Structural Estimation
Maximum Simulated Likelihood, Simulated Method of Moments & Indirect Inference

Abi Adams

HT 2018

Outline for Today

Aim: Understand how to connect theory and data using key 'structural' approaches

I Warm Up: OLS

II Maximum Simulated Likelihood

III Method of Simulated Moments

IV Indirect Inference

Warm Up: OLS

- There are a number of estimators for even a simple linear model

$$y_i = X_i\beta + \epsilon_i$$

with $E(X\epsilon_i) = 0$

  - 1. Basic Matrix Manipulation

$$\hat{\beta} = \left(X'X\right)^{-1} X'y$$

Warm Up: OLS

- 2. Maximum Likelihood

$$(\hat{\beta}, \hat{\sigma}) = \arg\max_{\beta} \quad \log \mathcal{L}(\beta, \sigma^2)$$

where

$$\log \mathcal{L}(\beta, \sigma^2) = -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)' (y - X\beta)$$

Warm Up: OLS

- 3. Generalised Method of Moments
  - Given mean independence of the error term

$$
\begin{aligned}
E\left[X\epsilon\right] &= E\left[X(y - X\beta)\right] \\
&= 0
\end{aligned}
$$

  - The sample moment conditions are then

$$
m = \frac{1}{N}\left[\sum_{i=1}^{N} X_i(y_i - X_i\beta)\right]
$$

Warm Up: OLS

- Fundamental idea is to pick $\hat{\beta}$ to make the sample moments as close to zero as possible

$$\hat{\beta} = \arg\min_{\beta} m(\beta) W m(\beta)'$$

  - W: weight matrix
  - Needs to be positive definite, e.g. identity matrix

Moving to More Complicated Models

- Fundamental ideas are no different in most approaches to mainstream structural estimation
- Use an economic model to define a likelihood function or a set of moment conditions
- Issue will be when these become complicated such that don't have closed form for likelihood function or moment conditions
- Will then need to use simulation based methods

Static Choice: Maximum Simulated Likelihood

- Tackle idea of maximum simulated likelihood in context of multinomial choice
- Choice between a set of options $j = \{0, ..., J\}$ with utility given by:

$$u_{ij} = \alpha_j + x_{ij}\beta + z_i\delta_j + \epsilon_{ij} \quad (1)$$

- Above is a standard random utility specification with additive errors

$$d_i = arg \max_j u_{ij} \quad (2)$$

Normalisations

- Now for identification, clear we will need location and scale normalisations (utility ordinal!)
- Considering the binary case, choose $j = 1$ if

$$u_{i1} > u_{i0} \tag{3}$$

$$\alpha_1 + x_{i1}\beta + z_i\delta_1 + \epsilon_{i1} > \alpha_0 + x_{i0}\beta + z_i\delta_0 + \epsilon_{i0} \tag{4}$$

$$(\alpha_1 - \alpha_0) + (x_{i1} - x_{i0})\beta + z_i(\delta_1 - \delta_0) + \epsilon_{i1} - \epsilon_{i0} > 0 \tag{5}$$

- Thus need to normalise:

$$\alpha_0 = 0 \tag{6}$$

$$\delta_0 = 0 \tag{7}$$

Normalisations

- Also need to restrict error distribution
- Can only identify difference, $\epsilon_{i1} - \epsilon_{i0} \sim (\mu, \sigma^2)$
- Location normalisation: $\mu = 0$
- Scale normalisation: either $\sigma = 1$ or restrict coefficient to equal 1

Multinomial Probit Example

- Interested in an individual's occupational choice: academic, banker, farmer, artist.
- $\epsilon_i \sim N(0, \Omega)$
- Choice probabilities are then given by:

$$P_{ij} = Pr(u_{ij} + \epsilon_{ij} > u_{ij'} + \epsilon_{ij'} \quad \text{for all } j \neq j') \tag{8}$$

$$= \int \mathbf{1}\,(u_{ij} + \epsilon_{ij} > u_{ij'} + \epsilon_{ij'} \quad \text{for all } j \neq j')\,\phi(\epsilon_i)\,\mathrm{d}\epsilon_i \tag{9}$$

- This integral does not have a closed form solution (unlike the multinomial logit) and so must be evaluated numerically through simulation

Identification: Caution!

- Restrictions on the covariance matrix are more complicated in the multinomial probit model than in the binary or multinomial logit cases
- Bunch & Kitamura (1989): published papers contain unidentified probit models
- Keane (1992): identification fragile and so should aim also to incorporate exclusion restrictions where possible
- Be careful!! See Train Textbook for method for checking whether restriction on covariance matrix sufficient for identification

Estimation

- The probability that person *i* chooses the alternative that they are observed choosing is:

$$\prod_j (P_{ij})^{y_{ij}} \tag{10}$$

- Assuming that each decision maker's choice is independent of that of all other decision maker's (conditional on observables), the likelihood and log-likelihood of each individual choosing the option we see them choosing is:

$$\mathcal{L}(\beta, \Omega) = \prod_i \prod_j (P_{ij}(\beta, \Omega))^{y_{ij}} \tag{11}$$

$$\log \mathcal{L}(\beta, \Omega) = \sum_i \sum_j y_{ij} \log(P_{ij}(\beta, \Omega)) \tag{12}$$

Estimation

- If you had a closed form for P_{ij}, you'd be ready to go! (see Matlab code)
- Issue with the multinomial probit is that we don't have a closed form for the choice probabilities (which might also be the case once you branch out into more complicated models)
- Most applications thus used a simulation-based estimation procedure (although simulation not necessary for low dimensional problems – could use quadrature methods to approximate the integral — but this works poorly for J>4 (see Train))
- Even if you have a closed form for your choice probabilities you might want to use simulation as a way to reduce the computational burden of estimation in some scenarios, e.g. consideration set models

Maximum Simulated Likelihood

- Basic idea: simulate model to evaluate outcome probabilities given a particular parameter vector
- Basic idea: use these simulated probabilities to evaluate the log-likelihood function
- Most basic, generally applicable method: (smoothed) Accept-Reject simulator (although also investigate GHK if your aim is estimation of the multinomial probit)

Accept-Reject Simulator

- Originally proposed by Manski & Lerman (1981)
- 1. Set a starting value for (β, Ω).....
- 2*. For each individual, simulate a J-dimensional vector of errors from $N(0, \Omega_0)$. Label this draw r=1
- 3. Using the value of these errors, calculate the utility of each alternative: $u_{ij}^r = v_{ij}(\beta_0) + \epsilon_{ij}^r$
- 4. Determine which alternative has the greatest utility given the starting parameter value and draw of the error term. Record $I^r = 1$ if $u_{ij}^r > u_{ij'}^r$ for all $j \neq j'$ ('accept') and $I^r = 0$ otherwise ('reject')

Accept-Reject Simulator

- 5. Repeat 2-4 R times, updating the value of r on each iteration
- 6. Simulated probability

$$\widetilde{P}_{ij}(\beta, \Omega) = \frac{1}{R} \sum_{r=1}^{R} I^r \tag{13}$$

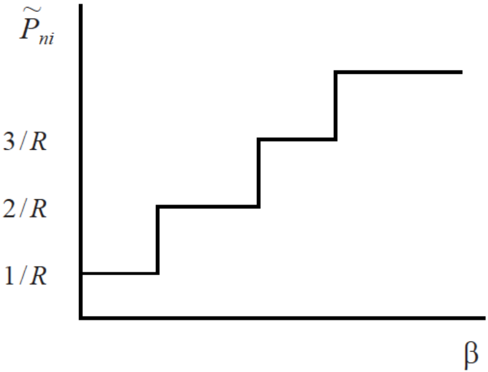
- Simulated likelihood function is then

$$\log \mathcal{L}(\beta, \Omega) = \sum_i \sum_j y_{ij} \log \left(\widetilde{P}_{ij} \right) \tag{14}$$

Issues with A-R Simulator

- An issue is that might be in a situation where $\widetilde{P}_{ij} = 0$ for any finite number of draws R — especially problematic with a large number of choices
- Simulated probabilities are not smooth in parameters (not twice differentiable)
- Numerical procedures to find the maximum of the likelihood typically rely on first and second derivatives
- Can result in poor performance of optimisation routines

Issues with A-R Simulator



Logit Smoothed A-R Simulator

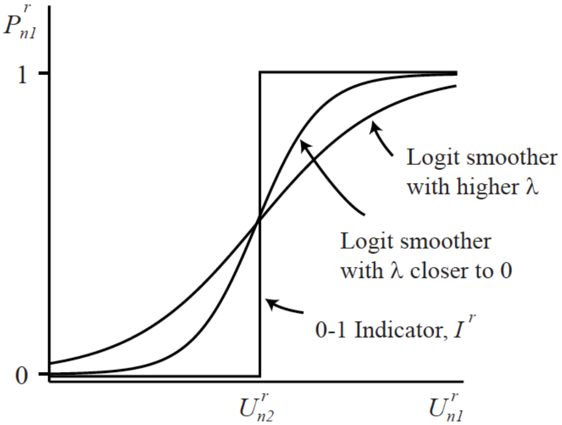
- Rather than use binary indicator variable, McFadden (1989) shows that can use any function for simulation P_{ij} so long as it is strictly positive, monotonic in u_{ij}^r and defined first and second derivatives with respect to u_{ij}^r (e.g. Logit)
- Just involves changing steps (4) and (6) above
- 4. After calculating utility of an option given error draw ϵ_i^r, calculate

$$S_i^r = \frac{exp(u_{ij}^r/\lambda)}{\sum_{j'} exp(u_{ij'}^r/\lambda)} \tag{15}$$

- 6. Simulated 'probability'

$$\widetilde{P}_{ij} = \frac{1}{R}\sum S_i^r \tag{16}$$

Logit Smoothed A-R Simulator



Maximum Simulated Likelihood: Summary

- Economic models might imply complicated structure to outcome probabilities
- Simulate from the model to construct these probabilities

$$\log \mathcal{L}(\beta, \Omega) = \sum_i \sum_j y_{ij} \log \left(\widetilde{P}_{ij} \right) \tag{17}$$

- See Keane and Wolpin (1995) as a classic example

"Matching Moments"

- Thus far we have discussed likelihood approaches to estimate structural models
- GMM, simulated method of moments and indirect inference form another class of popular methods
- These methods can be slightly less computationally burdensome in some scenarios and can also make the connection between identification and estimation more explicit

Standard GMM

- The standard GMM framework involves coming up with a set of moments

$$m(X_i, y_i, \theta) \tag{18}$$

for which

$$E\left(m(X_i, y_i, \theta_0)\right) = 0 \tag{19}$$

- The sample analogue comes from recognising that

$$\frac{1}{N}\sum_{i=1}^{N} m(X_i, y_i, \theta_0) \approx 0 \tag{20}$$

- More generally we are overidentified and so we choose $\widehat{\theta}$ to minimize:

$$\left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, y_i, \theta)\right]' W \left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, y_i, \theta)\right] \tag{21}$$

Relationship between GMM and MLE

- You can think of ML as a special case of GMM
- We have that

$$\theta_0 = arg \max E \left[\log \mathcal{L}(\theta; X_i, y_i) \right] \tag{22}$$

- As long as the problem is well behaved this implies

$$E \left(\frac{\partial \log \mathcal{L}(\theta; X_i, y_i)}{\partial \theta} \right) = 0 \tag{23}$$

- You can think of this like a moment condition!

Relationship between GMM and MLE

- Although caution: they are equivalent only when the log likelihood function is globally strictly concave
- Otherwise might have multiple solutions to the first order conditions
- Then only locally but not globally identical

Simulated Method of Moments

- Classic reference: McFadden (1989)
- First ignore observed covariates, X
- Take any function of the data, $g(y_i)$
- Then,

$$E(g(y_i)) = \int g\left(y(\epsilon;\theta_0)\right)\, \mathrm{d}F(\epsilon;\theta_0) \tag{24}$$

  where $y(\theta_0)$ and F represent the model/data generating process

- Thus do GMM with:

$$m(y_i,\theta) = g(y_i) - \int g\left(y(\epsilon;\theta)\right)\, \mathrm{d}F(\epsilon;\theta) \tag{25}$$

Simulated Method of Moments

- So in the sample we have:

$$\frac{1}{N}\sum_{i=1}^{N}\left[g(y_i) - \int g\left(y(\epsilon;\theta)\right)\, \mathrm{d}F(\epsilon;\theta)\right] \tag{26}$$

$$= \left[\frac{1}{N}\sum_{i=1}^{N} g(y_i)\right] - \int g\left(y(\epsilon;\theta)\right)\, \mathrm{d}F(\epsilon;\theta) \tag{27}$$

- We can approximate the right hand side term by simulating from the model
- If we simulate from the true value

$$\frac{1}{N}\sum_{i=1}^{N} g(y_i) - \frac{1}{R}\sum_{r=1}^{R} g(y(\epsilon_r;\theta_0)) \approx E(g(y_i)) - \int g\left(y(\epsilon;\theta_0)\right)\, \mathrm{d}F(\epsilon;\theta_0) \tag{28}$$

Simulated Method of Moments

- If we simulate from the true value

$$\frac{1}{N}\sum_{i=1}^{N} g(y_i) - \frac{1}{R}\sum_{r=1}^{R} g(y(\epsilon_r;\theta_0)) \approx E(g(y_i)) - \int g\left(y(\epsilon;\theta_0)\right)\,\mathrm{d}F(\epsilon;\theta_0) \tag{29}$$

- The nice thing is that we don't need *R* to be large for every individual, we just need *R* to be large for the one integral
- This can bring about computational benefits

Simulated Method of Moments

- Adding X's back in is straightforward

$$E(g(X_i, y_i)) = E\left[E\left(g(X_i, y_i)|X_i\right)\right] \quad (30)$$

$$= E\left[\int g\left(X_i, y(X_i, \epsilon; \theta_0)\right)\ \mathrm{d}F(\epsilon; \theta_0)\right] \quad (31)$$

- Now for the simulation we draw X from the empirical distribution of X_i and ϵ from F

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, y_i) - \frac{1}{R}\sum_{r=1}^{R} g(X_r, y(X_r, \epsilon_r; \theta_0)) \quad (32)$$

$$\approx E(g(X_i, y_i)) - E\left[\int g\left(X_i, y(X_i, \epsilon; \theta_0)\right)\ \mathrm{d}F(\epsilon; \theta_0)\right] \quad (33)$$

$$= 0 \quad (34)$$

Simulated Method of Moments

- In practise, minimise:

$$S'WS \quad \text{where} \quad S = \left[\frac{1}{N}\sum_{i=1}^{N} g(X_i, y_i) - \frac{1}{R}\sum_{r=1}^{R} g(X_r, y(X_r, \epsilon_r; \theta))\right] \tag{35}$$

Indirect Inference

- Extension of GMM idea — Gourieroux, Monrort & Renault (1993)
- Intuition for SMM: if I have the right DGP then taking the mean of the simulated data should give me the same answer as taking the mean of the real data

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, y_i) \approx \frac{1}{R}\sum_{r=1}^{R} g(X_r, y(X_r, \epsilon_r; \theta)) \tag{36}$$

- You can generalise this idea — the simulated data should 'look like' the real data
- Whatever I do to the real data, I should get the same answer if I do the same thing to simulated data

Indirect Inference

- Define a set of auxiliary parameters

$$\widehat{\beta} = arg \min_{\beta} F\left(\frac{1}{N}\sum_{i=1}^{N} g(X_i, y_i, \beta), \beta\right) \tag{37}$$

- Examples
  - Moments

$$\widehat{\beta} = arg \min_{\beta} \left(\frac{1}{N}\sum_{i=1}^{N} g(X_i, y_i) - \beta\right)^2 \tag{38}$$

  - Regression parameters

$$\widehat{\beta} = arg \min_{\beta} \frac{1}{N}\sum_{i=1}^{N} (y_i - X_i\beta)^2 \tag{39}$$

Indirect Inference

- Key thing is that the auxiliary model can be misspecified
- Creates a nice connection to reduced form results — can use 2SLS or Diff in Diff results as auxiliary parameters

Indirect Inference

- Define the population value of $\widehat{\beta}$

$$\beta_0 = arg \min_{\beta} F \left(E \left[g(X_i, y_i, \beta) \right], \beta \right) \tag{40}$$

$$= arg \min_{\beta} F \left(\int g(X_i, y(X_i, \epsilon; \theta_0), \beta) \, \mathrm{d}F(\epsilon; \theta_0), \beta \right) \tag{41}$$

- Define simulated auxiliary model as:

$$\widehat{B}(\theta) = \frac{1}{H} \sum_{h=1}^{H} arg \min_{\beta} F \left(\frac{1}{R} \sum_{r=1}^{R} g(X_{hr}, y_{hr}(\theta), \beta), \beta \right) \tag{42}$$

- H does not need to be large but as R gets large $\widehat{B}(\theta) \approx \beta_0$

Indirect Inference

- We then pick θ such that the parameters auxiliary model estimated off the real and simulated data are close

$$\widehat{\theta} = \mathit{arg}\min_{\theta} \left(\widehat{B}(\theta) - \widehat{\beta}\right)' \Omega \left(\widehat{B}(\theta) - \widehat{\beta}\right) \tag{43}$$

- It is optimal to choose Ω as the inverse of the covariance matrix for $\widehat{\beta}$
- In practise, however, many choose the diagonal matrix

Lockwood (2017)

- Especially clear structural application to assess the role of bequests in retirement and long-term care insurance choices
- Face significant uncertainty about length of life and the cost of future medical problems
- However, few retirees choose to insure themselves against these risks — in the US less than 5% buy life annuities and approx. 10% buy long term care insurance to cover the cost of nursing homes etc
- Can the motivation for bequests explain this lack of insurance?
- Would usually think no: would want to insure your bequests, but might play a role in which bequests are luxury goods

Long Term Care Use

- Use of formal care associated with
  - Being single
  - No children
  - Higher income
- If include couples in study, need to also include informal care....

Data

- Motivates restricting attention to the population of single retirees
- Data from the Health and Retirement Survey, longitudinal survey of a representative sample of the US population aged over 50 years old that occurs every two years
- Use data from the six waves 1998-2008
- Sample contains 3386 individuals (single, over 65 in 1998 and do not miss any of the interviews while alive)

Data

	Everyone 65+	Single retirees 65+
Female	0.58	0.78
Age	74.4	77.5
Wealth	$419,086	$238,643
Income	$33,891	$18,360
Own LTCI	0.10	0.09
Own annuity	0.06	0.06
Have children	0.91	0.85
Widowed	0.33	0.79
Never married	0.03	0.07
Importance of leaving a bequest		
Very	0.22	0.24
Somewhat	0.46	0.43
Not	0.32	0.32
Sample size	20,072	3,386

Table 1: Summary statistics of the sample of people aged 65 and older in the HRS and the subset of those who are single retirees (my sample). The statistics reported are means and are weighted by HRS respondent-level weights. The measure of wealth is at the household level and includes all non-annuity wealth and excludes annuitized wealth (e.g., the expected present value of future Social Security benefits). The measure of income is annual. The annuity ownership rate corresponds to annuities whose income stream continues as long as the individual lives. The values of all variables other than the importance of leaving a bequest come from the 1998 wave. The question about the importance of leaving a bequest was asked only in the 1992 wave, which primarily sampled cohorts younger than those in my sample. Among my sample of single retirees, less than nine percent answered this question.

Timing

1. Individual enters the period with wealth $w_t \geq 0$
2. Receive non-asset income y, realises acute medical care costs m_t and long term care costs ltc_t, and pays or receives any long term care insurance benefits due under her contract, $ltci_t$
3. Receives any social transfers
4. Decides how much to consume
5. Rate of return on savings and mortality realised
6. Individuals who die transfer remaining wealth to heirs – note, people cannot die in debt (no borrowing constraint)

Preferences

- Individuals maximise expected discounted utility from consumption and bequests

$$EU_t = u(c_t) + E_t \left\{ \sum_{a=t+1}^{T+1} \beta^{a-t} \left(\prod_{s=t}^{a-1} (1-\delta_s) \right) [(1-\delta_a)u(c_a) + \delta_a v(b_a)] \right\} \quad (44)$$

subject to a set of constraints to be discussed

  - t = current age; T: maximum age
  - δ_s: probability that an (s-1)-year old will die before age s

- Assume CRRA utility for consumption

$$u(c) = \frac{c^{1-\sigma} - 1}{1-\sigma} \quad (45)$$

Preferences

- Utility from bequests, with $\phi \in [0, 1]$

$$v(b) = \left(\frac{\phi}{1-\phi}\right)^{\sigma} \frac{\left(\frac{\phi}{1-\phi}c_b + b\right)^{1-\sigma}}{1-\sigma} \tag{46}$$

- Nests a set of functional forms found in prior literature
- c_b gives the level at which bequest motives 'kick in'
  - $c_b = 0$: preferences over consumption and bequests are homothetic
  - $c_b > 0$: bequests are a luxury & less risk averse over bequests than over consumption

Preferences

- Utility from bequests, with $\phi \in (0, 1)$

$$v(b) = \left(\frac{\phi}{1-\phi}\right)^{\sigma} \frac{\left(\frac{\phi}{1-\phi} c_b + b\right)^{1-\sigma}}{1-\sigma} \tag{47}$$

- Nests a set of functional forms found in prior literature
- ϕ is the marginal propensity to bequeath in a one-period problem of allocating wealth between consumption and an immediate bequest for people rich enough to consume at least c_b
    - Larger values of ϕ= leave larger share of wealth after buying at least c_b of consumption as bequests
    - As ϕ tends to 1, then bequest motive approaches a linear motive

States & Constraints

- Financial state variable: Net wealth before government transfers

$$\widehat{x}_t = w_t + y - m_t - ltc_t + ltci_t \tag{48}$$

- Social insurance programs put a floor under net wealth (which only kicks in if net wealth before transfers is below the relevant floor)

$$x_t = \max\{\widehat{x}_t, \bar{x}(h_t, ltci_t)\} \tag{49}$$
$$Pub_t = \mathbf{1}\left(\widehat{x}_t < \bar{x}(h_t, ltci_t)\right) \tag{50}$$

- There are five different health states, h_t: healthy (*he*), requiring home health care (*hhc*), assisted living facility (*alf*), nursing home care (*nh*) or dead (*d*)

States & Constraints

- Consumption is the product of spending decisions, $\widehat{c}_t \in [0, x_t]$, and the value of any long term care services received

$$c_t = \widehat{c}_t + c_m(h_t, Pub_t) \tag{51}$$

- Assets earn a rate of return r_t that is drawn from a distribution that depends on an individual's income quintile
- Next period wealth is then

$$w_{t+1} = (1 + r_t)(x_t - \widehat{c}_t) \geq 0 \tag{52}$$

- Conditional on living, next period net wealth before transfers is

$$\widehat{x}_{t+1} = (1 + r_t)(x_t - \widehat{c}_t) + y - m_{t+1} - ltc_{t+1} + ltci_{t+1} \tag{53}$$

Solution Method

- Solve by backwards induction from period T
- Long term care insurance is a fixed characteristic in every period other than the first purchasing period
- Other fixed characteristics are sex and retirement income y
- Time varying state variables are age (t), health (h_t) and net wealth before government transfers ($\widehat{x}_t$)
- For younger ages, discretise wealth into a fine grid and interpolate to evaluate value function between points

Solution Method

- Written in terms of the value function as:

$$V_t(\widehat{x}_t, h_t; ltci) = \max_{\widehat{c}_t \in [0, x_t]} \left\{ \begin{array}{l} u(\widehat{c}_t + c_m(h_t, Pub_t)) \\ +\beta Pr(h_{t+1} = d | h_t, t) E_t[v(b_{t+1})] \\ +\beta Pr(h_{t+1} \neq d | h_t, t) E_t \left[V_{t+1}(\widehat{x}_{t+1}, h_{t+1}; ltci) \right] \end{array} \right\} \tag{54}$$

also condition on sex and retirement income in practise

- Subject to

$$x_t = \max\{\widehat{x}_t, \bar{x}(h_t, ltci_t)\} \tag{55}$$

$$Pub_t = \mathbf{1}\left(\widehat{x}_t < \bar{x}(h_t, ltci_t)\right) \tag{56}$$

$$\widehat{x}_{t+1} = (1 + r_t)(x_t - \widehat{c}_t) + y - m_{t+1} - ltc_{t+1} + ltci_{t+1} \tag{57}$$

$$b_{t+1} = (1 + r_t)(x_t - \widehat{c}_t) \tag{58}$$

First Stage Parameter Values

- Estimation proceeds in two stages
  1. All parameters that can be identified outside the model are estimated or calibrated
  2. Remaining parameters estimated using SMM taking first stage parameters as given
- Second stage estimation to recover key preference parameters: strength and curvature of bequest motives (ϕ and c_b), consumption value of publicly financed care, discount factor and coefficient of relative risk aversion

Parameter	Source	Value in source	Value in this paper
Health states and transition probabilities: h_t, $Pr(h_{t+1} = h'\|h_t, t; s, y)$	Friedberg et al. (2014), De Nardi, French and Jones (2010), author's calculations	h_t: healthy (he), home health care (hhc), assisted living facility (alf), nursing home (nh), dead (d). $Pr(h_{t+1} = h'\|h_t, t; s, y)$: Friedberg et al. model for females	Adjust health transitions to match life expectancies by sex and income quintile in De Nardi, French and Jones (2010) (see Appendix A.1.2); maximum age (T) is 104
Long-term care costs: $ltc(h_t, t, s, q)$	MetLife Mature Market Institute (2002a,b), Friedberg et al. (2014), author's calculations	U.S. averages in 2002: $ltc(nh) = \$52,195$, $ltc(alf) = \$26,280$, $ltci(hhc, t) = \$37 * Q_s(t) + \$18 * Q_u(t)$, where $Q_s(t)$, $Q_u(t)$ from Friedberg et al. model[a]	Scale averages based on heterogeneity in NLTCS. Inflate values to reflect growth in spending, timing of care use. See Appendix A.1.3.
Acute medical care costs: $m_t \sim F_m(m; h_t, t, s, q)$	Author's estimates based on HRS	$m_t \sim \log N(\mu(h_t, t, s, q), \sigma(h_t, t, s, q)^2)$	Inflate values in source to reflect growth in spending, timing of care use. See Appendix A.1.4.
Long-term care insurance: $ltci_t(h_t, t; ltci)$	Brown and Finkelstein (2007): Typical contract, average load on policies held for life[b]	Pay premiums when healthy, receive benefits up to \$36,500 when require LTC (\$100/day benefit cap); 18% load	Inflate value in source to 2010 dollars: maximum benefit = \$44,350. Test robustness
Consumption value of privately-financed institutional care: c_{priv}	Author's judgment[c]	$c_{priv} = \$20,000$	Same as source, test robustness
Anticipated returns: $r \sim F_r(r; q)$	Author's estimates	$r \sim N(\mu_r(q), \sigma_r(q)^2)$	Same as source. See Appendix A.4.

Table 2: Baseline values of first-stage parameters. Notes:
(a) Q_s and Q_u are annual hours of skilled and unskilled home care, respectively. In the model, Medicare covers 35 percent of home health care spending (Robinson, 2002; Brown and Finkelstein, 2008), 25 percent of nursing home spending (Friedberg et al., 2014), and 0 percent of assisted living facilities.
(b) In calculating long-term care insurance premiums, future benefits and premiums are discounted at the risk-free interest rate, assumed to be 2 percent per year. The 18 percent load means that on average people receive 82 cents worth of benefits for each \$1 of premiums paid.
(c) As discussed in the text, the main effect of using different values of c_{priv} is to shift the estimated value of c_{pub} to maintain the same utility advantage, if any, of privately-financed care, $PCA = u(c_{priv}) - u(c_{pub})$. As discussed in Section 6.4, the results are robust to using alternative values of c_{priv}.

Second Stage Moments

- Two sets of moments
  1. Empirical wealth moments
     - Split sample into six five-year birth cohorts
     - Calculate the proportion in each cell in each wave with zero wealth, 50th and 75th percentiles
  2. Long term care insurance moments
     - Ownership rates at the four quantiles of the wealth distribution

Simulation Procedure

- For each candidate parameter vector, solve model separately for men and women and for those with or without long term care insurance
- Use the resulting value functions and choice rules to simulate the wealth path and long term care insurance status of each individual in a simulation sample
- Use the simulated data to calculate simulated moments using the same procedure used to calculate the empirical moments from the real data
- Evaluate goodness of fit of the simulated moments at this particular set of parameter values to the empirical moments using a classical minimum distance type estimator

Simulation Sample

- Draw 10,000 individuals with replacement from the HRS sample
- Use the resulting value functions and choice rules to simulate the wealth path and long term care insurance status of each individual in a simulation sample
- To ensure the simulation sample is representative of the population of US single retirees, the probability that an individual is drawn is proportional to their person-level survey weight
- X characteristics include: sex, retirement income, health status, portfolio shares

Estimation

- Select θ to minimise the distance between empirical and simulated moments

$$\min_{\theta} G' W G \tag{59}$$

- Baseline weighting matrix is the inverse of the estimated variance-covariance matrix of the second stage moments, $W = \widehat{\Omega}_g^{-1}$
- The more precisely a particular moment is estimated, and the less correlated it is with other moments in the estimation, the greater weight it receives in estimation

Estimation

- Duffie & Singleton (1993): The variance-covariance matrix of $\widehat{\theta}$ is

$$\Omega_\theta = \left(G'_\theta W C_\theta\right)^{-1} G'_\theta W \left[\Omega_g + \frac{N_d}{N_s}\Omega_g + G_\chi \Omega_\chi G'_\chi\right] W G_\theta \left(G'_\theta W G_\theta\right)^{-1} \tag{60}$$

where

- G_θ and G_χ are the gradient matrices of the moment conditions with respect to θ and χ (second and first stage parameters)
- Ω_θ and Ω_χ: variance-covariance matrices of the second and first stage parameter vectors
- N_d and N_s: sample size of the real and simulated samples

Over Identification Test

- If the model is correct, the statistic below converges in distribution to a chi-squared random variable with degrees of freedom equal to the number of second stage moments less the number of second stage parameters

$$\widehat{\varphi}(\widehat{\theta}, \chi_0)' R^{-1} \widehat{\varphi}(\widehat{\theta}, \chi_0) \tag{61}$$

where

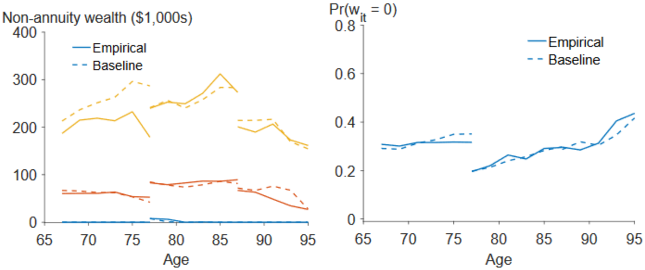
- $\widehat{\varphi}(\widehat{\theta}, \chi_0)$ are the moment conditions
- $R = \left(\frac{\Omega_g}{N_d} + \frac{\Omega_g}{N_s} + G_\chi \Omega_\chi G'_\chi \right)$ when $W = \Omega_g^{-1}$
- As calibrating many first stage parameters, treat χ as if it were known $G_\chi = 0$ — makes the second stage parameters appear more precise than they actually are and the fit of the model worse than it actually is

Results

	Baseline weighting matrix		Robust weighting matrix	
	Baseline model (1)	No bequest motive ($\phi = 0$) (2)	Baseline model (3)	No bequest motive ($\phi = 0$) (4)
Parameter estimates, $\hat{\theta}$				
$\hat{\phi}$: bequest motive	0.92	0	0.91	0
	(0.01)	-	(0.02)	-
$\hat{c}_b$: bequest motive ($1,000s)	20.3	0	15.0	0
	(1.1)	-	(1.2)	-
$\hat{c}_{pub}$: public care ($1,000s) ($c_{priv} = \$20,000$)	4.7	14.7	6.5	20.0
	(0.4)	(0.4)	(1.6)	(1.2)
$\hat{\underline{x}}_{comm}$: wealth floor in community ($1,000s)	7.4	11.1	7.8	10.2
	(0.2)	(0.2)	(0.2)	(0.2)
$\hat{\beta}$: discount factor (annual)	0.98	0.90	0.98	0.88
	(0.01)	(0.01)	(0.01)	(0.01)
$\hat{\sigma}$: risk aversion	2.0	10.6	2.0	11.3
	(0.03)	(0.1)	(0.1)	(0.1)
Goodness-of-fit				
χ^2 stat	81.9	137.2	86.7	147.6
p-value of model	0.22	1.6e-5	0.13	1.2e-6
p-value of no-bequest motive restriction	1.0e-12	-	5.8e-14	-
Simulated LTCI (%)	4.2	12.0	3.6	6.2
(Empirical LTCI = 5.6%)				

Table 3: Estimation results. Columns one and two are based on the baseline weighting matrix, the inverse of the estimated variance-covariance matrix of the second-stage moment conditions. Columns three and four are based on the "robust" weighting matrix, the inverse of the diagonal of the estimated variance-covariance matrix of the second-stage moment conditions (Pischke, 1995). The second and fourth columns come from estimating the nested version of the model without a bequest motive ($\phi = 0$). Standard errors are in parentheses.

Results



(a) 25th, 50th, and 75th percentiles

(b) Share with zero wealth

Figure 1: Empirical wealth moments (solid lines) and simulated wealth moments (dashed lines) for odd-numbered cohorts. (Even-numbered cohorts are excluded to avoid overlapping lines and are shown in Appendix A.6.) Panel (a) shows the 25th, 50th, and 75th percentiles of wealth; the 25th percentiles, most of which are zero, are not targeted by the estimation. Panel (b) shows the share with zero wealth. The x-axis shows the average age of surviving members of the cohort. The empirical and simulated wealth moments do not coincide in 1998 (the left-most points of each set of curves) due to sampling error from drawing a finite sample for the simulation.

Results

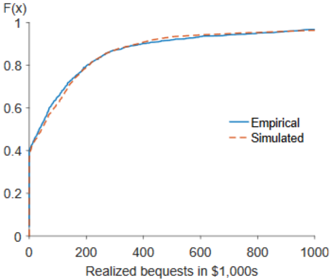


Figure A2: Cumulative distribution function of wealth in the last wave in which an individual is alive among individuals who die during the sample period. Wealth in the last period in which an individual is alive is a proxy for realized bequests that is better-measured than actual bequests. The simulated distribution is generated by the baseline model and parameter estimates.

Results

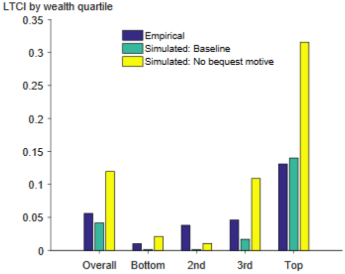


Figure 2: Simulated and empirical long-term care insurance ownership rates, overall and by wealth quartile. The simulated ownership rates are based on two sets of preferences: estimates from the baseline model and estimates from the model without bequest motives.

Results

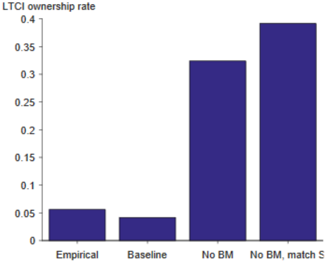
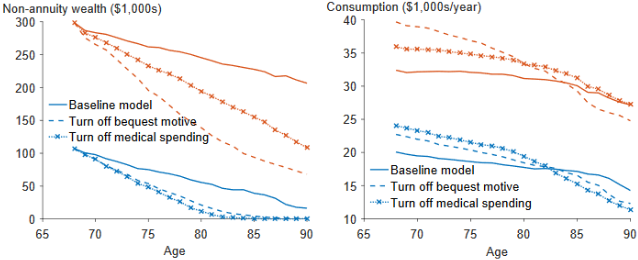


Figure 4: Simulated and empirical long-term care insurance ownership rates. The simulated ownership rates are based on three sets of preferences: the baseline estimates, the baseline estimates except with the bequest motive turned off ("No BM"), and estimates from the model without bequest motives fitted to the wealth moments ("No BM, match S").

Results



(a) Non-annuity wealth (b) Consumption spending

Figure 6: Simulated evolution of the median and 75th percentile of the distributions of non-annuity wealth and consumption spending among members of the first cohort (aged 65–69 in 1998) who remain alive for at least 23 years into their simulated future (at which time their average age is 90). The solid lines track wealth and consumption spending in the baseline model. The dashed lines track wealth and consumption spending if the bequest motive is turned off. The dotted lines with x markers track wealth and consumption spending if medical spending, including spending on both acute and long-term care, is shut down. Specifically, the decision rules come from a model without any long-term care or acute medical care costs ($ltc(h,t) = m_t = 0 \ \forall h, t$), but the simulation of wealth and consumption spending profiles includes medical costs. Differences in wealth and consumption spending therefore reflect only differences in behavior and not differences in realized medical expenses. Individuals in the simulation are assigned their reported (empirical) long-term care insurance ownership status.

Results

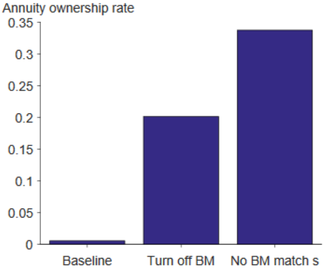
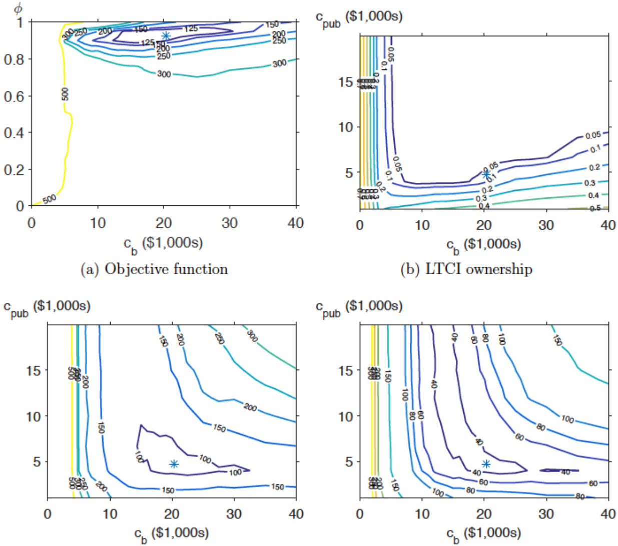


Figure 7: Simulated ownership rate of life annuities that pay the annuitant $5,000 of real income each year for life and have a ten percent load, typical of the U.S. private market (Brown, 2007). "Baseline" is the simulated ownership rate with the baseline estimates. "Turn off BM" is the simulated ownership rate with the baseline estimates, except without the bequest motive. "No BM match s" is the simulated ownership rate with the estimates based on the model without bequest motives estimated to match the wealth moments. Although it can be difficult to distinguish between annuities that do and do not offer lifespan insurance in the data, my best estimate is that in 1998 the share of single retirees aged 70–79 who owned a life annuity was 7.1 percent.

Identification



(a) Objective function

(b) LTCI ownership

(c) Objective function: wealth only

(d) Objective function: median wealth only

Figure A3: Panel (a): Contour plot of the objective function in (c_b, ϕ)-space with the other parameters held fixed at their baseline estimated values. Higher contours indicate greater mismatch