better, however, for some of the subsamples, which are marked by a "*".[9]

In practical terms, however, the differences are small. This is shown in Figure 1, which applies a variant of Pareto's law.[10] In particular, if health costs follow a Pareto distribution with cdf $G(.)$,

$$\ln(1 - G(hc)) = -\gamma[\ln hc - \ln hc_L], \tag{6}$$

so that the probability that health costs exceed hc is log-linear in hc. Figure 1 shows the curve for $\ln(1 - G(hc))$ generated by the data, which we refer to as the "Pareto curve". It also shows the curves generated by the truncated lognormal and Pareto models, described above, as well a "fitted" lognormal, described below. In the first panel, which shows curves for individuals over 79 years old, the empirical Pareto curve bends down at higher income levels; consistent with the Vuong statistic for that subsample, the lognormal distribution clearly provides a better fit. In the second panel of Figure 1, which shows curves for agents in the 65-79 age group, the empirical Pareto curve remains straight. The two competing models, however, appear to fit the data equally well; this, too, is consistent with the Vuong statistic, which assigns a relatively small advantage to the Pareto model. In both panels, however, the upper tail of the health cost distribution is well-approximated with a truncated lognormal distribution.

[8] The gap in the truncation points reflects the \$1,700 gap in average health costs for the two groups. When considering the married and unmarried, we use a floor of \$5,000. In addition, the truncation points for younger (age < 65) individuals with privately-purchased insurance are set to \$6,000 and \$10,000 for unmarried and married respondents, respectively. Because these individuals tend to have especially high health costs, the usual truncation points are too low to isolate the upper tail of their health cost distribution. The better fit of the lognormal model in this subsample is not a function of this higher truncation point. In fact, when the truncation points are \$4,000 or \$6,000, the lognormal model is more likely to outperform the Pareto model.

[9] Fitting the truncated normal distribution proved especially difficult for these subsamples as well; 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 arise because these data are in fact well approximated by the Pareto model.

[10] See Creedy (1977) for an interesting discussion of Pareto's derivation.

Subsample	N	$\widehat{\mu}$	$\widehat{\sigma^2}$	$L_N(\widehat{\mu}, \widehat{\sigma^2})$	$\widehat{\gamma}$	$L_N(\widehat{\gamma})$	P-value(D_N)
Entire Sample	4591	2.84	3.71	−1850.8	1.81	−1861.3	0.0047
Age < 65, Unmarried							
All*	579	−19.61	15.85	−225.8	1.84	−225.6	0.6954
Employer-Provided*	237	−16.07	12.04	−59.6	2.12	−59.3	0.8954
Privately-Purchased*	81	−15.57	11.98	−20.7	2.11	−20.5	0.9396
No Insurance	73	6.37	2.09	−43.3	1.49	−44.1	0.2009
Medicaid	34	7.15	1.44	−18.3	1.56	−18.8	0.2340
Medicare	54	5.29	3.49	−40.8	1.27	−41.2	0.2496
Ages 65-79, Unmarried							
All*	494	−20.10	16.39	−201.8	1.81	−200.9	0.9960
Employer-Provided*	110	−22.86	20.03	−57.5	1.61	−57.5	0.6883
Privately-Purchased*	201	−13.43	9.56	−26.9	2.39	−26.1	0.9872
No Insurance	5	8.63	1.31	−4.9	0.95	−5.2	0.2914
Medicaid*	40	−24.69	22.07	−22.2	1.56	−22.2	0.6872
Medicare*	138	−23.92	21.74	−78.5	1.54	−78.4	0.5695
Age > 79, Unmarried							
All	534	2.72	5.51	−402.5	1.27	−404.5	0.0848
Employer-Provided	117	6.51	3.40	−112.8	1.02	−114.7	0.1093
Privately-Purchased*	216	−24.05	21.59	−119.8	1.56	−119.6	0.8464
No Insurance	4	8.90	0.24	−1.9	1.41	−2.6	0.3060
Medicaid	56	8.25	1.14	−43.4	1.20	−46.0	0.0801
Medicare	141	6.86	2.35	−111.6	1.21	−113.9	0.0926
Age < 65, Married							
All	1449	7.58	1.04	−472.6	1.93	−492.8	0.0009
Employer-Provided	709	−2.70	5.68	−160.9	2.17	−161.3	0.3296
Privately-Purchased	248	5.45	1.68	−4.6	2.66	−5.2	0.2838
No Insurance	158	8.17	0.88	−69.2	1.71	−73.1	0.0461
Medicaid	13	−4.41	8.45	−6.2	1.69	−6.2	0.4593
Medicare	47	2.97	3.83	−20.1	1.77	−20.2	0.3673
Ages 65-79, Married							
All*	934	−12.82	9.46	−122.4	2.39	−121.0	0.9788
Employer-Provided*	258	−12.97	9.54	−35.3	2.38	−34.5	0.9772
Privately-Purchased	455	−7.82	6.46	−7.7	2.67	−7.8	0.4826
No Insurance	7	9.01	0.03	2.9	3.08	0.9	0.0443
Medicaid	30	−1.80	6.70	−13.5	1.73	−13.5	0.4393
Medicare*	184	−17.14	13.54	−58.2	1.98	−58.0	0.7962
Age > 79, Married							
All	233	−9.44	10.71	−96.7	1.79	−96.7	0.4127
Employer-Provided*	50	−16.64	13.00	−14.7	2.03	−14.7	0.6535
Privately-Purchased	127	8.20	0.60	−27.1	2.15	−30.0	0.0932
No Insurance	1	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
Medicaid	3	9.26	1.65	−3.4	0.80	−3.7	0.4032
Medicare	52	1.15	7.23	−40.2	1.25	−40.3	0.3799

Table 2: PARAMETER ESTIMATES AND LIKELIHOOD RATIOS

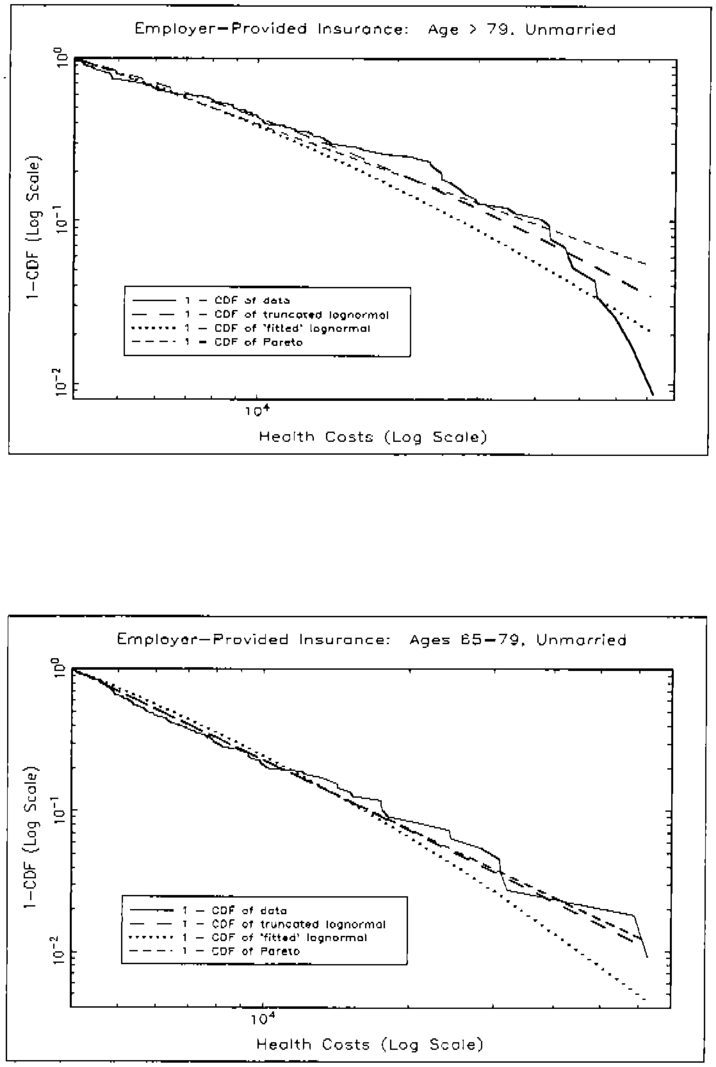


Figure 1: PARETO CURVES

Having concluded that the upper tail can be approximated by truncated lognormal, we turn to the entire cross-sectional distribution of health costs. A nice property of a lognormal specification for the entire distribution of health costs is that an ARMA process with lognormal innovations generates a lognormal unconditional/cross-sectional distribution.[11] Because

[11] If health costs follow a stationary stochastic process, the unconditional distribution of an individual's health costs should be equivalent to the cross-sectional distribution of the population.

the goal of this paper is to simultaneously model both the cross sectional and time series properties of health costs, we consider whether a lognormal can fit the entire distribution of health costs.

One problem with a lognormal distribution is that a large number of individuals have very low health expenses; 18% report less than $250 of health costs and 8% report no expenditures at all. If one includes these individuals,[12] the empirical distribution of logged health costs has a *lower* tail that is too fat for a lognormal distribution. This is illustrated in Figure 2, which shows the empirical cdf and the lognormal cdf generated by standard estimates for all unmarried individuals.[13]

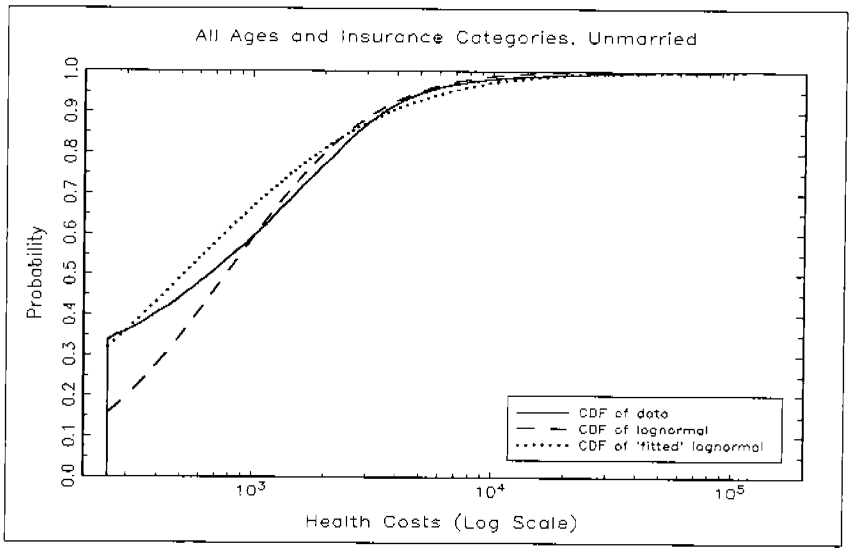


Figure 2: DISTRIBUTION OF LOGGED HEALTH COSTS FOR RESPONDENTS AGED 65-79

These considerations lead to 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

[12]For more interpretation of these low cost reports, see French and Kamboj (2002).
[13]To be consistent with the time series analyses below, health costs below $250 were recoded to $250.

costs. In particular, we pick values μ and σ^2 such that

$$e^{\mu+\sigma^2/2} = E(hc), \tag{7}$$

$$\Phi(\frac{\ln hc_{0.995} - \mu}{\sigma}) = 0.995, \tag{8}$$

where $E(hc)$ and $hc_{0.995}$ are the mean and the 99.5th quantile of the empirical cross section.[14] Defining $z_{0.995} = \Phi^{-1}(0.995)$ to be the 99.5th quantile of the standard normal distribution, we get

$$\sigma = z_{0.995} - \sqrt{z_{0.995}^2 - 2[\ln hc_{0.995} - \ln(E(hc))]}, \tag{9}$$

$$\mu = \ln hc_{0.995} - \sigma z_{0.995}. \tag{10}$$

Figure 1 shows that in addition to replicating average health costs, this "fitted" specification fits the far upper tails of the data fairly well. As Figure 2 shows, the weakness of this specification is that it dramatically undershoots the lower tail of the distribution. In practical terms, this is relatively a 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 3 presents the parameters of the "fitted" distributions, and the log-likelihoods that result when these distributions are applied to the upper tail. Table 3 also presents means and variances for each subsample. Comparing Table 3 to the truncated lognormal estimates shown in Table 2 reveals that while the parameters themselves differ greatly, the fits are for the most part quite similar.

[14]The GAUSS function we use to find the 99.5th quantile of the data interpolates between data points to get the exact quantile. Mechanically, this allows us to find the 99.5th quantile in subsamples with less than 200 observations. The precision of such estimates, of course, is quite low.

Subsample	$\widehat{E}(\ln hc)$	$\widetilde{Var}(\ln hc)$	$\widetilde{\mu}$	$\widetilde{\sigma^2}$	$L_N(\widetilde{\mu}, \widetilde{\sigma^2})$
Entire Sample	7.07	1.47	7.05	1.56	−1898.2
		Age < 65, Unmarried			
All	6.58	1.17	6.41	1.75	−235.4
Employer-Provided	6.65	0.96	6.62	1.16	−65.4
Privately-Purchased	7.64	0.96	7.43	1.31	−25.0
No Insurance	6.24	1.02	5.78	2.46	−43.3
Medicaid	5.93	0.74	5.60	2.04	−18.8
Medicare	6.68	1.58	6.50	2.38	−40.9
		Ages 65-79, Unmarried			
All	6.67	1.19	6.29	2.17	−223.6
Employer-Provided	6.71	1.02	6.40	1.90	−59.5
Privately-Purchased	7.46	0.78	7.30	1.10	−44.8
No Insurance	6.28	1.23	5.77	3.08	−5.3
Medicaid	5.99	0.81	5.39	2.72	−22.9
Medicare	6.51	1.11	5.92	2.67	−80.8
		Age > 79, Unmarried			
All	6.79	1.52	6.55	2.38	−405.0
Employer-Provided	6.86	1.62	6.51	2.90	−113.9
Privately-Purchased	7.58	0.99	7.37	1.49	−128.5
No Insurance	6.24	1.31	6.60	1.21	−2.8
Medicaid	6.13	1.13	5.53	3.16	−44.5
Medicare	6.56	1.32	6.36	2.30	−113.5
		Age < 65, Married			
All	7.47	1.34	7.56	1.08	−473.1
Employer-Provided	7.40	1.12	7.42	1.03	−167.6
Privately-Purchased	8.65	0.83	8.63	0.71	−18.2
No Insurance	7.12	1.70	7.32	1.34	−69.6
Medicaid	6.58	1.49	6.18	2.65	−6.5
Medicare	7.69	1.27	7.64	1.26	−20.5
		Ages 65-79, Married			
All	7.59	1.19	7.61	1.06	−171.5
Employer-Provided	7.46	1.05	7.40	1.17	−51.0
Privately-Purchased	8.28	0.62	8.22	0.68	−40.8
No Insurance	7.27	1.26	7.62	0.42	1.2
Medicaid	6.79	1.71	6.94	1.68	−13.7
Medicare	7.37	1.20	7.33	1.30	−63.2
		Age > 79, Married			
All	7.67	1.44	7.61	1.49	−103.7
Employer-Provided	7.52	1.19	7.28	1.66	−17.6
Privately-Purchased	8.34	0.76	8.35	0.62	−28.1
No Insurance	7.37	0.92	7.62	0.45	NA
Medicaid	6.35	1.25	5.44	4.04	−3.9
Medicare	7.52	1.49	7.45	1.86	−41.1

$\widehat{E}(\ln hc)$ and $\widetilde{Var}(\ln hc)$ correspond to the empirical mean and variance
$\widetilde{\mu}$ and $\widetilde{\sigma^2}$ correspond to "fitted" mean and variance

Table 3: PARAMETERS AND LIKELIHOOD RATIOS FOR FITTED DISTRIBUTIONS

4 The Persistence of Health Cost Shocks

In this section, we consider the time series properties of health costs. Health cost uncertainty depends not only on the probability of receiving a massive health cost shock, but also on the persistence of a massive health cost shock. It follows from the estimates presented in the first row of Table 3 that in any given year only 0.13% of households have health costs in excess of \$50,000. Given that 68% of households in the HRS-AHEAD sample have at least \$50,000 in assets, most households can self-insure against a \$50,000 shock by dissaving. However, if a household must spend \$50,000 per year for 10 years, its total medical expenditure is roughly \$500,000. Given that only 14% of all households have assets over \$500,000, this is a disastrous expense.

4.1 The Model

Feenberg and Skinner (1994) find that their health cost data are reasonably well represented by the sum of an AR(1) and a white noise process.[15] Simulating this process, they find that a \$1 shock to health costs raises future health costs by \$2.65. In order to re-examine their findings, we estimate the following error components model:

$$\ln hc_{it} = X'_{it}\beta + R_{it}, \tag{11}$$

$$R_{it} = f_i + a_{it} + u_{it}, \tag{12}$$

$$a_{it} = \rho a_{it-1} + \epsilon_{it}, \tag{13}$$

$$u_{it} = \psi_{it} + \theta\psi_{it-1}, \tag{14}$$

where $X'_{it}\beta$ represents average health costs given covariates X_{it}. We decompose the residual R_{it} into: f_i, a permanent person-specific component; a_{it}, an autoregressive component; and u_{it}, a moving average component. For tractability, we assume that all the components in equations (11) -(14) are mutually orthogonal[16] and that a_{it} is a stationary process. Note that

[15]Note that this sum can be written as an ARMA(1,1).
[16]Specifically, we assume $E[X_{it}R_{it}] = E[f_i a_{it}] = E[f_i u_{it}] = E[a_{it}u_{it}] = E[a_{it-1}\epsilon_{it}] = E[\psi_{it}\psi_{it-1}] = 0$.

in this model t denotes a two year period.

The estimation procedure has two stages. First, we regress log health costs on demographic and health insurance variables that households can use to forecast future health costs. The estimates are presented in Table 4. The omitted category is a single female younger than 65 who has no health insurance.[17]

	Coef.	Std. Err.
male	-0.072	0.017
married	0.851	0.016
age	0.033	0.008
age squared	-0.0002	0.0001
employer-provided $\times$ (age $<$ 65)	0.251	0.020
private $\times$ (age $<$ 65)	1.361	0.030
medicaid $\times$ (age $<$ 65)	-0.495	0.034
none or medicare $\times$ (age $\geq$65)	-0.042	0.029
employer-provided $\times$ (age $\geq$ 65)	0.137	0.029
private $\times$ (age $\geq$ 65)	0.898	0.031
medicaid $\times$ (age $\geq$ 65)	-0.586	0.034
wave dummies included		
$R^2 = 0.270$, $\sigma = 1.0336$, $N = 36,970$		

Table 4: REGRESSION OF LOG HEALTH COSTS

In the second step of the estimation procedure, we fit the model described in equations (12)-(14) to the empirical covariance matrix shown in Table 5. Empirical covariances are shown below the diagonal of this matrix, and empirical correlations are shown above the diagonal. Standard errors are in parentheses. Because the data are unbalanced, Table 5 also shows the number of observations in each cell (in brackets).

We fit several variants of the error components model to this covariance matrix. Details of the estimation procedure are in appendix A. Table 6 shows parameter estimates and values of Newey's (1985) overidentification test statistic. 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.

[17]Most of the parameter estimates are of the expected sign. The one surprising finding is that those with no health insurance have lower health costs than those with employer-provided insurance. This is particularly surprising given that employers on average contribute \$2,700 towards employee's health insurance plans (EBRI, 1999). See French and Kamboj (2002) for more on how to interpret the estimates in Table 4.

Empirical Covariances and Correlations of the Residuals of log Health Costs, Waves 2-5				
	Wave 2	Wave 3	Wave 4	Wave 5
Wave 2	1.0993 (0.0232) [3957]	0.3948	0.3643	0.3590
Wave 3	0.4585 (0.0199) [3366]	1.2267 (0.0192) [9722]	0.4058	0.3319
Wave 4	0.3778 (0.0187) [3380]	0.4446 (0.0147) [6999]	0.9784 (0.0144) [11369]	0.4691
Wave 5	0.3776 (0.0192) [3528]	0.3688 (0.0142) [6590]	0.4655 (0.0117) [9362]	1.0066 (0.0147) [11415]
Standard errors in parentheses, number of observations in brackets.				

Table 5: EMPIRICAL COVARIANCE MATRIX

The first column of Table 6 shows the results for a simple AR(1) model. This model is overwhelmingly rejected by the data; the overidentification test statistic is 367, for a p-value of 0. The reason for this failure can be seen in Table 5, which shows that while there is large decline from the variance to the first autocovariance, the decline between the first and second (and second and third) autocovariance is much smaller. An AR(1) model, which generates a geometrically declining series of autocovariances, cannot replicate this progression.

The above reasoning suggests adding a moving average component to the AR(1) model. We begin with the simplest case, setting $\theta = 0$, so that the moving average component is white noise, and assuming this white noise component of health costs, ψ_{it}, is homoskedastic across waves. This leads to the following covariance restrictions:

$$Var(R_{it}) = Var(a) + Var(\psi), \tag{15}$$

$$Cov(R_{it}, R_{it-k}) = \rho^k Var(a). \tag{16}$$

Note that $Var(a)$ and $Var(\psi)$ are not time specific. Estimates from this model are reported in the second column of Table 6. The overidentification test statistic is 99, implying a

	Model					
Parameter	1	2	3	4	5	6
$Var(a_{it})$	1.026	0.520	0.687	0.361	0.520	0.425
	(0.009)	(0.017)	(0.14)	(0.038)	(0.017)	(0.039)
$Var(\epsilon_{it})$	0.739	0.127	0.675	−0.011	0.127	0.043
	(0.014)	(0.045)	(0.012)	(0.111)	(0.045)	(0.106)
ρ	0.529	0.870	0.134	1.015	0.870	0.949
	(0.008)	(0.019)	(0.017)	(0.047)	(0.019)	(0.041)
$Var(u_{it})$		0.532		0.691	0.557	−0.683
		(0.017)		(0.050)	(0.035)	(0.078)
$Var(\psi_{it})$		0.532		0.678	0.557	−0.683
		(0.017)		(0.034)	(0.023)	(8.741)
θ				0.134		0.078
				(0.025)		(0.027)
$Var(f_i)$			0.365			
			(0.014)			
χ^2-statistic	432.6	131.1	124.2	123.7	8.8	6.3
Degrees of freedom	8	7	7	6	4	2
p-value	0.000	0.000	0.000	0.000	0.068	0.042

Notes
Standard errors in parentheses
Estimates obtained using a diagonal weighting matrix
$Var(a_{it})$ = variance of transitory component of health costs
$Var(\epsilon_{it})$ = variance of the innovation in autoregressive component
ρ = autoregressive coefficient of health costs
$Var(u_{it})$ = average variance of moving average component
$Var(\psi_{it})$ = average variance of the innovation to the moving average component
θ = moving average coefficient of health costs
$Var(f_i)$ = variance of permanent person-specific component

Table 6: PARAMETER ESTIMATES

considerably better fit than the AR(1). Given that we have only 7 degrees of freedom, however, the model is still strongly rejected.

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. Columns 3 and 4 of Table 6 show estimates from these two models. These models do not fit the data much better than the AR(1) with white noise. Moreover, parameter estimates from the MA(1) model produce estimated variances that are negative.[18]

[18] Adding a white noise component also produced negative variances.

What does seem to improve goodness of fit, however, is allowing the variance of the white noise component, ψ_{it}, to differ across waves. The questions used to generate the health cost measure vary from wave to wave. Such heteroskedasticity could reflect variation in survey questions. Results from this model are shown in column 5 of Table 6. 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 9.7. Although the model is still rejected at the 10% level, there are no large systematic differences between the data and the fitted model. The variances and covariances of the fitted model are shown in Table 7. Note how closely the fitted variances and covariances match the empirical variances and covariances 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	1.0993			
Wave 3	0.4525	1.2267		
Wave 4	0.3935	0.4525	0.9784	
Wave 5	0.3421	0.3935	0.4525	1.0066

Table 7: 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 a MA(1) with heteroskedastic innovations. The model does fit the data slightly better, but introduces two additional parameters, leaving us with only two degrees of freedom. Fortunately, the parameters ρ, $Var(a_{it})$, and $Var(u_{it})$[19] seem stable across most specifications. 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. These implications are discussed below.

An important limitation of our data is that our health cost measure refers to average health costs over the previous two years. In most studies of household decision-making the relevant time period is one year. We use a simulation model, described in Appendix B, to calibrate annualized white noise and AR(1) processes that are consistent with the two-year

[19]Note that $Var(u_{it}) = Var(\psi_{it}) + \theta^2 Var(\psi_{it-1})$.

estimates presented in Table 6. Table 8 presents the parameters that result when the annual data are modeled as the sum of an AR(1) and homoskedastic white noise processes.

Also included in Table 8 are estimates of $Var(a)$, $Var(u)$, and ρ from Feenberg and Skinner (1994) and Hubbard, Skinner, and Zeldes (1994), both of which are measured at 1-year frequencies. Feenberg and Skinner's estimates of both $Var(a)$ and $Var(u)$ 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 Skinner likely underestimate the variance of health costs.[20] Third, their sample consists only of individuals whose health 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 $Var(u)$ is much smaller than ours, and assuming that measurement error is transitory, differences in the pervasiveness of measurement error across datasets will be reflected in differences in $Var(u)$.

Hubbard, Skinner, and Zeldes (1994, 1995) use data from the 1977 National Health Care Expenditures Survey to estimate the cross sectional variance of health costs. Their estimated cross-sectional variance is slightly smaller than our estimate. We are not sure what causes this discrepancy, although their data are 20 years older than our data. The variance of health costs has potentially grown over time. Alternatively, their data might be measured more accurately than our data, leading to lower variance from measurement error.[21] Given that Hubbard, Skinner and Zeldes take their estimates of the shares of variance of $Var(a)$ and $Var(u)$ from Feenberg and Skinner, they attribute more of the cross-sectional variance to the autoregressive component, a_{it}, than do we.

[20] When restricting our sample to those who have non-missing health costs in all waves, the variance of health costs drops by 5%.

[21] Hubbard, Skinner and Zeldes delete all zero-cost observations from their data. When we use their bottom coding rules, our cross sectional variance increases. Therefore, bottom coding decisions are not the source of the discrepancy.

Frequency	$Var(a)$	$Var(u)$	ρ
2 year	.520	.532	.870
1 year	.515	.972	.954
Feenberg and Skinner	.269	.099	.896
Hubbard, Skinner, Zeldes	.920	.220	.901

Table 8: PARAMETER ESTIMATES, 1 AND 2 YEAR FREQUENCIES

4.2 Lifetime Health Cost Risk

With the parameters described immediately above, we can estimate the health cost risk that households face. To evaluate this risk, we run the following experiment. First, we simulate a large number of health cost histories. Next, we simulate the health cost histories again, but with the shocks set to zero in initial period. The differences between the two sets of histories give the effects of the shocks on lifetime health costs.

In order to analyze the lifetime incidence of health cost shocks, we combine information on the variance and persistence of shocks from Tables 3 and 8. We take both the autoregressive parameter ρ and the share of cross-sectional variance explained by the autoregressive component, $\frac{Var(a)}{Var(a)+Var(u)}$, from Table 8, using the parameters for annual frequencies. As discussed above, however, our preferred cross-sectional model matches the mean and the 99.5 percentile of the cross-sectional distribution, rather than the cross sectional variance. Therefore, we pick the cross sectional variance parameter, $\sigma^2 = Var(a) + Var(u)$, from the fitted estimates in Table 3, using the value that applies to individuals that are ages 65-79, married and covered by employer-provided insurance. We rescale $Var(a)$ and $Var(u)$ so that the sum $Var(a) + Var(u)$ is equivalent to this variance parameter. For the experiments below, this led us to set mean log health costs to 7.153 and the variance to 1.652.[22]

Using the parameters just described, we simulate 30-year health cost sequences for 1 million households. Each household begins at age 64 with a draw of (a_{i64}, u_{i64}), and then

[22]We want our value of σ^2 when measured at a two-year frequency to equal 1.17, which is the value of $\widetilde{\sigma^2}$ reported in Table 3 for married households aged 65-79 with employer-provided insurance. Note that the measure of total variance shown in Table 8, measured at a two-year interval, is 1.05. Therefore, we scale up the one-year values of $Var(a)$ and $Var(u)$ by $\frac{1.17}{1.05}$, so that the adjusted values we use in the experiments are $Var(a) = 0.515 \times \frac{1.17}{1.05} = 0.574$ and $Var(u) = 0.972 \times \frac{1.17}{1.05} = 1.083$. To keep the overall mean the same, we lower $\widetilde{\mu}$ from 7.395, which is reported in Table 3, to 7.153.

realizes a 30-year sequence of the innovations $\{\epsilon_{it}, u_{it}\}$. The resulting health cost sequence is then discounted back to age 65, using a discount factor that combines an annual interest rate of 3% with age-specific mortality adjustments. We then redo the sequence with 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.

This Monte Carlo exercise allows us to construct the distribution of lifetime health cost shocks. When the AR(1) and white noise innovations are considered together, lifetime health cost shocks have a standard deviation of $9,300. This considerably larger than the contemporaneous (age-65) standard deviation of $5,400, indicating that persistence in health costs is quite important. Turning to catastrophic shocks, we find that 1% of the population will receive a lifetime shock of at least $34,000, and 0.1% will receive a shock of at least $80,000. The AR(1) innovation, ϵ_{it}, has a lifetime standard deviation of $7,700 and a contemporaneous standard deviation of $1,450. While the AR(1) innovation is only responsible for 7% of the contemporaneous variance, it is responsible for 67% of the lifetime variance. Transitory shocks generate most of the cross-sectional and short-term variance, but persistent shocks, reflecting chronic conditions, generate most of the lifetime risk.

The amount of health cost risk implied by our estimates is higher than that found by Feenberg and Skinner (1994) or Hubbard, Skinner and Zeldes (1994). Redoing the Monte Carlo exercise with Feenberg and Skinner's parameter values, we find that lifetime health cost shocks have a standard deviation of $3,500, of which $3,400 is attributable to the AR(1) innovation. When Hubbard, Skinner and Zeldes' parameter values are used, the standard deviation of lifetime health cost shocks rises to $7,000 (versus our value of $9,300).

Feenberg and Skinner report that a $1 health cost shock today leads to $2.65 in future health costs. In contrast, we find that a $1 health cost shock today only leads to an additional $.40 in health costs in the future. Most of this reflects differences in parameters. Given that our error components model attributes a much smaller fraction of the variance to autoregressive component than theirs, we would expect the effects of health cost shocks to be much less persistent in our model. Using Feenberg and Skinner's values of $Var(a)$ and

$Var(u)$ (from their ARMA(1,1) model) and our simulation methods, a \$1 health cost shock today leads to an additional \$1.84 in health costs in the future. Using Hubbard, Skinner and Zeldes' values, the corresponding figure is 2.56.[23]

4.3 Lognormality Revisited

We conclude this section by re-examining the lognormality assumption. When estimating the model described in this section, we made no assumptions about the distribution of innovations. However, in order to make inference about lifetime uncertainty over health costs, we assumed that innovations to the health cost process are lognormally distributed. Below we provide descriptive evidence about this assumption. An important implication of lognormal innovations is that there should be a linear relationship between the log of the current health costs (more precisely, the log health cost residuals) and the log of lagged health costs. Figure 3 provides evidence on this. The left panel of Figure 3 shows the relationship between health costs over the last two years and health costs for the same households over the preceding two years. The center and right panel show the relationship between health costs over the past two years and preceding four and six years, respectively. The figures also show the fitted regression lines implied by the AR(1) plus white noise model.[24] Note that in all three panels, the relationship is roughly linear.

put figures in fig3.zip here

Figure 3: RELATIONSHIP BETWEEN HEALTH COST RESIDUALS AND LAGGED HEALTH COST RESIDUALS

[23]Our measure was constructed as follows. For each simulated individual, we divided the discounted health cost stream caused by an age-65 health cost shock by the age-65 health cost shock itself. The numbers reported in the text correspond to the median of this ratio. We took the median of this ratio 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.

[24]Note that this is $\frac{Cov(R_{it},R_{it-k})}{Var(R_{it})} = \frac{\rho^k Var(a)}{Var(a)+Var(u)}$ for $k \geq 1$.

5 Numerical Applications

Health cost uncertainty is potentially an important determinant of savings and retirement behavior. It follows that a useful test of our model of health costs is to see whether it generates accurate predictions within dynamic models of saving and retirement. Since these models are often solved using value function iteration, an important prediction is the expectation:

$$\int V(A - e^{\ln hc}) f(\ln hc) d \ln hc, \tag{17}$$

where $V(.)$ is a value function, A is assets and $f(.)$ is the pdf of logged health costs—we work with logs for convenience.[25]

Our goal in this section of the paper is two-fold. First, we provide evidence on whether a lognormal model of health costs allows us to accurately compute the integral in equation (17). In particular, we see whether integration using the density function implied by our statistical model generates the same expectation as an integration using the unrestricted empirical density. Second, because the integral in equation (17) must often be solved numerically, we provide a practical guide on how to utilize our estimates with numerical integration.

The first approach to computing the integral in equation (17) is to find the sample mean:

$$\int V(A - e^{\ln hc}) f(\ln hc) d \ln hc \approx \frac{1}{N} \sum_{n=1}^{N} V(A - e^{\ln hc_n}), \tag{18}$$

where n indexes observations and N is the sample size. This approach places few restrictions on $f(.)$, and the approximation becomes exact as N approaches infinity. But in exercises where equation (17) must be calculated repeatedly, solving the value function N times is not computationally feasible. Nonetheless, computing the sample mean as in equation (18)

[25] Ideally, the expectation should be conditioned on current and lagged health costs; we should consider $E(V_{t+1}|hc_t, hc_{t-1}, ...)$ rather than $E(V_{t+1})$. An important part of this section, however, is to calculate the integral in equation (17) with an empirical density that has not been restricted to fit any analytical model of health costs. Given that we have a relatively short panel, we cannot compute the conditional density $f_N(\ln hc_{t+1} | \ln hc_t, \ln hc_{t-1}, ...)$, nor can we infer this using $f_N(\ln hc_{t+1} | \ln hc_t)$. Because health costs are not Markov, a researcher cannot disentangle the persistent and transitory components of $\ln hc_t$ without making distributional assumptions.

provides a useful benchmark for evaluating other approaches.

A more practical approach is to approximate the expected value function by quadrature, i.e, by evaluating the value function at a small number of points:

$$\int V(A - e^{\ln hc}) f(\ln hc) d \ln hc \approx \sum_{i=1}^{I} V(A - e^{\ln hc_i}) w(\ln hc_i), \tag{19}$$

where $\ln hc_i$ is a quadrature node and $w_i = w(\ln hc_i)$ is a quadrature weight, with $\sum_i w_i = 1$. In order for quadrature to yield an accurate approximation, one must not only choose the correct distribution for $f(.)$, but good values of nodes and weights as well.

A promising approach is Gauss-Hermite quadrature, where the nodes and weights are picked so that when $f(.)$ is the standard normal pdf and $V(.)$ is a low-order polynomial in $\ln hc$, the approximation is exact.[26] Gauss-Hermite quadrature need not work well, however, when $V(A - e^{\ln hc})$ is not well-approximated by a low-order polynomial. We thus consider an alternative approach. First, we constrain x_i to lie in $[x_L, x_H]$, with

$$x_i = x_L + [x_H - x_L] \left(\frac{i-1}{I-1} \right)^{\nu}, \tag{20}$$

with $\nu > 0$ controlling the spacing of the grid points. As ν gets smaller, the grid becomes increasingly fine at the upper end; conversely, $\nu = 1$ implies evenly-spaced nodes. Let $\Phi(.)$ denote the standard normal distribution. If μ and σ^2 are the parameters of a lognormally-distributed health cost distribution, one can approximate the integral with:

$$\ln hc_i = \sigma x_i + \mu, \tag{21}$$

$$w_i = \Phi([x_{i+1} + x_i]/2) - \Phi([x_i + x_{i-1}]/2), \quad i = 2, ..., I-1, \tag{22}$$

$$= \Phi([x_2 + x_1]/2), \quad i = 1, \tag{23}$$

$$= 1 - \Phi([x_I + x_{I-1}]/2), \quad i = I. \tag{24}$$

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

Note that as I gets large, this converges to the Stieltjes integral, $\int V(A - e^{\ln hc})dF(\ln hc)$, where $F(.)$ is the cdf of logged health costs. We will refer to this as (logarithmically) "even grid" quadrature.[27] In the exercise below, we let x_i lie in $[-2, 4.2]$ and set ν to 0.75.[28]

In order to evaluate different integration techniques, we specialize the value function as

$$V(A - hc) = \frac{1}{1-\rho}[max\{A - hc, A_{min}\}]^{1-\rho}, \quad \rho \geq 0, \ \rho \neq 1, \tag{25}$$

$$= \ln(max\{A - hc, A_{min}\}), \quad \rho = 1, \tag{26}$$

which combines a CRRA function with an asset floor given by A_{min}.[29] We measure the importance of health cost uncertainty by computing the equivalent differential (EQD), the decrease in assets that would leave a consumer facing no uncertainty no better off than a consumer facing uncertainty:

$$V(A_{CE} - EQD) \equiv E(V(max\{A - hc, A_{min}\})) \tag{27}$$

where $A_{CE} = \frac{1}{N}\sum_n max\{A - hc_n, A_{min}\}$ denotes the average post-health-cost asset balance.

Figure 4 plots the equivalent differentials generated under each of the integration methods, using the empirical and fitted distributions for the entire health cost sample. In Figure 4, the asset floor is fixed at \$25,000, ρ is set to 2 or 5, and the number of nodes is fixed at 14.[30] What varies is the level of pre-health-cost assets, A; Figure 4 plots the functions $EQD_N(A)$, $EQD_{GH}(A)$, and $EQD_{EG}(A)$, the equivalent differentials using the empirical distribution, Gauss-Hermite Quadrature, and the even grid, respectively.

When $\rho = 2$, so that the value function is not too curved, Gauss-Hermite quadrature better matches the data integral. Note, however, when the even grid quadrature fails, the

[27] It is also worth noting that with certain combinations of nodes and intervals, even grid quadrature can be quite similar to Gauss-Hermite quadrature.

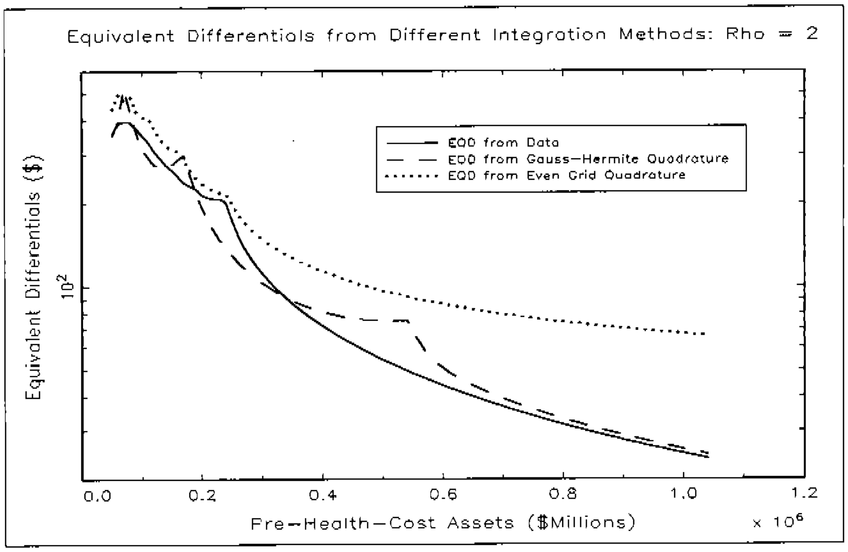
[28] The motivation for the upper bound—beyond performance—is that in the data we use below, the largest observed value of logged health costs lies roughly 4.2 standard deviations above the mean.

[29] As a practical matter, this function is usually not well-defined for $A_{min} \leq 0$; a more generous floor could be attributed to social insurance.

[30] The correct value of A_{min} is not obvious. While the government-provided consumption floor is well below \$25,000, it is not clear how consumption floors translate into an asset floor in such a stylized setting.

error in the equivalent differential is less than \$100. When $\rho = 5$, and the curvature is more severe, even grid quadrature better matches the data. The reason why is that when $\rho = 5$ the equivalent differential generated by the data integral, EQD_N, is driven by the largest observed health expenditure, hc_{max}. While hc_{max} in the data is of catastrophic size (roughly \$200,000), it is nonetheless finite, and the risk effects it generates eventually shrink as pre-health-cost assets, A, continue to grow. In particular, the second panel of Figure 4 shows that once A reaches \$250,000, EQD_N declines steadily.[31] Since the largest node of the even grid generates a health cost roughly equivalent to hc_{max}, the even grid function EQD_{EG} has a similar shape. But the largest two nodes under 14-point Gauss-Hermite quadrature generate bigger expenses, and thus lead to higher equivalent differentials for agents with high asset levels.

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



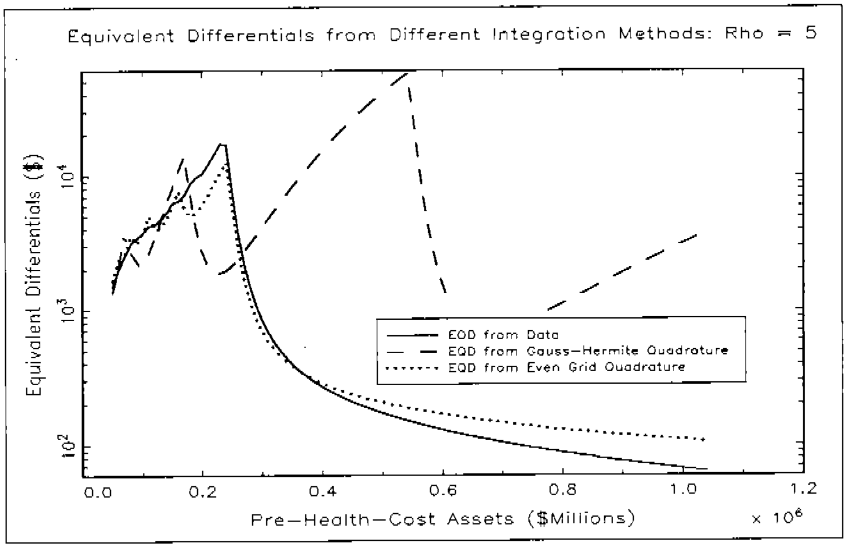


Figure 4: Equivalent Differentials from Different Integration Methods

The interpretation of these results depends on whether one really believes that health costs have no upper bound. Figure 5 shows the equivalent differentials generated by even grid quadrature with an interval of $[-2, 6]$ and 500 nodes. Such an interval allows for health costs well in excess of \$1 million.[32] Note that as A grows, the maximum effective health cost shock, $A - A_{min}$, and thus the total amount of health cost risk, grows as well. On the other hand, higher values of A imply lower levels of absolute risk aversion. Figure 5 shows that when $\rho = 2$, the latter effect dominates, and EQD_{EG} declines in the same manner as EQD_N. When $\rho = 2$, the question of whether health costs are bounded is unimportant. But when $\rho = 5$, the two effects offset, and EQD_{EG} is flat. In this case, a finite sample with a finite largest element may greatly understate the degree of health cost risk that wealthy agents actually face.

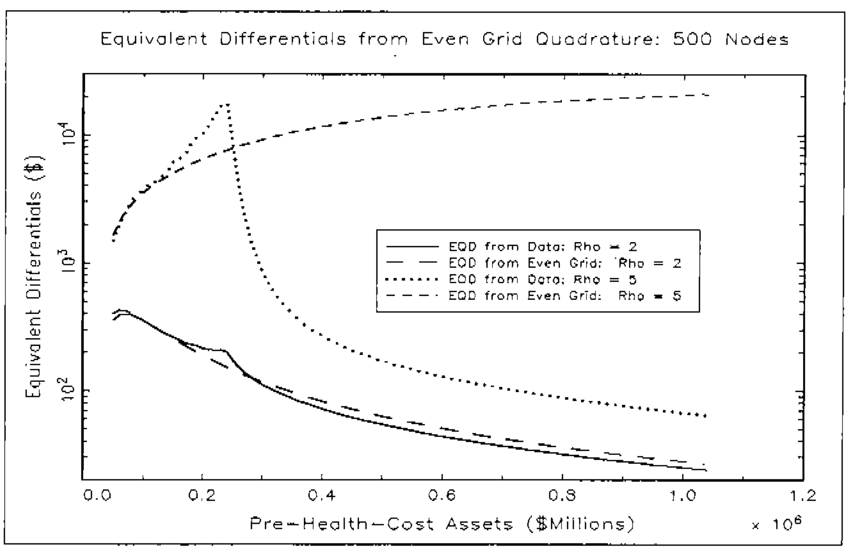


Figure 5: EQUIVALENT DIFFERENTIALS FROM EVEN GRID QUADRATURE

[32]The first line of Table 3 shows that the fitted lognormal distribution for the entire sample has a mean of 7.05 and a standard deviation of 1.25.

If finite samples understate the degree of health cost risk, the version of even grid quadrature with a maximum possible health cost shock of $200,000, as shown in Figure 4, is inadequate as well. Ideally, we would like to account for even more severe health cost shocks. As discussed above, the upper nodes in 14-node Gauss-Hermite quadrature do correspond to extremely high health costs. Observing the lower panel of Figure 4, however, shows that when $\rho = 5$ the equivalent differential generated by Gauss-Hermite quadrature is erratic. This erratic behavior is casued by the small number of evaluations of high health costs. 500-node quadrature, on the other hand, is clearly impractical. Figure 6 presents another alternative, even grid quadrature with an interval of $[-2, 6]$ and 14 nodes. When Figure 5 provides the benchmark, this variant of even grid quadrature is less successful than Gauss-Hermite quadrature in handling low levels of risk-aversion, but much more successful when $\rho = 5$.[33] With $\nu = 0.75$, the even grid nodes are more concentrated at the upper end than the Gauss-Hermite nodes. When $\rho = 5$, and agents are more risk-averse, the additional emphasis on large health costs leads to a better approximation.

On the other hand, if one believes that the largest cost in our 37,000-observation sample estimates the upper bound of health cost distribution, the even grid methods illustrated in Figures 5 and 6 are inappropriate. To better reflect the upper bound, one can return to the even grid used in Figure 4; the upper bound of that grid closely matches the upper bound of the data. It turns out, however, that the largest node under 8-node Gauss-Hermite quadrature matches the data maximum as well. Not surprisingly, 8-node Gauss-Hermite quadrature approximates the data integral very well. This is shown in Figure 7.

In short, when consumers are risk-averse, the most successful quadrature methods are the ones that best reflect the extreme upper tail of the cost distribution. If this upper tail has a finite bound, the most accurate quadrature methods are the ones that match this bound with their highest nodes. In the case at hand, 8-node Gauss-Hermite quadrature matched the data integral—which by definition has a finite upper bound—very well. This implies that

[33] It is also worth noting that even though the approximation error for $\rho = 2$ is quite large in relative terms, in absolute terms it is around $100, compared to average health costs of around $2,000.

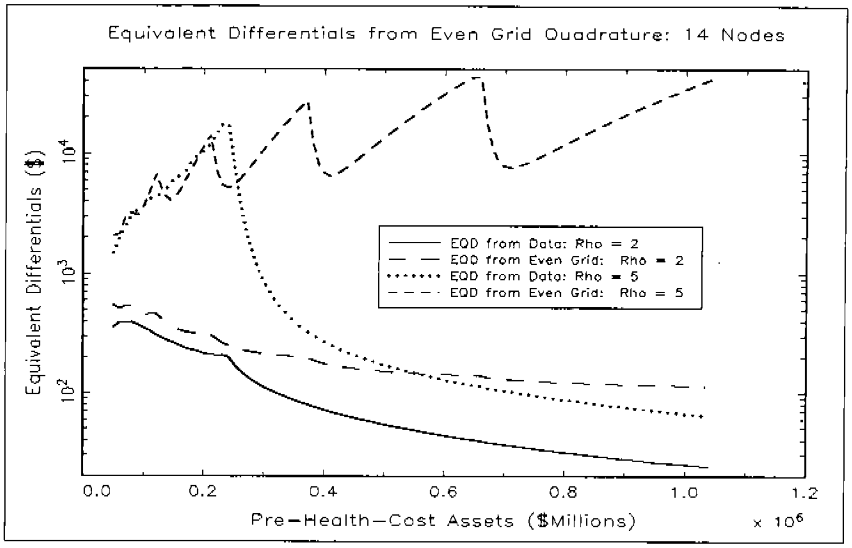


Figure 6: Equivalent Differentials from Even Grid Quadrature

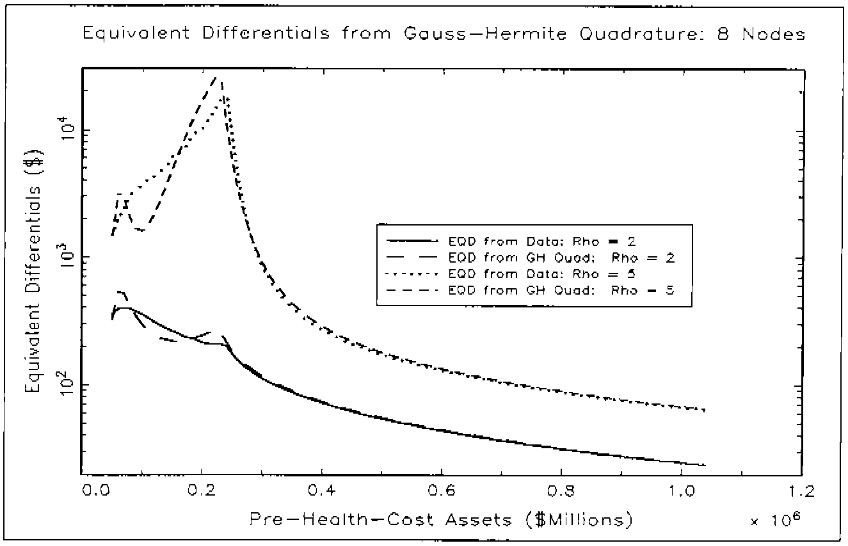


Figure 7: Equivalent Differentials from Gauss-Hermite Quadrature

not only that our fitted lognormal distribution works well, but that low-order quadrature can be accurate. In fact, 8-node Gauss-Hermite quadrature substantially outperformed 14-node Gauss-Hermite quadrature. This may seem paradoxical, as more nodes typically imply better accuracy, but the highest node in the 8-node scheme matches the upper bound of the data, while the highest node in the 14-node scheme exceeds the upper bound. On the other hand, if there is no upper bound, the most accurate quadrature methods are the ones that concentrate their nodes at extremely high values. In our exercise, the best method appears to a variant of 14-node even grid quadrature.

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 costs. When matching both the mean and the 99.5 percentile of the empirical health cost distribution to a lognormal distribution, we find that we match the upper 20% of the distribution very well. Matching the mean and variance of health costs potentially leads to erroneous inference about the right tail of the distribution of health costs. Similar to Feenberg and Skinner (1994), we find that the data generating process for health costs is well represented by an ARMA(1,1). Contrary to Feenberg and Skinner, however, we find that most of the cross-sectional variation in health costs comes from transitory variation. A $1 health cost shock today only leads to an additional $.40 in future health costs. Nevertheless, in any given year .1% of our sample suffers from a health cost shock that costs at least $80,000 over one's lifetime. This estimate is larger than previous estimates. Lastly, we present evidence on the accuracy of numerical solutions when integrating over health costs.

Before concluding, we note four important caveats to our analysis. First, there are some non-trivial measurement problems with our data. It is possible that some of the transitory variation in health costs is merely measurement error. This might lead us to overstate the amount of variability in health 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 the variability in health costs. The other three problems are more conceptual. The second problem is that the quantity of health care services consumed is to some extent a choice. This means that low income, low wealth households can reduce health costs by reducing medical services consumed.[34]. Third, low income, low wealth households have access to Medicaid, making health care services very inexpensive. Lastly, those with high health costs often die shortly after their health cost shock. Because they die shortly after the health cost shock, people who suffer from massive health cost shocks face little risk of being financially destitute (Pauly, 1990). Most of these caveats suggest that our estimates tend to overstate the amount of health cost uncertainty that households face.

Appendix A: Distribution of Estimators

This appendix describes estimation procedures for the model described in Section 4. The procedure is the same as Abowd and Card (1989) or Pischke (1995), except the procedure 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 5. Defining T as the number of years of data, we have $L = T(T+1)/2$ moment conditions, which are the unique elements of the covariance matrix in Table 5. Define $m_{il}(\theta)$ as the contribution of individual i to moment condition $l \in \{1, ..., L\}$, where θ is the vector of parameters. This object is merely the difference between an individual contribution to a variance or covariance and the variance or covariance implied by the model in equations (12)-(14). For example, assuming that $u_{it} = \psi_{it}$ is homoskedastic white noise, an individual contribution to a variance is

$$m_{il}(\theta) = R_{it}^2 - (Var(a) + Var(\psi)). \tag{28}$$

Let N is the number of individuals observed in any wave, N_l is the number of observations in

[34]See Cutler and Zeckhauser (2000) for estimates of income elasticities

the lth moment condition, and $\frac{N_l}{N}$ is the share of all individuals ever observed that contribute to the lth moment condition. Moreover, let that share be constant as N grows.

Because the data are unbalanced, we treat observations that are not observed as missing moment contributions. Therefore, the sample moment condition corresponding to equation (28) is

$$m_{Nl} = (\theta)\frac{N_l}{N}\Big(\frac{1}{N_l}\sum_{i}^{N_t} R_{it}^2 - (Var(a) + Var(\psi))\Big). \tag{29}$$

Denoting $m_N(\theta)$ as the $L \times 1$ vector of all sample moment conditions, we minimize

$$Nm_N(\theta)'Wm_N(\theta). \tag{30}$$

Denoting $\hat{\theta}$ as the estimated vector of coefficients and θ_0 as the true vector of coefficients, the estimator has a sampling distribution of $\hat{\theta}$ is

$$\sqrt{N}(\hat{\theta} - \theta_0) \rightsquigarrow_D N(0, V), \tag{31}$$

$$V = (D'WD)^{-1}D'W\Phi WD(D'WD)^{-1}, \tag{32}$$

$$D = \frac{\partial m(\theta)}{\partial \theta}\,|_{\theta=\theta_0}, \tag{33}$$

and Φ is the fourth moment matrix of the data. We estimate D and Φ using their sample analogs. For example, if moment condition l represents a variance, $\Phi_{l,l} = \frac{1}{N}\sum_{i=1}^{N_l}\Big(R_{it}^2 - E[R_{it}^2]\Big)^2$ where $E[R_{it}^2]$ is estimated using $\frac{1}{N_l}\sum_{i=1}^{N_l} R_{it}^2$. Assuming that the model is properly specified, Newey (1985) shows that

$$Nm(\hat{\theta})Q^{-1}m(\hat{\theta}) \tag{34}$$

is distributed $\chi^2_{L-rank(D)}$ if the model is properly specified. Q^{-1} is the generalized inverse of Q and

$$Q = P\Phi P, P = I - D(D'WD)^{-1}D'W. \tag{35}$$

Again, to estimate Q, we replace the objects in equation (35) with their sample analogs.

We use two different weighting schemes in our analysis. First, we use an "optimal" weighting matrix, i.e. $W = \Phi^{-1}$. Assuming optimal weighting, $V = (D'WD)$ and $Q = \Phi$. Although the optimal weighting matrix is asymptotically efficient, it can be severely biased in the small sample. The fatter the tails of the empirical distribution, the more severe the bias (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 had only a small effect on the results.

Appendix B: Calibrating the Health Cost Process at a One Year Frequency

This appendix describes the procedure to calibrate $Var(a), Var(u)$, and ρ. Defining a_{1it}, u_{1it}, ρ_1 as the autoregressive component, white noise component, and autoregressive coefficient at a one year frequency and a_{2it}, u_{2it}, ρ_2 as the autoregressive component, white noise component, and autoregressive coefficient at a two year frequency, we choose $Var(a_1), Var(u_1)$, and ρ_1 to match the variance and the first two autocovariances of health costs, measured at a two year frequency:

$$Var\Big(\ln[e^{a_{1it}+u_{1it}} + e^{a_{1it+1}+u_{1it+1}}]\Big) = Var(a_2) + Var(u_2), \tag{36}$$

$$Cov\Big(\ln[e^{a_{1it}+u_{1it}} + e^{a_{1it+1}+u_{1it+1}}], \ln[e^{a_{1it+2}+u_{1it+2}} + e^{a_{1it+3}+u_{1it+3}}]\Big) = \rho_2 Var(a_2) \tag{37}$$

$$Cov\Big(\ln[e^{a_{1it}+u_{1it}} + e^{a_{1it+1}+u_{1it+1}}], \ln[e^{a_{1it+4}+u_{1it+4}} + e^{a_{1it+5}+u_{1it+5}}]\Big) = (\rho_2)^2 Var(a_2) \quad (38)$$

where $e^{a_{1it}+u_{1it}}$ is the annual health cost residual, $a_{1it} \sim N(0, Var(a_1)), u_{1it} \sim N(0, Var(u_1))$, and $a_{1it+k} = \rho^k a_{1it} + \sum_{j=0}^{k-1} \rho^j \epsilon_{1it-j}$. We compute the variances and covariances of the left hand side objects of equations (36)-(38) by simulation. We have three equations and three unknowns, so the parameters are exactly identified.

References

[1] **Abowd, J., and David Card,** "On the Covariance Structure of Earnings and Hours Changes," *Econometrica,* 57(2), 1989, 411-445.

[2] **Altonji, J., and L. Segal**, "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics,* 14(3), 1996, 353-366.

[3] **Baker, M.**, "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics,* , 15(2), 1997, 338-375.

[4] **Blau, D. and D. Gilleskie "A Dynamic Structural Model of Health Insurance and Retirement," mimeo, 2000.**

[5] **Creedy, J.**, "Pareto and the Distribution of Income," *Review of Income and Wealth,* 23(4), 1977, 405-411.

[6] **Cutler, D. and R. Zeckhauser**, "The Anatomy of Health Insurance," in A. Culyer and J. Newhouse, eds., *Handbook of Health Economics* , 2000, 563-643.

[7] **Eichner, M., McClellan, M., and D. Wise**, "Insurance or Self-Insurance? Variation, Persistence, and Individual Health Accounts," in D. Wise, ed., *Inquiries in the Economics of Aging* , 1998, 19-45.

[8] **Employee Benefit Research Institute**, *Health Benefits Databook*, EBRI-ERF, 1999.

[9] **Feenberg, D., and J. Skinner**, "The Risk and Duration of Catastrophic Health Care Expenditures," *The Review of Economics and Statistics* , 76(4), 1994, 633-647.

[10] **French, E., and J. Jones**"The Effects of Health Insurance and Self-Insurance on Retirement Behavior," manuscript, 2002.

[11] **French, E., and K. Kamboj** " The Effect of Health Insurance on Health Costs and Health Care Utilization of the Elderly," *Economic Perspectives,* forthcoming.

[12] **Hubbard, R., J. Skinner, and S. Zeldes,** "The Importance of Precautionary Motives in Explaining Individual and Aggregate Saving", *Carnegie-Rochester Series on Public Policy,* 1994, 59-125.

[13] **Hubbard, R., J. Skinner, and S. Zeldes,** "Precautionary Saving and Social Insurance", *Journal of Political Economy,* 103(2), 1995, 360-399.

[14] **Judd, K.,** *Numerical Methods in Economics,* MIT Press, 1998.

[15] **Kimball, M.,** "The Importance of Precautionary Motives in Explaining Individual and Aggregate Saving: A Comment", *Carnegie-Rochester Series on Public Policy,* 1994, 127-132.

[16] **National Center for Health Statistics,** *Health and Aging Chartbook* , Hyattsville, MD, 1999.

[17] **Newey, W.,** "Generalized Method of Moments Specification Testing", *Journal of Econometrics,* 29(3), 1985, 229-256.

[18] **Palumbo, M.** "Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle," *Review of Economic Studies,* 66(2), 1999, 395-421.

[19] **Pauly, M.** "The Rational Non-Purchase of Long Term-Care Insurance," *Journal of Political Economy,* 98(1), 1990, 153-168.

[20] **Pischke, J-S**, "Measurement Error and Earnings Dynamics: Some Estimates From the PSID Validation Study," *Journal of Business & Economics Statistics*, 13(3), 1995, 305-314.

[21] **Roos, N., E. Shapiro, and R. Tate**, "Does a Small Minority of Elderly Account for a Majority of Health Care Expenditures? A Sixteen Year Perspective," *Milbank Memorial Fund*, 67(3-4), 1989, 347-369.

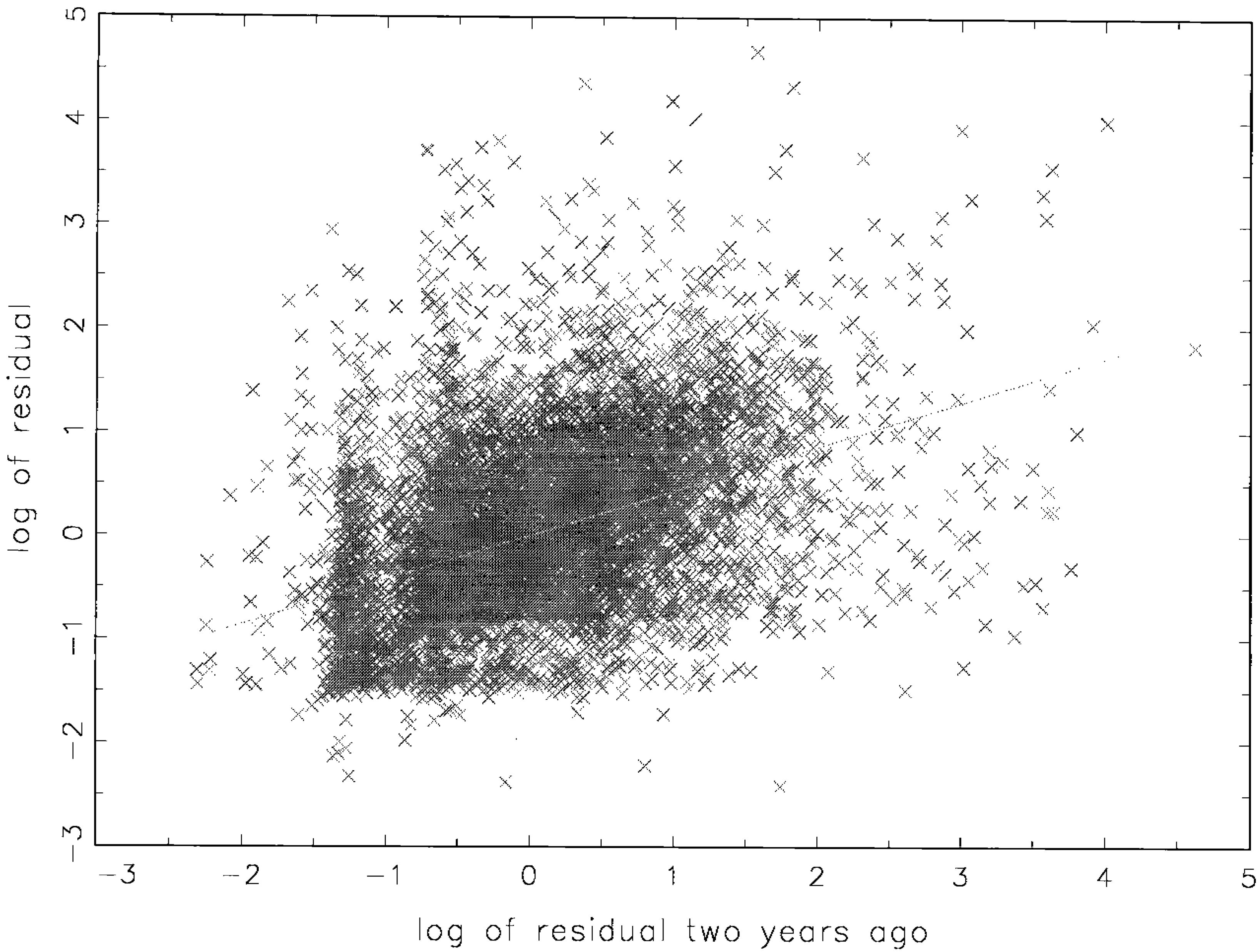
[22] **Rust, J. and C. Phelan,** "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica,* 65(4), 1997, 781-831.

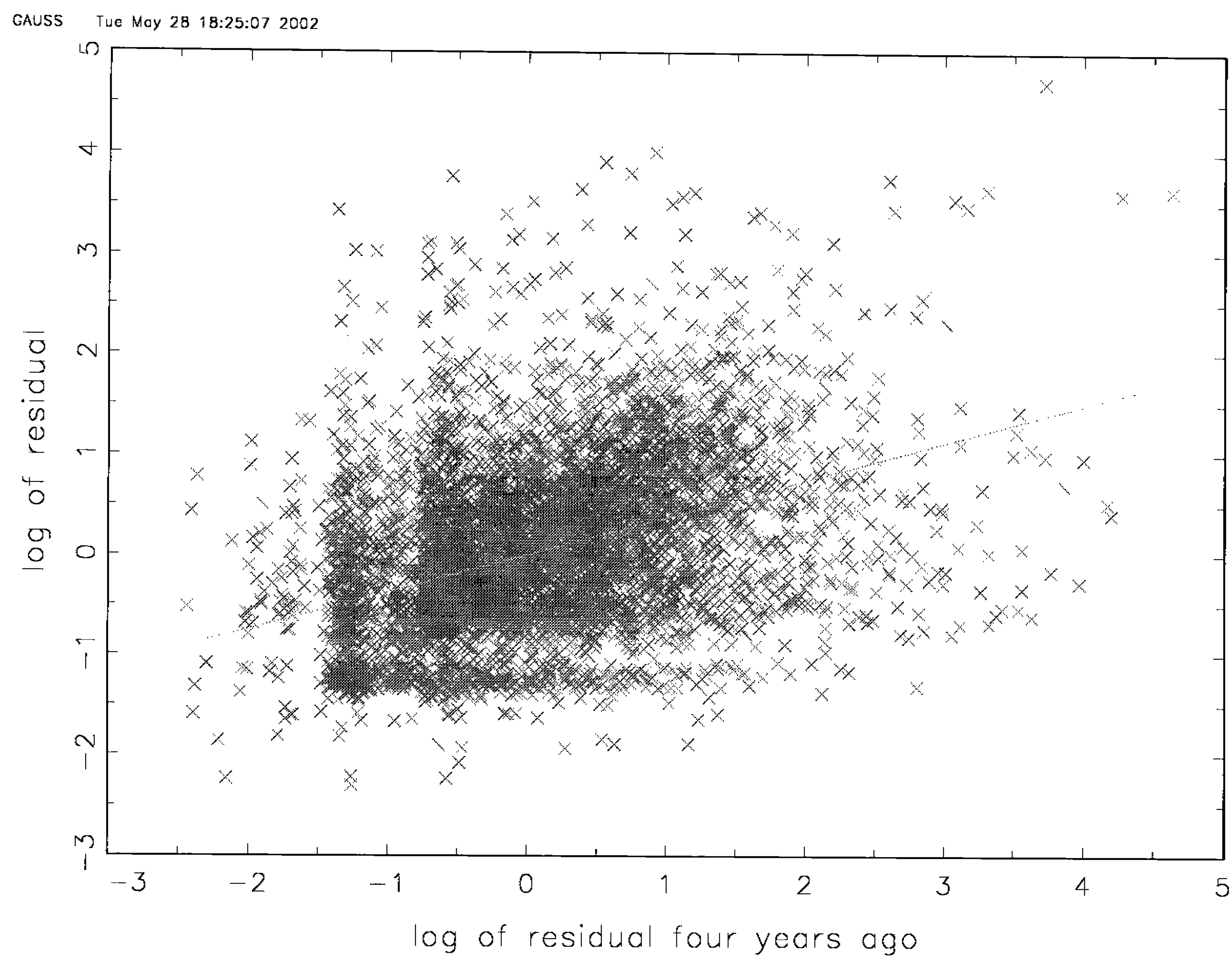
[23] **Selden, T., K. Levit, J. Cohen, S. Zuvekas, J. Moeller, D. McKusick, and R. Arnett,** "Reconciling Medical Expenditure Estimates from the MEPS and NHA, 1996," *Health Care Financing Review,* 23(1), 2001, 161-178.

[24] **United States,** *"Older Americans 2000: Key Indicators of Well-Being* , 2000.

[25] **Vuong, Q.,** "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica,* 57(2), 1989, 307-333.

GAUSS Tue May 28 18:24:58 2002




GAUSS Tue May 28 18:25:07 2002
log of residual
log of residual four years ago
-3
-2
-1
0
1
2
3
4
5

GAUSS Wed May 22 14:31:05 2002