MICHIGAN RETIREMENT AND DISABILITY RESEARCH CENTER
UNIVERSITY OF MICHIGAN

Promoting research on retirement, disability, and Social Security policy

On the Distribution and Dynamics of Medical Expenditure Among the Elderly

Karolos Arapakis, Eric French, John Bailey Jones, and Jeremy McCauley

MRDRC WP 2021-436

UM19-11

On the Distribution and Dynamics of Medical Expenditure Among the Elderly

Karolos Arapakis
University College London

John Bailey Jones
Federal Reserve Bank of Richmond

Eric French
University of Cambridge, University College London and Institute for Fiscal Studies

Jeremy McCauley
University of Bristol

November 2021

Michigan Retirement and Disability Research Center, University of Michigan, P.O. Box 1248. Ann Arbor, MI 48104, mrdrc.isr.umich.edu, (734) 615-0422

Acknowledgements

The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Retirement Research Consortium through the University of Michigan Retirement Research Center Award RDR18000002. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA, the Federal Reserve System, or any agency of the federal government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States government or any agency thereof.

On the Distribution and Dynamics of Medical Expenditure Among the Elderly

Abstract

Using data from the Health and Retirement Study linked to administrative Medicare and Medicaid records along with the Medical Expenditure Panel Survey, we estimate the stochastic process for total and out-of-pocket medical spending. By focusing on dynamics, we consider not only the risk of catastrophic expenses in a single year, but also the risk of moderate but persistent expenses that accumulate into a catastrophic lifetime cost. We assess the reduction in out-of-pocket medical spending risk provided by public insurance schemes such as Medicare or Medicaid. We find that although Medicare and Medicaid pay the majority of medical expenses, households at age 65 will incur, on average, $59,000 in out-of-pocket costs with 10% of households incurring more than $121,000 in out-of-pocket expenses over their remaining lives.

Citation

Arapakis, Karolos, Eric French, John Bailey Jones, and Jeremy McCauley. 2021. "On the Distribution and Dynamics of Medical Expenditure Among the Elderly." Ann Arbor, MI. University of Michigan Retirement and Disability Research Center (MRDRC) Working Paper; MRDRC WP 2021-436. https://mrdrc.isr.umich.edu/publications/papers/pdf/wp436.pdf

Authors' acknowledgements

We are very grateful to Naoki Aizawa, Irena Dushi, Bryan Tysinger, and Lauren Nicholas for helpful discussions and to Jack Light for sharing code. We are very grateful to Emily Emick, Rachel Rogers, Pauline Sow, and Yidi Wu for excellent research assistance. Funding from the Social Security Administration through the Michigan Retirement Research Center, the Economic and Social Research Council (grants (RES-544-28-50001) and (ES/P001831/1)) and Norface (grant TRISP 462-16-120) for this work is gratefully acknowledged.

M RDRC

1 Introduction

Despite nearly universal enrollment in Medicare, most elderly Americans still face the risk of catastrophic medical expenses. This is because Medicare does not pay for long hospital and nursing home stays and requires co-payments for many other treatments. Medicaid fills many of these gaps, but only for households that pass a means test. Medical spending is thus a major financial concern among elderly households. In a recent survey, more affluent individuals were worried about rising health care costs than about any other financial issue (Merrill Lynch Wealth Management 2012).

In this paper we document patterns of medical spending among older households, distinguishing between spending covered by public insurance programs such as Medicare or Medicaid and the out-of-pocket expenses borne by the households themselves. Even though numerous papers have estimated the medical spending risks that older Americans face in any given year, very few studies have estimated the distribution of cumulative lifetime spending. These lifetime totals, however, are critical when assessing the income and savings adequacy of older households. Households care not only about the risk of catastrophic expenses in a single year, but also about the risk of moderate but persistent expenses that accumulate into catastrophic lifetime costs. We use new data and methods to improve the measurement and assessment of this risk. In particular, we make three contributions.

First, to the best of our knowledge, this is the first paper to estimate the dynamic process for total spending by all payors among the population aged 65+. Our main dataset is the Health and Retirement Study (HRS) which has high quality information on out-of-pocket medical spending, linked to Medicare and Medicaid administrative data. To this we add data from the Medical Expenditure Panel Survey (MEPS), which allow us to impute private insurance and other payments not measured directly in the HRS. This yields an estimate of medical spending for HRS households that accounts for all payors, giving us the first long panel measures of comprehensive medical spending for the age 65+ population.

Second, we estimate the stochastic process for medical spending with a specification far more flexible than those used in previous studies. Specifically, we use the framework developed by Arellano, Blundell, and Bonhomme (2017), which allows for non-linear persistence and non-normal shocks. Using this specification allows us to more accurately predict the distribution of medical spending, both annual and cumulative, after age 65. This allows us to better understand the risks facing elderly households over the entirety of their remaining life.

Third, we model the share of medical spending paid for by Medicare and Medicaid, and calculate the extent to which Medicare and Medicaid reduce lifetime medical spending. Using detailed data and an advanced methodology allow us to better understand who benefits from Medicaid and Medicare.

In our framework medical spending depends on household structure and health, among other factors.

We estimate dynamic models of these variables. Simulating our estimated models allows us to construct household histories and compute the distribution of total lifetime medical spending. Thus we can calculate the share of people who face catastrophic medical spending over the course of their lives.

We find that lifetime medical spending during retirement is high and uncertain. Over their remaining lives, households at age 65 will incur, on average, \$272,000 in total medical spending, of which \$59,000 will be paid out of pocket. At the top tail, 10% of households will incur more than \$563,000 in total medical spending of which \$121,000 will be paid out-of-pocket. The level and the dispersion of remaining lifetime spending diminishes only slowly with age. For example, a household alive at age 90 will on average spend more than \$99,000 in total and \$21,000 out-of-pocket before they die. The reason for this is that as households age, surviving individuals on average have fewer remaining years of life, but are also more likely to live to extremely old age when medical spending is very high. Although initial health, and initial marital status have large effects on this spending, much of the dispersion in lifetime spending is due to events realized in later years. We find that Medicare and Medicaid cover a large amount of lifetime medical spending, substantially lowering the risk of catastrophic medical bills. However, on average, households still pay for 22% of all medical spending out of pocket.

The rest of the paper is organized as follows. Section 2 contains a literature review. In section 3 we discuss some key features of the data sets that we use in our analysis, the HRS and MEPS, and describe how we construct our measure of medical spending. In section 4 we introduce our model and describe our simulation methodology. We discuss our results in section 5 and conclude in section 6.

2 Previous Literature

The life-cycle implications of health and medical spending, especially at older ages, are an area of considerable interest. Several papers (De Nardi, French, and Jones 2010, Kopecky and Koreshkova 2014, Ameriks et al. 2020, De Nardi et al. 2021) show that health care costs that rise with age and income can explain much of the U.S. elderly's saving behavior. Other work suggests that medical spending risk is important in explaining cross-country differences in the consumption (Banks et al. 2019) and savings decisions (Nakajima and Telyukova 2018) of elderly households. Still other studies have considered the role of medical expenses in: bankruptcy (Livshits, MacGee, and Tertilt 2007); the adequacy of savings at retirement (Scholz, Seshadri, and Khitatrakun 2006, Skinner 2007); and annuitization (Lockwood 2012, Pashchenko 2013, Reichling and Smetters 2015).

Relative to the work exploring the consequences of medical spending risk, the number of papers attempting to measure this risk is limited. Most studies in this literature analyze out-of-pocket spending (e.g., Feenberg and Skinner 1994, French and Jones 2004, Fahle, McGarry, and Skinner 2016, Hurd, Michaud, and Rohwedder 2017); the studies analyzing total spending in the U.S. use data from individual insurance plans (Alemayehu and Warner 2004, Hirth et al. 2016) or short panels of two years or less (Pashchenko and Porapakkarm 2016).

Several studies have found that out-of-pocket medical spending shocks are reasonably well-described as the sum of a persistent AR(1) process and a white noise shock (Feenberg and Skinner 1994, French and Jones 2004, Jones et al. 2018, De Nardi et al. 2021). French and Jones (2004) 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. However, they also note that their procedure must be applied with care. They point out, for example, that estimates are sensitive to bottom coding decisions, in part because of left skewness in the data that the normal distribution does not capture. De Nardi, French, and Jones (2010, 2016) extend French and Jones's (2004) spending model to account for additional covariates, and embed the medical spending model in a model of lifecycle savings decisions, but consider only singles and do not control for end-of-life events.[1] Jones et al. (2018) and De Nardi et al. (2021) extend the model further to account for the risk of end-of-life spending and the joint spending of couples. Our paper differs from these studies in two ways. First, it analyzes spending by all payors, as opposed to just out-of-pocket spending. This allows us to assess the extent to which Medicare and Medicaid reduce the medical spending risk facing older households. Second, we replace the common specification, which assumes a Gaussian AR(1)-plus-white noise, with a more advanced semi-parametric framework developed by Arellano, Blundell, and Bonhomme (2017). This framework allows for non-Gaussian shocks and persistence that varies with age and the value of the shock.[2]

3 Data and Descriptive Statistics

We use HRS data matched with administrative Medicare and Medicaid records. We use MEPS data to impute payments for payors missing in the HRS. The result is a version of HRS medical spending data that is representative of all payors. We use data from 1999-2012, when we also have the Medicare and Medicaid records.

3.1 The HRS

The HRS is a representative biennial survey of the population ages 51 and older, and their spouses. To focus on the Medicare population, we focus on those ages 65 and older, using data starting in 1999.

[1] See French et al. (2006) and Poterba, Venti, and Wise (2017) on the importance of these events.

[2] Yet another model of out-of-pocket spending appears in Webb and Zhivan (2010). They estimate a rich model of stochastic morbidity and mortality with multiple health indicators and assume that medical expenditures are a function of these health conditions, along with a collection of socioeconomic indicators. In their framework, all of the variation in medical spending is due to variation in these controls; there are no residual shocks.

Although drawn from the non-institutionalized population when first interviewed, these individuals are tracked and reinterviewed as they enter nursing homes and other institutions. Consistency with our demographic model also leads us to drop a small number of households who, for example, are "partnered" or whose partner reports conflicting marital status. Furthermore, we drop households who do not consent to provide their Medicare and Medicaid information or there appear to be problems with the matching. This leaves us with 4,391 households comprising 39,002 household year observations. Appendix B documents our sample selection criteria.

The HRS conducts interviews every other year. Households are followed until both members die; attrition for other reasons is low. When the respondent for a household dies, in the next wave an "exit" interview with a knowledgeable party – usually another family member – is conducted. This allows the HRS to collect data on end-of-life medical conditions and spending (including burial costs). Fahle, McGarry, and Skinner (2016), compare the medical spending data from the "core" and exit interviews, show that out-of-pocket spending rises significantly in the last year of life.

The HRS has a variety of health indicators. We assign individuals to the nursing home state if they were in a nursing home at least 120 days since the last interview or if they spent at least 60 days in a nursing home before the next scheduled interview and died before that scheduled interview. We assign the remaining individuals a health status of "good" if their self-reported health is excellent, very good or good and a health status of "bad" if their self-reported health is fair or poor.

The HRS collects data on all out-of-pocket medical expenses, including private insurance premia and nursing home care. The HRS medical spending measure is backward-looking: medical spending at any wave is measured as total out-of-pocket spending over the preceding two years. French, Jones, and McCauley (2017) compare out-of-pocket medical spending data from the HRS, MEPS, and the Medicare Current Beneficiary Survey (MCBS). They find that the HRS matches up well with the MEPS for items that MEPS covers, but that the HRS is more comprehensive than the MEPS in terms of the items covered.

To control for socioeconomic status, we follow De Nardi et al. (2021) and construct a measure of lifetime earnings or "permanent income" (PI). We first find each household's "non-asset" income, a pension measure that includes Social Security benefits, defined benefit pension benefits, veterans benefits and annuities. Because there is a roughly monotonic relationship between lifetime earnings and these pension variables, post-retirement non-asset income is a good measure of lifetime permanent income. Since these income sources tend to change only in response to inflation and death of a family member, we assume this income is a deterministic function of a household specific effect (which is a measure of PI) and household composition. We then use fixed effects regression to convert non-asset income, which depends on age and household composition as well as lifetime earnings, to a scalar measure comparable across all households. In particular, we assume that the log of non-asset income for household i at age t follows

$$\ln y_{it} = \alpha_i + \kappa(t, f_{it}) + \omega_{it}, \tag{1}$$

where: α_i is a household-specific effect; $\kappa(t, f_{it})$ is a flexible function of age and family structure f_{it} (i.e., couple, single man, or single woman); and ω_{it} represents measurement error. The percentile ranks of the estimated fixed effects, $\widehat{\alpha}_i$, form our measure of permanent income, $\widehat{I}_i$.

3.2 Administrative Medicare and Medicaid Records

The Centers for Medicare and Medicaid Services (CMS) have confidential administrative spending records for Medicare (Parts A, B and D) and Medicaid, that we link to the survey responses of consenting HRS respondents. We have Medicare data for each year between 1991 and 2016. These records include reimbursement amounts for inpatient, skilled nursing facility, home health, and hospice claims made under Medicare Part A, as well as outpatient, carrier (non-institutional medical care providers such as individual or group practitioners, non-hospital labs, and ambulances), and durable-medical-equipment payments made under Medicare Part B. To this total we add drug-related spending made under Medicare Part D, which began in 2006.

As with the Medicare records, we link restricted Medicaid data for those who give permission, allowing us to measure Medicaid spending for each year between 1999 and 2012. The Medicaid files contain information on enrollment, service use, and spending. Appendix A describes the Medicare and Medicaid data in more detail. Linking these data to the HRS results in a broad set of spending measures for the years 1999-2012. These will be the years used in our main analysis.

3.3 Imputations Using MEPS Data

While the HRS contains accurate measures of out-of-pocket medical spending and can be linked to Medicare and Medicaid records, it does not contain information on the payments made by private and smaller public insurers (such as the Veterans Administration or state and local health departments). To circumvent this issue, we use data from the 1996-2017 waves of the Medical Expenditure Panel Survey (MEPS).

The MEPS is a nationally representative survey of non-institutionalized households. MEPS respondents are interviewed up to 5 times over a 2 year period, forming short panels. We aggregate the data to an annual level. MEPS respondents are asked about their (and their spouse's) health status, health insurance, and the health care expenditures paid out-of-pocket, by Medicaid, by Medicare, private insurance and by other sources. The survey responses are matched to medical spending information provided

Total Spending Percentiles	Total Spending Average Exp.	Total Spending Pct. of Total	Percent paid by Out-of-Pocket	Percent paid by Medicare	Percent paid by Medicaid	Percent paid by Other
All	24,000	100.0	21.6	64.0	10.7	3.7
95-100%	138,700	28.9	24.7	61.0	11.2	3.1
90-95%	74,200	15.5	17.9	63.9	16.2	2.0
70-90%	39,200	32.7	16.8	68.7	11.6	2.9
50-70%	16,000	13.4	23.3	64.7	6.0	6.1
0-50%	4,600	9.5	32.2	56.3	4.0	7.6

Notes: Total spending is the sum of annual household Medicare, Medicaid, out-of-pocket and other spending, age 65+. Other includes private insurance and other government payors. Expenditures are expressed in 2014 dollars.

Table 1: Medical Spending by Total Expenditures and Payor

by health care providers. Although the MEPS does not capture certain types of medical expenditures, such as nursing home expenditures, it captures sources of medical spending extremely well.[3]

To impute medical spending not captured in the HRS, we proceed in two steps. First, we use the MEPS data to regress these payments on a set of observable variables found in both data sets. Variables include household income, a fourth order age polynomial, labor force participation status, education, marital status, doctor and hospital visits, race indicators, health measures, out-of-pocket spending and interactions. This regression has an R^2 statistic of 0.13 for private insurance payments and 0.02 for other payors. Second, we impute these expenses in the HRS data using a conditional mean-matching procedure, a procedure very similar to hot-decking. Applying the MEPS regression coefficients to the HRS data yields predicted values for each HRS household, to which we add residuals drawn from MEPS households with similar levels of predicted spending. We describe our approach in more detail in Appendix C.

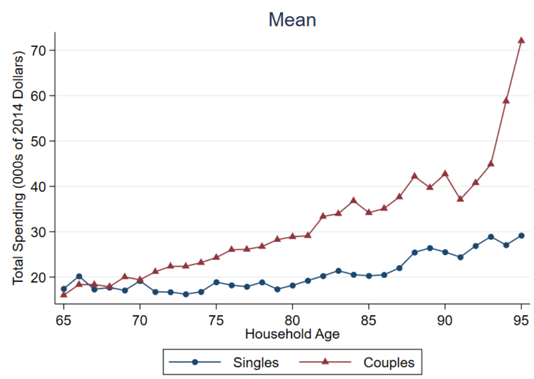
3.4 Our Medical Spending Measure

Table 1 summarizes the medical spending data contained in our HRS sample. The first two columns of the table show total spending by all payors for older households, sorted by total spending percentile, while the remaining 4 columns decompose this total by payor. Table 1 shows that Medicare and especially Medicaid benefits are concentrated among the top half of the total spending distribution, helping to ensure against catastrophic expenses. Another way to see this outcome is to note that at the bottom half of the spending distribution, about 32% of medical expenses are paid out of pocket, while at the top, out-of-pocket expenditures comprise between about 17% and 25% of the total.

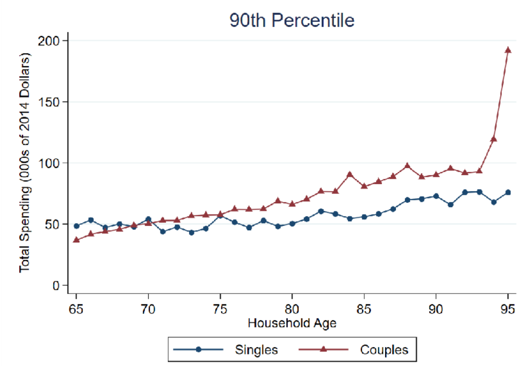
[3] Pashchenko and Porapakkarm (2016) compare MEPS data to the aggregate statistics and show that MEPS captures types of spending very well.

In Appendix D we compare our medical spending measures against the MCBS. French, Jones, and McCauley (2017) finds that the data on out-of-pocket and insurance premia in the two surveys match up well. We extend their exercise by comparing Medicare and Medicaid payments in the HRS restricted data to Medicare and Medicaid payments in the MCBS.

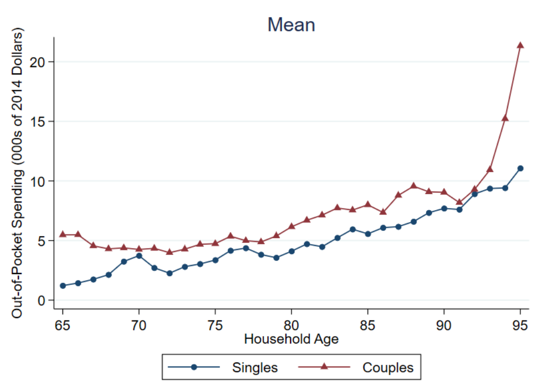
Figure 1 shows how the mean and 90^{th} percentile of both total and out-of-pocket medical spending change with the age of the household head. Medical spending rises rapidly with age for both singles and couples. As noted in Table 1, medical spending is very concentrated. Thus, it should come as no surprise that the 90^{th} percentile is significantly higher than the mean. Spending for couples is higher than for singles, but in most cases is much less than double the spending of singles. It bears noting that as households age, they tend to transition from couples to singles; at older ages, the number of couples is small, leading to erratic profiles.



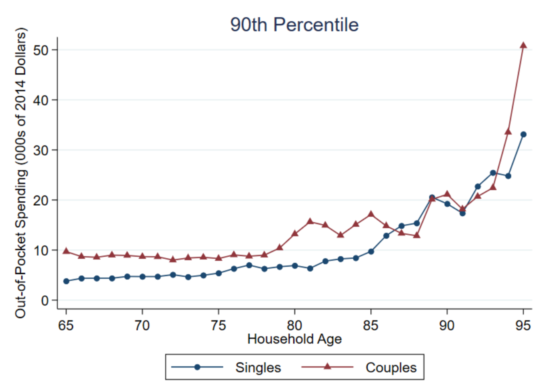
(a) Total Spending: Means



(b) Total Spending: 90^{th} Percentile



(c) Out-of-Pocket: Means



(d) Out-of-Pocket: 90^{th} Percentile

Figure 1: Mean and 90^{th} Percentile of Household Medical Spending by Age

4 The Model

We estimate the distribution of lifetime medical spending in five steps. In the first step we estimate the log of total medical spending as a function of age, health, family structure and PI, using Ordinary Least Squares (OLS). In the second step, we estimate the stochastic process for the unexplained component of medical spending – the residuals from the first step regression – using the methodology developed by Arellano, Blundell, and Bonhomme (2017). In the third step, we estimate the mapping from total medical spending to out-of-pocket medical spending. In the fourth step we estimate a Markov Chain model of health and mortality. In the final step, we use the estimated models to simulate health, mortality, and lifetime medical spending.

4.1 Total Medical Spending

4.1.1 Underlying framework

Let $M_{i,t}$ denote total medical spending for household i at time t ,and let $m_{i,t}$ denote its logarithm net of the observed variables contained in the vector $X_{i,t}$:

$$\ln M_{i,t} = X_{i,t}\gamma + m_{i,t}. \tag{2}$$

We assume that $m_{i,t}$ can be expressed as the sum of the persistent first-order Markov component $\eta_{i,t}$ and the transitory component $\varepsilon_{i,t}$:

$$m_{i,t} = \eta_{i,t} + \varepsilon_{i,t}, \quad \forall i \in \{1, ..., N\}, \forall t \in \{1, ..., T\}. \tag{3}$$

We assume the transitory component is i.i.d., but require only that it be zero-mean and satisfy the regularity conditions set forth in Arellano, Blundell, and Bonhomme (2017). In practice the vector $X_{i,t}$ includes an age polynomial, health indicators, household structure and death-year indicators, PI percentile, and interactions among the aforementioned variables.

4.1.2 A common specification

Many previous studies, following Feenberg and Skinner (1994) and French and Jones (2004), have used the following specification for the shocks:

$$\eta_{i,t} = \phi\eta_{i,t-1} + \zeta_{i,t}, \tag{4}$$

$$\eta_{i,init} \overset{iid}{\sim} N\Big(0, \sigma^2_{\eta_{init}}\Big), \zeta_{i,t} \overset{iid}{\sim} N\Big(0, \sigma^2_{\zeta}\Big), \varepsilon_{i,t} \overset{iid}{\sim} N(0, \sigma^2_{\varepsilon}). \tag{5}$$

Thus, the persistent component $\eta_{i,t}$ is an AR(1) with the innovation $\zeta_{i,t}$ independent of $\eta_{i,t-1}$, whereas the transitory component is homoskedastic white noise. We will refer to this specification as the "standard model".

Equations (4) - (5) impose three types of restrictions.

1. Linearity of the process for the persistent component. In this context linearity means that the conditional expectation of $\eta_{i,t-1}$ is linear in its lagged value $\eta_{i,t-1}$. Moreover, it is straightforward to show that the expectation of $m_{i,t}$ given $m_{i,t-1}$ is linear, $m_{i,t} = \delta m_{i,t-1} + e_{i,t}$, with $E(e_{i,t} \mid m_{i,t-1}) = 0$.

2. Stationarity of the distributions of $\zeta_{i,t}$, $\varepsilon_{i,t}$ and $e_{i,t}$. The distribution of $\zeta_{i,t}$, $\varepsilon_{i,t}$, is unconditionally stationary, and the distribution of $e_{i,t}$ is stationary conditional on t.

3. Normality, not only of the shocks $\zeta_{i,t}$ and $\varepsilon_{i,t}$, but also of the forecast error $e_{i,t}$.

See Appendix E for details.

We assess the implications of the standard model in Figure 2, following the graphical analysis of earnings found in De Nardi, Fella, and Paz-Pardo (2020). Figure 2 shows moments of the medical spending residuals conditional on their lagged percentile ranks. To construct the residual $m_{i,t}$, we first regress the log of medical spending on age, household composition, and health variables as in equation (2). We next regress each household's time-t residual on its lagged value, estimating $m_{i,t} = \delta m_{i,t-1} + e_{i,t}$ and calculating the "second-order" residual $e_{i,t}$.

Figure 2 plots the mean, standard deviation, skewness and kurtosis of $e_{i,t}$ as a function of the percentile rank of $m_{i,t-1}$. The top left panel plots means. This graph shows that at the 95th percentile of lagged medical spending, $m_{i,t-1}$, the current value of $e_{i,t}$ is on average -0.1. Put differently, if last period's medical spending is high, this period's value is also high, but it is 10% lower than what would be implied by a linear regression on lagged medical spending. The data thus reject the first restriction, linearity, of the standard model.

The top right panel of Figure 2 plots the standard deviation of medical spending. This is U-shaped, declining until the 40th percentile of last period's spending and increasing after the 60th percentile. In other words, dispersion in medical spending is highest for those with low or high medical spending. The data thus reject the prediction that the standard deviation will be constant in $m_{i,t-1}$.

The bottom left panel shows skewness. Normality implies that skewness is zero across the distribution, but in the data skewness is positive for low values of spending and is below -0.5 for high values. This means for people with high medical spending last period, this period's medical spending tends to fall a lot when it falls. The bottom right panel shows kurtosis: normality would imply that this is three, but in the data it is often higher, implying a distribution with tails fatter than those found in the normal distribution. In short, Figure 2 shows that data rejects all three predictions of the standard model.

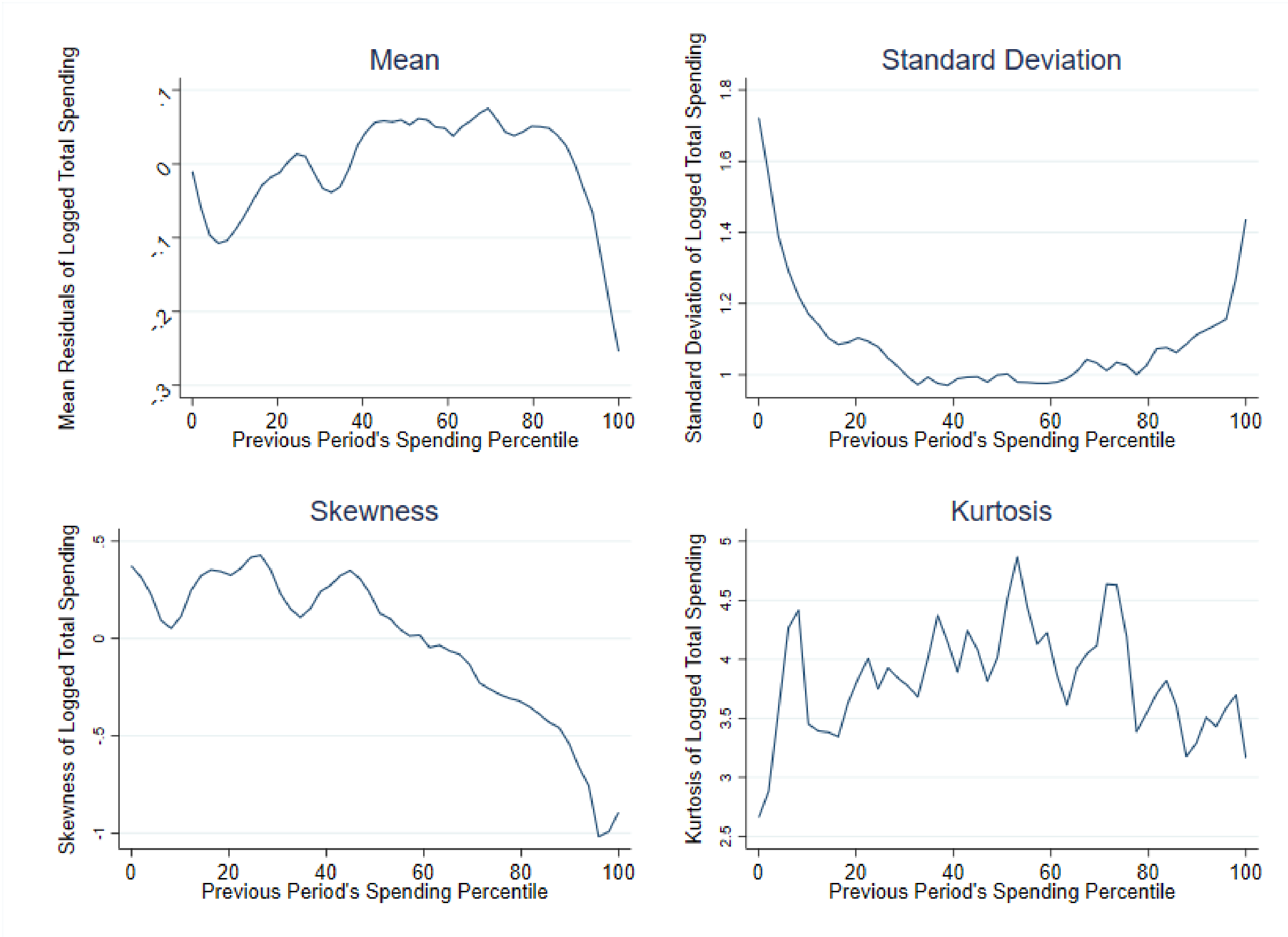


Figure 2: Moments of the Medical Spending Distribution

4.1.3 The Arellano, Blundell, and Bonhomme (2017) estimator

Our finding that the data are strongly at odds with the standard model leads us to seek a more general framework that relaxes the previous three restrictions yet fits within our two-component structure. Thus we use the quantile-based panel data model proposed by Arellano, Blundell, and Bonhomme (2017) and extended by Arellano et al. (2021).

To apply Arellano, Blundell, and Bonhomme's (2017) framework, rewrite the conditional distribution for the persistent component $\eta_{i,t}$ as:

$$\eta_{i,t} = Q_\eta(\nu_{i,t} \,|\, \eta_{i,t-1}, a_{i,t}), \quad \nu_{i,t} \overset{iid}{\sim} U[0,1], \tag{6}$$

where $Q_\eta(\nu \,|\, \eta_{i,t-1}, a_{i,t})$ denotes the νth quantile of $\eta_{i,t}$ conditional on its lagged value and age ($a_{i,t}$). The quantile function Q_η maps $\eta_{i,t}$'s conditional rank, $\nu_{i,t}$, into a value of $\eta_{i,t}$ itself. To fix ideas, if we draw $\nu_{i,t} = 0.1$, the realized value of $\eta_{i,t}$ will equal the 10th percentile of the conditional distribution of $\eta_{i,t}$ at age $a_{i,t}$, given $\eta_{i,t-1}$. As a rank, $\nu_{i,t}$ is distributed uniformly over the $[0,1]$ interval.

In the standard model, it follows from equation (4) that the quantile function takes the linearly separable form $\eta_{i,t} = Q_\eta(\nu_{i,t} \,|\, \eta_{i,t-1}, a_{i,t}) = \phi\eta_{i,t-1} + \sigma_\zeta \Phi^{-1}(\nu_{i,t})$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function and σ_ζ is the standard deviation of $\zeta_{i,t}$. (Conversely, with $\nu_{i,t} \sim U[0,1]$, we have $\sigma_\zeta \Phi^{-1}(\nu_{i,t}) \sim N(0, \sigma_\zeta^2)$). Age-independence, normality, and linearity can thus be expressed as restrictions on the quantile function in equation (6).

In its most unrestricted form, this specification allows for a great degree of flexibility. One way to see this is to construct the persistence measure

$$\phi_\tau(\eta_{i,t-1}, a_{i,t}) = \frac{\partial Q_\eta(\tau \,|\, \eta_{i,t-1}, a_{i,t})}{\partial \eta_{i,t-1}}, \tag{7}$$

with τ denoting the conditional rank of interest. $\phi_\tau(\eta_{i,t-1}, a_{i,t})$ measures the effect of $\eta_{i,t-1}$ on the τth conditional quantile of $\eta_{i,t}$. Persistence can vary by rank (τ), age ($a_{i,t}$) and prior realization ($\eta_{i,t-1}$). In contrast, in the standard model persistence always equals the constant ϕ.

In estimation we parametrically approximate the conditional quantile function by low-order Hermite polynomials. Let $h_k^\eta(\cdot)$ denote the kth Hermite polynomial used in the approximation of $\eta_{i,t}$, with $\{h_k^\eta(\cdot)\}_{k=0}^{K_\eta}$ forming the polynomial basis for the approximation. $Q_\eta(\tau \,|\, \eta_{i,t-1}, a_{i,t})$ is thus a linear combination of the K_η Hermite polynomials, with the coefficients on the polynomials, $\{\beta_k^\eta(\tau)\}_{k=0}^{K_\eta}$ themselves functions of the quantile rank τ. We thus have

$$Q_\eta(\tau \,|\, \eta_{i,t-1}, a_{i,t}) = \sum_{k=0}^{K_\eta} \beta_k^\eta(\tau) \cdot h_k^\eta(\eta_{i,t-1}, a_{i,t}), \quad \tau \in (0,1]). \tag{8}$$

The distributions of the initial shock $\eta_{i,1}$ and the transitory shocks $\{\varepsilon_{i,t}\}_t$ are handled in ways analogous to how we handle the persistent component:

$$Q_1(\tau \,|\, a_{i,1}) = \sum_{k=0}^{K_1} \beta_k^1(\tau) \cdot h_k^1(a_{i,1}), \quad \tau \in (0,1), \tag{9}$$

$$Q_\varepsilon(\tau \,|\, , a_{i,t}) = \sum_{k=0}^{K_\varepsilon} \beta_k^\varepsilon(\tau) \cdot h_k^\varepsilon(a_{i,t}), \quad \tau \in (0,1). \tag{10}$$

For these distributions we do not condition on $\eta_{i,t-1}$ but only only on age.

Each of the coefficient functions $\left(\{\beta_k^\eta(\tau)\}_{k=0}^{K_\eta}, \{\beta_k^1(\tau)\}_{k=0}^{K_1}, \{\beta_k^\varepsilon(\tau)\}_{k=0}^{K_\varepsilon}\right)$ in equations (8)-(10) is modelled with a set of polynomial splines defined over the intervals $\{[\tau_{\ell-1}, \tau_\ell]\}_{\ell=1}^{L}$, along with two low-dimensional tail functions defined over $(0, \tau_1]$ and $[\tau_L, 1)$. It is the parameters for these weighting functions that we must estimate.

As both the persistent and transitory shocks are unobserved, we cannot estimate the parameters of the weighting functions directly using quantile regressions. Furthermore, our data are unbalanced because sample members die. We therefore follow the extension of the *E-M* algorithm described in and applied by Arellano et al. (2021).

- In the *E-step* we find the posterior distribution of the unobserved persistent shocks ($\{\eta_{i,t}\}_t$) implied by the data and the current parameterization of the model. In particular, we use the coefficients of the Hermite polynomials $\left(\{\beta_k^{\eta}(\tau)\}_{k=0}^{K_\eta}, \{\beta_k^{1}(\tau)\}_{k=0}^{K_1}, \{\beta_k^{\varepsilon}(\tau)\}_{k=0}^{K_\varepsilon}\right)$, which fully determine the distributions of the shocks, and a Monte Carlo method to simulate draws from the distributions of the initial shock $\eta_{i,1}$ and the subsequent shocks $\{\eta_{i,t}\}_t$.[4] This part of the procedure is a special case of the Sequential Monte Carlo methods described in greater detail in Creal (2012).

- In the *M-step* we use quantile regressions to update the coefficient functions for the Hermite polynomials, using the distribution of $\{\eta_{i,t}\}_t$ found in the *E-step*. Once the coefficients have been updated, we return to the *E-step* and simulate new draws.

We iterate between the *E* and *M* steps until the parameters converge. See Appendix F for a more detailed description of the methodology.

4.2 Health and Mortality

Let $hs_{i,g,t}$ denote the health of member $g \in \{h, w\}$ in household i at age t. Health has four mutually exclusive possible values: dead; in a nursing home; in bad health; or in good health. We assume that the transition probabilities for an individual's health depend on his or her current health, age, permanent income I, and gender g.[5] It follows that the elements of the health transition matrix are given by

$$\pi_{q,r}(t, I_i, g) = \Pr\left(hs_{i,g,t+1} = r \quad hs_{i,g,t} = q; t, I_i, g\right), \tag{11}$$

with the transitions covering a one-year interval. Although the HRS interviews every other year, we adopt the approach in De Nardi, French, and Jones (2016), who fit annual models of health to the HRS data for

[4]This approach takes advantage of the Markovian structure of the model and has been shown to perform well in low-dimensional models.

[5]We do not allow health transitions to depend on medical spending. The empirical evidence on whether medical spending improves health, especially at older ages, is surprisingly mixed (De Nardi, French, and Jones 2016). Likely culprits include reverse causality – sick people have higher expenditures – and a lack of insurance variation – almost every retiree gets baseline insurance through Medicare. We also do not allow health transitions to depend on marital status. De Nardi et al. (2021) find that after controlling for income and past health, marital status has little added predictive power.

singles. We extend their approach to account for the dynamics of two-person households. We estimate health/mortality transition probabilities by fitting the transitions observed in the HRS to a multinomial logit model.[6] See Appendix G for further details.

5 Results

5.1 Health and Longevity

Table 2 shows the life expectancies implied by our demographic model for those still alive at age 65. The first panel of the table shows the life expectancies for singles under different configurations of gender, PI percentile, and age-65 health. The healthy live longer than the sick, the rich (higher PI) live longer than the poor, and women live longer than men. For example, a single man at the 10th PI percentile in a nursing home expects to live only 4.14 more years, while a single woman at the 90th percentile in good health expects to live 22.5 more years. The second panel of the table shows life expectancies for married *households*, that is, the average length of time that at least one member of the household is still alive or, equivalently, the life expectancies for the oldest survivors. While wives generally outlive husbands, a non-trivial fraction of the oldest survivors are men, and the life expectancy for a married household is roughly three years longer than that of a married woman with same initial health and PI quantile.

The results shown in Table 2 are disaggregated by PI and initial health. When we average over all these factors, we find that a man alive at age 65 will on average live an additional 17.50 years, while a woman alive at age 65 will on average live an additional 20.96 years.[7]

Another key statistic for our analysis is the probability that a 65 year old will spend significant time (a stay of more than 120 days) in a nursing home before he or she dies. We estimate this probability to be 24.4% for men and 36.9% for women. Nursing home incidence differs relatively modestly across the PI distribution. Although high-income people are less likely to be in a nursing home at any given age, they live longer, and older individuals are much more likely to be in a nursing home. The nursing home risk also varies relatively little with initial health status (good or bad), for similar reasons.

Although all households in the HRS are initially non-institutionalized, the HRS does a good job of tracking individuals as they enter in nursing homes. French and Jones (2004) show that the HRS sample matches very well the aggregate statistics on the share of the elderly population in a nursing home by 1999 when the sample begins. In 1999 roughly 2.8% of men and 6.1% of women in the data had entered

[6] We do not control for cohort effects. Instead, our estimates are a combination of period (cross-sectional) and cohort probabilities. This may lead us to underestimate the lifespans expected by younger cohorts as they age. Nevertheless, lifespans have increased only modestly over the sample period. Accounting for cohort effects would have at most a modest effect on our estimates.

[7] We construct these distributions with bootstrap draws of people aged 63-67 in HRS.

Permanent Income Percentile	Men Nursing Home	Men Bad Health	Men Good Health	Women Nursing Home	Women Bad Health	Women Good Health	All*
Individuals							
10	4.14	15.25	17.49	7.32	18.70	20.22	18.06
50	4.89	16.99	19.37	8.71	20.52	22.04	20.68
90	4.91	17.35	19.86	8.83	20.97	22.52	20.96
Couples (oldest survivors)†							
10	9.35	21.95	23.15				22.50
50	11.27	24.04	25.17				24.98
90	11.46	24.56	25.69				25.52
All Men							17.50
All Women							20.96
All Couples (oldest survivor)							24.83

* Averages taken over initial health found in the data. Results indexed by PI percentile are taken over the associated PI quintile.

† Health-specific results for couples based on the assumption that the spouses have the same health at age 65.

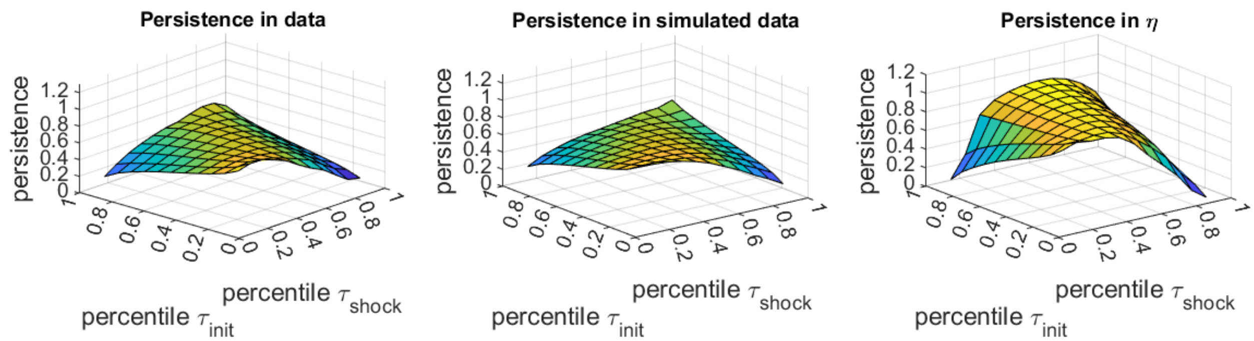
Table 2: Life expectancy in years, conditional on reaching age 65

nursing homes. We also understate the number of nursing home visits because we exclude short-term visits: as Friedberg et al. (2014) and Hurd, Michaud, and Rohwedder (2017) document, many nursing home stays last only a few weeks and are associated with lower expenses. We focus only on the longer and more expensive stays faced by households.

5.2 The Cross Sectional Distribution of Medical Spending and Parameter Estimates

The Persistence of Medical Spending: Panel (a) of Figure 3 shows the estimated persistence of medical spending in the data. Formally, it is the derivative of the conditional quantile function, defined in equation (7), averaged across age ($a_{i,t}$). It shows how persistent next period's medical spending is as a function of next period's shock (indexed by its percentile rank τ_{shock}) and the percentile rank of today's medical spending (τ_{init}).

Panel (a) shows that persistence is higher near the middle of the distribution. This means that those in the middle may face moderate but persistent expenses that can accumulate into a catastrophic lifetime cost. In contrast, persistence is quite low (less than 0.1) for large medical shocks among those with low initial medical spending. It is much higher, almost 0.6, for those with high initial spending and large shocks (the top point on the graph). What this means is that catastrophic medical spending is particularly persistent. Persistence is also higher for those with low initial spending and small shocks (the bottom



(a) HRS data (b) Model simulation (c) Persistent component

Figure 3: Persistence of medical spending.

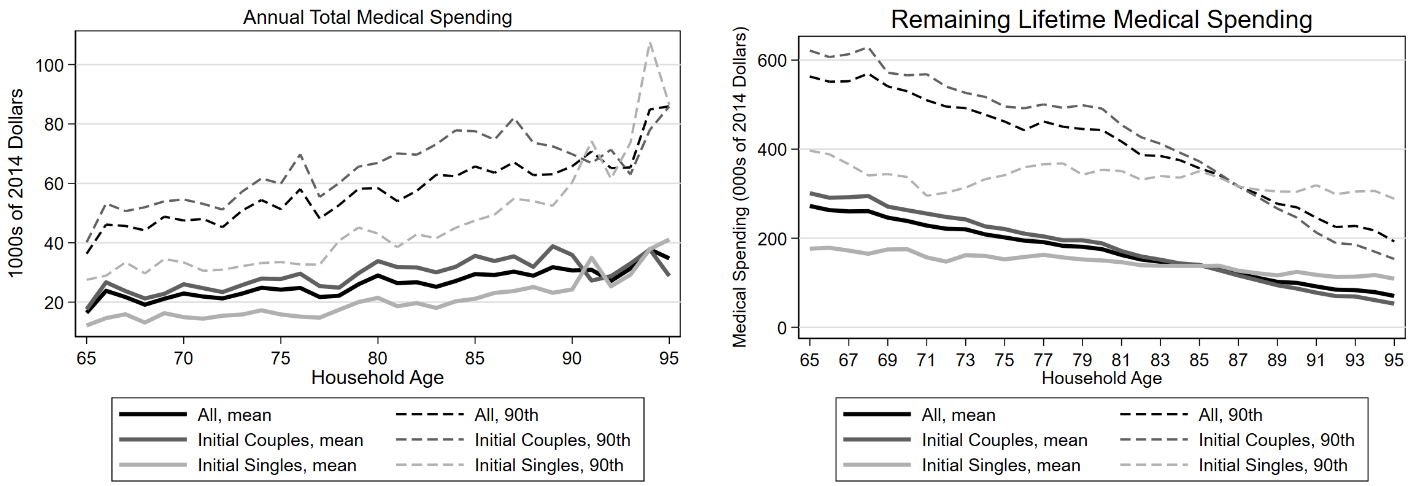
point on the graph). In short, when medical spending is low, it is likely to stay low, and when medical spending is high, it is likely to stay high. However, it follows from Figures 2 and 3 that when medical spending is high, it can fall a lot (as the conditional distribution is left-skewed when current spending is high) and when medical spending is low, it can rise a lot (as the conditional distribution is right-skewed when current spending is low). Recall that the AR(1)-plus-white noise specification often assumed in the literature implies that persistence would be constant. Panel (b) shows persistence in the simulated data. The similarity of Panels (a) and (b) shows that the model matches the persistence of medical spending observed in the data.

Panel (c) shows persistence in the permanent component of medical spending η. The persistence of η is much higher than for medical spending overall, as the latter includes the transitory shock ε. Over much of its distribution, the persistence of η is close to 1, indicating that for severe health conditions, such as dementia, medical spending is very persistent. It is these shocks, which cause high spending for many periods, that can drain a family's resources.

Budget sets: Table 1 shows that the mapping from total expenditures to the out-of-pocket expenditures paid by households is approximately linear. We therefore assume that everyone pays 21.6% of total costs out of pocket.

5.3 The Distribution of Lifetime Medical Spending

Panel (a) of Figure 4 shows the model-predicted mean and 90th percentile of annual health care spending for surviving households (including those who died during the year). Medical spending rises rapidly with age. For example, mean spending rises from \$22,900 per year at age 70 to \$34,700 at age 95. The upper tail rises even more rapidly, with the 90th percentile increasing from \$47,200 to \$85,200. These patterns are broadly similar to the raw data in Figure 1. One difference is that the medical spending gap between singles and couples is smaller at older ages than in Figure 1. This is because here we are conditioning



(a) (b)

Figure 4: Mean and 90^{th} Percentile of Annual (Panel a) and Remaining Lifetime (Panel b) Total Medical Spending for Surviving Households, Initial Singles and Initial Couples.

on initial household structure whereas Figure 1 conditions on current household structure. Virtually all initial couples have lost a spouse and are single if either household member is still alive at age 90. Thus it is unsurprising that medical spending of initial singles and initial couples is similar at this age. Here we focus on initial household structure so we can calculate the present value of lifetime spending.

Panel (b) of Figure 4 plots our main variable of interest, lifetime total spending. At each age, we calculate the present discounted value of remaining medical spending from that age forward, using an annual real discount rate of 3 percent. These lifetime totals are considerable. At age 65, households will incur, on average, over $272,000 of medical spending over the remainder of their lives. Those at the 90th percentile of the spending distribution will incur spending of $563,000. One might expect the lifetime totals to fall rapidly as households age and near the ends of their lives. This is not the case. A household alive at age 90 will on average spend more than $99,000 before they die. The slow decline of lifetime costs is due mostly to the tendency of medical costs to rise with age. Households that live to older ages have shorter remaining lives but higher annual spending rates.

A number of papers have considered whether most of the rise in medical spending with age is due rising mortality and thus end-of-life expenses: see the discussion in De Nardi et al. (2016). In our spending model we allow for for the jump in medical spending prior to death as well as age related medical spending.

5.4 Out-of-Pocket Medical Spending

Our baseline measure is total medical spending, which is the sum of payments made out-of-pocket, by Medicare, by Medicaid, and by other private and public payors. Figure 5 shows the distributions of out-of-pocket spending predicted by the model. To construct these values, we take the total medical spending amounts, then calculate the out-of-pocket amount using the budget sets discussed in Section 5.2. Parallel to Figure 4, Panel (a) of Figure 5 shows the distribution of annual out-of-pocket spending and Panel (b) shows remaining lifetime out-of-pocket medical spending.

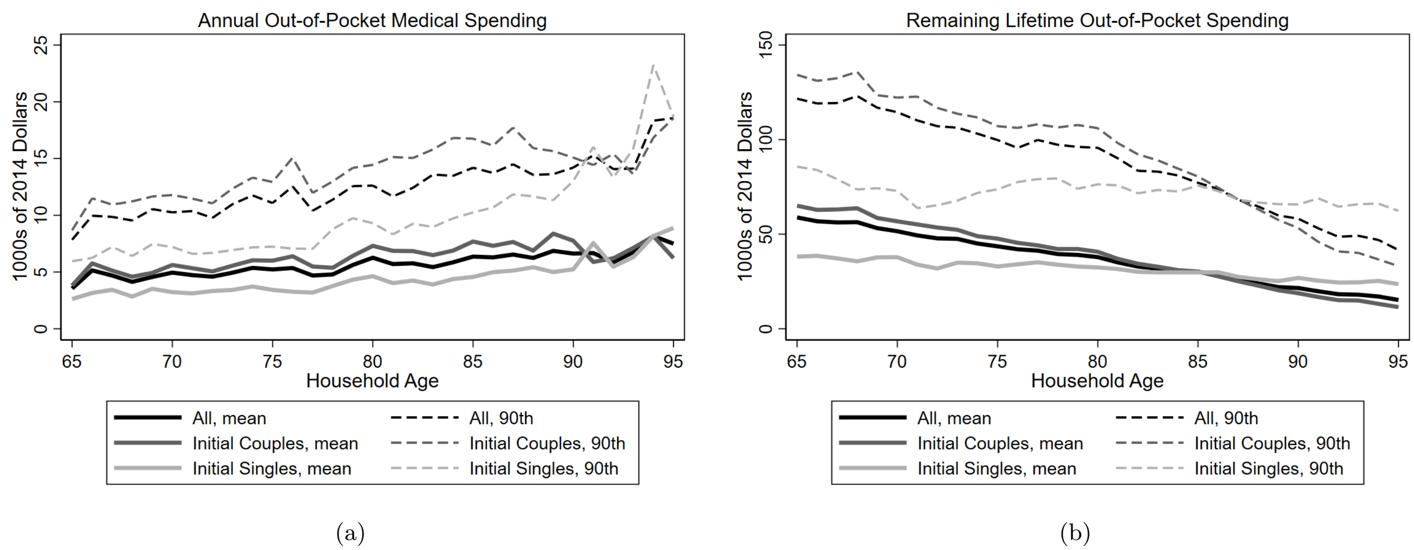


Figure 5: Mean and 90^{th} Percentile of Annual (Panel a) and Remaining Lifetime Out-of-Pocket Medical Spending for Surviving Households (Panel b), Initial Singles and Initial Couples.

Predicted out-of-pocket spending in Figure 5 mirrors the patterns seen in Figure 4, although the levels are lower. At age 65, households will on average incur over \$59,000 of medical spending over the remainder of their lives. Those at the 90th percentile of the spending distribution will incur spending of \$121,000. A household alive at age 90 will on average spend more than \$21,000 out-of-pocket before they die, again highlighting the slow decline of expected lifetime costs as households age.

A number of recent papers have argued that Medicaid significantly reduces the out-of-pocket spending risk faced by older households. Brown and Finkelstein (2008) conclude that Medicaid crowds out private long-term care insurance for about two-thirds of the wealth distribution. De Nardi, French, and Jones (2016) find that most single retirees, including those at the top of the income distribution, value Medicaid at more than its actuarial cost. While both of these papers model Medicaid formally, they lack data to estimate the underlying distributions and thus rely heavily on functional form assumptions. Furthermore, previous papers have tended to abstract away from Medicare payments, in large part because Medicare spending data have not been available. We find that despite these payors, medical spending risk is high in old age.

6 Discussion and Conclusions

In this paper we use health and spending models to simulate the distribution of lifetime medical expenditures as of age 65, adding to the handful of studies on this topic. We also assess the importance of Medicare and Medicaid in reducing lifetime medical spending risk. The simulations show that lifetime medical spending is high and uncertain, and that the level and the dispersion of this spending diminish only slowly with age. Most of these expenditures will be covered by Medicare, Medicaid, or other private and public insurers. Our data suggest that out-of-pocket spending rises more or less linearly with total spending, with households covering about 22% of their total expenditure.

We find that at age 65, households will on average incur over $59,000 of medical spending over the remainder of their lives. Those at the 90th percentile of the spending distribution will incur spending of $121,000. A household alive at age 90 will on average spend more than $21,000 out-of-pocket before they die, highlighting the slow decline of expected lifetime costs as households age.

We conclude by pointing out some caveats to our analysis. We assume, as do many other empirical papers, that medical spending is exogenous, while in reality it is a choice variable. Although the demand for some medical goods and services is extremely inelastic, the demand for others might be elastic. Nursing home care, for example, is a bundle of medical and non-medical commodities, and the latter can vary greatly in quality, with the choice between a single and a shared room being just one example. Thus households can reduce their medical spending risk by purchasing fewer medical goods. While our assumption of exogenous spending arguably leads us to overstate out-of-pocket spending risk, our assumption that the effective co-pay rate is always 22% arguably leads us to understate the risk. For example, most households lack private nursing home insurance, Medicare coverage is extremely limited, and Medicaid coverage is subject to means-testing. Thus many households (such as those in nursing homes whose wealth is too high to qualify for Medicaid) pay more than 22% and others (who are in hospital for modest periods of time) will spend less than 22%. This leaves many households potentially exposed to significant long-term care expenditure risk. Our approach can be extended to consider these types of risks.

References

Alemayehu, Berhanu and Kenneth E. Warner. 2004. "The Lifetime Distribution of Health Care Costs." *Health Services Research* 39 (3):627–642.

Ameriks, John, Joseph S Briggs, Andrew Caplin, Matthew D Shapiro, and Christopher Tonetti. 2020. "Long-Term-Care Utility and Late-in-Life Saving." *Journal of Political Economy* 128 (6):2375–2451.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework." *Econometrica* 85 (3):693–734. URL `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13795`.

Arellano, Manuel, Richard Blundell, Stéphane Bonhomme, and Jack Light. 2021. "Heterogeneity of Consumption Responses to Income Shocks in the Presence of Nonlinear Persistence." .

Banks, James, Richard Blundell, Peter Levell, and James P Smith. 2019. "Life-Cycle Consumption Patterns at Older Ages in the United States and the United Kingdom: Can Medical Expenditures Explain the Difference?" *American Economic Journal: Economic Policy* 11 (3):27–54.

Brown, Jeff and Amy Finkelstein. 2008. "The Interaction of Public and Private Insurance: Medicaid and the Long Term Care Insurance Market." *American Economic Review* 98 (5):837–880.

Creal, Drew. 2012. "A survey of sequential Monte Carlo methods for economics and finance." *Econometric reviews* 31 (3):245–296.

De Nardi, Mariacristina, Giulio Fella, and Gonzalo Paz-Pardo. 2020. "Nonlinear household earnings dynamics, self-insurance, and welfare." *Journal of the European Economic Association* 18 (2):890–926.

De Nardi, Mariacristina, Eric French, and John Bailey Jones. 2010. "Why Do the Elderly Save? The Role of Medical Expenses." *Journal of Political Economy* 118 (1):39–75.

———. 2016. *American Economic Review* 106 (11):3480–3520.

De Nardi, Mariacristina, Eric French, John Bailey Jones, and Jeremy McCauley. 2016. "Medical Spending on the US Elderly." *Fiscal Studies* 37 (3-4):717–747.

De Nardi, Mariacristina, Eric French, John Bailey Jones, and Rory McGee. 2021. "Why Do Couples and Singles Save During Retirement?" Tech. rep., National Bureau of Economic Research.

Doucet, Arnaud, Adam M Johansen et al. 2009. "A tutorial on particle filtering and smoothing: Fifteen years later." *Handbook of nonlinear filtering* 12 (656-704):3.

Fahle, Sean, Kathleen McGarry, and Jonathan Skinner. 2016. "Out-of-Pocket Medical Expenditures in the United States: Evidence from the Health and Retirement Study." *Fiscal Studies* 37 (3-4):785–819.

Feenberg, Daniel and Jonathan Skinner. 1994. "The Risk and Duration of Catastrophic Health Care Expenditures." *The Review of Economics and Statistics* 76 (4):633–647.

French, Eric, Mariacristina De Nardi, John Bailey Jones, Olyesa Baker, and Phil Doctor. 2006. "Right before the end: Asset decumulation at the end of life." *Economic Perspectives* 30 (3):2–13.

French, Eric and John Bailey Jones. 2004. "On the distribution and dynamics of health care costs." *Journal of Applied Econometrics* 19 (6):705–721. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.790.

French, Eric, John Bailey Jones, and Jeremy McCauley. 2017. "The Accuracy of Economic Measurement in the Health and Retirement Study." *Forum for Health Economics and Policy* 20 (2).

Friedberg, Leora, Wenliang Hou, Wei Sun, Anthony Webb, and Zhenyu Li. 2014. "New Evidence on the Risk of Requiring Long-term Care." Center for Retirement Research at Boston College Working Paper 2014-12.

Hirth, Richard A., Sebastian Calonico, Teresa B. Gibson, Helen Levy, Jeffrey Smith, and Anup Das. 2016. "Long-Term Health Spending Persistence among the Privately Insured in the US." *Fiscal Studies* 37 (3-4):749–783.

Hurd, Michael D., Pierre-Carl Michaud, and Susann Rohwedder. 2017. "Distribution of Lifetime Nursing Home use and of Out-of-Pocket Spending." *Proceedings of the National Academy of Sciences* 114 (37):277–297.

Jones, John Bailey, Mariacristina De Nardi, Eric French, Rory McGee, and Justin Kirschner. 2018. "The Lifetime Medical Spending of Retirees." Working Paper 24599, National Bureau of Economic Research. URL http://www.nber.org/papers/w24599.

Koenker, Roger and Gilbert Bassett Jr. 1978. "Regression quantiles." *Econometrica: journal of the Econometric Society* :33–50.

Kopecky, Karen and Tatyana Koreshkova. 2014. "The Impact of Medical and Nursing Home Expenses on Savings and Inequality." *American Economic Journal: Macroeconomics* 6 (3):29–72.

Livshits, Igor, James MacGee, and Michele Tertilt. 2007. "Consumer bankruptcy: A fresh start." *American Economic Review* 97 (1):402–418.

Lockwood, Lee M. 2012. "Bequest motives and the annuity puzzle." *Review of economic dynamics* 15 (2):226–243.

Merrill Lynch Wealth Management. 2012. "Merrill Lynch Affluent Insights Survey Fact Sheet." URL `http://www.ml.com/Publish/Content/application/pdf/GWMOL/FactSheet_ML_AIS_February2012.pdf`.

Nakajima, Makoto and Irina A. Telyukova. 2018. "Medical Expenses and Saving in Retirement: The Case of U.S. and Sweden." Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute.

Nicholas, Lauren Hersch, Kenneth M Langa, Theodore J Iwashyna, and David R Weir. 2011. "Regional variation in the association between advance directives and end-of-life Medicare expenditures." *Jama* 306 (13):1447–1453.

Pashchenko, Svetlana. 2013. "Accounting for non-annuitization." *Journal of Public Economics* 98:53–67.

Pashchenko, Svetlana and Ponpoje Porapakkarm. 2016. "Medical spending in the US: facts from the Medical Expenditure Panel Survey Data Set." *Fiscal Studies* 37 (3-4):689–716.

Poterba, James M., Steven F. Venti, and David A. Wise. 2017. "The Asset Cost of Poor Health." *Journal of the Economics of Ageing* 9:172–184.

Reichling, Felix and Kent Smetters. 2015. "Optimal Annuitization with Stochastic Mortality Probabilities." *American Economic Review* 105 (11):3273–3320.

Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun. 2006. "Are Americans Saving Optimally for Retirement?" *Journal of Political Economy* 114:607–643.

Skinner, Jonathan. 2007. "Are you sure you're saving enough for retirement?" *Journal of Economic Perspectives* 21 (3):59–80.

Webb, Anthony and Natalia A. Zhivan. 2010. "How Much is Enough?: The Distribution of Lifetime Health Care Costs." Center for Retirement Research at Boston College Working Paper 2010-1.

A Our Medicare and Medicaid Data

Medicare

We link restricted Medicare fee-for-service (Parts A and B), and Part D data for the years 1999-2012 (2006 was the first year of Medicare Part D and thus our Part D data begins then) to our HRS survey data for respondents who consent to allow their Medicare data to be linked to their survey responses (approximately 64.7% percent of persons in our study population). These records have enrollment information and data on reimbursement amounts for inpatient, skilled nursing facility, home health, and hospice claims (Medicare Part A), as well as outpatient, carrier (non-institutional medical care providers such as individual or group practitioners, non-hospital labs, and ambulances), and durable medical equipment claims for Medicare Part B.

We use the Beneficiary Annual Summary File (BASF), which summarizes information from the micro-level claims records. The BASF contains annual information for each individual on the number of months of enrollment in Medicare Part A, Part B, and non-fee-for-service plans. The BASF has information on Medicare fee-for-service (FFS) claims. Almost all claims for services used by non-FFS Medicare patients are not observed in these data, so all analyses exclude an individual in a given year if they were enrolled in a non-FFS Medicare plan for more than half the year.

Medicare Part D is the prescription drug benefit. We calculate the Medicare Part D payment using the Part D event files. For the Part D Medicare contribution we subtract from the gross drug cost the payments paid by the beneficiary, family, or friends.

Medicaid

As with the Medicare data, we are able to link restricted Medicaid data (CMS Medicaid Analytic eXtract, or "MAX" files) for those in the HRS who gave permission, allowing us to measure Medicaid expenditures for the Medicaid beneficiaries in our data set for the years 1999-2012. The MAX files contain personal summaries (which contain eligibility, enrollment, and demographic information) and claims data across four service categories (inpatient, long-term care, prescription drugs, and other services). Other services include a variety of services (e.g., physician services and lab work) that do not fit under the other three service categories. The inpatient, long-term care, prescription drugs, and other services files contain the primary variable of interest, "Medicaid Payment Amount," which is the total amount of money paid by Medicaid for a particular service. We sum over all the claims for all the different service categories for a particular individual in each year.

B Sample Selection

We drop households for various reasons. These include disagreement between household members, problematic mortality transitions, households with multiple members of the same sex, and refusal to provide Medicare and Medicaid spending records. Table 3 below denotes the starting public HRS sample size and the sample size after every drop.

Drop reason	Post-drop sample size
	26,598
One spouse dead and the other claims never married or divorced	26,596
One spouse claims married and the other something else	25,343
Two or more members of the same sex	24,897
One spouse claims never married an the other is not missing	24,818
One spouse claims divorced and the other widowed	24,545
People who are partnered	23,466
People who "died", then came back to life	23,161
Transitions with widowed or both widowed	23,055
Transitions to got married or divorced	22,719
Household split etc. based on sub-household identifier	22,695
Households that appear once and have missing marital status	22,683
No domestic partner in sample but married in some wave	21,903
No living member the year household joined HRS	21,713
Households dead or under 65 in the estimation period (1999-2012)	9,969
2+ years post death Medicare and Medicaid spending	6,783
Incomplete Medicare and Medicaid records	4,391

Notes: Sample size refers to the number of unique households.

Table 3: Sample Selection

C Imputing Missing Medical Expenditures

Our goal is to measure all medical spending: the variable M_{it} in equation (2) of the main text is defined to include out-of-pocket spending, Medicare and Medicaid payments, and private and other public (such as Veterans Administration benefits, and care provided by local and state health departments) insurance payments. While the HRS includes information on out-of-pocket spending and can be linked to Medicare and Medicaid payments, it does not include Medicare Part C, private, or other public insurance payments. In this appendix we describe how we use data from the Medical Expenditure Panel Survey (MEPS) to impute these payments in the HRS. Although the MEPS has extremely high quality information on all payors for all household members, it lacks the long panel dimension of the HRS. Our imputation procedures allow us to exploit the best of both data sets.

Our imputation procedure has two steps. First, we use the MEPS to infer private and other public insurance payments, conditional on variables observed in both datasets. Second, we impute private and other public insurance payments in the HRS data using a conditional mean matching procedure (which is a procedure very similar to hot-decking).

First Step of Imputation Procedure

We use the MEPS to infer payments of other payors, conditional on the observable variables that exist in both the MEPS and the HRS datasets.

Let i index individuals in the HRS and j index individuals in the MEPS. Define M_{it}^{obs} as out of pocket, Medicaid, and Medicare (Part A, B, and D, but not Part C) payments observed in both the HRS and MEPS data sets, M_{it}^{miss} as the components of medical spending missing in the HRS but observed in the MEPS, and $M_{it} = M_{it}^{miss} + M_{it}^{obs}$ as total medical spending. To impute M_{it}^{miss}, which is missing in the HRS, we follow David et al. (1986), French and Jones (2011), and De Nardi et al. (2020) and use a predictive mean-matching regression approach. There are two steps to our procedure. First, we use the MEPS data to regress M_{it}^{miss} on observable variables that exist in both data sets. Second, we impute M_{it}^{miss} in the HRS data using a conditional mean-matching procedure, a procedure very similar to hot-decking.

First, for every member of the the MEPS sample, we regress the variable of interest M^{miss} on the vector of observable variables z_{jt}, yielding $M_{jt}^{miss} = z_{jt}\beta + \varepsilon_{jt}$. Second, for each individual j in the MEPS we calculate the predicted value $\widehat{M^{miss}}_{jt} = z_{jt}\hat{\beta}$, and for each member of the sample we calculate the residual $\widehat{\varepsilon}_{jt} = M_{jt}^{miss} - \widehat{M^{miss}}_{jt}$. Third, we sort the predicted value $\widehat{M^{miss}}_{jt}$ into deciles and keep track of all values of $\hat{\varepsilon}_{jt}$ within each decile. We use this procedure separately to impute Medicare Part C benefits, private payments, and other payments.

In practice we include in z_{jt} a fourth-order age polynomial, marital status, gender, self-reported health (=1 if self reported health is good, very good, or excellent), race, visiting a medical practitioner (doctor, hospital or dentist), out-of-pocket medical spending, education of head (high school, some college, college), death of an individual, and total household income. We estimate this regression two times: once for the privately insured, and once for other payors.

Because the measure of medical spending in the HRS is medical spending over two years, we divide HRS out-of-pocket medical spending by 2 and assume that medical spending is equal across the two years.

Second Step of Imputation Procedure

For every observation in the HRS sample with a positive Medicaid indicator, we impute $\widehat{Med}_{it} = z_{it}\widehat{\beta}$, using the values of $\widehat{\beta}$ estimated from the MEPS. Then we impute ε_{it} for each observation of this subsample by finding a random observation in the MEPS with a value of $\widehat{Med}_{jt}$ in the same decile as $\widehat{Med}_{it}$, and setting $\hat{\varepsilon}_{it} = \hat{\varepsilon}_{jt}$. The imputed value of Med_{it} is $\widehat{Med}_{it} + \hat{\varepsilon}_{it}$.

As David et al. (1986) point out, our imputation approach is equivalent to hot-decking when the "z" variables are discretized and include a full set of interactions. The advantages of our approach over hot-decking are two-fold. First, many of the "z" variables are continuous. Second, to improve goodness of fit we use a large number of "z" variables. We find that adding extra variables are very important for improving goodness of fit when imputing payments. Because hot-decking uses a full set of interactions, this would result in a large number of hot-decking cells relative to our sample size. Thus, in this context, hot decking is too data intensive.

D Validating the Administrative Medical Spending Data

Here, we examine in greater detail the accuracy of the administrative medical spending data, as well as the out-of-pocket spending found in the Assets and Health Dynamics of the Oldest Old (AHEAD) cohort of the HRS, comparing them to data from the MCBS. See De Nardi, French, and Jones (2016) and De Nardi et al. (2016) for more details of the MCBS data and (for example) Nicholas et al. (2011) for details of the HRS linked data.

The MCBS is a nationally representative survey of Medicare beneficiaries, consisting of Disability Insurance recipients and Medicare recipients aged 65 and older. The survey contains an over-sample of beneficiaries older than 80 and disabled individuals younger than 65. Respondents are asked about health status, health insurance, and health care spending (from all sources). The MCBS data are matched to Medicare records, and medical spending data are created through a reconciliation process that combines information from survey respondents with Medicare administrative files. As a result, the survey is thought to give extremely accurate data on Medicare payments and fairly accurate data on out-of-pocket and Medicaid payments. As in the HRS survey, the MCBS survey includes information on those who enter a nursing home or die. Respondents are interviewed up to 12 times over a 4 year period. We aggregate the data to an annual level. In both samples, we applied only modest sample selection restrictions. The key sample selection issue shown in Table 3 is that in the HRS we drop households with missing or erroneous Medicare or Medicaid records.

Here we compare distributions of total, out-of-pocket, Medicare, and Medicaid payments between the MCBS are the HRS data. Medical spending the HRS is measured at an individual level (rather than household) to be comparable with the MCBS. The comparison can be seen in Table 4. Medical spending is higher in our HRS sample than in the MCBS sample. Furthermore, this higher level of spending is driven by higher out-of-pocket spending, Medicare, and Medicaid spending. These differences potentially are an advantage of the HRS data since, as noted in De Nardi et al. (2016), the MCBS clearly understates aggregate Medicare and especially Medicaid spending, potentially due to the issue that the MCBS does not have administrative data on Medicaid spending, and thus relies heavily on imputation.

The next set of benchmarking exercises that we perform is for out-of-pocket medical spending, Medicaid recipiency and income between the AHEAD cohort of the HRS and MCBS. For both the HRS and MCBS, we restrict the sample to singles (over the sample period) who meet the HRS/AHEAD age criteria (at least 70 in 1994, 72 in 1996, ...) and who are not working over the sample period. Because the MCBS sample lacks spousal information, for this analysis we focus only on singles. We use De Nardi, French, and Jones's (2016) measure of permanent income and construct a measure of permanent income, which is the percentile rank of total income over the period we observe these individuals (the MCBS asks only about total income). The first four columns of Table 5 show sample statistics from the full HRS/AHEAD sample while the final three columns of the table show sample statistics from the MCBS sample. The first statistics we compare are income. Total income in the HRS/AHEAD data (including asset and other non-annuitized income) lines up well with total income in the MCBS data, although income in the top quintile of the MCBS is higher than in the HRS/AHEAD. Next, we compare out-of-pocket medical spending in the MCBS and HRS/AHEAD. Out-of-pocket medical spending (including insurance payments) averages \$2,360 in the bottom PI quintile and \$6,340 in the top quintile in the HRS/AHEAD. In comparison, the same numbers in the MCBS data are \$3,540 and \$7,020. Overall, out-of-pocket medical spending in the MCBS and HRS/AHEAD are similar, which may be surprising given that the two surveys each have their own advantages in terms of survey methodology.[8] The share of the population receiving Medicaid transfers is also very similar in the HRS/AHEAD and MCBS. 61% and 70% of those in the bottom PI quintile are on Medicaid in the the HRS/AHEAD and MCBS, respectively. In the top quintile, 3% of people are on Medicaid in the HRS/AHEAD whereas 5% are in the MCBS.

E Plotting the Moments of Medical Spending

This appendix shows how we plot the moments of medical spending in Figure 2. We follow closely De Nardi, Fella, and Paz-Pardo's (2020) approach.

Assume that the log of medical spending for household (or person) i at period t is given by equation (2), which we replicate here:

$$\ln M_{i,t} = X_{i,t}\gamma + m_{i,t}, \tag{12}$$

where $X_{i,t}$ is a vector of conditioning variables such as age, household composition, and health. As long

[8]There are more detailed questions underlying the out-of-pocket medical expense questions in the HRS, including the use of "unfolding brackets". Respondents can give ranges for medical expense amounts, instead of a point estimate or "don't know" as in the MCBS. The MCBS has the advantage that forgotten medical out-of-pocket medical expenses will be imputed if Medicare had to pay a share of the health event.

	Total Spending				OOP			
Total Spending	HRS		MCBS		HRS		MCBS	
Percentiles	Average Exp.	Pct. Total	Average Exp.	Pct. Total	Average Exp.	Pct. Total	Average Exp.	Pct. Total
All	17,091	100	14,120	100	3,825	100	2,740	100
95-100%	114,238	33.4	97,880	34.6	45,643	59.6	26,930	49.1
90-95%	59,000	17.3	48,890	17.3	8,619	11.3	6,700	12.2
70-90%	26,870	31.4	20,540	29.1	3,480	18.2	2,920	21.3
50-70%	9,025	10.6	7,750	11	1,394	7.3	1,360	9.9
0-50%	2,502	7.3	2,250	8	178	3.6	420	7.6

	Medicare				Medicaid			
Total Spending	HRS		MCBS		HRS		MCBS	
Percentiles	Average Exp.	Pct. Total	Average Exp.	Pct. Total	Average Exp.	Pct. Total	Average Exp.	Pct. Total
All	11,343	100	7,720	100	1,896	100	1,320	100
95-100%	85,268	37.6	67,560	43.7	33,773	89.1	24,980	94.7
90-95%	41,731	18.4	28,370	18.4	4,092	10.8	1,360	5.2
70-90%	17,251	30.4	10,280	26.6	230	0.2	10	0.1
50-70%	5,031	8.9	2,980	7.7	0	0	0	0
0-50%	1,076	4.7	550	3.5	0	0	0	0

Table 4: Individual Medical Spending Percentiles: HRS versus MCBS

Income Quintile	Total Income	HRS/AHEAD Data Annuity Income	Out-of-pocket Expenses	Medicaid Recipiency	MCBS Data Total Income	Out-of-pocket Expenses	Medicaid Recipiency
1	7,740	4,820	2,550	60.9	6,750	4,050	69.9
2	10,290	8,270	4,270	28.1	10,020	5,340	41.8
3	15,500	10,900	5,050	11.0	13,740	6,470	15.5
4	19,290	14,390	6,360	5.6	19,710	7,300	8.0
5	33,580	26,300	7,000	3.0	44,150	8,020	5.4

Notes:1996-2010, for those age 72 and older in 1996.

Table 5: Income, Out-of-pocket Spending, and Medicaid Recipiency Rates: HRS versus MCBS

as we impose no restrictions on $m_{i,t}$, equation (12) is an identity that is true by construction. It is often commonly assumed that $m_{i,t}$ is independent of $X_{i,t}$, but this is not necessary for our purposes. Likewise, we can construct the autoregression

$$m_{i,t} = \delta m_{i,t-1} + e_{i,t}. \tag{13}$$

Our procedure is as follows:

1. Using OLS, estimate the coefficient γ in equation (12) and construct $\{m_{i,t}\}_{i,t}$.
2. Find $r_{i,t}$, the percentile rank of $m_{i,t}$, using a single ranking across all periods.
3. Using OLS, estimate the coefficient δ in equation (13) and construct $\{e_{i,t}\}_{i,t}$.
4. Construct the variable $y_{i,t} = f(e_{i,t})$, where $f(e)$ is a function that can be used to construct summary statistics.
   (a) For means we use $f(e) = e$.
   (b) For variances we use $f(e) = e^2$. (To get standard deviations, we need to find the square root of the predicted variances.) Let $\widehat{\sigma}^2$ denote the estimated variance, the average value of e^2.
   (c) For skewness we use $f(e) = e^3/\widehat{\sigma}^3$, where $\widehat{\sigma}^3$ utilizes the standard deviation from step (b).
   (d) For kurtosis we use $f(e) = e^4/\widehat{\sigma}^4$.
5. For each summary statistic, regress the appropriate $y_{i,t}$ against $r_{i,t-1}$, and plot the predicted values. In practice, we use a kernel-weighted local polynomial regression.

If the medical spending residuals follow an AR(1), then equation (13) is equivalent to equation (4) and the residuals $\{\hat{e}_{i,t}\}_{i,t}$ from equation (13) will deliver a consistent estimates of the residuals $\{\zeta_{i,t}\}_{i,t}$ from equation (4). Given that $\zeta_{i,t}$ is assumed to be iid and normally distributed (see equation (5)), it will at any value of $m_{i,t-1}$ have a mean of zero, constant variance, no skewness, and kurtosis of three.

In the more complex case where medical spending residuals follow an AR(1) plus white noise, the probability limit of $\hat{\delta}$ from equation (13) will be:

$$
\begin{aligned}
\text{plim } \hat{\delta} &= \frac{Cov(m_{i,t}, m_{i,t-1})}{Var(m_{i,t-1})} \\
&= \frac{Cov(\eta_{i,t}, \eta_{i,t-1})}{Var(\eta_{i,t-1}) + Var(\varepsilon_{i,t-1})} \\
&= \phi \frac{Var(\eta_{it-1})}{Var(\eta_{it-1}) + Var(\varepsilon_{i,t-1})},
\end{aligned}
\tag{14}
$$

so that

$$
\begin{aligned}
e_{i,t} &= m_{i,t} - \delta m_{i,t-1} \\
&= \eta_{i,t} + \varepsilon_{i,t} - \delta(\eta_{i,t-1} + \varepsilon_{i,t-1}) \\
&= \phi \eta_{i,t-1} + \zeta_{i,t} + \varepsilon_{i,t} - \delta(\eta_{i,t-1} + \varepsilon_{i,t-1}) \\
&= (\phi - \delta)\eta_{i,t-1} + \zeta_{i,t} + \varepsilon_{i,t} - \delta \varepsilon_{i,t-1}.
\end{aligned}
$$

All the objects above are normally-distributed and zero-mean, and thus unconditionally e_{it} is also normally-distributed and zero-mean. This is also true conditional on $m_{i,t-1}$, as

$$
\begin{aligned}
\mathbb{E}[e_{i,t} \,|\, m_{i,t-1}] &= \mathbb{E}[(\phi - \delta)\eta_{i,t-1} + \zeta_{i,t} + \varepsilon_{i,t} - \delta \varepsilon_{i,t-1} \,|\, m_{i,t-1}] \\
&= (\phi - \delta)\,\mathbb{E}[\eta_{i,t-1} \,|\, m_{i,t-1}] - \delta\,\mathbb{E}[\varepsilon_{i,t-1} \,|\, m_{i,t-1}] \\
&= (\phi - \delta)\Big(\frac{Cov(\eta_{i,t-1}, m_{i,t-1})}{Var(m_{i,t-1})} m_{i,t-1}\Big) - \delta\Big(\frac{Cov(\varepsilon_{i,t-1}, m_{i,t-1})}{Var(m_{i,t-1})} m_{i,t-1}\Big) \\
&= (\phi - \delta)\frac{Var(\eta_{i,t-1})}{Var(m_{i,t-1})} m_{i,t-1} - \delta \frac{Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})} m_{i,t-1} \\
&= \phi \frac{Var(\eta_{i,t-1}) Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})^2} m_{i,t-1} - \phi \frac{Var(\eta_{i,t-1}) Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})^2} m_{i,t-1} \\
&= 0,
\end{aligned}
$$

with the fifth line following from equation (14), which implies that

$$
\phi - \delta = \phi\Big(1 - \frac{Var(\eta_{i,t-1})}{Var(\eta_{i,t-1}) + Var(\varepsilon_{i,t-1})}\Big) = \phi \frac{Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})}.
$$

The conditional variance of $e_{i,t}$ is given by:

$$
\begin{aligned}
Var(e_{i,t}\,|\,m_{i,t-1}) &= \mathbb{E}\left[e_{i,t}^2\,|\,m_{i,t-1}\right] - \mathbb{E}[e_{i,t}\,|\,m_{i,t-1}]^2 \\
&= \mathbb{E}\left[\left((\phi-\delta)\eta_{i,t-1} + \zeta_{i,t} + \varepsilon_{i,t} - \delta\varepsilon_{i,t-1}\right)^2\,|\,m_{i,t-1}\right] \\
&= (\phi-\delta)^2\,\mathbb{E}\left[\eta_{i,t-1}^2\,|\,m_{i,t-1}\right] + \mathbb{E}\left[\zeta_{i,t}^2\,|\,m_{i,t-1}\right] + \mathbb{E}\left[\varepsilon_{i,t}^2\,|\,m_{i,t-1}\right] \\
&\quad +\delta^2\,\mathbb{E}\left[\varepsilon_{i,t-1}^2\,|\,m_{i,t-1}\right] - 2\delta(\phi-\delta)\,\mathbb{E}\left[\eta_{i,t-1}\varepsilon_{i,t-1}\,|\,m_{i,t-1}\right] \\
&= (\phi-\delta)^2\left(Var(\eta_{i,t-1}\,|\,m_{i,t-1}) + \mathbb{E}[\eta_{i,t-1}\,|\,m_{i,t-1}]^2\right) + \sigma_\zeta^2 + \sigma_\varepsilon^2 \\
&\quad +\delta^2\left(Var(\varepsilon_{i,t-1}\,|\,m_{i,t-1}) + \mathbb{E}[\varepsilon_{i,t-1}\,|\,m_{i,t-1}]^2\right) - 2\delta(\phi-\delta)\,\mathbb{E}\left[\eta_{i,t-1}\varepsilon_{i,t-1}\,|\,m_{i,t-1}\right] \\
&= (\phi-\delta)^2 Var(\eta_{i,t-1})\left(1 - \frac{Var(\eta_{i,t-1})}{Var(m_{i,t-1})}\right) + \delta^2 Var(\varepsilon_{i,t-1})\left(1 - \frac{Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})}\right) \\
&\quad +(\phi-\delta)^2\frac{Var(\eta_{i,t-1})^2}{Var(m_{i,t-1})^2}m_{i,t-1}^2 + \delta^2\frac{Var(\varepsilon_{i,t-1})^2}{Var(m_{i,t-1})^2}m_{i,t-1}^2 \\
&\quad -2\delta(\phi-\delta)\frac{Var(\eta_{i,t-1})Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})^2}m_{i,t-1}^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2 \\
&= (\phi-\delta)^2 Var(\eta_{i,t-1})\frac{Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})} + \delta^2 Var(\varepsilon_{i,t-1})\frac{Var(\eta_{i,t-1})}{Var(m_{i,t-1})} \\
&\quad +\phi^2\frac{Var(\eta_{i,t-1})^2 Var(\varepsilon_{i,t-1})^2}{Var(m_{i,t-1})^4}m_{i,t-1}^2 + \phi^2\frac{Var(\eta_{i,t-1})^2 Var(\varepsilon_{i,t-1})^2}{Var(m_{i,t-1})^4}m_{i,t-1}^2 \\
&\quad -2\phi^2\frac{Var(\eta_{i,t-1})^2 Var(\varepsilon_{i,t-1})^2}{Var(m_{i,t-1})^4}m_{i,t-1}^2 + \sigma_\zeta^2 + \sigma_\varepsilon^2 \\
&= ((\phi-\delta)^2 + \delta^2)\frac{Var(\eta_{i,t-1})Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})} + \sigma_\zeta^2 + \sigma_\varepsilon^2.
\end{aligned}
$$

Hence, the mean of $e_{i,t}$ conditional on $m_{i,t-1}$ is zero and the variance of $e_{i,t}$ conditional on $m_{i,t-1}$ is constant. For the conditional skewness and kurtosis, note that the distribution of $e_{i,t}$ conditional on $m_{i,t-1}$ is normal, as it is a linear combination of variables that themselves follow a multivariate normal distribution, conditional on $m_{i,t-1}$ which is normal. It follows that the skewness and kurtosis are constant.

F Estimation methodology

We estimate the model using the extension of the *E-M* algorithm employed by Arellano, Blundell, and Bonhomme (2017). Recall from equations (8)-(10) that the functions $Q_\eta(\tau\,|\,\eta_{i,t-1}, a_{i,t})$, $Q_1(\tau\,|\,a_{i,1})$ and $Q_\varepsilon(\tau\,|\,,a_{i,t})$ are constructed from Hermite polynomials $\left(\{h_k^\eta(\cdot)\}_{k=0}^{K_\eta}, \{h_k^1(\cdot)\}_{k=0}^{K_1}, \{h_k^\varepsilon(\tau)\}_{k=0}^{K_\varepsilon}\right)$, using the coefficient functions $\{\beta_k^\eta(\tau)\}_{k=0}^{K_\eta}$, $\{\beta_k^1(\tau)\}_{k=0}^{K_1}$, and $\{\beta_k^\varepsilon(\tau)\}_{k=0}^{K_\varepsilon}$. The coefficient functions are in turn modeled with a set of polynomial splines defined over the intervals $\left\{[\tau_{\ell-1}, \tau_\ell]\right\}_{\ell=1}^{L}$, along with two tail functions for $(0, \tau_1]$ and $[\tau_L, 1)$. It is the parameters for these weighting functions that we must estimate.

Define θ as the vector of all parameters (the β parameters) in equations (8)-(10). The procedure to estimate θ is as follows. Starting with the vector $\hat{\theta}^{(0)}$ we iterate between the following two steps until $\hat{\theta}^{(j)}$ converges:

1. *Stochastic E-Step*: For each observation i, draw S values of $\eta_i^{(s)} = (\eta_{i1}^{(s)}, ..., \eta_{iT}^{(s)})$ from $f_i(\,.\,;\hat{\theta}^{(j)})$ $\big($derived from $Q_1^{(j)}(\cdot)$, $Q_\eta^{(j)}(\cdot)$ and $Q_\varepsilon^{(j)}(\cdot)\big)$.

2. *M-step*: Find

$$\underset{\beta_{\ell 0}^{\eta},...,\beta_{\ell K_\eta}^{\eta}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{s=1}^{S}\sum_{t=2}^{T}\rho_{\tau_\ell}\left(\eta_{it}^{(s)} - \sum_{k=1}^{K_\eta}\beta_{\ell k}^{\eta}\, h_k^{\eta}(\eta_{i,t-1}^{(s)}, a_{it})\right), \quad \ell = 1, ..., L.$$

   We use $\rho_\tau(\cdot)$ to denote Koenker and Bassett Jr's (1978) quantile "check" function. To identify the full set of splines, this function is minimized at each point ℓ on the grid over τ. The coefficients for $\varepsilon_{i,t}$ and $\eta_{i,1}$ likewise solve

$$\underset{\beta_{\ell 0}^{\varepsilon},...,\beta_{\ell K_\varepsilon}^{\varepsilon}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{s=1}^{S}\sum_{t=1}^{T}\rho_{\tau_\ell}\left(m_{i,t} - \eta_{it}^{(s)} - \sum_{k=1}^{K_\varepsilon}\beta_{\ell k}^{\varepsilon}\, h_k^{\varepsilon}(a_{it})\right), \quad \ell = 1, ..., L,$$

$$\underset{\beta_{\ell 0}^{1},...,\beta_{\ell K_1}^{1}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{s=1}^{S}\rho_{\tau_\ell}\left(\eta_{i1}^{(s)} - \sum_{k=1}^{K_1}\beta_{\ell k}^{1}\, h_k^{1}(a_{i1})\right), \quad \ell = 1, ..., L.$$

   There are also moment conditions related to the tails of the distribution: See Arellano, Blundell, and Bonhomme (2017). These estimates give us $\hat{\theta}^{(j+1)}$.

For longer panels, settings with unbalanced data, or when estimating more complicated models the *E-step* can perform poorly when using standard samplers (e.g., Metropolis-Hastings). We therefore employ the sequential Monte-Carlo (SMC) approach implemented by Arellano et al. (2021). Comprehensive surveys of these methods can be found in Doucet, Johansen et al. (2009) and Creal (2012).

We will use the Gaussian analogues to equations (8), (9) and (10) as importance distributions.

Step 1: SMC Stochastic E-Step to sample from $f(\eta_{i,1}, ..., \eta_{i,T}|Y_i^T, a_i^T)$. For $i = 1, ..., N$:

At $t = 1$:

1. Sample S particles $\eta_1^{(s)} \sim g(\eta_1|y_1)$, where $g(\cdot)$ is the closed form posterior from the Gaussian model.
2. Compute the weights $w_1(\eta_1^{(s)})$ and apply a self-normalization to obtain $W_1^{(s)} \propto w_1(\eta_1^{(s)})$.
3. If $Var(W^{(s)})$ exceeds some threshold, re-sample $\{W_1^{(s)}, \eta_1^{(s)}\}$ to obtain S equally weighted particles.

At $t > 1$:

1. Sample S particles $\eta_t^{(s)} \sim g(\eta_t|\eta_{t-1}, y_t)$, where $g(\cdot)$ is the closed-form posterior from the Gaussian model.
2. Compute the weights $w(\eta_{1:t}^{(s)})$ and apply a self-normalization to obtain $W_t^{(s)} \propto w_t(\eta_{1:t}^{(s)})$.
3. If $Var(W^{(s)})$ exceeds some threshold, re-sample $\{W_t^{(s)}, \eta_t^{(s)}\}$ to obtain S equally weighted particles.
4. If $t = T$, sample P particles to be used in the *M-Step*. (We set $P = 1$).

Step 2: M-Step

1. Update quantile regressions for equations (8), (9) and (10) .
2. Update Laplace parameters for the tail functions.
3. Update parameters for Gaussian proposal distributions.

G Demographic Transition Probabilities in the HRS

Let $hs_{i,j,t} \in \{0,1,2,3\}$ denote death ($hs_{i,j,t} = 0$) and the 3 mutually exclusive health states of the living (nursing home = 1, bad = 2, good = 3, respectively) of household member j, household i, time t. Let $x_{i,j,t}$ be a vector that includes a constant, age, permanent income, gender, and powers and interactions of these variables, and indicators for previous health and previous health interacted with age. Our goal is to construct the likelihood function for the transition probabilities.

Using a multivariate logit specification, we have, for $q \in \{1,2,3\}$, $r \in \{0,1,2,3\}$, we rewrite equation (11) as

$$\begin{aligned}
\pi_{q,r,t} &= \Pr(hs_{i,g,t+1} = r|\, hs_{i,g,t} = q; x_{i,g,t}) \\
&= \gamma_{qr} \Big/ \sum_{s\in\{0,1,2,3\}} \gamma_{qs}, \\
\gamma_{qs} &\equiv 1, \quad s = 0 \\
\gamma_{qs} &= \exp\left(x_{i,g,t}\beta_s\right), \quad s \in \{1,2,3\},
\end{aligned}$$

where $\{\beta_s\}_{s=1}^{3}$ are coefficient vectors for each future state s and $x_{i,g,t}$ is the explanatory variable vector which depends on the current state q.

The formulae above give 1-period-ahead transition probabilities, whereas what we observe in the HRS data set are 2-period ahead probabilities, $\Pr(hs_{i,g,t+2} = r|\, hs_{i,g,t} = q; x_{i,g,t})$. The two sets of probabilities are linked, however, by

$$\begin{aligned}\Pr(hs_{i,g,t+2} = r \mid hs_{i,g,t} = q; x_{i,g,t}) &= \sum_s \Pr(hs_{i,g,t+2} = r \mid hs_{i,g,t+1} = s; x_{i,g,t}) \Pr(hs_{i,g,t+1} = s \mid hs_{i,g,t} = q; x_{i,g,t}) \\ &= \sum_s \pi_{sr,t+1} \pi_{qs,t},\end{aligned}$$

imposing $\pi_{00,t+1} = 1$. This allows us to estimate $\{\beta_k\}$ directly from the data using maximum likelihood.