THE RISK AND DURATION OF
CATASTROPHIC HEALTH CARE EXPENDITURES

Daniel Feenberg

Jonathan Skinner

THE RISK AND DURATION OF CATASTROPHIC HEALTH CARE

ABSTRACT

Catastrophic medical expenses are an important economic risk facing the elderly. Little is known about the persistence of such out-of-pocket medical costs. We measure the time-series property of medical costs using information on medical deductions from a panel of tax returns. During the period of analysis, 1968-73, taxpayers could deduct medical expenses above 3 percent of income. We correct for the resulting censoring bias using multivariate Tobit estimated with a variant of the smoothed simulated maximum likelihood (SSML) method. The data suggest that the burden of out-of-pocket medical expenses is substantially larger for lower income families. Furthermore, the estimated coefficients suggest substantial time-persistence in out-of-pocket medical care costs; a $1 increase in out-of-pocket medical spending is predicted to increase future spending by an additional $2.80. These results may shed light both on the social value of catastrophic health insurance as well on aggregate saving behavior.

Daniel Feenberg
N.B.E.R.
1050 Massachusetts Ave.
Cambridge, MA 02138

Jonathan Skinner
N.B.E.R.
1050 Massachusetts Ave.
Cambridge, MA 02138

I. Introduction

Long-term illness threatens the financial health as well as the physical health of elderly families. Despite existing coverage provided by Medicare, Medicaid, and private health insurance, the risk of a catastrophic medical expense is large. Out-of-pocket medical expenses for the elderly in 1988 were estimated to be $2,394 per person, or roughly 18 percent of average per capita income for the elderly (US Congress, 1989).

There is some evidence from cross-sectional data about the financial risk of medical expenses in a given year. But there is little evidence on the extent of lifetime out-of-pocket medical expenses. If catastrophic medical expenses are highly correlated over time, then even modest levels of health costs -- in any given year -- can gradually deplete a family's entire nest-egg of retirement saving. Furthermore, the greater the persistence of medical expenses, the greater the value of catastrophic medical insurance.[1]

We estimate the time-series properties of health expenditures using panel data on taxpayer returns collected by the IRS. We use data from the late 1960s and early 1970s when taxpayers who itemized could deduct medical expenses above a floor of 3 percent of adjusted gross income (AGI). The reporting of medical expenses over a six-year period 1968-73 allows for accurate measures of long-term persistence, as well as providing the information about the catastrophic costs likely to have the largest impact on the financial health of the elderly.

---

[1]   It is straightforward to show that for a concave utility function, the value of actuarially fair insurance per dollar of premium is greatest for low-probability/high-cost events.

We observe medical expenses only for households that itemize and that experience expenses above 3% of AGI. Hence we must account for these selection criteria in estimating the dynamic model. The estimation problem is a classic tobit problem, albeit with a variable lower limit and in higher dimensions. Because of the difficulty of evaluating the implied four-dimensional normal distribution function we use the ingenious smoothed simulated maximum likelihood (SSML) approach developed in Borsch-Supan and Hajivassiliou (1992), and based in turn on the Geweke-Hajivassiliou-Keane simulation method.[2]

The SSML approach evaluates multi-dimensional distribution functions (and hence the multidimensional tobit) by the use of simulated error terms drawn multiple times for each observation. By averaging over the realizations of the simulated draws, a "simulated" likelihood function can be calculated.[3] Given that the same simulated draws are used at each iteration of the likelihood function, one may then use standard hill-climbing techniques to solve for the maximum likelihood coefficient estimates. We seek to further improve the sampling properties of the SSML by using a method of pseudorandom sampling of error terms due to Halton (1960).

---

[2] See Hajivassiliou, McFadden, and Ruud (1991), Geweke (1989), Keane (1990), Deak (1980a), and Gassmann (1988). Our primary reference is Borsch-Supan and Hajivassiliou (1992), which contains a clear exposition of the simulated maximum likelihood method and contrasts it with alternative approaches.

[3] This approach differs from the method of simulated moments (MSM) that minimizes the weighted "distance" between the moments of the simulated data and the implied moments of the parameter vector (McFadden, 1989; Pakes and Pollard, 1989; Stern, 1992).

Empirical results from the tax data suggest that out-of-pocket medical expenses are income inelastic; the burden of such costs is substantially higher for lower to middle income elderly families than for higher income families. The estimated coefficients also indicate persistence in medical care expenses; a one-dollar unexpected increase in current medical spending signals an overall discounted future increase of $2.80 in *additional* out-of-pocket expenses. Furthermore, the annual innovation in out-of-pocket expenses is substantial; a one-standard deviation positive shock to medical expenses is estimated to increase spending by $3000 in present value terms. Given that the median family holds only $6600 in liquid assets (Venti and Wise, 1991), these results suggest that many elderly families are at considerable financial risk from health costs.

In recent years, there has been increasing attention paid by politicians and the public to catastrophic medical expenses (Haas, 1988; Kosterlitz, 1988). While the risk of large-scale medical expenses have been documented for a single year, the more relevant measure of risk -- lifetime risk -- has not previously been measured. Using the estimated measures of medical care risk, we show that over a five-year period, persistent medical expenses can easily account for more than 20 percent of average annual income for an elderly family with an annual income of $15,000. Higher income families do not experience the same risk because out-of-pocket medical spending for elderly families is income inelastic. In light of this finding, the public outcry over federal catastrophic health insurance legislation in 1988 may have been a consequence of the highest income groups bearing the largest burden of the additional tax, but experiencing the least risk to disposable income.

The next section examines measures of existing out-of-pocket medical risk. Section III develops a simple ARMA model of medical spending, and describes how the model is estimated using the simulated likelihood estimator approach applied to a quadravariate tobit. Results are presented in Section IV, with simulations of the results shown in Section V. The paper concludes in Section VI.

Section II: Cross-Sectional Measures of Medical Care Spending

There have been recent efforts using cross-section data to gauge the importance of out-of-pocket medical expenses. The National Medical Expenditure Study (NMES) sampled roughly 14,000 households in 1977 to determine total and out-of-pocket expenses.[4] The distribution of out-of-pocket (OOP) medical expenses (including insurance premiums) in excess of 3 percent of family income are shown in Figure 1 for populations aged less than 65 and for those over age 65. We restrict the sample to families with incomes over $5000 (in $1987) to emphasize that high ratios are not the consequence of low income levels.[5]

The difference in OOP spending for elderly and non-elderly families is not dramatically different except for catastrophic expenses. For the population under age 65, 4.3 percent of the population spent at least 20 percent of their income on out-of-pocket medical expenses. By contrast, 7.9 percent of elderly families, and 14

---

[4] A more recent survey from 1987 exists, but the expenditure data has not yet been prepared for public use.

[5] The sample is further restricted to those who provided full survey information on insurance premiums, income, and other demographic data. A total of 9658 households remained in the data, of which 1963 were above age 64 (and with income above $5000).

percent of families with income less than $10,000, spent more than 20 percent of income on medical expenses (Figure 1).

Statistics from a study by the House Select Committee on Aging suggested that in 1977, *average* spending on out-of-pocket medical care was 12.3 percent. Why are the NMES statistics so much lower than the Congressional estimates? One reason is that the NMES neglects perhaps the most important uninsured medical care risk -- nursing home costs. Nursing homes are rarely covered by private insurance, and Medicare typically covers at most the first month. Many nursing home patients subsequently become eligible for Medicaid, but this is only after they have entirely spent down nearly all (nonhousing) wealth. The survey excluded all nursing home residents entirely, and costs for nursing homes were excluded even for respondents who had spent part of the year in a nursing home.

Nursing home costs are extensive and largely uncovered by any insurance. Cohen, Tell, and Wallack (1986) calculate that an individual at age 65 has a 45 percent chance of being admitted to a nursing home at least once before his or her death. Rivlin and Weiner (1988) estimate that single elderly men have a 30 percent chance of being admitted to a nursing home in any give year. While fewer than one-third of all admissions are for more than one year, the average (undiscounted) cost per admission is $36,125 assuming a daily cost of care equal to $85 (e.g., Rivlin and Weiner, 1988).[6] The advantage of the tax data is that, at least theoretically, uninsured

---

[6] Although most patients are discharged within 3 months, enough remain for lengthy stays (e.g., more than 5 years) that the average length of stay exceeds one year.

hospital, physician, and nursing home costs should be reflected in the declared medical expenses.

Consider first the sample of taxpayers age 65 and over who were exogenous itemizers (i.e., they would have itemized even if medical expenses had been zero) and who were either single or filing jointly. Figure 2 shows the distribution of medical expenses expressed as a fraction of AGI for elderly itemizing taxpayers with taxable income in excess of $2000, for tax years 1970 and 1988. Because the limit for medical expenses in 1988 was 7.5 percent of AGI, we focus only on medical expenses in excess of 7.5 percent in both years.[7] Note that these figures may not be entirely consistent with the NMES survey since social security payments are largely untaxed and would not be reflected in AGI. Still, the figure implies that 21.6 percent of elderly families in 1970 spent more than 7.5% of their income on out-of-pocket medical expenses, with 5.8 percent having spent more than 20 percent. By 1988, 25.4 percent of families spent more than 7.5% of income on medical expenses, with 7.8 percent having spent more than 20 percent. A more dramatic change in spending occurred for elderly families with AGI between $10,000 and $25,000 (expressed in 1987 dollars). Figure 2 shows, for example, that the percent of families in this income class spending between 20-40% of AGI on medical costs rose from 6.1 percent in 1970 to 11.5 percent in 1988. That is, the largest burden of increased out-of-pocket medical expenses appears to be on lower-middle income elderly families.

---

[7] In 1970, 18.5 percent declared expenses between 3-5% of AGI, and 10.4 percent declared expenses between 5-12.5% of AGI.

The overall level of medical spending in the tax data appears low relative to statistics from US Congress (1989). One reason is the higher income levels of those who itemize; in 1970, AGI for elderly itemizers, $35,055, (in $1987) was nearly double AGI for the elderly sample as a whole (see Tables 1 and 2). Calculations from the NMES indicate that the average propensity to spend income on out-of-pocket medical services (including insurance premiums) falls dramatically as income rises, from an average of 11.2 percent for all households with income between $5-10,000, to 4.8 percent for households with income between $40-60,000. Hence it is likely that the sample of itemizers will, on average, report lower medical expenses as a fraction of income than the representative household.

While the cross-sectional studies provide a measure of the pervasiveness of medical expenses in any given year, and trends in medical care spending over time, they cannot provide a measure of the lifetime risk of catastrophic medical care costs. To address this issue, we must turn next to a more structural model of medical care costs.

III. Time-Series Structure of Out-of-Pocket Medical Expenses

Health care expenditures reflect, but may not perfectly measure, objective changes in health status. As is well established, medical care expenses are the outcome of a complicated endogenous decision involving current health, investments in future health (e.g., Grossman, 1972), income, and relative prices determined by insurance coverage. Let the unobservable "health status" for individual i in period t be written ψ_{it}; log medical spending h^*_{it} is assumed to be

$$h^*_{it} = g(\psi_{it}, v_i, X_{it}) \tag{1}$$

where g is a function mapping into h^*_{it} intrinsic health status ψ_i, individual tastes v_i (which may also include unobservable permanent characteristics such as private insurance coverage), and observable characteristics X_{it} such as age and the tax-price of medical care.

Taking a first-order approximation around population means yields

$$h^*_{it} \approx [g(0,0,\overline{X}) + g_3(X_{it} - \overline{X})] + [g_1\overline{\psi}_i + g_2 v_i] + [g_1(\psi_{it} - \overline{\psi}_i)]$$

(2)

where g_j is the derivative of the function g with respect to the jth element, $\overline{\psi}_i$ is the individual-specific level of permanent health status, $\overline{X}$ is the population mean of the observable characteristics, and we normalize the population mean of v and ψ to be 0.

Log health spending is expressed in (2) as the sum of three bracketed terms. The first is the variation in health spending as a function of observable characteristics such as age, marital status, the tax price of medical deductions, and income. The second bracketed term measures the "fixed effect" of an individual family owing to either permanent differences in health status $g_1\psi_i$ or permanent differences in tastes towards medical spending $g_2 v_i$, perhaps as a consequence of differences in health insurance coverage. The final term measures time-specific differences in medical spending as a consequence of differences over time in health status. Of course, if there are time-specific changes in the function $g(\cdot)$, such changes would also be reflected in this final, time-varying error term.

Letting ϕ_1 be the fixed effect in the second brackets of (1), and v_{it} the time-specific variation in medical costs (shown in the third bracket of (2)), we may write the model as

$$h^*_{it} = X_{it}\beta + \phi_i + v_{it} \qquad (2)$$

where v_{it} is assumed to be an ARMA(p,q) error process. Assuming for the moment an ARMA(1,2) process, v_{it} can be written

$$v_{it} = \rho v_{it-1} + m_1 \varepsilon_{it-1} + m_2 \varepsilon_{it-2} + \varepsilon_{it}$$

where ρ is the autoregressive term, m_1 and m_2 the moving average terms, and ε_{it} is the current white-noise shock assumed to be distributed normally. The overall error term is denoted $u_{it} = \phi_i + v_{it}$. This general structure can capture serial correlation which decays geometrically over time (corresponding to $\rho > 0$ and $m_1 = m_2 = 0$), a "random walk" combined with "transitory" errors in medical expenses ($\rho \approx 1$ and $m_1 < 0$), or a model in which most errors simply reflect permanent individual differences in health status ($\sigma^2_\phi >> \sigma^2_v$).

In the typical case when all data are observed, this model may be solved by first estimating β using OLS, and then estimating the ARMA model from the covariance structure of the residuals using standard minimum distance estimators (e.g., Abowd and Card, 1989). The drawback of the taxpayer panel is that information on medical spending is available only for those who both (i) itemized and (ii) spent more than 3 percent of adjusted gross income on medical care.

The problem of estimating an unbiased parameter vector $\theta = \{\beta, \sigma^2_\psi, \rho, m_1, m_2, \sigma^2_\varepsilon\}$ is that medical spending will be underreported when it is low (i.e., less than 3 percent of AGI) or when other deductions are sufficiently low that even high medical expenses are not sufficient to justify itemizing. That is, writing the model of medical expenses as

$$H^* = X_T\beta + U$$

where H^* is the true Tx1 vector of medical spending in the T years, X_T is the (stacked) vector of independent variables for T years, and

U is the Tx1 vector of error terms. Observed log medical spending is H, where the element of H is defined as

$$
\begin{aligned}
h_{it} &= h_{it}^* \quad if \quad h_{it}^* \geq \ln(.03 \times Y_{it}) \\
&\qquad\qquad and \quad d_{it} \geq S_t \\
h_{it} &= 0 \qquad\qquad otherwise
\end{aligned}
\tag{4}
$$

That is, observed medical spending is zero unless (i) total itemized deductions d_{it} exceeds the standard deduction in year t, S_t, and (ii) log medical spending h_{it}^* exceeds the (log) limit equal to 3 percent of AGI.

The covariance matrix $\Omega \equiv E\{uu'\}$ is a function of the time-series parameters of the dynamic model. For example, in the ARMA(1,2) model considered above, the diagonal elements of Ω are written

$$
\sigma_u^2 = \sigma_\phi^2 + \frac{\left[1 + m_1^2 + m_2^2 + 2[\rho m_1 + m_2(m_1 + \rho)]\right]\sigma_\epsilon^2}{1 - \rho^2}
\tag{5}
$$

It is clear from the model above that we cannot simply restrict our attention to those who itemize, since the selection criterion for whether medical expenses are observed depends on total deductions d_{it}, of which medical spending is an element. Put another way, an unusually large error term in h_{it}^* will increase the probability that the taxpayer itemizes, which in turn biases the estimated coefficients. To correct for this potential bias, we restrict the sample to taxpayers who are "exogenous" itemizers: in the absence of medical expenses, they would have itemized anyway. This means that the error term u_{it} will have no impact on the selection of the data. Of course, some efficiency is lost by restricting the sample to itemizers. A more

complete model would include an equation explaining the decision to itemize, a topic beyond the scope of this paper.

Another issue is whether tobits are an appropriate estimation approach in light of the restriction that the structural estimates predict both the probability of observing positive medical expenses and the amount of medical expenses conditional on being observed. This assumption is clearly violated if, for example, there are large fixed costs to recording medical expenses above 3 percent of AGI, in which case a multidimensional version of the Heckman (1979) procedure might be preferred. Medical deductions, however, are one of the few cases where tobit estimators may be appropriate. In the sample of taxpayers, individuals are itemizing already, so there is no additional cost of keeping, say, state tax records. Furthermore, there is rarely a threshold cost of spending more than 3 % of AGI (assuming that individuals even know what their AGI will be). The only case in which there could be fixed costs of filing for medical care is if record keeping is costly.[8]

The estimation of even this simple model is nontrivial. Letting T be the number of years in the panel, there are 2^T possible combinations of being above, or below, the 3 percent cutoff point. For example, consider the case in which T = 4 and the first two years are observed but the last two years are not; the likelihood of observing M given X and the vector of parameters θ is

---

[8]   Another possible shortcoming of the model is that some taxpayers may not itemize even if they are eligible because of the compliance cost of record-keeping (Pitt and Slemrod, 1989).

$$P(H|X,\Omega) \sim f(u_1,u_2|\Omega)F(a_3-x_3\beta,a_4-x_4\beta|\Omega,u_1,u_2) \qquad (6)$$

where *f* is the density and F the distribution function. Solving for the likelihood function is straightforward but tedious when the distribution function F (conditional on u_1 and u_2) is of dimensionality less than 3 (in the case above, it is bivariate). The problem comes when the distribution function F is in four or more dimensions, making numerical evaluations infeasible (Deak, 1980b).[9]

To address this problem, Borsch-Supan and Hajivassiliou (1992) use the SSML method to integrate out the n-dimensional distribution functions using simulation methods. (A discussion of this and other approaches to Monte Carlo simulation methods is provided in Appendix A.) Their method is easiest to see when medical expenses in all years lie below 3 percent of AGI. Then we known that

$$U < A - X\beta \equiv A^* \qquad (7)$$

That is, the unobserved error terms are in a 4-dimensional region bounded from above by the vector $\{a_t\} = \ln(.03 \times Y_t) - X_t\beta$ where Y_t measures adjusted gross income for individual i in year t (individual subscripts have been dropped for simplicity).

The trick in evaluating the four-dimensional distribution function in (7) is to note that it can also be expressed as the product of conditional univariate distribution functions $\Phi(\cdot)$;

$$F(A^*) \sim \Phi(a_t|u_{t-1},u_{t-2},u_{t-3})\,\Phi(a_{t-1}|u_{t-2},u_{t-3})\,\Phi(a_{t-2}|u_{t-3})\,\Phi(a_{t-3})$$

---

[9] Although see Hausman and Wise (1978).

13

By appropriate sampling over the unobservable error terms u_k, it is possible to approximate $F(A^*)$ by simulation methods. But it is difficult to sample randomly from a truncated arbitrary covariance structure. To do this, first take the Cholesky decomposition, L, that satisfies $L'L = \Omega$. Then one can express samples of correlated random terms $\hat{u}$ as $\hat{u} = L\hat{e}$, where $Var(\hat{u}) = \Omega$ and $\hat{e}$ is a vector of independent standard normal random draws. A slightly harder problem is to draw random error terms from the truncated region in (7). The SSML approach does this by exploiting the lower triangular structure of L. Because

$$u_t - \sum_{k-1}^{t} \ell_{tk} e_k \qquad (8)$$

where ℓ_{ij} is the (lower triangular) element of L, the inequality constraint (7) can be rewritten as

$$e_1 - \frac{u_1}{\ell_{11}} \leq \frac{a_1^*}{\ell_{11}} = \alpha_1$$

$$e_2 - \frac{u_2 - \ell_{21} e_1}{\ell_{22}} \leq \frac{a_2^* - \ell_{21} e_1}{\ell_{22}} = \alpha_2$$

$$e_3 - \frac{u_3 - \ell_{31} e_1 - \ell_{32} e_2}{\ell_{33}} \leq \frac{a_3^* - \ell_{31} e_1 - \ell_{32} e_2}{\ell_{33}} = \alpha_3$$

(9)

.
.
.

That is, the distribution function $F(A^*|\Omega)$ is transformed to $F(a^*|I)$ where I is the (4x4) identity matrix. It is easier to simulate this latter distribution function; one simply samples univariate truncated normal error terms by use of the appropriate inverse cumulative density function.[10] The recursive nature of L implies a sequential process for random sampling; first e_1 from $\Phi(a_1)$, then e_2 from $\Phi(a_2)$, and so forth. Given R random draws of the Tx1 vectors, the probability of the observed data given X and θ are

$$P(H|X,\Theta) = \frac{1}{R} \sum_{k-1}^{R} \left[\prod_{s-1}^{T} \Phi(\alpha_s) \right] \tag{10}$$

Preliminary results from Hajivassiliou, McFadden, and Ruud (1991) suggest that this approach yields accurate approximations relative to other simulation methods.

The model is easily extended to the case of a tobit. Suppose that medical spending is observed only in the second year. Conditional on the first draw, we can derive the implied (nonrandom) error term e_2:

$$e_2 = \left[\frac{u_2 - \ell_{21}e_1}{\ell_{22}} \right] \tag{11}$$

where u_2 has been calculated from the observed data conditional on θ. Note that e_2 still varies over the R random draws because for a

---

[10] In the case considered, Borsch-Supan and Hajivassiliou draw on results by McFadden and Ruud to show that setting
$$e_t = F^{-1}(F(A_t)e_{ut}),$$
where e_{ut} is uniform on (0,1), ensures a random normal iid draw from the truncated distribution.

given u_2, e_2 depends on the random draw e_1. Since u_2 is known, the distribution function $\Phi(\sigma_2)$ is replaced with the density function $(\sqrt{(2\pi)}\sigma^*)^{-1} f(u_2/\sigma^*)$ where σ^* is the standard error of u_2 *conditional* on e_1.[11] Of course, if all T years are observed, then the contribution to the likelihood function reduces to a standard T-dimensional density.

The standard approach to simulated methods of estimation is to use Monte Carlo or antithetic variance methods of sampling error terms. The crude Monte Carlo approach is generally not the most efficient means of numerical approximation because it does not prevent the "bunching" of draws in a particular region of the function, especially when the number of draws is small. As a consequence, Monte Carlo sampling converges slowly at a rate of $1/\sqrt{N}$. Antithetic variance methods, in which e_j (with mean zero) is randomly chosen but the next draw $e_{j+1} = -e_j$, improves the efficiency somewhat. But there are a number of more efficient procedures that provide a faster convergence rate for a broad class of well-behaved functions. We adopt the classical Hammersly points which Halton (1960) has shown has a worst-case convergence rate on the order of $(\ln n)^{d-1}/n$ where d is the dimensionality of the integral. In the average case, however, the core convergence is much faster (Wozniakowski, 1991).

Halton (1960) sequences are sometimes referred to as "pseudorandom" because they are not randomly drawn sequences,

---

[11] In general, the conditional variance is

$$\sigma_t^* = \sigma_t\sqrt{1-R_{1,\ldots,t-1}}$$

where $R_{1,\ldots,t-1}$ is the partial R^2 of the tth error term conditional on knowing the previous t-1 error terms.

but are predetermined so as to provide an efficient span of the probability space. However, they are designed to be independent both within a particular sequence and across sequences. Each Halton sequence is calculated based on a different prime number. In our case, we require a 3-dimensional vector of random draws, so we construct Halton sequences for prime numbers 3, 5, and 7. The sequences are shown in appendix B along with the GAUSS computer program to generate them. Gassmann (1988) presents evidence suggesting that a similar technique for improving the simulation of a multinomial can lead to substantial efficiency gains relative to the traditional Monte Carlo or anitithetic variance approach.

IV. Data and Estimation

The data are a panel of 21,343 tax returns for the tax years 1967-73, although we can only use the four years (1968, 1970, 1972, and 1973) in which medical deductions are reported. Membership in the panel is determined by the last four digits of the primary taxpayer's Social Security number (two four-digit numbers were chosen, leading to a 2-in- 10,000 sample). With this design, taxpayers may enter or leave the sample in any year, unlike the usual panel that experiences only attrition. In particular, women leave the sample upon marriage and re-enter if widowed or divorced. The information available for each taxpayer includes virtually all the entries on the 1040 form plus many items from supporting forms. These data were matched by Charles Clotfelter to social security data on the primary taxpayer's age.

The data from the IRS has advantages and disadvantages relative to survey data. One disadvantage (aside from the censoring problem) is that taxpayers may overstate medical expenses, since

there are clear financial incentives to do so. On the other hand, there are also clear financial penalties for overstating medical expenses. Relative to survey data, we expect to find an upward bias in the level of out-of-pocket medical expenses, but also lower measurement error. Medical expenses allowed by the IRS encompass a wider range of services than standard cross-sectional surveys. However, this more comprehensive measure of the cost of medical illness may be more appropriate for measuring the financial impact of long-term illness.

The balanced sample is limited to those who filed for each of the four years. We miss households who did not need to file for any of the years, or households who filed very late. Losing nonfilers will bias the sample towards higher income families. However, for many of the elderly in the sample, we will also tend to understate total income since social security was not included in AGI at the time.

Married couples filing separately are excluded, as are households with age or sex missing, or with adjusted gross income less than zero. All dollar amounts are converted to $1987 using the GNP consumption deflator. The dummy variable for sex is equal to one if the taxpayer is an unmarried female, and the variable for marital status is equal to one if married filing jointly or a surviving widow filing in that year. The excluded variable is a male single head of household.

Summary statistics are provided in Table 1 for the entire sample of households aged 55 or older in 1968. Average income declined during the period from $25,178 in 1968 to $21,248, probably as a consequence of retirement among some of the sample. Perhaps the most notable change during the period of analysis is the sharp drop in itemized deductions between 1970 and 1972. As Peele (1991) has noted, this is the consequence of an increase in the

standard deduction from $1000 to $2000 in the 1972 tax year. This contributed to a decline in taxpayers taking a medical deduction (from 53 percent of all taxpayers in 1970 to 34 percent in 1972) and an increase in the value of medical deductions conditional on taking the deduction.

Table 2 presents summary statistics for the balanced sample; those who are exogenous itemizers (i.e., they would have filed even if medical deductions had been zero) and with AGI in excess of $2000 for each of the four years. Average income for itemizers is nearly double that for the general population of older filers. Roughly three-quarters of these itemizers declared medical expenses in each year.

As a first step, we run univariate tobit regressions for each year. This allows us to include a larger sample of taxpayers because we need not exclude observations if they were missing from other years. Table 3 presents the coefficient estimates from the univariate tobits for taxpayers who were exogenous itemizers and reported AGI in excess of $2000. The income elasticity of medical expenses is usually quite low; in all years expect 1970, the coefficient estimate is less than 0.35. The net of tax price for these same three years is significant and greater than one in absolute value. The coefficient in 1968, for example, can be interpreted as implying that a 10 percentage point increase in the marginal tax rate will lead to a 14 percent increase in medical spending.[12]

One difficulty with including the marginal tax rate is that it is essentially a nonlinear function of AGI, so that the tax price may be

---

[12] The date of death (through 1976) was also tried in these regressions, but the estimated coefficients were insignificant.

reflecting nonlinear effects of income (Feenberg, 1987). The divergence in the income and price elasticities between 1970 and other years, for example, could be the consequence of instability in coefficient estimates for two highly correlated variables. However, similar regressions that exclude the tax price yield the same pattern; the income elasticity in 1970 is considerably higher than in other years.

Chesher and Irish (1987) and others have suggested that synthetic residuals from limited dependent variables may be used to test for omitted variables and distributional assumptions. While such residuals cannot be used to estimate unbiased covariance structures, they are still useful for assessing graphically the extent of covariance over time in medical expenses. As before, suppose that u_{it} is the residual for family i in year t conditional on medical expenses being reported. The residual for those who did not report medical expenses can be imputed as the expected value of the error term;

$$\hat{u}_{it} = \frac{f(a_{it}^* - x_{it}\beta)}{1 - F(a_{it}^* - x_{it}\beta)}$$

Using the coefficients from the univariate tobits, we compare the actual or imputed residuals in 1972 and 1973 using Figure 3a. Both the horizontal and vertical axis are in log scales, and the diamonds are the residuals for families that did not declare any medical expenses in either year. The empty circles measure positive medical spending in 1972, the crosses positive medical spending in 1973, while the filled circles represent positive medical spending in both years. It is apparent from Figure 3a that there is a very strong year-to-year correlation in medical expenses conditioning on the

exogenous factors such as income, age, and filing status. Relatively few households declared medical expenses in one year but not in the other.[13]

How persistent are these medical expenses for the 5-year span 1968-73 provided by the panel? Figure 3b graphs both actual and imputed residuals for the years 1968 and 1973. While the covariance structure is not nearly as pronounced (i.e., there are more residuals in the northwest and southeast quadrants), there is still evidence of long-term correlation.

Table 4 presents results from the full quadravariate tobit estimation with the pooled panel data, using 10 draws of the error term.[14] Column [1] begins with the regression equation similar to the single-year tobits in Table 3, although individual year dummy variables are included. Rather than estimating the ARMA model, we simply estimate the unrestricted covariance matrix. The income elasticity, 0.06 is very small (and insignificant), while the tax price effect is large and significant. While the age coefficients are not significant individually, they predict a generally rising profile of medical costs. Finally, presence of a spouse increases log medical expenses by 0.416 (or a more than 50 percent increase), while the

---

[13] Note that the correlation between 1972 and 1973 for the diamonds (those who did not report medical expenses in either year) is a consequence of the time-series correlation in X, the independent variables, rather than the correlation in the error term u.

[14] We drew from the 11-20th elements of the Halton sequences to avoid a potential problem with the first few elements of the sequences (see Appendix A). Increasing the number of draws, R, to 50 had no effect on the coefficient estimates.

difference in medical spending between a male-headed and female-headed (single) household is estimated to be only 2 percent.

The implied covariance structure of the model is shown in Figure 4 by the symbol $\mathbf{X}$, ranging from the log variance (at year 0) to the 5-year covariance (at year 5).[15] There is considerable correlation over time; the correlation coefficient in out-of-pocket medical spending between 1968 and 1973, for example, is 0.40. This contrasts with similar calculations by Newhouse et al (1989) for total medical spending using data from the Rand Health Insurance Experiment. The overall 5-year correlation coefficient was only 0.045, although the equivalent correlation coefficient for outpatient services was 0.45. Their results are somewhat consistent with ours if outpatient services are largely uninsured, and hence paid out-of-pocket. (Also, the Rand data was drawn from the entire population rather than being limited to those over age 54.)

We begin by estimating an general ARMA(1,2) model without a fixed effect, with coefficients reported in Column [2] of Table 4. The estimated coefficients suggest considerable persistence; the AR(1) component is 0.908 (and significantly different from both 0 and 1). The coefficients m_1 and m_2 are -.401 and -.122 respectively, implying that roughly half of a health shock in year t is dissipated within two years. The estimates also imply a relatively large annual innovation in medical expenses; the log standard deviation of ε is 0.435.

---

[15] The coefficients (and t-statistics), ranging from $E(u_t u_t)$ to $E(u_t u_{t+5})$ are .367 (17.37); .251 (11.58); .208 (10.25); .189 (8.71); .194 (8.78); .146 (6.52).

To see whether this specification captures the covariance structure of the error terms, we compare in Figure 4 the predicted covariance terms (the diamond) from column [2] with the unrestricted covariance (the X). Except for the 4th and 5th year, the correspondence is nearly exact. Alternative specifications, such as an ARMA(2,0) with a fixed effect, yields lower log likelihood values. Hence we restrict our attention to the ARMA(1,2) specification.[16]

As noted above, it is not entirely possible to jointly estimate the tax price and the income elasticity of medical spending. Column [3] therefore excludes the tax price, yielding an income elasticity of .38, which is highly significant; other coefficients were unaffected.

The question remains whether the large degree of estimated persistence in the ARMA model is just a consequence of permanent differences across individuals. In Column [4], the full model with individual-specific variation σ_ϕ^2 is introduced, although to ensure that the estimated variance of ϕ is positive, we actually calculate $\log(\sigma_\phi^2)$. The point estimate of the individual-specific variation is $e^{-13.7}$, and its exclusion has no effect on the log likelihood function (i.e., column [3] vs [4]). There is no evidence suggesting the presence of a fixed effect in medical spending.

A more general model of health status and medical expenses would include the joint determination of medical spending and income. Costly illness may precipitate early retirement, or depletion

---

[16] The potential gain from expanding the ARMA(p,q) model is limited by the difference in the likelihood function between [1] and [2], 2.6. Hence any additional autoregressive or moving average term is unlikely to reject a χ^2 test. We also resisted more complicated ARMA models, such as an ARMA(2,2), because of the exceedingly complicated and nonlinear covariance terms.

of existing assets for medical expenses could then reduce subsequent income. Testing for such endogeneity is difficult given that the truncated nature of the data. Still, we can allow for this potential endogeneity by simply excluding annual income, and allowing differences in permanent income to be reflected in the individual effect. The estimated coefficients, presented in Column [5] of Table 4, suggests even greater persistence in medical expenses, with the AR coefficient equal to 0.936. The individual-specific effect, however, is still essentially zero. Finally, Column [6] presents results for the sample of families with a household head at least 65 in 1968. Out-of-pocket medical spending rises past age 65, and unmarried single women are predicted to spend more on medical expenses than their unmarried male counterparts. The ARMA parameters are quite similar to those estimated for the full sample, with only somewhat less persistence in medical expenses.

To see the implied time-series patterns of medical care spending implied by these estimates, consider a shock in the innovation ε_t equal to one standard deviation. Figure 4 traces out the incremental effect of this shock on u_{t+j}, $j > 0$, for the estimated ARMA coefficients in column [4], the asterisk, and column [5], the empty box. Note that the vertical axis now corresponds to the log error term u_t (rather than the log variance, as in the discussion above). Both sets of coefficients imply substantial short-term fluctuations in medical spending; for the coefficients from column [4] the log standard deviation of the innovation ε is 0.435. Evaluated at the predicted mean (log) medical spending for a married couple age 64 with an income of \$25,178 (in 1968, from Table 1), the one-standard-deviation range in predicted next-period annual spending is

between $962 and $2,295. For any given shock to spending, roughly one-quarter remains after 5 years.

There are two other ways of quantifying the degree of persistence in medical spending. One is to note that a $1 shock in medical spending is predicted to cause a $2.80 increase in additional medical spending in the future, discounted at a 3% interest rate. Alternatively, a one-standard-deviation positive "shock" in the innovation ε generates, on average, a more than $3000 increase in the present value of out-of-pocket medical expenses (also discounted at 3%). Since the median family at ages 60-65 own only $6,600 in liquid financial wealth (Venti and Wise, 1991), it is clear that the persistence of out-of-pocket medical spending represents a substantial risk to retirement saving by the elderly.

The next section uses the estimated coefficients from Table 4 to compare the risk of medical expenditures -- over a 5-year period -- for lower and higher income families. The differences in both the level and the risk of out-of-pocket medical expenses among income groups could explain why 1988 catastrophic health care bill met with such resistance.

V. Annual and Lifetime Medical Spending Risk

This section simulates the lifetime risk of out-of-pocket expenses using coefficients from column [4] in Table 4 for two income groups; those with AGI of $15,000 and those with $30,000. We assumed that each household was married with a male head. To calculate medical expenses over a 5-year period, we used the monte-carlo program @RISK to trace out the distribution of medical costs, averaged over the five-year period. The random error term chosen in the initial age, assumed to be 66, was from the unconditional

distribution, while subsequent error terms were chosen from the distribution conditional on the previous year.

Figure 5a shows the distribution of the predicted average medical spending over 5 years, expressed as a ratio of the assumed constant income level of $15,000 (in $1987). Average spending on medical services is 11 percent, with 9.8 percent of families spending, on *average*, more than 20 percent of income on out-of-pocket medical expenses. By contrast, average out-of-pocket medical spending for those with an income of $30,000, shown in Figure 5b, is only 7 percent. Furthermore, the distribution is much narrower, with less than 2 percent of families spending, on average, more than 20 percent of their income on out-of-pocket medical expenses.

The differences in both the mean and the variance of spending may provide one explanation for why the 1988 catastrophic health care legislation was revoked only months after being passed. The tax structure was progressive, with higher income families paying up to $1600 per year in premiums. Given that such families are at less risk from catastrophic medical expenses, relative to their incomes, the insurance value of catastrophic medical care may not have justified their increased tax bill.[17]

VI. Conclusion

Catastrophic medical expenses are perhaps the most important economic risk facing the elderly. This paper addresses the question of whether out-of-pocket medical expenses are persistent

---

[17] The catastrophic health bill also did not cover long-term care, perhaps the source of the greatest financial risk to upper-income families.

over time. We use information on medical deductions drawn from a survey of taxpayer returns during 1968-73 supplied by the IRS. Medical expenses can only be deducted in excess of 3 percent during the period of our study, so we use a quadravariate tobit, estimated using a smoothed simulated maximum likelihood technique, to correct for censoring bias. We find considerable persistence in medical expenditures; a one dollar shock in annual medical expenses will, on average, be followed by an increase in medical spending of $2.80 in the future.

There are some indications that the multivariate tobits may be missing a substantial fraction of the catastrophic health spending because such families may not file taxes in such years or be excluded from the sample as endogenous itemizers. For example, in the sample of 354 exogenous itemizers, the maximum out-of-pocket medical expense in any year was $14,096, while the maximum payment among endogenous itemizers was $43,868. Hence we may be truncating the top tail of the distribution.

The impact of uninsured risk affects more than the demand for catastrophic health care. Kotlikoff (1986) developed theoretical models suggesting that health risk could generate precautionary saving. Levin (1990) found evidence suggesting that families who were not well-insured against health risks also tended to maintain higher levels of precautionary saving. Palumbo (1991) has used the Panel Study of Income Dynamics to estimate empirically a model of consumption and saving subject to health risk. Finally, Hubbard, Skinner, and Zeldes (1992) have attempted to use parameters of health expense risk and earnings risk to explain the skewed distribution of wealth holdings. The estimated parameters of health risk estimated from tax data may provide a better picture of the true

dynamic risk faced by the current elderly as well as the future financial risks of currently younger families.

Bibliography

Abowd, John M., and David Card, "On the Covariance Structure of Earnings and Hours Changes," Econometrica 57 (March 1989): 411-445.

Borsch-Supan, Axel, and Vassilis Hajivassiliou, "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," Journal of Econometrics (1992, forthcoming).

Chesher, Andrew, and Margaret Irish, "Residual Analysis in the Grouped and Censored Normal Linear Model," Journal of Econometrics 34 (1987): 33-61.

Cohen, Marc A., Eileen J. Tell, and Stanley S. Wallack, "The Lifetime Risks and Costs of Nursing Home Use Among the Elderly," Medical Care 24 (December 1986): 1161-72.

Deák, I, "Computation of Multiple Normal Probabilities," in P. Kall and A. Prekopa (eds.) Recent Results in Stochastic Programming. New York: Springer-Verlag (1980a).

Deák, I, "Three Digit Accurate Multiple Normal Probabilities," Num. Math. 35 (1980b): 369-80.

Feenberg, Daniel, "Are Tax Price Models Really Identified: The Case of Charitable Giving," The National Tax Journal 40(4) (December 1987): 629-633.

Gassmann, H., "Conditional Probability and Conditional Expectation of a Random Vector." in Yu. Ermoliev and R. J-B. Wets (eds) Numerical Techniques for Stochastic Optimization. New York: Springer-Verlag (1988).

Geweke, J., "Efficient Simulation from the Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evaluation of Constraint Probabilities," mimeo, Duke University (1989).

Grossman, Michael, <u>The Demand For Health: A Theoretical and Empirical Investigation</u>. Occasional Paper No. 119. New York: Columbia University Press for NBER (1972).

Haas, Lawrence J., "Fiscal Catastrophe," <u>National Journal</u> (October 7, 1988): pp. 2453-56.

Hajivassiliou, Vassilis, Daniel McFadden, and Paul Ruud, "Simulation of Multivariate Normal Orthant Probabilities: Methods and Programs," mimeo (1991).

Halton, J., "On the Efficiency of Certain Quasirandom Sequences of Points in Evaluating Multidimensional Integrals," <u>Num Math</u> (1960) : 84-90.

Hausman, Jerry A., and David Wise, "Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," <u>Econometrica</u> 47 (1978): 403-426.

Heckman, James J., "Sample Selection as a Specification Error," <u>Econometrica</u> (1979): 153-61.

Hubbard, R. Glenn, Skinner, Jonathan, and Stephen P. Zeldes, "Social Insurance and Precautionary Saving," mimeo (March 1992).

Johnson, Norman L., and Samuel Kotz, <u>Distributions in Statistics: Continuous Multivariate Distributions</u>. New York: John Wiley and Sons (1972).

Keane, Michael, "A Computationally Practical Simulation Estimator for Panel Data, with Applications to Estiamting Temporal Dependence in Employment and Wages," mimeo, University of Minnesota (July 1990).

Kosterlitz, Julie, "Catastrophic Coverage a Catastrophe?" <u>National Journal</u> (November 19, 1988) pp. 2949-50.

Kotlikoff, Laurence J., "Health Expenditures and Precautionary Saving," NBER Working Paper No. 2008 (August 1986).

Levin, Lawrence, "On the Demand for Health Insurance Among the Elderly," mimeo, Santa Clara University (May 1989).

McFadden, Daniel, "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," Econometrica 57 (September 1989): 995-1026.

Newhouse, Joseph P., Willard G. Manning, Emmett B. Keeler, and Elizabeth M. Sloss, "Adjusting Capitation Rates Using Objective Health Measures and Prior Utilization," Health Care Financing Review 10 (3) (Spring 1989): pp. 41-53.

Pakes, Ariel, and David Pollard, "Simulation and the Asymptotics of Optimization Estimators," Econometrica 57 (September 1989): 1027-1058.

Palumbo, Michael, "Health Risk and Consumption of the Elderly," mimeo, University of Virginia (1991).

Peele, Pamela, "A Review of 'The Risk and Duration of Catastrophic Health Care Expenditures'", mimeo, VPI & SU, Blacksburg, VA (1990).

Pitt, Mark, and Joel Slemrod, "The Compliance Cost of Itemizing Deductions: Evidence From Individual Returns," American Economic Review 79 (December 1989): 1224-1232.

Rivlin, Alice M., and Joshua M. Wiener, Caring for the Disabled Elderly: Who Will Pay? Washington DC: Brookings Institution (1988).

Stern, Steven, "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models," Econometrica (1992, forthcoming).

US Congress, "Health Care Costs for America's Elderly, 1977-88," Select Committee on Aging, House of Representatives 101-712 (March 1989).

Venti, Steven F., and David A. Wise, "Aging and the Income Value of Housing Wealth," Journal of Public Economics 44 (1991): 371-97.

Wozniakowski, H., "Average Case Complexity of Multivariate Integration," Bulletin of the American Mathematical Society 24 (1991): 185-194.

Table 1: Summary Statistics: All Households
Age 55 and Above in 1968

Variable	1968	1970	1972	1973
Average Adjusted Gross Income (AGI)	25,178 (21,367)	23,550 (20,180)	22,496 (20,693)	21,248 (20,593)
Median AGI	20,533	18,504	16,851	15,242
Tax Rate	20.78 (8.54)	21.26 (9.80)	18.71 (9.67)	17.54 (10.09)
Age	62.5 (7.0)	64.5 (7.0)	66.5 (7.0)	67.5 (7.0)
Itemizers	0.588 (0.49)	0.772 (0.42)	0.384 (0.49)	0.364 (0.48)
% Medical Expenses > 0	0.485 (0.50)	0.533 (0.50)	0.336 (0.47)	0.322 (0.47)
Aver. Med. Expenses [1]	1906 (2046)	1896 (2289)	2400 (3126)	2526 (3700)
Number of Dependents	0.36 (0.80)	0.27 (0.70)	0.20 (0.59)	0.16 (0.52)

Notes: N = 1675. All dollar amounts normalized to $1987.

[1] Average of medical expenses conditional on itemizing (includes "endogenous" itemizers)

Table 2: Summary Statistics: Exogenous Itemizers
Age 55 and Above in 1968

Variable	1968	1970	1972	1973
Average Adjusted Gross Income (AGI)	44,140 (31,067)	41,148 (28,569)	41,115 (30,203)	39,468 (31,583)
Median AGI	35,656	35,055	32,799	30,774
Tax Rate	26.27 (10.12)	27.12 (11.05)	26.02 (10.14)	24.90 (10.94)
Age	61.0 (6.1)	63.0 (6.1)	65.0 (6.1)	66.0 (6.1)
% Medical Expenses > 0	0.754 (0.43)	0.788 (0.41)	0.799 (0.40)	0.819 (0.39)
Aver. Med. Expenses [1]	2331 (1788)	2121 (1519)	2157 (1613)	2050 (1280)
Number of Dependents	0.61 (1.01)	0.46 (0.89)	0.32 (0.74)	0.27 (0.65)

Notes: N = 354. All dollar amounts normalized to $1987. Balanced sample limited to those who itemized in each of the 4 years and with AGI > $2000 in each year.
[1] Average of medical expenses conditional on itemizing.

Table 3: Univariate Tobit Regressions

Variables	1968	1970	1972	1973
Log (Adjusted Gross Income)	0.338 (5.92)	1.179 (18.86)	0.227 (3.35)	0.152 (1.99)
Age	-0.037 (0.93)	-0.018 (0.44)	-0.065 (1.58)	0.023 (0.54)
Age^2 (x10^{-2})	0.034 (1.16)	0.016 (0.51)	0.059 (1.93)	-0.009 (0.29)
Sex	0.256 (3.17)	0.375 (4.22)	0.093 (0.89)	0.109 (0.97)
Marital Status	0.516 (5.99)	0.490 (5.33)	0.375 (3.73)	0.432 (3.94)
Log (Net of Tax Price)	-1.405 (4.16)	-0.390 (1.15)	-1.308 (4.00)	-1.464 (4.00)
Constant	3.540 (2.41)	-5.439 (3.46)	5.818 (3.75)	3.658 (2.25)
σ_u^2	0.489 (43.88)	0.974 (47.28)	0.341 (36.25)	0.402 (36.99)
N	1328	2262	910	945

Notes: Dependent variable is log of medical care expenses. All dollar amounts normalized to $1987. Absolute value of t-statistics in parentheses. Net of tax price is (minus) the first-dollar tax on medical expenditures.

Table 4: Quadravariate Tobit Regressions

Variables	[1] Age > 54	[2] Age > 54	[3] Age > 54	[4] Age > 54	[5] Age > 54	[6] Age > 64
Log (Income)	0.061 (0.96)	0.065 (1.24)	0.384 (13.33)	0.384 (15.23)		0.349 (7.29)
Age $(\times 10^{-2})$	-0.899 (0.38)	-1.356 (0.46)	0.540 (3.86)	0.102 (0.31)	-2.180 (0.40)	-16.650 (1.27)
Age^2 $(\times 10^{-4})$	2.392 (0.66)	2.014 (0.95)	0.702 (3.28)	1.026 (3.04)	2.776 (0.71)	12.574 (1.48)
Sex	0.021 (0.19)	0.012 (0.12)	0.064 (0.60)	0.064 (0.65)	0.018 (0.15)	0.416 (2.35)
Marital Status	0.416 (4.61)	0.404 (5.25)	0.257 (3.26)	0.257 (3.41)	0.332 (3.94)	0.502 (3.93)
Net of Tax Price	-2.295 (5.68)	-2.300 (6.80)				
$\ln(\sigma^2_\psi)$	*			-13.678 (1.16)[1]	-15.550 (0.67)[1]	-76.68 (1.36)[1]
ρ	*	0.916 (39.83)	0.908 (34.01)	0.908 (50.53)	0.936 (49.26)	0.892 (15.79)
m_1	*	-0.401 (5.74)	-0.440 (6.53)	-0.440 (7.33)	-0.358 (4.80)	-0.598 (4.71)
m_2	*	-0.122 (1.72)	-0.058 (0.64)	-0.058 (1.33)	-0.165 (1.86)	0.073 (0.53)
σ^2_e	*	0.180 (18.56)	0.189 (17.99)	0.189 (21.72)	0.193 (15.82)	0.208 (11.29)
Constant	8.306 (4.25)	8.083 (7.48)	2.509 (8.23)	2.657 (8.23)	7.318 (4.02)	8.855 (1.78)
Year = 1970	-0.099 (2.82)	-0.099 (3.09)	-0.067 (1.65)	-0.067 (3.05)	-0.081 (2.23)	-0.098 (1.40)
Year = 1972	-0.047 (1.17)	-0.048 (1.25)	-0.041 (0.94)	-0.040 (3.03)	-0.062 (1.46)	-0.021 (0.25)

|						
Year = 1973	-0.031 (0.67)	-0.034 (0.82)	-0.034 (0.60)	-0.033 (2.68)	-0.071 (1.52)	-0.122 (1.35)
Log Likelihood	- 1014.7	- 1017.3	- 1033.6	-1033.6	- 1107.3	-235.2

Notes to Table 4

N = 354 for age > 54 and N = 79 for age > 64. The dependent variable is log of medical care expenses. All dollar amounts are in $1987. Absolute value of t-statistics in parentheses. Net of tax price is (minus) the first-dollar tax on medical expenditures.
* Unrestricted covariance shown in Figure 4.
1 t-statistic of the log variance.

Figure 1: Distribution of
Out-Of-Pocket Medical Expenses
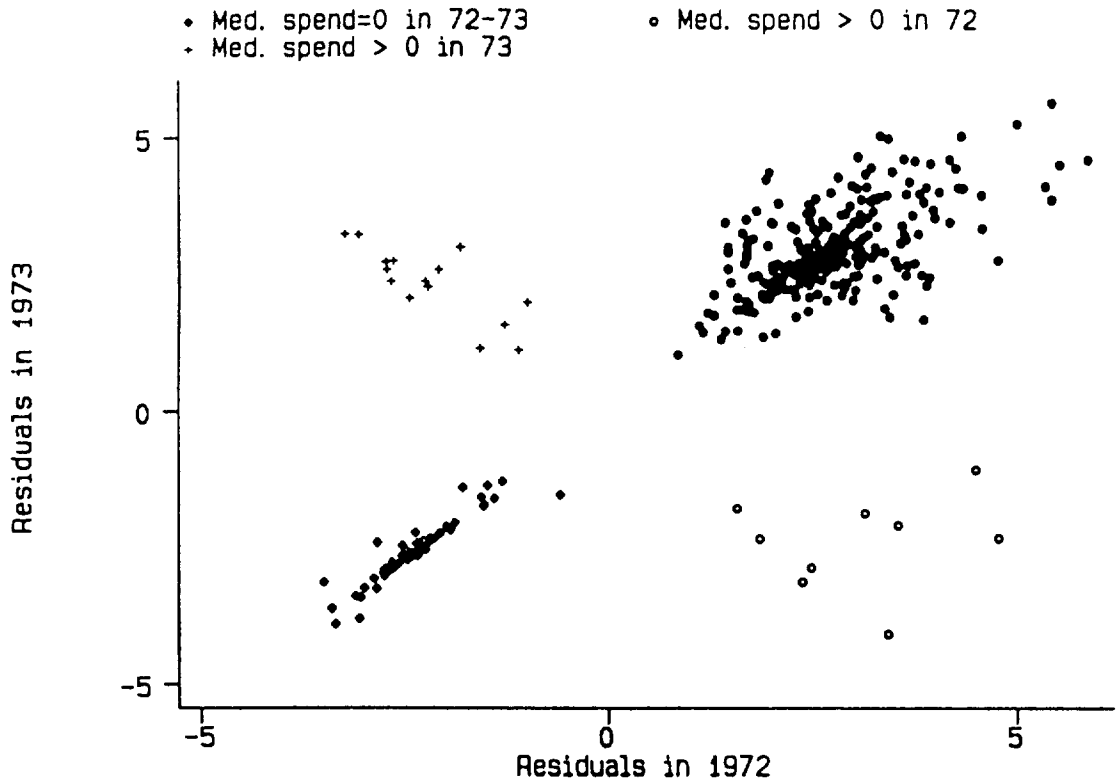(Greater than 3% of Income)
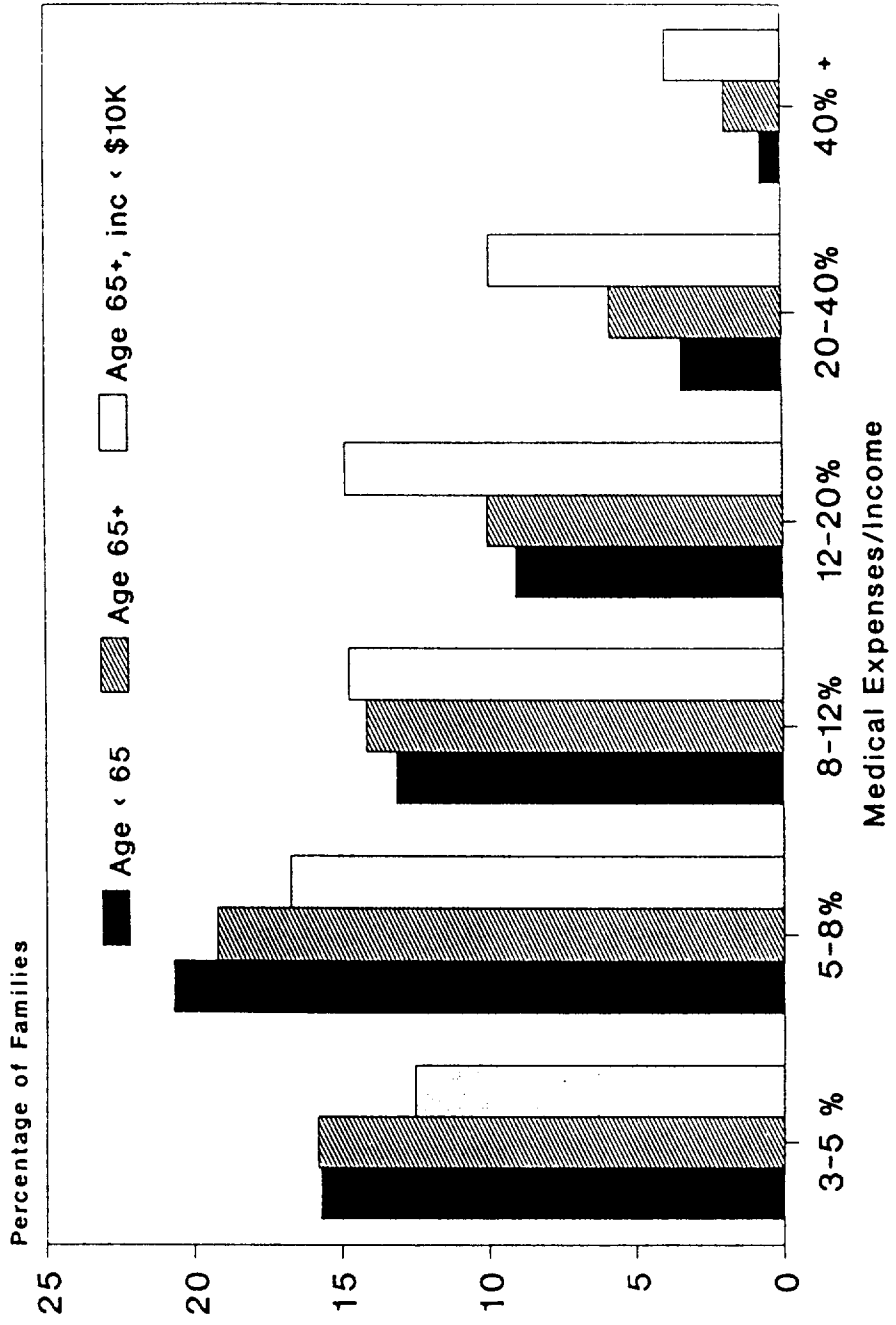
Source: NMCES 1977

Figure 2: Out-of-Pocket Medical
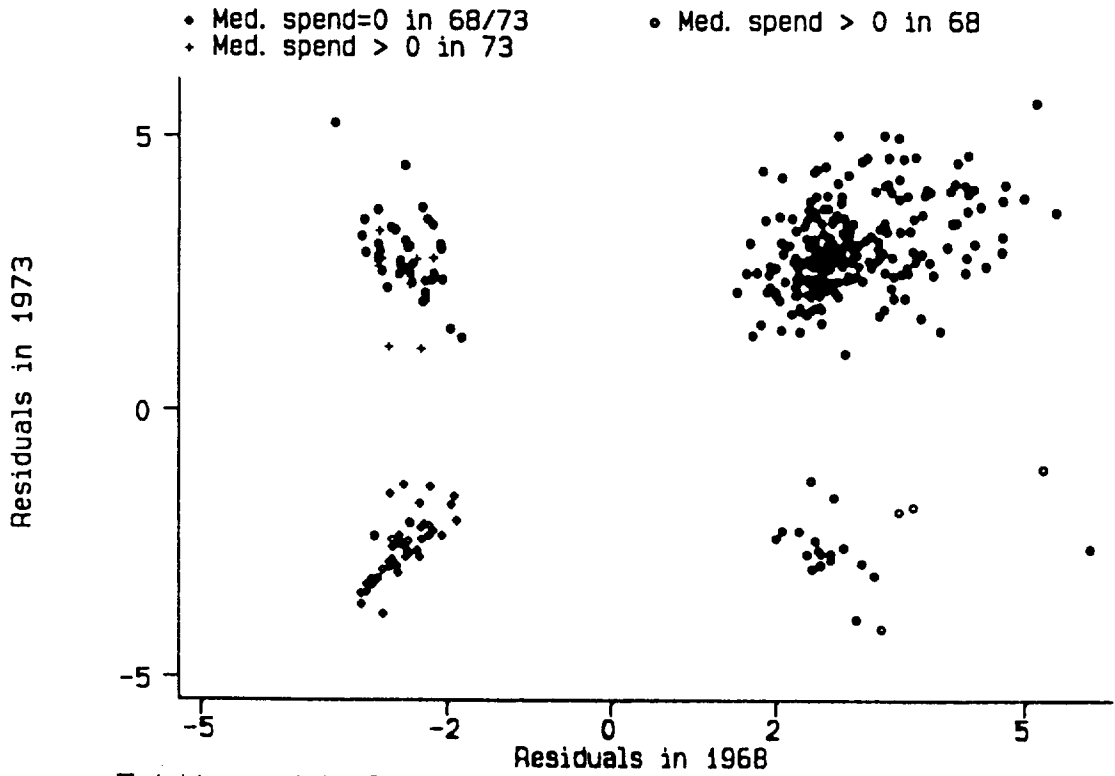Expenses: 1970 and 1988 Tax Returns
[Ages 65 and Above]

Percentage of Families

Legend:
- 1988 (All)
- 1970 (All)
- 1988 ($10-25,000)
- 1970 ($10-25,000)

Medical Expenses/Income

7.5-12.5%   12.5-20%   20-40%   40% +
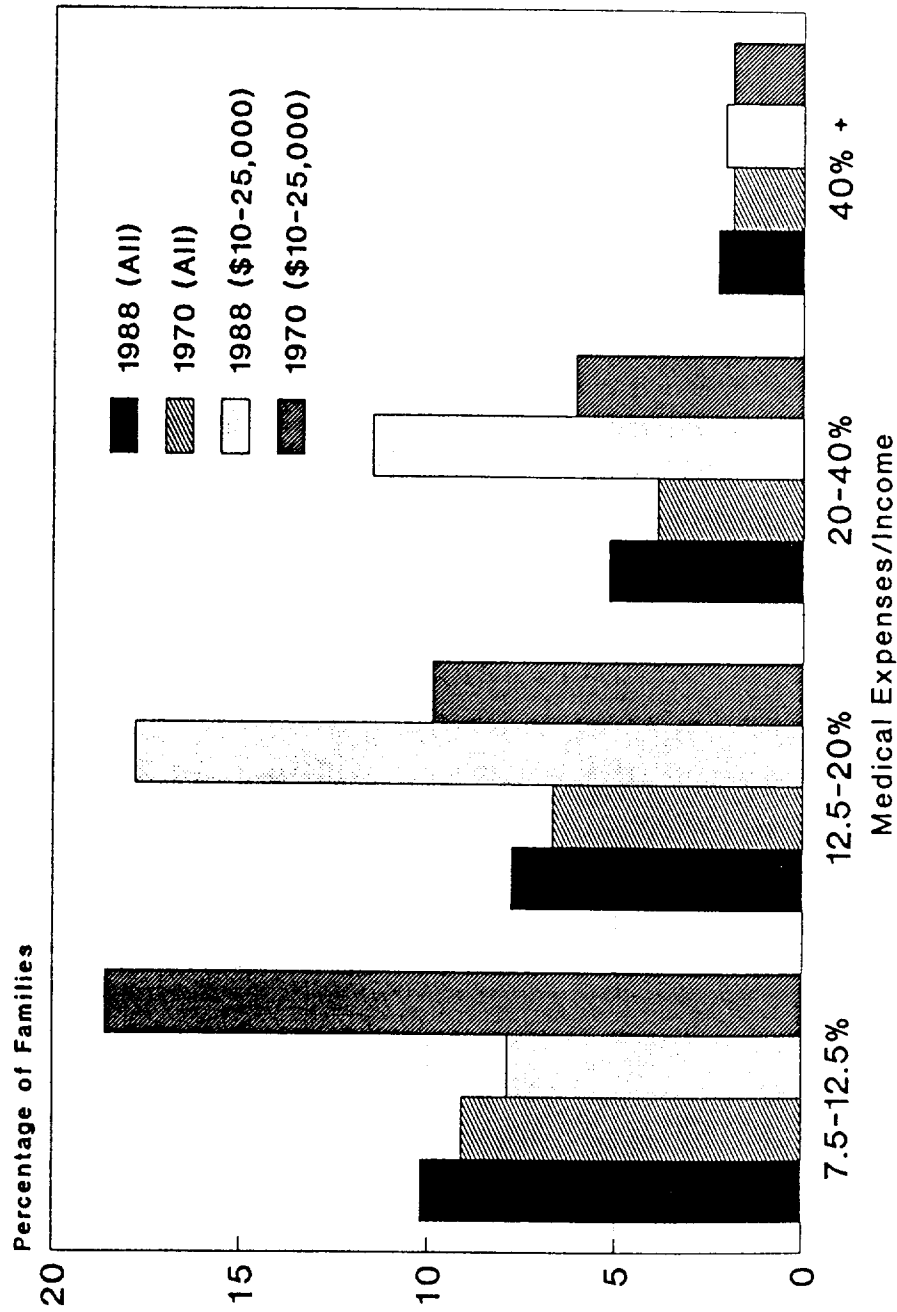
Source: IRS/Michigan Public Use Tapes

• Med. spend=0 in 72-73   ○ Med. spend > 0 in 72
+ Med. spend > 0 in 73

Tobit residuals: Log Medical Spending, 1972/73

Figure 3b

• Med. spend=0 in 68/73   ○ Med. spend > 0 in 68
+ Med. spend > 0 in 73

Tobit residuals: Log Medical Spending, 1968/73

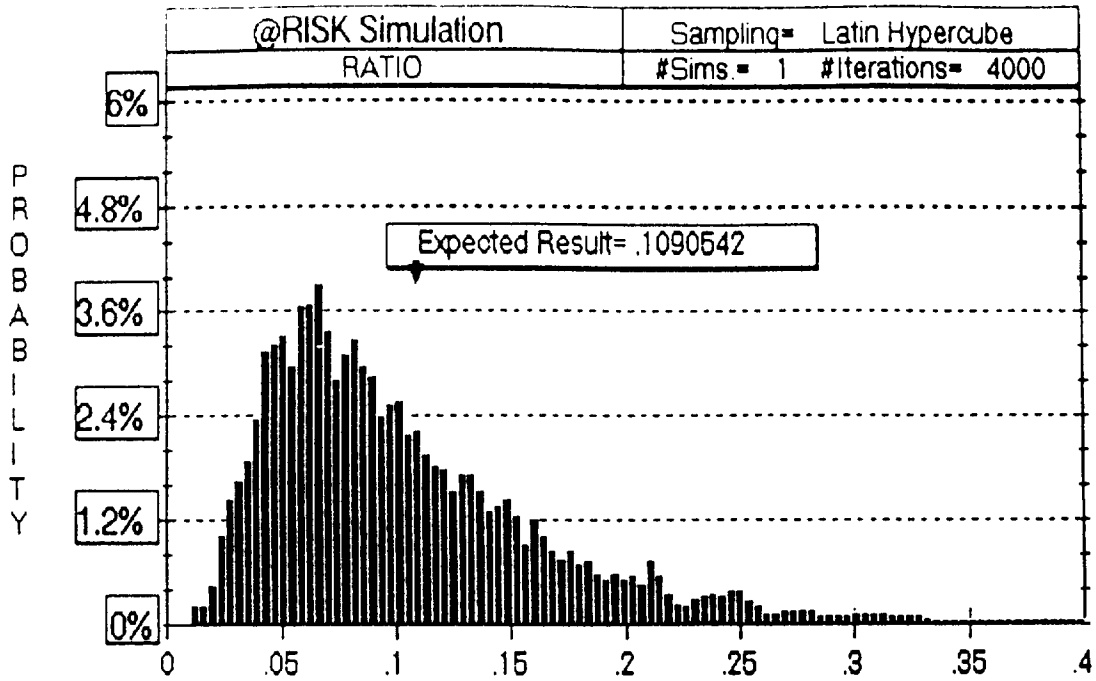Figure 4: Covariance Structure
and Impulse Decay of ARMA(1,2)
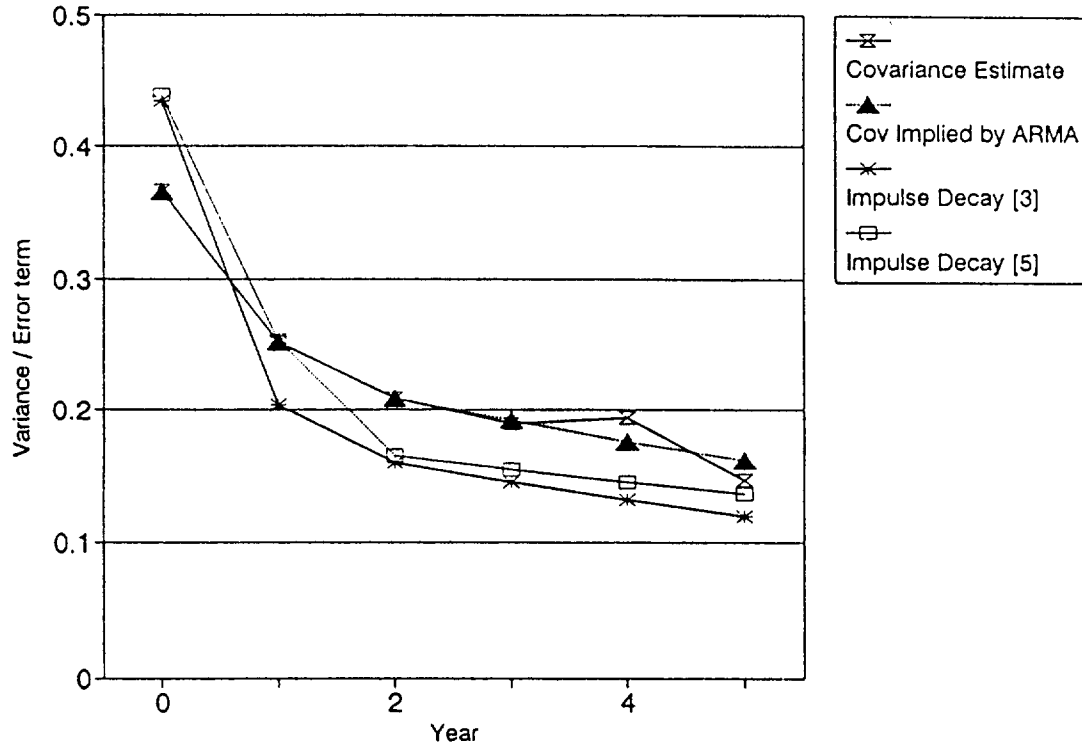
Figure 5a: Ratio of Average Medical
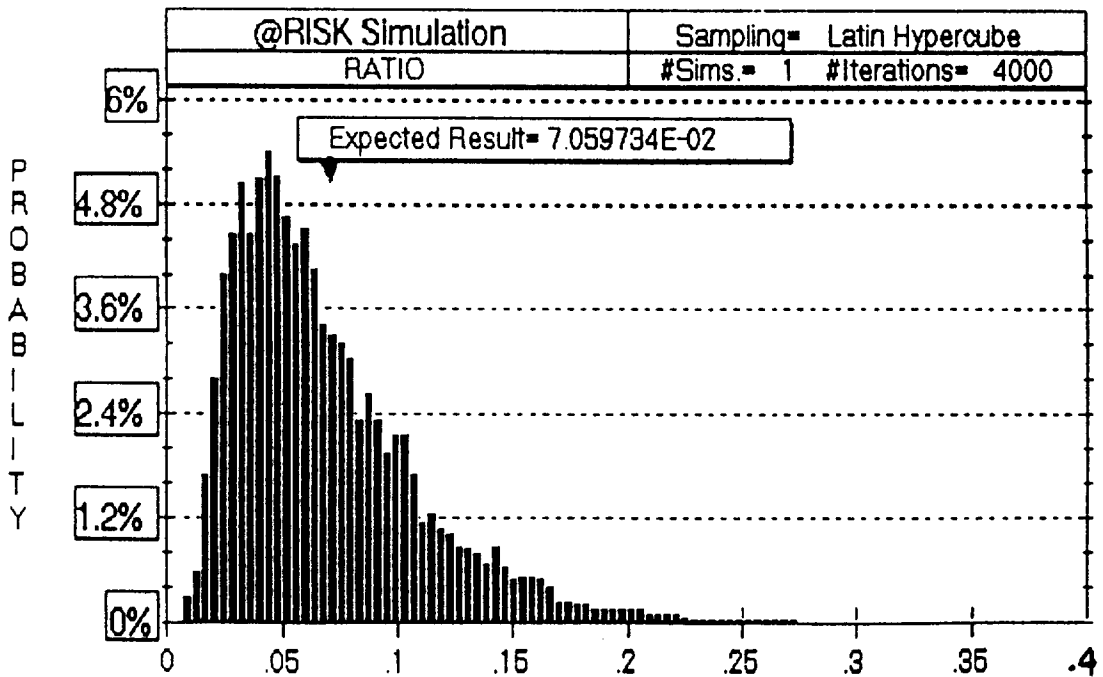Expenses / Income, 5-year period:
Income = $15,000

Figure 5b: Ratio of Average Medical
Expenses / Income, 5-Year Period:
Income = $30,000

Appendix A: An Informal Introduction
to Simulated Probability Estimators

This appendix is intended to provide a very simple introduction to the use of Monte-Carlo methods to integrate the multidimensional normal density functions often necessary in estimating limited dependent variable models. The discussion is informal so that, for example, we restrict our attention to one or two-dimensional integrals rather than treating the n-dimensional problem in full generality.

All Monte-Carlo integration methods are based on the observation that over the unit interval:

$$\int_0^1 f(u)\,du \approx \frac{1}{R}\sum_{j=1}^{R} f(u_j) \qquad\qquad (u_j \sim U[0,1]) \qquad (A.1)$$

That is, the integral of $f(u) = f(u|X,\Omega)$ on a unit interval can be approximated by the average value of the function over a random sample of u_i uniformly distributed over that same interval. The accuracy of this approximation improves at the rate $R^{\frac{1}{2}}$, independent of the dimensionality of u. Quadrature techniques can only be expected to converge at a rate inversely proportional to the dimensionality of f, necessarily slow for four or more dimensions.

The generalization to indefinite integrals can be easily accomplished by the use of weighted samples of random numbers distributed across the region of integration. Consider the one-sided indefinite integral from $-\infty$ to a. Suppose p(u) is an arbitrary density function chosen by the experimenter with P(a) (=prob(u<a))) known. Then the sample of u_is are drawn from p(u) (conditional on $u_i < a$), and the numerical approximation becomes

$$\int_{-\infty}^{a} f(u)\,du \approx \frac{1}{R}\sum_{j=1}^{R} \frac{f(u_j)P(a)}{p(u_j)} \qquad (A.2)$$

Effectively the integral is approximated by the experimental mean of f(u) -- weighted by the sampling distribution p(u), with an adjustment for the conditional nature of the sample. There are a few restrictions on p(u) but broadly, it needs to be non-zero where f(u) is non-zero. For efficiency, the function p should be chosen by the experimenter so that it is highly correlated with f(u); by inspection of A.2, when the sampling function is highly correlated with the integrand the variance of the individual elements is minimized, thereby increasing the accuracy of the estimate of the sum (There are other methods as well to reduce the variance of the simulator such as the Halton sequence or antithetic variates procedures that induce negative correlations among the draws). Such variance reduction techniques constitute the bulk of a vast literature on Monte-Carlo methods, but here we take note only of methods proposed by Lerman and Manski (1981), Stern (1992), and the smoothed simulated maximum likelihood (SSML) of Borsch-Supan and Hajavissiliou (1992). These methods are specific to integrating the normal density function.

An early attempt to simulate the multivariate normal distribution function was by Lerman and Manski (1981). For p(u) L&M picked the multivariate normal — this is attractive because it is easy to sample from this distribution (using the Cholesky transformation outlined below) and because it makes the variance of f(u)/p(u) equal to zero, conditional on knowing P(a). This is of course too good to be true, since one is trying to estimate the

value of P(a)!   So L&M set the region of integration to $[-\infty, +\infty]$ and set the integrand to zero outside the region of interest, i.e.:

$$prob(u < a) - \int_{-\infty}^{a} f(u,x,\Omega)\, du - \frac{1}{R}\sum_{j-1}^{R} \frac{f(u,x,\Omega) * i(u_j < a)}{f(u,x,\Omega)} - \frac{1}{R}\sum_{j-1}^{R} i(u_j < a) \tag{A.3}$$

where $f(u,x,\Omega)$ is the multivariate normal density and the indicator function $i(\cdot)$ returns zero for false and one for true. This is called *simple frequency simulation*, and it is particularly inefficient to use in evaluating likelihood functions, for two reasons. First, the relative accuracy of the integral estimate is poor when a is small, including the real possibility that the probability estimate will be zero if no draws fall within the relevant region. Since the likelihood function evaluates the log of the probability, assigning a zero probability to an option that is actually chosen in the sample will cause the likelihood function to explode. Second, the estimate is not continuous in the parameter space, so that a small change in x or Ω does not necessarily change the (finite) likelihood function. Specialized and not necessarily efficient hill-climbing techniques are required to optimize a function that is flat almost everywhere.

McFadden (1989) discussed a number of alternative smoothing weights to sidestep the discontinuities of the simulated function, including a *smooth unbiased simulator* equivalent to A.2. One of the more promising approaches was developed by Stern (1992) who developed a smoothed simulator that appears to be quite efficient in approximating the true function with a limited number of draws.   He suggests dividing u up into the sum of two random normal deviates, w_1 and w_2, where $w_1 \sim N(0,D)$ (with D a diagonal matrix) and $w_2 \sim N(x,\Omega-D)$. (The optimal choice of D is

discussed in detail in Stern). Denoting the density of w_1 by $f(u)$ and that of w_2 by $f(u,x,\Omega)$ this allows us to write:

$$\text{prob}(w_1 + w_2 < a) = \int_{-\infty}^{+\infty} \text{prob}(w_1 < a - w_2) f(w_2, x, \Omega - D) \, dw_2$$

$$= \int_{-\infty}^{+\infty} \Phi(a - w_2) f(w_2, x, \Omega - D) \, dw_2 \tag{A.4}$$

$$\approx \frac{1}{R} \sum_{j=1}^{R} \frac{\Phi(a - w_2) f(w_2, x, \Omega - D)}{f(w_2, x, \Omega - D)}$$

$$= \frac{1}{R} \sum_{j=1}^{R} \Phi(a - w_2)$$

where draws of w_2 are from $N(x, \Omega - D)$. The resulting sum varies smoothly with changes in Ω or x, and is never zero for any set of draws.

This is an example of a more general technique of variance reduction, namely splitting the integrand into the product of an easily integrable function (in this case a standard normal density) and a density function (in this case a general normal density). By sampling out of the same density function, an easily integrated function remains.

The simulated likelihood estimator developed by Borsch-Supan, Geweke, Hajavissiliou, and Keene use this strategy to go one step further. They depend on a remarkable property of the multivariate normal distribution, namely that if e_1, e_2 are independent normals with unit variance, and $L = \{\ell_{ij}\}$ is the Cholesky decomposition of Ω, then:

$$u_1 - \ell_{11} e_1$$

$$u_2 - \ell_{21} e_1 + \ell_{22} e_2$$

(A.5)

defines (u_1, u_2) as multivariate normals with covariance matrix $\Omega = L'L$ (see the text for more details.) Solving for e:

$$e_1 - u_1 / \ell_{11}$$

$$e_2 - (u_2 - \ell_{21} e_1) / \ell_{22}$$

(A.6)

Consider a single observation from a two-equation simultaneous Probit estimation with both dependents observed zeroes:

$$u_1 < x_1 \beta - a_1$$

$$u_2 < x_2 \beta - a_2$$

(A.7)

The likelihood of this particular realization is:

$$\text{prob}(u_1 < a_1 \text{ and } u_2 < a_2)$$

(A.8)

$$- \text{prob}(e_1 < a_1 / \ell_{11} \text{ and } e_2 < (a_2 - \ell_{21} e_1 / \ell_{22}))$$

(A.9)

$$- \int_{-\infty}^{a_1/\ell_{11}} \int_{-\infty}^{(a_2-\ell_{21})/\ell_{22}} f(e_1) f(e_2) de_2 de_1$$

(A.10)

$$- \int_{-\infty}^{a_1/\ell_{11}} f(e_1) \Phi((u_2 - \ell_{21} e_1)/\ell_{22}) de_2$$

(A.11)

Like (A.1) this integral can also be approximated by a weighted average equivalent to (A.2):

$$\frac{1}{R} \sum_{j=1}^{R} \frac{f(e_{1j}) \Phi((a_2 - \ell_{21} e_{1j})/\ell_{22})}{f(e_{1j})/\Phi(a_1/\ell_{11})}$$

(A.12)

where the e_{1i} are sampled from a normal distribution conditional on $e_n < a_1$. Here f is selected as the sampling function (playing the role of p(x) in (A.2)) especially for its ability to reduce the variance of the individual terms in the sum; i.e., (A.11) simplifies to:

$$\Phi(a_1/\ell_{11})\frac{1}{R}\sum_{j=1}^{R}\Phi((a_2-\ell_{21}e_{1j})/\ell_{22}) \qquad (A.13)$$

As shown in (A.13), the ingeneous choice of the weighting matrix p(u) that is both continuous in the parameter space and that appears effective at reducing the variance of the simulated terms may be one reason why the method appears to simulate accurately relative to alternative approaches (Hajivassiliou, McFadden, and Ruud, 1991).

Appendix B: Halton Sequences

The following Gauss program generates an nx1 Halton sequence using a prime number q; the first 15 outcomes are reported in Table A.1. See Halton (1960) for a schematic diagram of the sequences.

```
proc halton(n,q);
local phi,i,y,x;
phi = reshape(0,n + 1,1);
i = 1;
do while i < n + 1;
  i = i + 1;
  y = 1/q;
  x = 1-phi[i-1];
  do while ((x) le (y + 1.e-11));
    y = y/q;
  endo;
  phi[i] = phi[i-1] + (q + 1)*y-1;
endo;
retp(phi[2:n + 1]);
endp;
```

Table A.1: Example of Halton Sequences

Sequence No.	q = 3	q = 5	q = 7
1	1/3	1/5	1/7
2	2/3	2/5	2/7
3	1/9	3/5	3/7
4	4/9	4/5	4/7
5	7/9	1/25	5/7
6	2/9	6/25	6/7
7	5/9	11/25	1/49
8	8/9	16/25	8/49
9	1/27	21/25	15/49
10	10/27	2/25	22/49
11	19/27	7/25	29/49
12	4/27	12/25	36/49
13	13/27	17/25	43/49
14	22/27	22/25	2/49
15	7/27	3/25	9/49