Solution Methods for Microeconomic Dynamic Stochastic Optimization Problems

2024-05-25

Christopher D. Carroll[1]

Note: The GitHub repo `SolvingMicroDSOPs` associated with this document contains python code that produces all results, from scratch, except for the last section on indirect inference. The numerical results have been confirmed by showing that the answers that the raw python produces correspond to the answers produced by tools available in the Econ-ARK toolkit, more specifically those in the `HARK` which has full documentation. The MSM results at the end have have been superseded by tools in the `EstimatingMicroDSOPs repo`.

Abstract

These notes describe tools for solving microeconomic dynamic stochastic optimization problems, and show how to use those tools for efficiently estimating a standard life cycle consumption/saving model using microeconomic data. No attempt is made at a systematic overview of the many possible technical choices; instead, I present a specific set of methods that have proven useful in my own work (and explain why other popular methods, such as value function iteration, are a bad idea). Paired with these notes is Python code that solves the problems described in the text.

Keywords	Dynamic Stochastic Optimization, Method of Simulated Moments, Structural Estimation, Indirect Inference
JEL codes	E21, F41

Powered by Econ-ARK

PDF:	https://github.com/llorracc/SolvingMicroDSOPs/blob/master/SolvingMicroDSOPs.pdf
Slides:	https://github.com/llorracc/SolvingMicroDSOPs/blob/master/SolvingMicroDSOPs-Slides.pdf
Web:	https://llorracc.github.io/SolvingMicroDSOPs
Code:	https://github.com/llorracc/SolvingMicroDSOPs/tree/master/Code
Archive:	https://github.com/llorracc/SolvingMicroDSOPs

(Contains LaTeX code for this document and software producing figures and results)

---

[1]Carroll: Department of Economics, Johns Hopkins University, Baltimore, MD, ccarroll@jhu.edu

Contents

1 Introduction

These lecture notes provide a gentle introduction to a particular set of solution tools for the canonical consumption-saving/portfolio allocation problem. Specifically, the notes describe and solve optimization problems for a consumer facing uninsurable idiosyncratic risk to nonfinancial income (e.g., labor or transfer income), first without and then with optimal portfolio choice,[1] with detailed intuitive discussion of various mathematical and computational techniques that, together, speed the solution by many orders of magnitude. The problem is solved with and without liquidity constraints, and the infinite horizon solution is obtained as the limit of the finite horizon solution. After the basic consumption/saving problem with a deterministic interest rate is described and solved, an extension with portfolio choice between a riskless and a risky asset is also solved. Finally, a simple example shows how to use these methods (via the statistical 'method of simulated moments' (MSM for short)) to estimate structural parameters like the coefficient of relative risk aversion (*a la* Gourinchas and Parker (2002) and Cagetti (2003)).

2 The Problem

The usual analysis of dynamic stochastic programming problems packs a great many events (intertemporal choice, stochastic shocks, intertemporal returns, income growth, the taking of expectations, time discounting, and more) into a complex decision in which the agent makes an optimal choice simultaneously taking all these elements into account. For the dissection here, we will be careful to break down everything that happens into distinct operations so that each element can be scrutinized and understood in isolation.

We are interested in the behavior a consumer who begins period t with a certain amount of 'capital' $\mathbf{k}_t$, which is immediately rewarded by a return factor R_t with the proceeds deposited in a **b**ank **b**alance:

$$\mathbf{b}_t = \mathbf{k}_t \mathsf{R}_t. \tag{1}$$

Simultaneously with the realization of the capital return, the consumer also receives noncapital income $\mathbf{y}_t$, which is determined by multiplying the consumer's 'permanent income' $\mathbf{p}_t$ by a transitory shock θ_t:

$$\mathbf{y}_t = \mathbf{p}_t \theta_t \tag{2}$$

whose whose expectation is 1 (that is, before realization of the transitory shock, the consumer's expectation is that actual income will on average be equal to permanent income $\mathbf{p}_t$).

The combination of bank balances **b** and income **y** define's the consumer's 'market

---

[1] See Merton (1969) and Samuelson (1969) for a solution to the problem of a consumer whose only risk is rate-of-return risk on a financial asset; the combined case (both financial and nonfinancial risk) is solved below, and much more closely resembles the case with only nonfinancial risk than it does the case with only financial risk.

resources' (sometimes called 'cash-on-hand,' following Deaton (1992)):

$$\mathbf{m}_t = \mathbf{b}_t + \mathbf{y}_t, \tag{3}$$

available to be spent on consumption $\mathbf{c}_t$ for a consumer subject to a liquidity constraint that requires $\mathbf{c} \leq \mathbf{m}$ (though we are not imposing such a constraint yet - see subsection 6.7).

The consumer's goal is to maximize discounted utility from consumption over the rest of a lifetime ending at date T:

$$\max \ \mathbb{E}_t \left[\sum_{n=0}^{T-t} \beta^n \mathrm{u}(\mathbf{c}_{t+n}) \right]. \tag{4}$$

Income evolves according to:

$$\begin{aligned} \mathbf{p}_{t+1} &= \mathcal{G}_{t+1}\mathbf{p}_t && \text{– permanent labor income dynamics} \\ \log \ \theta_{t+n} &\sim \ \mathcal{N}(-\sigma_\theta^2/2, \sigma_\theta^2) && \text{– lognormal transitory shocks } \forall \ n > 0. \end{aligned} \tag{5}$$

Equation (5) indicates that we are allowing for a predictable average profile of income growth over the lifetime $\{\mathcal{G}\}_0^T$ (to capture typical career wage paths, pension arrangements, etc).[2] Finally, the utility function is of the Constant Relative Risk Aversion (CRRA), form, $\mathrm{u}(\bullet) = \bullet^{1-\rho}/(1-\rho)$.

It is well known that this problem can be rewritten in recursive (Bellman) form:

$$\mathrm{v}_t(\mathbf{m}_t, \mathbf{p}_t) = \max_{\mathbf{c}} \ \mathrm{u}(\mathbf{c}) + \beta \, \mathbb{E}_t[\mathrm{v}_{t+1}(\mathbf{m}_{t+1}, \mathbf{p}_{t+1})] \tag{6}$$

subject to the Dynamic Budget Constraint (DBC) implicitly defined by equations (1)-(3) and to the transition equation that defines next period's initial capital as this period's end-of-period assets:

$$\mathbf{k}_{t+1} = \mathbf{a}_t. \tag{7}$$

3 Normalization

The single most powerful method for speeding the solution of such models is to redefine the problem in a way that reduces the number of state variables (if at all possible). In the consumption context, the obvious idea is to see whether the problem can be rewritten in terms of the ratio of various variables to permanent noncapital ('labor') income $\mathbf{p}_t$ (henceforth for brevity, 'permanent income.')

In the last period of life T, there is no future value, $\mathbf{v}_{T+1} = 0$, so the optimal plan is

---

[2]For expositional and pedagogical purposes, this equation assumes that there are no shocks to permanent income (though they are trivial to add). A large literature finds that, in reality, permanent (or at least extremely highly persistent) shocks exist and are quite large; such shocks therefore need to be incorporated into any 'serious' model (that is, one that hopes to match and explain empirical data), but the treatment of permanent shocks clutters the exposition without adding much to the intuition, so permanent shocks are omitted from the analysis until the last section of the notes, which shows how to match the model with empirical micro data. For a full treatment of the theory including permanent shocks, see Carroll (2023b).

to consume everything:

$$\mathbf{v}_T(\mathbf{m}_T, \mathbf{p}_T) = \frac{\mathbf{m}_T^{1-\rho}}{1-\rho}. \tag{8} \quad \text{\{eq:levelTm1\}}$$

Now define nonbold variables as the bold variable divided by the level of permanent income in the same period, so that, for example, $m_T = \mathbf{m}_T/\mathbf{p}_T$; and define $\mathrm{v}_T(m_T) = \mathrm{u}(m_T)$.[3] For our CRRA utility function, $\mathrm{u}(xy) = x^{1-\rho}\mathrm{u}(y)$, so (8) can be rewritten as

$$\begin{aligned} \mathbf{v}_T(\mathbf{m}_T, \mathbf{p}_T) &= \mathbf{p}_T^{1-\rho}\frac{m_T^{1-\rho}}{1-\rho} \\ &= \mathbf{p}_{T-1}^{1-\rho}\mathcal{G}_T^{1-\rho}\mathrm{v}_T(m_T). \end{aligned} \tag{9} \quad \text{\{eq:vT\}}$$

Now define a new optimization problem:

$$\begin{aligned} \mathrm{v}_t(m_t) = \max_{c_t} \quad &\mathrm{u}(c_t) + \beta\,\mathbb{E}_t[\mathcal{G}_{t+1}^{1-\rho}\mathrm{v}_{t+1}(m_{t+1})] \\ \text{s.t.} \quad & \\ a_t &= m_t - c_t \\ k_{t+1} &= a_t \\ b_{t+1} &= \underbrace{(\mathsf{R}/\mathcal{G}_{t+1})}_{\equiv\mathcal{R}_{t+1}} k_{t+1} \\ m_{t+1} &= b_{t+1} + \theta_{t+1}, \end{aligned} \tag{10} \quad \text{\{eq:vNormed\}}$$

where division by $\mathcal{G}$ in second-to-last equation yields a normalized return factor $\mathcal{R}$ which is the consequence of the fact that we have divided $t+1$ level variables by $\mathbf{p}_{t+1} = \mathcal{G}_{t+1}\mathbf{p}_t$.

```
# This is a snippet of code that constructs mass points
# for the equiprobable representation of the problem
```

Then it is easy to see that for $t = T - 1$, we can write boldface (nonnormalized) $\mathbf{v}$ as a function of v (normalized value) and permanent income:

$$\mathbf{v}_t(\mathbf{m}_t, \mathbf{p}_t) = \mathbf{p}_t^{1-\rho}\mathrm{v}_t(m_t), \tag{11} \quad \text{\{eq:vLvlFromvFunc\}}$$

and so on back to all earlier periods. Hence, if we solve the problem (10) which has only a single state variable m_t, we can obtain the levels of the value function from (11), and of consumption and all other variables from the corresponding permanent-income-normalized solution objects by multiplying each by $\mathbf{p}_t$, e.g.

$$\mathrm{c}_t(\mathbf{m}_t, \mathbf{p}_t) = \mathbf{p}_t\mathrm{c}_t(\overbrace{\mathbf{m}_t/\mathbf{p}_t}^{m_t}).$$

We have thus reduced the problem from two continuous state variables to one (and thereby enormously simplified its solution).

For future reference it is useful to write (10) in the traditional way, by substituting $b_{t+1}, k_{t+1},$ and a_t into m_{t+1}:

$$\mathrm{v}_t(m_t) = \max_c \quad \mathrm{u}(c) + \beta\,\mathbb{E}_t[\mathcal{G}_{t+1}^{1-\rho}\mathrm{v}_{t+1}(\overbrace{(m_t - c)(\mathsf{R}/\mathcal{G}_{t+1}) + \theta_{t+1}}^{m_{t+1}})]. \tag{12} \quad \text{\{eq:vusual\}}$$

---

[3]Nonbold value is bold value divided by $\mathbf{p}^{1-\rho}$ rather than $\mathbf{p}$.

4 Notation

4.1 Periods, Stages, Steps

The problem so far assumes that the agent has only one decision problem to solve in any period. But it is increasingly common to model agents who have multiple choice stages per period; a problem might have, say, a consumption decision (call it the c stage), a labor supply stage (call it ℓ) and a choice of what proportion ς of their assets to invest in a risky asset (the portfolio-choice stage).

The modeler might well want to explore whether the order in which the stages are solved makes any difference, either to the substantive results or to aspects of the computational solution like speed and accuracy.

If, as in section 2, we hard-wire into the solution code for each stage an assumption that its successor stage will be something in particular (say, the consumption stage assumes that the portfolio choice is next), then if we want to change the order of the stages (say, labor supply after consumption, followed by portfolio choice), we will need to re-hard-wire each of the stages to know particular things about its new successor (for example, the specifics of the distribution of the rate of return on the risky asset must be known by whatever stage precedes the portfolio choice stage).

But one of the cardinal insights of Bellman's (1957, "Dynamic Programming") original work is that *everything that matters* for the solution to the current problem is encoded in a 'continuation-value function.' Using Bellman's insight, we describe here a framework for isolating the stage problems within a period from each other, and the period from its successors in any future period; the advantage of this is that the isolated stage and period problems will then be 'modular': We can solve them in any order *without changing any code* (only transitions need to be rewired). After considering the stage-order $[\ell, c, \varsigma]$, the modeler can costlessly reorder the stages to consider, say, the order $[\ell, \varsigma, c]$.[4]

4.2 Steps

The key is to distinguish, within each stage's Bellman problem, three steps:

1. **Arrival**: Incoming state variables (e.g., k) are known, but any shocks associated with the period have not been realized and decision(s) have not yet been made

2. **Decision**: All exogenous variables (like income shocks, rate of return shocks, and predictable income growth $\mathcal{G}$) have been realized (so that, e.g., m's value is known) and the agent solves the optimization problem

3. **Continuation**: After all decisions have been made, their consequences are measured by evaluation of the continuing-value function at the values of the 'outgoing' state variables (sometimes called 'post-state' variables).

---

[4]As long as the beginning-of-stage and end-of-stage value functions for the stages all depend on the same state variables; see the discussion in section 8.

When we want to refer to a specific step in the stage we will do so by using an indicator which identifies that step. Here we use the consumption stage problem described above to exemplify the usage:

Step	Indicator	State	Usage	Explanation
Arrival	$\leftarrow$	k	$v_{\leftarrow}(k)$	value at entry to stage (before shocks)
Decision(s)	(blank)	m	$v(m)$	value of stage-decision (after shocks)
Continuation	$\rightarrow$	a	$v_{\rightarrow}(a)$	value at exit (after decision)

Notice that the value functions at different steps of the stage have distinct state variables. Only k is known at the beginning of the stage, and other variables take on their values with equations like $b = k\mathcal{R}$ and $m = b + \theta$. We will refer to such within-the-stage creation of variables as 'evolutions.' So, the consumption stage problem has two evolutions: from k to m and from m to a.

```
# make empty model
# MNW, SB, AL to substitute their ideas here
# Using : as the python prompt

## create empty two-period-model object
>>> mdl2prd = HARK2.modelMake(modelKind=finiteHorizon,
                              [other configs])

## what does the solution object look like?
### - there is only one interval filled in, which is a blank slate:

>>> mdl2prd.solution

[
    interval
]

>>> prdT=0 # To make the stuff below a bit easier to follow
>>> mdl2prd.solution[prdT].interval

(returns whatever info is created for the default interval)

# Now let's build what is in the interval
## Assume there is a yml file 'vEndTerm' for constructing v(S)=0 for all states:

>>> mdl2prd.solution[prdT].interval.add_vEndPrd(
    parse_file_model='vEndTerm_plato.yml' # Platonic description
    parse_file_setup='mdl2prd_config.yml' # like, simulation_method=monte_carlo
    )

# probably we need to put the stages in a sublist ...
>>> mdl2prd.solution[prdT].interval.make_empty_stage_list()

# The last stage is added first
## Period T is a special case in which no solving is necessary since we know the
    solution is c=m

>>> mdl2prd.solution[prdT].interval.append_stage(
    stage_name='c',
    parse_file_model='consume_all_m.yml',
    parse_file_stage='consume_all_m_stagefile.yml'
  ) # not sure whether parse_file_stage is needed ...

# Now add a portfolio stage
## append_stage will test whether there are already any other stages in existence
## If so it will
```

```
### - prepend to the list of stages an empty earlier stage
### - set stg[new_stg].vEndStg = stg[successor_stg].vBegStg

>>> mdl2prd.solution[prdT].interval.append_stage(
    stage_name='Shr',
    parse_file_model='portfolio_shr_solve.yml',
    parse_file_model='portfolio_shr_stage.yml' # again not sure if needed
    )

# We are finished adding stages; finish by adding vBeg
>>> mdl2prd.solution[prdT].interval.add_vBeg() #
>>> mdl2prd.solution[prdT].interval.solution
    [vBeg, [ Shr, c ], vEnd]
```

4.3 Transitions

In the backward-induction world of Bellman solutions, to solve the problem of a particular period we must start with an end-of-period (continuation) value function, which we designate by explicitly including the period indicator in the subscript (the := symbol denotes that the object on the right hand side is assigned to the object on the left hand side):

$$\mathrm{v}_{t_{\rightarrow}}(a) := \beta \mathrm{v}_{\leftarrow(t+1)}(\overbrace{a}^{=k}), \tag{13}$$

and we are not done solving the problem of period t until we have constructed a beginning-of-period value function $\mathrm{v}_{\leftarrow t}(k)$.

Similarly, in order to solve the problem of any stage, we must endow it with an end-of-stage continuation-value function. For the last stage in a period, the end-of-stage function is taken to be end-of-period value function; in our case where there is only one stage, this can be written cleanly as:

$$\mathrm{v}_{\rightarrow}(a) := \mathrm{v}_{t_{\rightarrow}}(a). \tag{14}$$

4.4 The Decision Problem in the New Notation

From 'inside' the decision stage, the Decision problem can now be written much more cleanly than in equation (10):

$$\mathrm{v}(m) = \max_{c} \; \mathrm{u}(c) + \mathrm{v}_{\rightarrow}(\overbrace{m-c}^{a}) \tag{15}$$

5 The Usual Theory, and a Bit More Notation

For reference and to illustrate our new notation, we will now derive the Euler equation and other standard results for the problem described above. Since

$$\mathrm{v}_{\rightarrow}(a) := \mathrm{v}_{t_{\rightarrow}}(a) := \beta \mathrm{v}_{\leftarrow(t+1)}(a) = \beta \, \mathbb{E}_{\leftarrow(t+1)}[\mathcal{G}_{t+1}^{1-\rho} \mathrm{v}_{t+1}(\overbrace{a(\mathsf{R}/\mathcal{G}_{t+1}) + \theta_{t+1}}^{m_{t+1}})], \tag{16}$$

given m_t, the first order condition for (12) with respect to a is

$$\mathrm{u}^c(m_t - a) = \mathrm{v}_{t_\rightharpoonup}^a(a) = \mathbb{E}_{\leftharpoonup(t+1)}[\beta\mathcal{R}_{t+1}\mathcal{G}_{t+1}^{1-\rho}\mathrm{v}_{t+1}^m(m_{t+1})]$$
$$= \mathbb{E}_{\leftharpoonup(t+1)}[\beta\mathsf{R} \quad \mathcal{G}_{t+1}^{-\rho}\mathrm{v}_{t+1}^m(m_{t+1})] \tag{17}$$

and because the `Envelope` theorem tells us that

$$\mathrm{v}_t^m(m_t) = \mathbb{E}_{\leftharpoonup(t+1)}[\beta\mathsf{R}\mathcal{G}_{t+1}^{-\rho}\mathrm{v}_{t+1}^m(m_{t+1})] \tag{18}$$

we can substitute the LHS of (18) for the RHS of (17) to get

$$\mathrm{u}^c(c_t) = \mathrm{v}_t^m(m_t) \tag{19}$$

and rolling forward one period,

$$\mathrm{u}^c(c_{t+1}) = \mathrm{v}_{t+1}^m(a_t\mathcal{R}_{t+1} + \theta_{t+1}) \tag{20}$$

so that substituting the LHS in equation (17) finally gives us the Euler equation for consumption:

$$\mathrm{u}^c(c_t) = \mathbb{E}_{t_\rightharpoonup}[\beta\mathsf{R}\mathcal{G}_{t+1}^{-\rho}\mathrm{u}^c(c_{t+1})]. \tag{21}$$

We can now restate the problem (12) with our new within-stage notation:

$$\mathrm{v}(m) = \max_c \quad \mathrm{u}(c) + \mathrm{v}_\rightharpoonup(m - c) \tag{22}$$

whose first order condition with respect to c is

$$\mathrm{u}^c(c) = \mathrm{v}_\rightharpoonup^a(m - c) \tag{23}$$

which is mathematically equivalent to the usual Euler equation for consumption.

We will revert to this formulation when we reach section 6.8.

6 Solving the Next-to-Last Period

To reduce clutter, we now temporarily assume that $\mathcal{G}_t = 1$ for all t, so that the $\mathcal{G}$ terms from the earlier derivations disappear, and setting $t = T$ the problem in the second-to-last period of life can now be expressed as

$$\mathrm{v}_{(t-1)}(m) = \max_c \quad \mathrm{u}(c) + \mathrm{v}_{(t-1)_\rightharpoonup}(\overbrace{m - c}^{a}) \tag{24}$$

where

$$\mathrm{v}_{(t-1)_\rightharpoonup}(a) := \beta\mathrm{v}_{\leftharpoonup t}(a) \equiv \beta\,\mathbb{E}_{\leftharpoonup t}\left[\mathrm{v}_t(\underbrace{a\mathcal{R}_t + \theta_t}_{m_t})\right]$$

Using (0) $t = T$; (1) $\mathrm{v}_t(m) = \mathrm{u}(m)$; (2) the definition of $\mathrm{u}(m)$; and (3) the definition of the expectations operator,

$$\mathrm{v}_{\leftharpoonup t}(a) = \int_0^\infty \frac{(a\mathcal{R}_t + \vartheta)^{1-\rho}}{1-\rho}d\mathcal{F}(\vartheta) \tag{25}$$

where $\mathcal{F}(\theta)$ is the cumulative distribution function for θ.

```
# A 'raw' solution to the T-1 solution works, but is very slow:
```

This maximization problem implicitly defines a 'local function' $c_{t-1}(m)$ that yields optimal consumption in period $t-1$ for any specific numerical level of resources like $m = 1.7$.

But because there is no general analytical solution to this problem, for any given m we must use numerical computational tools to find the c that maximizes the expression. This is excruciatingly slow because for every potential c to be considered, a definite integral over the interval $(0, \infty)$ must be calculated numerically, and numerical integration is *very* slow (especially over an unbounded domain!).

6.1 Discretizing the Distribution

Our first speedup trick is therefore to construct a discrete approximation to the lognormal distribution that can be used in place of numerical integration. That is, we want to approximate the expectation over θ of a function $g(\theta)$ by calculating its value at set of n_θ points θ_i, each of which has an associated probability weight w_i:

$$\mathbb{E}[g(\theta)] = \int_{\underline{\theta}}^{\overline{\theta}} (\vartheta) d\mathcal{F}(\vartheta)$$

$$\approx \sum_{\theta=1}^{n} w_i g(\theta_i)$$

(because adding n weighted values to each other is enormously faster than general-purpose numerical integration).

Such a procedure is called a 'quadrature' method of integration; Tanaka and Toda (2013) survey a number of options, but for our purposes we choose the one which is easiest to understand: An 'equiprobable' approximation (that is, one where each of the values of θ_i has an equal probability, equal to $1/n_\theta$).

We calculate such an n-point approximation as follows.

Define a set of points from $\sharp_0$ to $\sharp_{n_\theta}$ on the $[0, 1]$ interval as the elements of the set $\sharp = \{0, 1/n, 2/n, \ldots, 1\}$.[5] Call the inverse of the θ distribution $\mathcal{F}^{-1}$, and define the points $\sharp_i^{-1} = \mathcal{F}^{-1}(\sharp_i)$. Then the conditional mean of θ in each of the intervals numbered 1 to n is:

$$\theta_i \equiv \mathbb{E}[\theta | \sharp_{i-1}^{-1} \le \theta < \sharp_i^{-1}] = \int_{\sharp_{i-1}^{-1}}^{\sharp_i^{-1}} \vartheta \, d\mathcal{F}(\vartheta), \tag{26}$$

and when the integral is evaluated numerically for each i the result is a set of values of θ that correspond to the mean value in each of the n intervals.

The method is illustrated in Figure 1. The solid continuous curve represents the "true" CDF $\mathcal{F}(\theta)$ for a lognormal distribution such that $\mathbb{E}[\theta] = 1$, $\sigma_\theta = 0.1$. The short vertical

---

[5]These points define intervals that constitute a partition of the domain of $\mathcal{F}$.
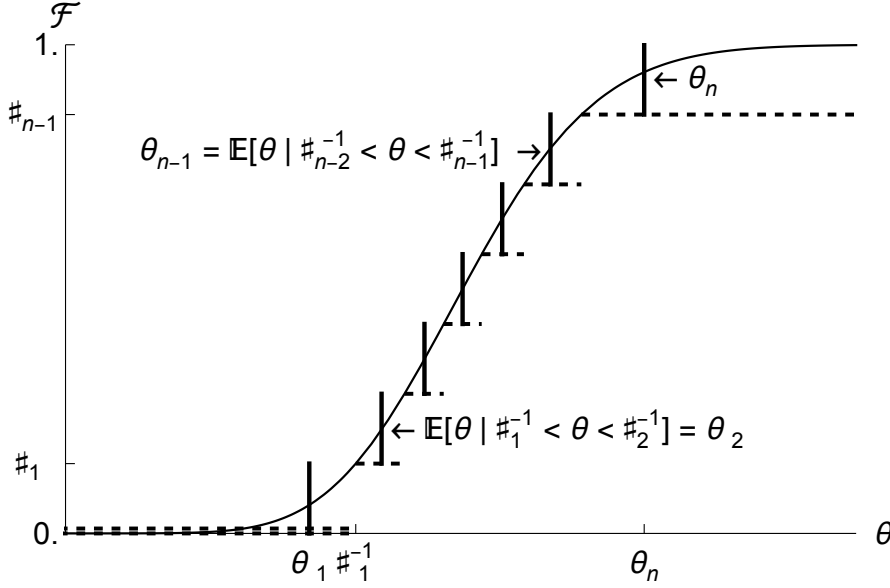
Figure 1 Equiprobable Discrete Approximation to Lognormal Distribution $\mathcal{F}$

{fig:discreteapprox}

line segments represent the n_θ equiprobable values of θ_i which are used to approximate this distribution.[6]

Because one of the purposes of these notes is to connect the math to the code that solves the math, we display here a brief snippet from the notebook that constructs these points.

```
# This is a snippet of code that constructs mass points
# for the equiprobable representation of the problem
```

$$\mathrm{v}_{t-1\rightarrow}(a) = \beta \left(\frac{1}{n_\theta}\right) \sum_{i=1}^{n_\theta} \frac{(\mathcal{R}_t a + \theta_i)^{1-\rho}}{1-\rho} \qquad (27)$$

{eq:vDiscrete}

We now substitute our approximation (27) for $\mathrm{v}_{(t-1)\rightarrow}(a)$ in (24) which is simply the sum of n_θ numbers and is therefore easy to calculate (compared to the full-fledged numerical integration (25) that it replaces).

```
# The code that corresponds to evaluation of the discretized max problem is
```

6.2 The Approximate Consumption and Value Functions

Given any particular value of m, a numerical maximization tool can now find the c that solves (24) in a reasonable amount of time.

The notebook code responsible for computing an estimated consumption function begins in "Solving the Model by Value Function Maximization," where a vector con-

---

[6]More sophisticated approximation methods exist (e.g. Gauss-Hermite quadrature; see Kopecky and Suen (2010) for a discussion of other alternatives), but the method described here is easy to understand, quick to calculate, and has additional advantages briefly described in the discussion of simulation below.
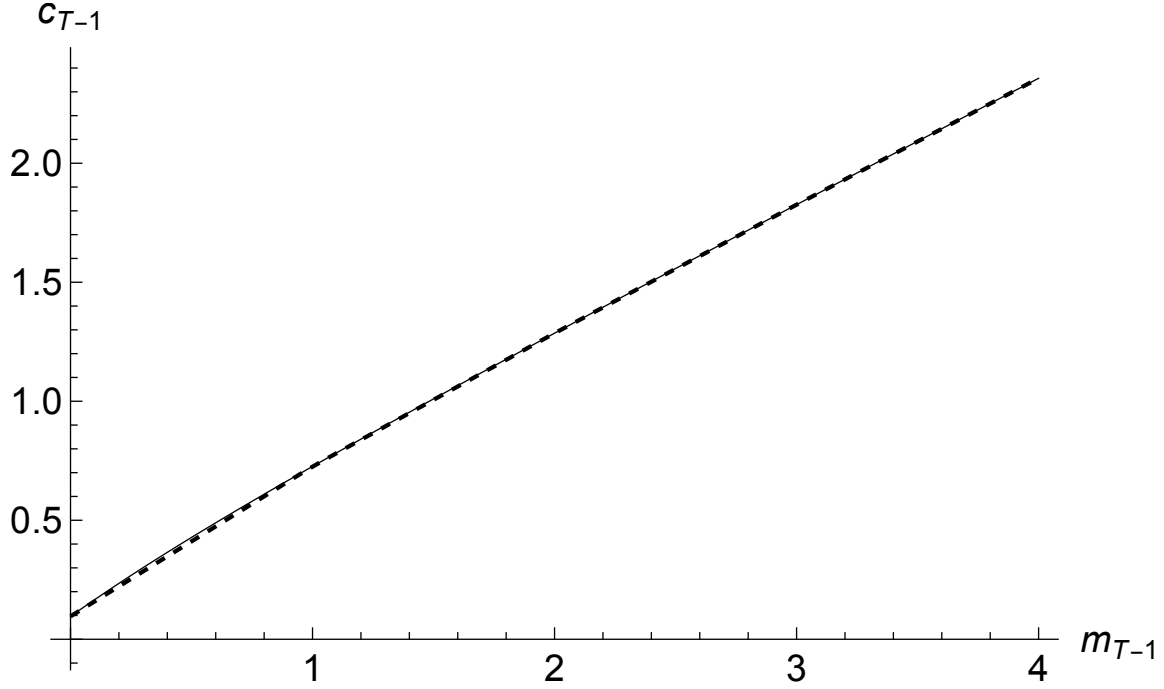
Figure 2 $c_{T-1}(m)$ (solid) versus $\grave{c}_{T-1}(m)$ (dashed)

taining a set of possible values of market resources m is created (in the code, various m vectors have names beginning `mVec`; in these notes we will use boldface italics to represent vectors, so we can refer to our collection of m points as $\boldsymbol{m}$ with values indexed by brackets: $\boldsymbol{m}[1]$ is the first entry in the vector, up to a last entry $\boldsymbol{m}[-1]$; we arbitrarily (and suboptimally) pick the first five integers as our five `mVec` gridpoints (in the code, `mVec_int`$= \{0., 1., 2., 3., 4.\}$)).

6.3 An Interpolated Consumption Function

This is accomplished in "An Interpolated Consumption Function," which generates an interpolating function that we designate $\grave{c}_{(t-1)}(m)$.

Figures 2 and 3 show plots of the constructed $\grave{c}_{t-1}$ and $\grave{v}_{t-1}$. While the $\grave{c}_{t-1}$ function looks very smooth, the fact that the $\grave{v}_{t-1}$ function is a set of line segments is very evident. This figure provides the beginning of the intuition for why trying to approximate the value function directly is a bad idea (in this context).[7]

6.4 Interpolating Expectations

Piecewise linear 'spline' interpolation as described above works well for generating a good approximation to the true optimal consumption function. However, there is a

---

[7]For some problems, especially ones with discrete choices, value function approximation is unavoidable; nevertheless, even in such problems, the techniques sketched below can be very useful across much of the range over which the problem is defined.
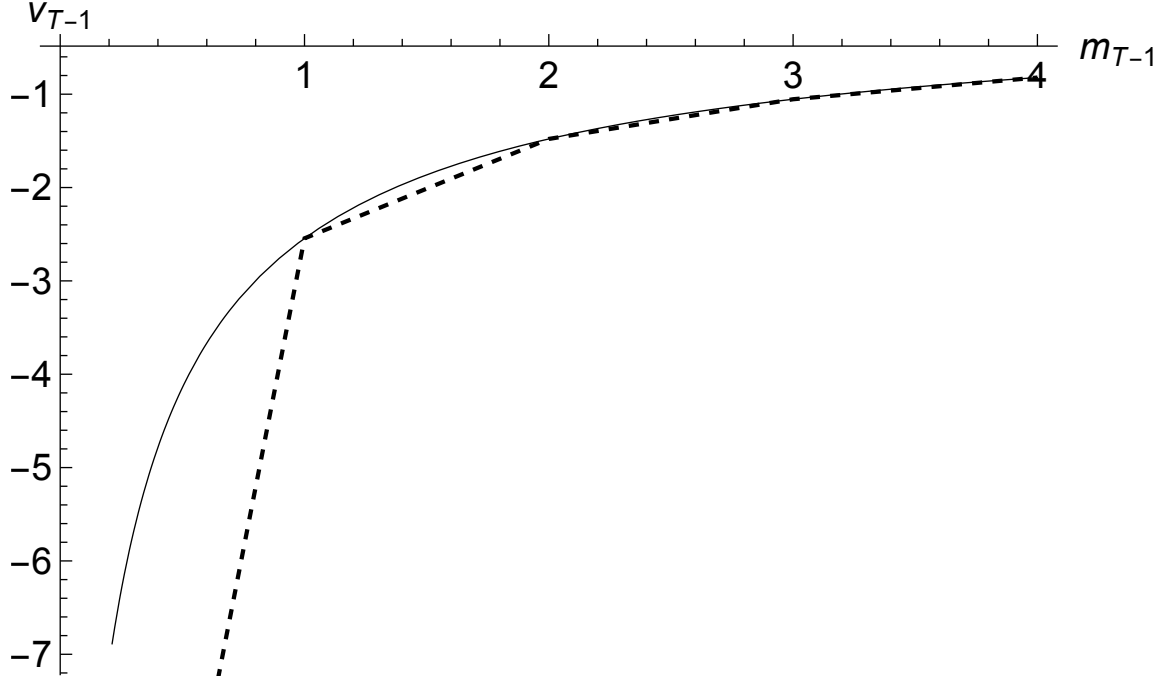
Figure 3 v_{T-1} (solid) versus $\grave{v}_{T-1}(m)$ (dashed)

clear inefficiency in the program: Since it uses equation (24), for every value of m the program must calculate the utility consequences of various possible choices of c (and therefore a_{t-1}) as it searches for the best choice.

For any given index j in $\boldsymbol{m}[j]$, as it searches for the corresponding optimal a, the algorithm will end up calculating $v_{(t-1)_\rightarrow}(\tilde{a})$ for many $\tilde{a}$ values close to the optimal a_{t-1}. Indeed, even when searching for the optimal a for a *different* m (say $\boldsymbol{m}[k]$ for $k \neq j$) the search process might compute $v_{(t-1)_\rightarrow}(a)$ for an a close to the correct optimal a for $\boldsymbol{m}[j]$. But if that difficult computation does not correspond to the exact solution to the $\boldsymbol{m}[k]$ problem, it is discarded.

The notebook section "Interpolating Expectations," now interpolates the expected value of *ending* the period with a given amount of assets.[8]

Figure 4 compares the true value function to the approximation produced by following the interpolation procedure; the approximated and exact functions are of course identical at the gridpoints of $\boldsymbol{a}$ and they appear reasonably close except in the region below $m = 1$.

Nevertheless, the consumption rule obtained when the approximating $\grave{v}_{(t-1)_\rightarrow}(a_{t-1})$ is used instead of $v_{(t-1)_\rightarrow}(a_{t-1})$ is surprisingly bad, as shown in figure 5. For example, when m goes from 2 to 3, $\grave{c}_{t-1}$ goes from about 1 to about 2, yet when m goes from 3 to 4, $\grave{c}$ goes from about 2 to about 2.05. The function fails even to be concave, which is distressing because Carroll and Kimball (1996) prove that the correct consumption function is strictly concave in a wide class of problems that includes this one.

In all figs, replace gothic h with notation corresponding to the lecture notes.

---

[8]What we are doing here is closely related to 'the method of parameterized expectations' of den Haan and Marcet (1990); the only difference is that our method is essentially a nonparametric version.

Figure 4 End-Of-Period Value $v_{(t-1)\rightarrow}(a_{t-1})$ (solid) versus $\grave{v}_{(T-1)\rightarrow}(a_{T-1})$ (dashed)
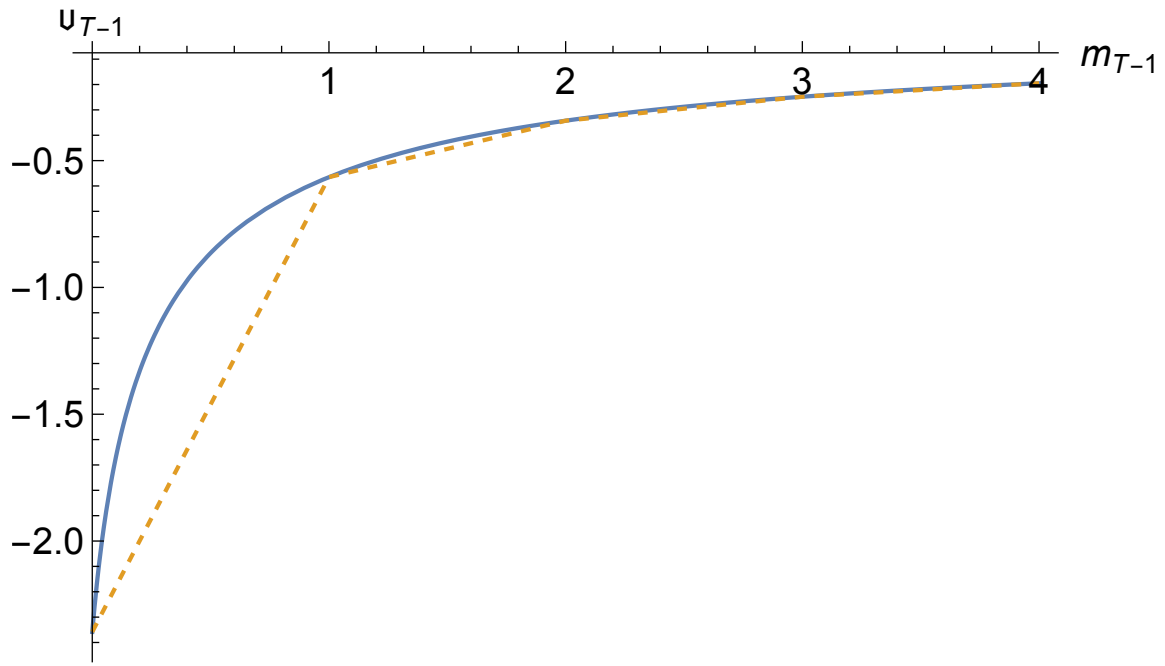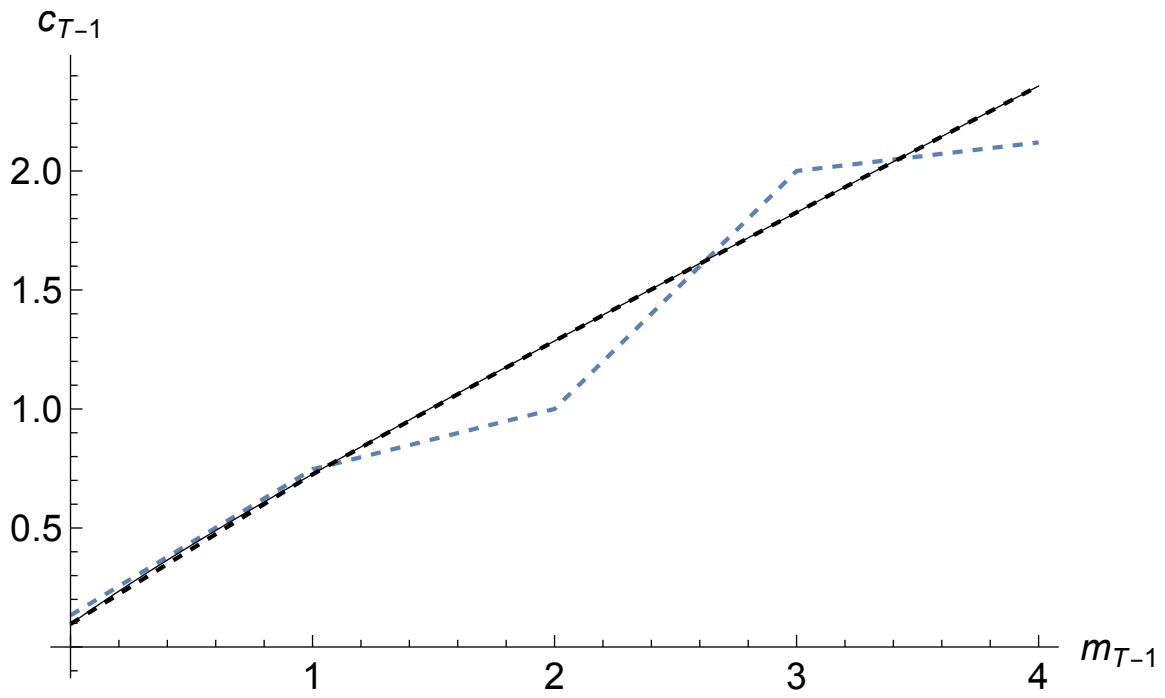
{fig:PlotOTm1Raw}

Figure 5 $c_{T-1}(m)$ (solid) versus $\grave{c}_{T-1}(m)$ (dashed)

{fig:PlotComparecT

6.5 Value Function versus First Order Condition

Loosely speaking, our difficulty reflects the fact that the consumption choice is governed by the *marginal* value function, not by the *level* of the value function (which is the object that we approximated). To understand this point, recall that a quadratic utility function exhibits risk aversion because with a stochastic c,

$$\mathbb{E}[-(c - \not{c})^2] < -(\mathbb{E}[c] - \not{c})^2 \tag{28}$$

(where $\not{c}$ is the 'bliss point' which is assumed always to exceed feasible c). However, unlike the CRRA utility function, with quadratic utility the consumption/saving *behavior* of consumers is unaffected by risk since behavior is determined by the first order condition, which depends on *marginal* utility, and when utility is quadratic, marginal utility is unaffected by risk:

$$\mathbb{E}[-2(c - \not{c})] = -2(\mathbb{E}[c] - \not{c}). \tag{29}$$

Intuitively, if one's goal is to accurately capture choices that are governed by marginal value, numerical techniques that approximate the *marginal* value function will yield a more accurate approximation to optimal behavior than techniques that approximate the *level* of the value function.

The first order condition of the maximization problem in period $T - 1$ is:

$$\mathrm{u}^c(c) = \beta\, \mathbb{E}_{\rightarrow(T-1)}[\mathsf{R}\mathrm{u}^c(c_t)]$$

$$c^{-\rho} = \mathsf{R}\beta \left(\frac{1}{n_\theta}\right) \sum_{i=1}^{n_\theta} \left(\mathsf{R}(m - c) + \theta_i\right)^{-\rho}. \tag{30}$$

The downward-sloping curve in Figure 6 shows the value of $c^{-\rho}$ for our baseline parameter values for $0 \leq c \leq 4$ (the horizontal axis). The solid upward-sloping curve shows the value of the RHS of (30) as a function of c under the assumption that $m = 3$. Constructing this figure is time-consuming, because for every value of c plotted we must calculate the RHS of (30). The value of c for which the RHS and LHS of (30) are equal is the optimal level of consumption given that $m = 3$, so the intersection of the downward-sloping and the upward-sloping curves gives the (approximated) optimal value of c. As we can see, the two curves intersect just below $c = 2$. Similarly, the upward-sloping dashed curve shows the expected value of the RHS of (30) under the assumption that $m = 4$, and the intersection of this curve with $\mathrm{u}^c(c)$ yields the optimal level of consumption if $m = 4$. These two curves intersect slightly below $c = 2.5$. Thus, increasing m from 3 to 4 increases optimal consumption by about 0.5.

Now consider the derivative of our function $\grave{\mathrm{v}}_{(t-1)}(a_{t-1})$. Because we have constructed $\grave{\mathrm{v}}_{(t-1)}$ as a linear interpolation, the slope of $\grave{\mathrm{v}}_{(t-1)}(a_{t-1})$ between any two adjacent points $\{\boldsymbol{a}[i], \boldsymbol{a}[i+1]\}$ is constant. The level of the slope immediately below any particular gridpoint is different, of course, from the slope above that gridpoint, a fact which implies that the derivative of $\grave{\mathrm{v}}_{(t-1)\rightarrow}(a_{t-1})$ follows a step function.

The solid-line step function in Figure 6 depicts the actual value of $\grave{\mathrm{v}}^a_{(t-1)\rightarrow}(3 - c)$. When we attempt to find optimal values of c given m using $\grave{\mathrm{v}}_{(t-1)\rightarrow}(a_{t-1})$, the numerical
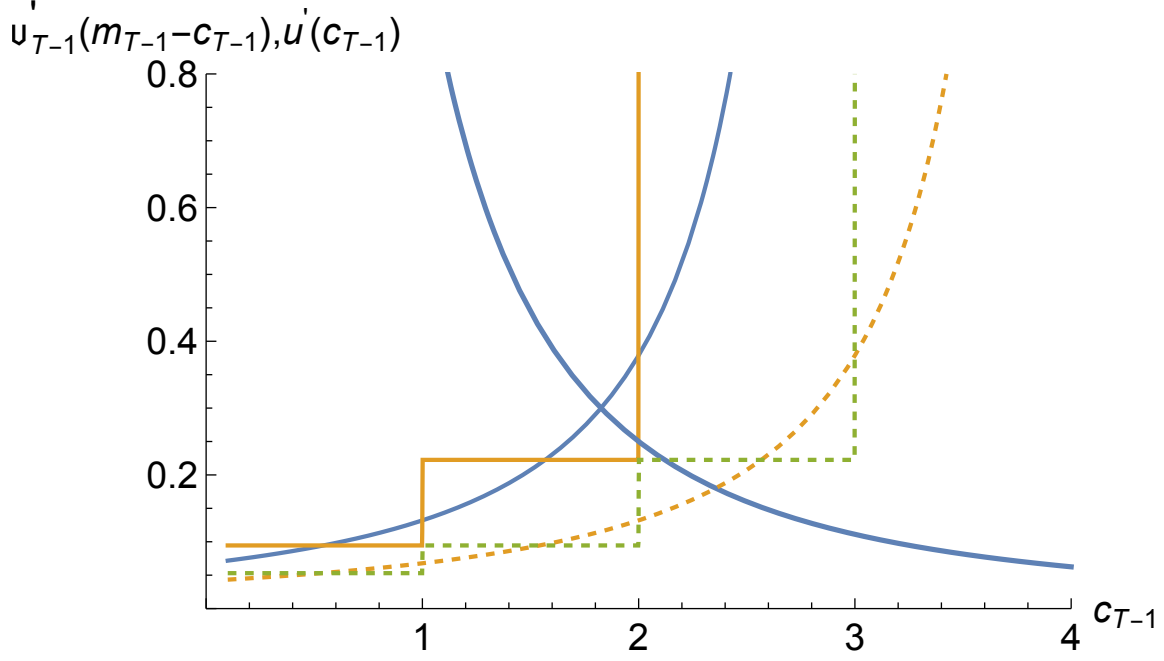
Figure 6 $u^c(c)$ versus $v^a_{(T-1)\rightarrow}(3-c), v^a_{(T-1)\rightarrow}(4-c), \grave{v}^a_{(T-1)\rightarrow}(3-c), \grave{v}^a_{(T-1)\rightarrow}(4-c)$

{fig:PlotuPrimeVSO

optimization routine will return the c for which $u^c(c) = \grave{v}^a_{(t-1)\rightarrow}(m - c)$. Thus, for $m = 3$ the program will return the value of c for which the downward-sloping $u^c(c)$ curve intersects with the $\grave{v}^a_{(t-1)\rightarrow}(3-c)$; as the diagram shows, this value is exactly equal to 2. Similarly, if we ask the routine to find the optimal c for $m = 4$, it finds the point of intersection of $u^c(c)$ with $\grave{v}^a_{(t-1)\rightarrow}(4 - c)$; and as the diagram shows, this intersection is only slightly above 2. Hence, this figure illustrates why the numerical consumption function plotted earlier returned values very close to $c = 2$ for both $m = 3$ and $m = 4$.

We would obviously obtain much better estimates of the point of intersection between $u^c(c)$ and $v^a_{(t-1)\rightarrow}(m - c)$ if our estimate of $\grave{v}^a_{(t-1)\rightarrow}$ were not a step function. In fact, we already know how to construct linear interpolations to functions, so the obvious next step is to construct a linear interpolating approximation to the *expected marginal value of end-of-period assets function* at the points in $\boldsymbol{a}$:

$$v^a_{(t-1)\rightarrow}(\boldsymbol{a}) = \beta R \left(\frac{1}{n_\theta}\right) \sum_{i=1}^{n_\theta} (\mathcal{R}_t\boldsymbol{a} + \theta_i)^{-\rho} \tag{31}$$

{eq:vEndPrimeTm1

yielding $\boldsymbol{v}^a_{(t-1)\rightarrow}$ (the vector of expected end-of-period-$(T - 1)$ marginal values of assets corresponding to $\mathtt{aVec}$), and construct $\grave{v}^a_{(t-1)\rightarrow}(a_{t-1})$ as the linear interpolating function that fits this set of points.

The results are shown in Figure 7. The linear interpolating approximation looks roughly as good (or bad) for the *marginal* value function as it was for the level of the value function. However, Figure 8 shows that the new consumption function (long dashes) is a considerably better approximation of the true consumption function (solid)
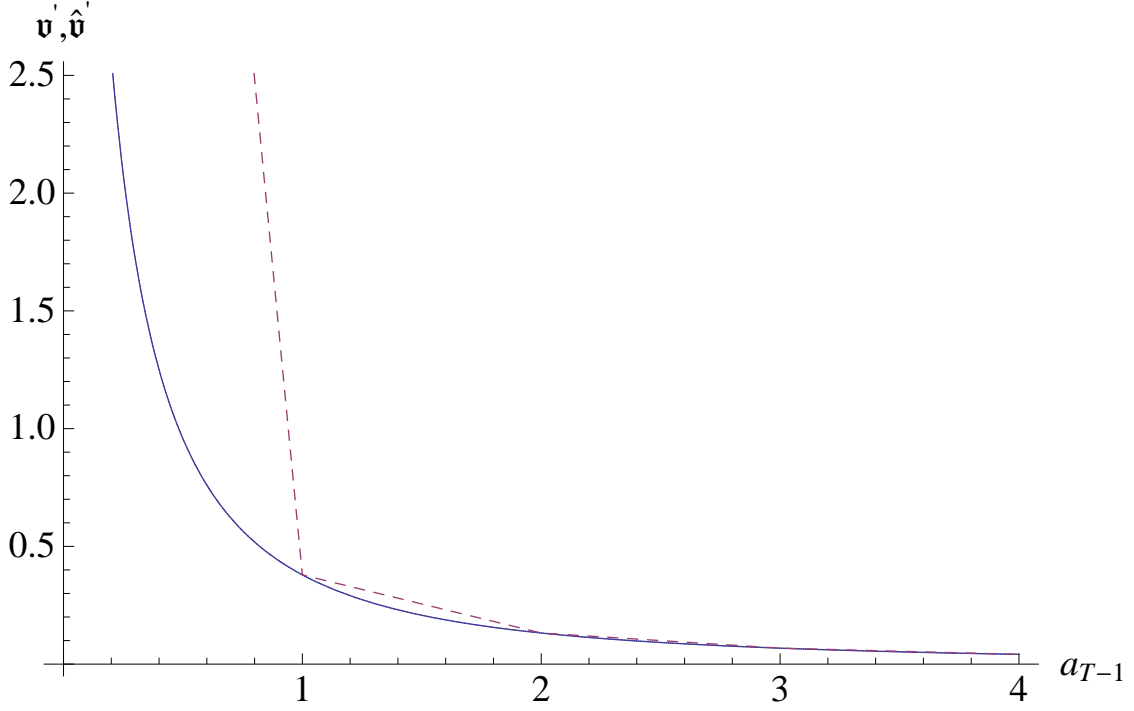
Figure 7 $\mathrm{v}^a_{(t-1)\to}(a_{t-1})$ versus $\grave{\mathrm{v}}^a_{(t-1)\to}(a_{t-1})$

{fig:PlotOPRawVSI

than was the consumption function obtained by approximating the level of the value function (short dashes).

6.6 Transformation

{subsec:transformat

Even the new-and-improved consumption function diverges notably from the true solution, especially at lower values of m. That is because the linear interpolation does an increasingly poor job of capturing the nonlinearity of $\mathrm{v}^a_{(t-1)\to}$ at lower and lower levels of a.

This is where we unveil our next trick. To understand the logic, start by considering the case where $\mathcal{R}_t = \beta = \mathcal{G}_t = 1$ and there is no uncertainty (that is, we know for sure that income next period will be $\theta_t = 1$). The final Euler equation (recall that we are still assuming that $t = T$) is then:

$$c_{t-1}^{-\rho} = c_t^{-\rho}. \tag{32}$$

In the case we are now considering with no uncertainty and no liquidity constraints, the optimizing consumer does not care whether a unit of income is scheduled to be received in the future period t or the current period $t-1$; there is perfect certainty that the income will be received, so the consumer treats its PDV as equivalent to a unit of current wealth. Total resources available at the point when the consumption decision is made is therefore comprised of two types: current market resources m and 'human

15
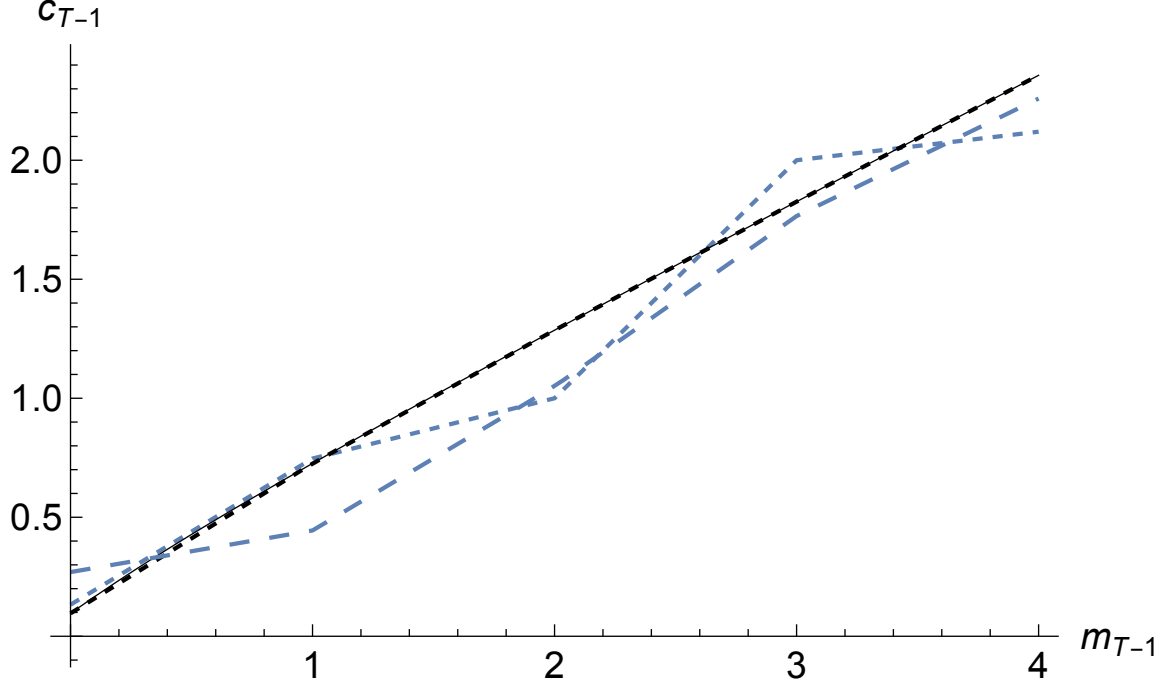
Figure 8 $c_{t-1}(m)$ (solid) Versus Two Methods for Constructing $\grave{c}_{t-1}(m)$

wealth' (the PDV of future income) of $h_{t-1} = 1$ (because it is the value of human wealth as of the end of the period, there is only one more period of income of 1 left).

$$\mathrm{v}^{m}_{(t-1)}(m) = \left(\frac{m+1}{2}\right)^{-\rho}. \tag{33}$$

Of course, this is a highly nonlinear function. However, if we raise both sides of (33) to the power $(-1/\rho)$ the result is a linear function:

$$\left[\mathrm{v}^{m}_{(t-1)}(m)\right]^{-1/\rho} = \frac{m+1}{2}. \tag{34}$$

This is a specific example of a general phenomenon: A theoretical literature discussed in Carroll and Kimball (1996) establishes that under perfect certainty, if the period-by-period marginal utility function is of the form $c_t^{-\rho}$, the marginal value function will be of the form $(\gamma m_t + \zeta)^{-\rho}$ for some constants $\{\gamma, \zeta\}$. This means that if we were solving the perfect foresight problem numerically, we could always calculate a numerically exact (because linear) interpolation.

To put the key insight in intuitive terms, the nonlinearity we are facing springs in large part from the fact that the marginal value function is highly nonlinear. But we have a compelling solution to that problem, because the nonlinearity springs largely from the fact that we are raising something to the power $-\rho$. In effect, we can 'unwind' all of the nonlinearity owing to that operation and the remaining nonlinearity will not be nearly so great. Specifically, applying the foregoing insights to the end-of-period value function

$v_{(t-1)}^a(a)$, we can define an 'inverse marginal value' function

$$\Lambda_{t\rightarrow}^a(a) \equiv \left(v_{t\rightarrow}^a(a)\right)^{-1/\rho} \tag{35}$$

{eq:cGoth}

which would be linear in the perfect foresight case.[9] We then construct a piecewise-linear interpolating approximation to the Λ_t^a function, $\lambda_{t\rightarrow}^a(a_t)$, and for any a that falls in the range $\{\boldsymbol{a}[1], \boldsymbol{a}[-1]\}$ we obtain our approximation of marginal value from:

$$\dot{v}_t^a(a) = [\lambda_t^a(a)]^{-\rho} \tag{36}$$

The most interesting thing about all of this, though, is that the Λ_t^a function has another interpretation. Recall our point in (23) that $u^c(c_t) = v_{\rightarrow}^a(m_t - c_t)$. Since with CRRA utility $u^c(c) = c^{-\rho}$, this can be rewritten and inverted

$$
\begin{aligned}
(c_t)^{-\rho} &= v_{\rightarrow}^a(a_t) \\
c_t &= \left(v_{t\rightarrow}^a(a)\right)^{-1/\rho}.
\end{aligned}
\tag{37}
$$

What this means is that for any given a, if we can calculate the marginal value associated with ending the period with that a, then we can learn the level of c that the consumer must have chosen if they ended up with that a as the result of an optimal unconstrained choice. This leads us to an alternative interpretation of Λ^a. It is the function that reveals, for any ending a, how much the agent must have consumed to (optimally) get to that a. We will therefore henceforth refer to it as the 'consumed function:'

$$\dot{c}_{t\rightarrow}(a_t) \equiv \lambda_{t\rightarrow}^a(a_t). \tag{38}$$

{eq:consumedfn}

Thus, for example, for period $t-1$ our procedure is to calculate the vector of $\boldsymbol{c}$ points on the consumed function:

$$\boldsymbol{c} = c_{(t-1)\rightarrow}(\boldsymbol{a}) \tag{39}$$

{eq:consumedfnvecs

with the idea that we will construct an approximation of the consumed function $\dot{c}_{(t-1)\rightarrow}(a)$ as the interpolating function connecting these $\{\boldsymbol{a}, \boldsymbol{c}\}$ points.

6.7 The Natural Borrowing Constraint and the a_{t-1} Lower Bound

{subsec:LiqConstrS

This is the appropriate moment to ask an awkward question: How should an interpolated, approximated 'consumed' function like $\dot{c}_{(t-1)\rightarrow}(a_{t-1})$ be extrapolated to return an estimated 'consumed' amount when evaluated at an a_{t-1} outside the range spanned by $\{\boldsymbol{a}[1], ..., \boldsymbol{a}[n]\}$?

For most canned piecewise-linear interpolation tools like scipy.interpolate, when the 'interpolating' function is evaluated at a point outside the provided range, the algorithm extrapolates under the assumption that the slope of the function remains constant beyond its measured boundaries (that is, the slope is assumed to be equal to the slope of nearest piecewise segment *within* the interpolated range); for example, if the bottommost gridpoint is $a_1 = \boldsymbol{a}[1]$ and the corresponding consumed level is $c_1 = c_{(t-1)\rightarrow}(a_1)$ we could

---

[9]There is a corresponding inverse for the value function: $\Lambda_{t\rightarrow}(a_t) = ((1-\rho)v_{t\rightarrow})^{1/(1-\rho)}$, and for the marginal marginal value function etc.

calculate the 'marginal propensity to have consumed' $\varkappa_1 = \grave{c}^a_{(t-1)\rightarrow}(a_1)$ and construct the approximation as the linear extrapolation below $\boldsymbol{a}[1]$ from:

$$\grave{c}_{(t-1)\rightarrow}(a) \equiv c_1 + (a - a_1)\varkappa_1. \tag{40}$$

To see that this will lead us into difficulties, consider what happens to the true (not approximated) $\mathrm{v}^a_{(t-1)\rightarrow}(a_{t-1})$ as a_{t-1} approaches a quantity we will call the 'natural borrowing constraint': $\underline{a}_{t-1} = -\underline{\theta}\mathcal{R}_t^{-1}$. From (31) we have

$$\lim_{a\downarrow\underline{a}_{t-1}} \mathrm{v}^a_{(t-1)\rightarrow}(a) = \lim_{a\downarrow\underline{a}_{t-1}} \beta\mathsf{R}\left(\frac{1}{n_\theta}\right)\sum_{i=1}^{n_\theta}(a\mathcal{R}_t + \theta_i)^{-\rho}. \tag{41}$$

But since $\underline{\theta} = \theta_1$, exactly at $a = \underline{a}_{t-1}$ the first term in the summation would be $(-\underline{\theta} + \theta_1)^{-\rho} = 1/0^\rho$ which is infinity. The reason is simple: $-\underline{a}_{t-1}$ is the PDV, as of $t-1$, of the *minimum possible realization of income* in t ($\mathcal{R}_t\underline{a}_{t-1} = -\theta_1$). Thus, if the consumer borrows an amount greater than or equal to $\underline{\theta}\mathcal{R}_t^{-1}$ (that is, if the consumer ends $t-1$ with $a_{t-1} \leq -\underline{\theta}\mathcal{R}_t^{-1}$) and then draws the worst possible income shock in period t, they will have to consume zero in period t, which yields $-\infty$ utility and $+\infty$ marginal utility.

As Zeldes (1989) first noticed, this means that the consumer faces a 'self-imposed' (or, as above, 'natural') borrowing constraint (which springs from the precautionary motive): They will never borrow an amount greater than or equal to $\underline{\theta}\mathcal{R}_t^{-1}$ (that is, assets will never reach the lower bound of $\underline{a}_{t-1}$). The constraint is 'self-imposed' in the precise sense that if the utility function were different (say, Constant Absolute Risk Aversion), the consumer might be willing to borrow more than $\underline{\theta}\mathcal{R}_t^{-1}$ because a choice of zero or negative consumption in period t would yield some finite amount of utility.[10]

This self-imposed constraint cannot be captured well when the $\mathrm{v}^a_{(t-1)\rightarrow}$ function is approximated by a piecewise linear function like $\grave{\mathrm{v}}^m_{(t-1)\rightarrow}$, because it is impossible for the linear extrapolation below $\underline{a}$ to correctly predict $\mathrm{v}^a_{(t-1)\rightarrow}(\underline{a}_{t-1}) = \infty$.

So, the marginal value of saving approaches infinity as $a \downarrow \underline{a}_{t-1} = -\underline{\theta}\mathcal{R}_t^{-1}$. But this implies that $\lim_{a\downarrow\underline{a}_{t-1}} c_{(t-1)\rightarrow}(a) = (\mathrm{v}^a_{(t-1)\rightarrow}(a))^{-1/\rho} = 0$; that is, as a approaches its 'natural borrowing constraint' minimum possible value, the corresponding amount of worst-case c must approach *its* lower bound: zero.

The upshot is a realization that all we need to do to address these problems is to prepend each of the $\boldsymbol{a}_{\mathsf{t}-1}$ and $\boldsymbol{c}_{\mathsf{t}-1}$ from (39) with an extra point so that the first element in the mapping that produces our interpolation function is $\{\underline{a}_{t-1}, 0.\}$. This is done in section "The Self-Imposed 'Natural' Borrowing Constraint and the a_{t-1} Lower Bound" of the notebook.

Figure 9 shows the result. The solid line calculates the exact numerical value of the consumed function $c_{(t-1)\rightarrow}(a)$ while the dashed line is the linear interpolating approximation $\grave{c}_{(t-1)\rightarrow}(a)$. This figure illustrates the value of the transformation: The true function

---

[10]Though it is very unclear what a proper economic interpretation of negative consumption might be – this is an important reason why CARA utility, like quadratic utility, is increasingly not used for serious quantitative work, though it is still useful for teaching purposes.
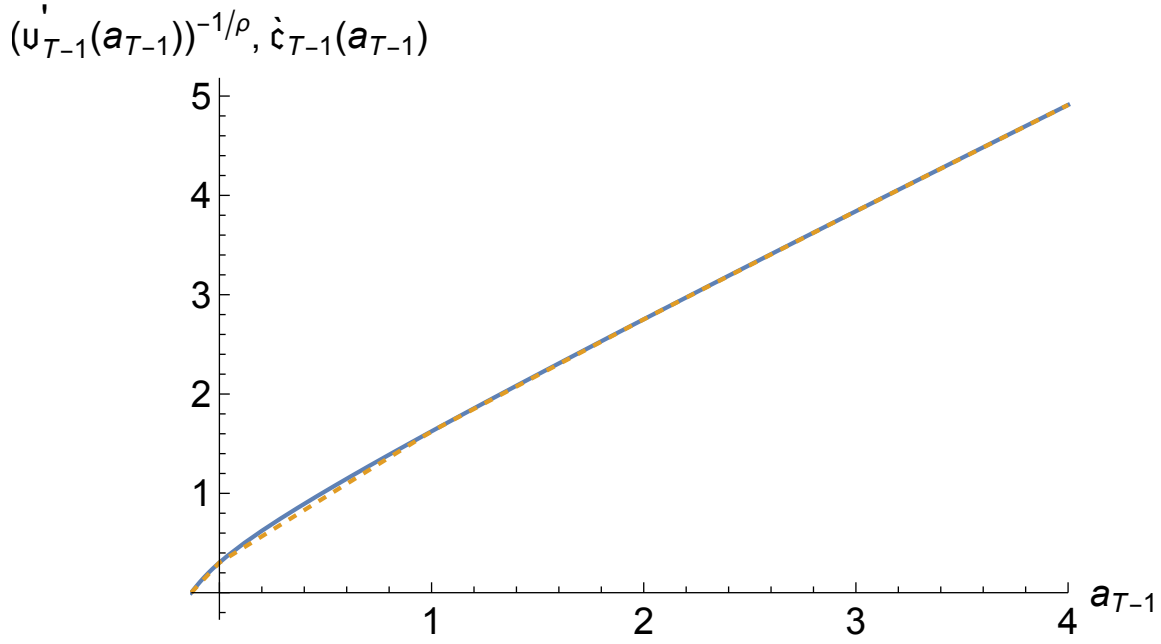
$(\mathsf{u}'_{T-1}(a_{T-1}))^{-1/\rho}, \grave{\mathsf{c}}_{T-1}(a_{T-1})$

Figure 9 True $\Lambda^a_{(t-1)_{\rightarrow}}(a)$ vs its approximation $\grave{\lambda}^a_{(t-1)_{\rightarrow}}(a)$

{fig:GothVInvVSGo}

is close to linear, and so the linear approximation is almost indistinguishable from the true function except at the very lowest values of a.

Figure 10 similarly shows that when we generate $\grave{\mathsf{v}}^a_{(t-1)_{\rightarrow}}(a)$ using our augmented $[\grave{\mathsf{c}}_{(t-1)_{\rightarrow}}(a)]^{-\rho}$ (dashed line) we obtain a *much* closer approximation to the true marginal value function $\mathsf{v}^a_{(t-1)_{\rightarrow}}(a)$ (solid line) than we obtained in the previous exercise which did not do the transformation (Figure 7).[11]

fix the problem articulated in the footnote

6.8 The Method of Endogenous Gridpoints ('EGM')

{subsec:egm}

The solution procedure we articulated above for finding $\mathsf{c}_{t-1}(m)$ still requires us, for each point in $\boldsymbol{m}_{\mathtt{t-1}}$, to use a numerical rootfinding algorithm to search for the value of c that solves $\mathsf{u}^c(c) = \mathsf{v}^a_{(t-1)_{\rightarrow}}(m-c)$. Though sections 6.6 and 6.7 developed a highly efficient and accurate procedure to calculate $\grave{\mathsf{v}}^a_{(t-1)_{\rightarrow}}$, those approximations do nothing to eliminate the need for using a rootfinding operation for calculating, for an arbitrary m, the optimal c. And rootfinding is a notoriously computation-intensive (that is, slow!) operation.

Fortunately, it turns out that there is a way to completely skip this slow rootfinding step. The method can be understood by noting that we have already calculated, for a set of arbitrary values of $\boldsymbol{a} = \boldsymbol{a}_{\mathtt{t-1}}$, the corresponding $\boldsymbol{c}$ values for which this $\boldsymbol{a}$ is optimal.

---

[11] The vertical axis label uses $\mathfrak{v}'$ as an alternative notation for what in these notes we designate as $\mathsf{v}^a_{(t-1)_{\rightarrow}}$). This will be fixed.
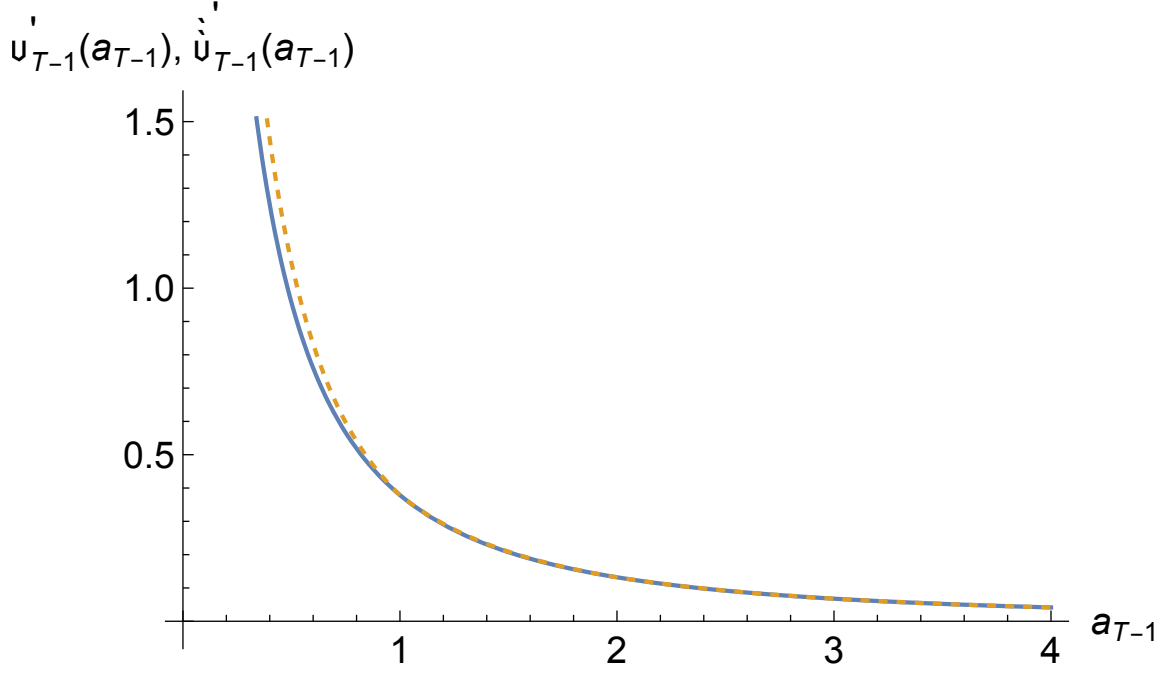
Figure 10 True $v^a_{(t-1)_\to}(a)$ vs. $\grave{v}^a_{(t-1)_\to}(a)$ Constructed Using $\grave{c}_{(t-1)_\to}(a)$

{fig:GothVVSGothC

But with mutually consistent values of $\boldsymbol{c}_{\mathtt{t-1}}$ and $\boldsymbol{a}_{\mathtt{t-1}}$ (consistent, in the sense that they are the unique optimal values that correspond to the solution to the problem), we can obtain the $\boldsymbol{m}_{\mathtt{t-1}}$ vector that corresponds to both of them from

$$\boldsymbol{m}_{\mathtt{t-1}} = \boldsymbol{c}_{\mathtt{t-1}} + \boldsymbol{a}_{\mathtt{t-1}}. \tag{42}$$

These m gridpoints are "endogenous" in contrast to the usual solution method of specifying some *ex-ante* (exogenous) grid of values of $\boldsymbol{m}$ and then using a rootfinding routine to locate the corresponding optimal consumption vector $\boldsymbol{c}$.

This routine is performed in the "Endogenous Gridpoints" section of the notebook. First, the `gothic.C_Tminus1` function is called for each of the pre-specfied values of end-of-period assets stored in `aVec`. These values of consumption and assets are used to produce the list of endogenous gridpoints, stored in the object `mVec_egm`. With the $\boldsymbol{c}$ values in hand, the notebook can generate a set of $\boldsymbol{m}_{\mathtt{t-1}}$ and $\boldsymbol{c}_{\mathtt{t-1}}$ pairs that can be interpolated between in order to yield $\grave{c}_{(t-1)}(m)$ at virtually zero computational cost![12]

One might worry about whether the $\{m, c\}$ points obtained in this way will provide a good representation of the consumption function as a whole, but in practice there are good reasons why they work well (basically, this procedure generates a set of gridpoints that is naturally dense right around the parts of the function with the greatest non-linearity). Figure 11 plots the actual consumption function c_{t-1} and the approximated consumption function $\grave{c}_{t-1}$ derived by the method of endogenous grid points. Compared

Rename gothic class, maybe to: EndPrd. Also, harmonize the notation in the notebook in the paper - for example, everwhere in the text we use cNrm for normalized consumption, but for some reason it is capital C in the gothic function.
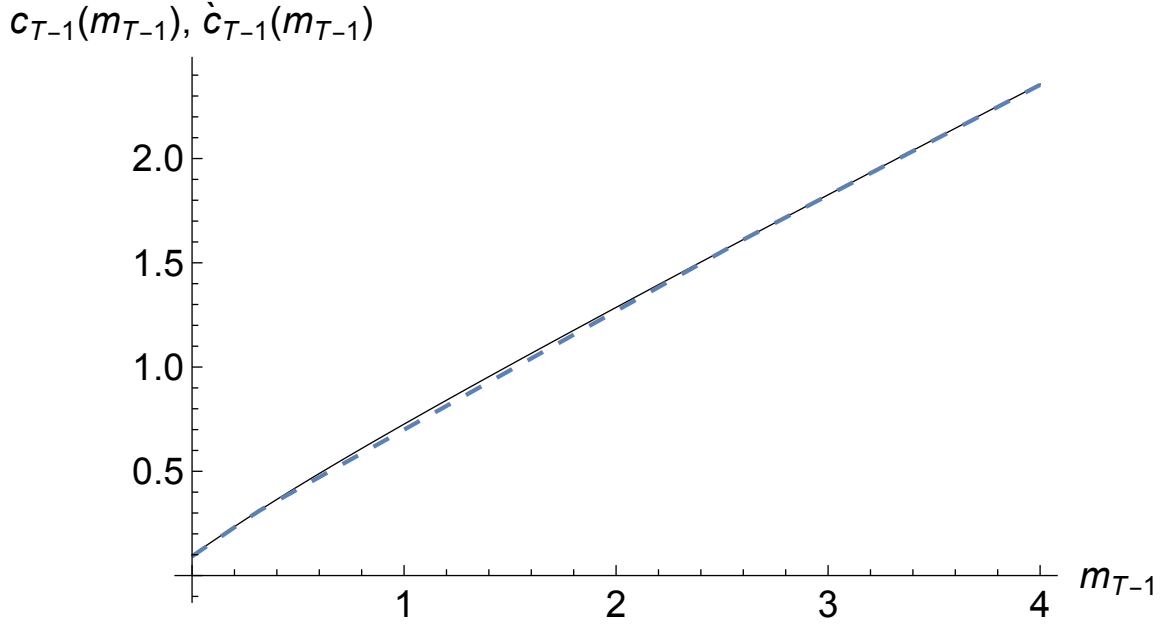
---

[12]This is the essential point of Carroll (2006).

Figure 11 $c_{t-1}(m)$ (solid) versus $\grave{c}_{t-1}(m)$ (dashed)

to the approximate consumption functions illustrated in Figure 8, $\grave{c}_{t-1}$ is quite close to the actual consumption function.

6.9 Improving the a Grid

Thus far, we have arbitrarily used a gridpoints of $\{0., 1., 2., 3., 4.\}$ (augmented in the last subsection by $\underline{a}_{t-1}$). But it has been obvious from the figures that the approximated $\grave{c}_{(t-1)\rightarrow}$ function tends to be farthest from its true value at low values of a. Combining this with our insight that $\underline{a}_{t-1}$ is a lower bound, we are now in position to define a more deliberate method for constructing gridpoints for a – a method that yields values that are more densely spaced at low values of a where the function is more nonlinear.

A pragmatic choice that works well is to find the values such that (1) the last value *exceeds the lower bound* by the same amount $\bar{a}$ as our original maximum gridpoint (in our case, 4.); (2) we have the same number of gridpoints as before; and (3) the *multi-exponential growth rate* (that is, $e^{e^{e^{\cdots}}}$ for some number of exponentiations n – our default is 3) from each point to the next point is constant (instead of, as previously, imposing constancy of the absolute gap between points).

Section "Improve the $\mathbb{A}_{grid}$" begins by defining a function which takes as arguments the specifications of an initial grid of assets and returns the new grid incorporating the multi-exponential approach outlined above.

Notice that the graphs depicted in Figures 12 and 13 are notably closer to their respective truths than the corresponding figures that used the original grid.

$$(\mathrm{\upsilon}'_{T-1}(\boldsymbol{a}_{T-1}))^{-1/\rho}, \grave{\mathrm{c}}_{T-1}(\boldsymbol{a}_{T-1})$$
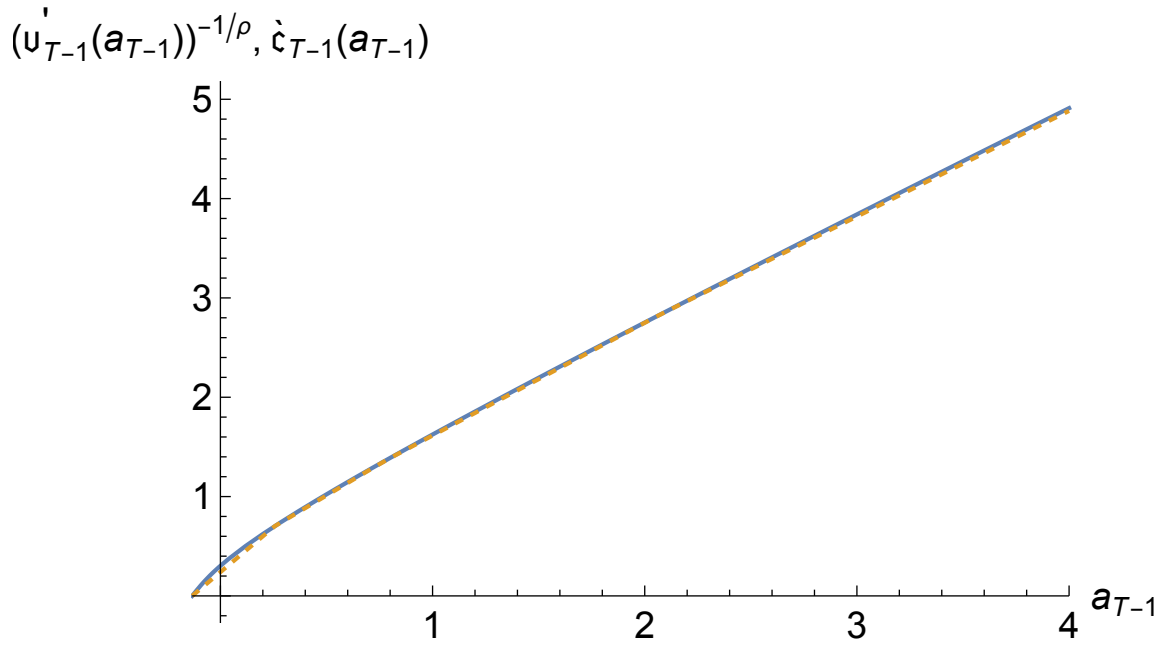
Figure 12 $\mathrm{c}_{(t-1)_{\rightarrow}}(a)$ versus $\grave{\mathrm{c}}_{(t-1)_{\rightarrow}}(a)$, Multi-Exponential `aVec`

{fig:GothVInvVSG}

$$\mathrm{\upsilon}'_{T-1}(\boldsymbol{a}_{T-1}), \grave{\grave{\mathrm{\upsilon}}}'_{T-1}(\boldsymbol{a}_{T-1})$$
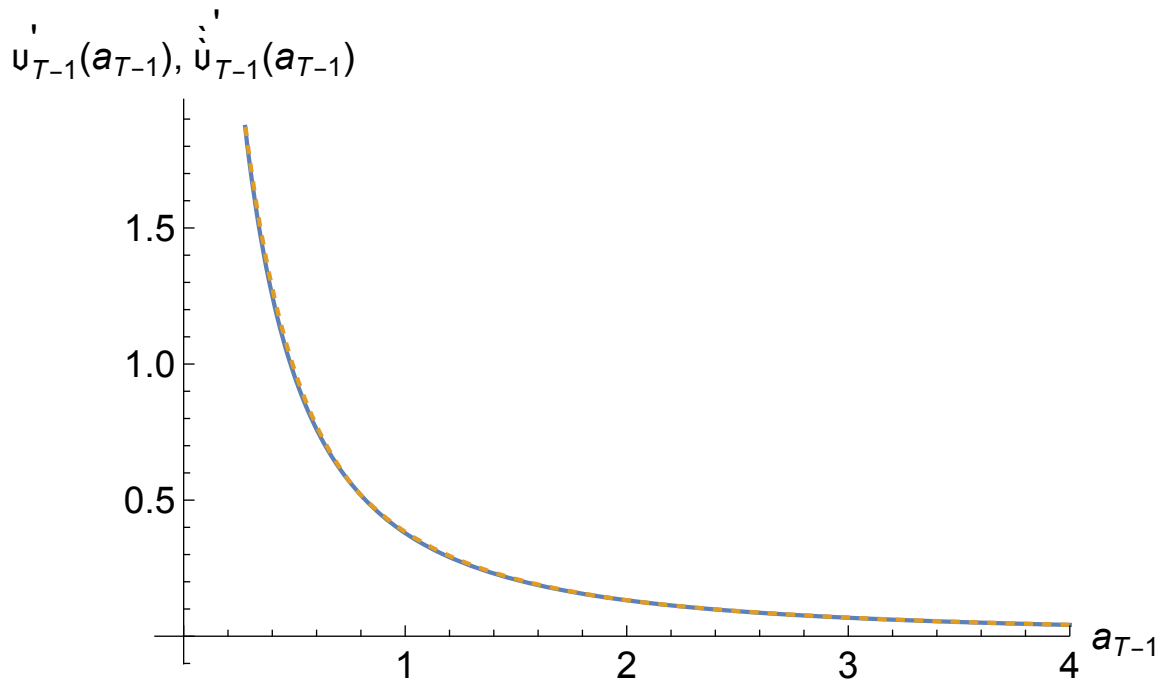
Figure 13 $\mathrm{v}^{a}_{(t-1)_{\rightarrow}}(a)$ vs. $\grave{\mathrm{v}}^{a}_{(t-1)_{\rightarrow}}(a)$, Multi-Exponential `aVec`

{fig:GothVVSGothC}

6.10 Program Structure

In section "Solve for $c_t(m)$ in Multiple Periods," the natural and artificial borrowing constraints are combined with the endogenous gridpoints method to approximate the optimal consumption function for a specific period. Then, this function is used to compute the approximated consumption in the previous period, and this process is repeated for some specified number of periods.

The essential structure of the program is a loop that iteratively solves for consumption functions by working backward from an assumed final period, using the dictionary `cFunc_life` to store the interpolated consumption functions up to the beginning period. Consumption in a given period is utilized to determine the endogenous gridpoints for the preceding period. This is the sense in which the computation of optimal consumption is done recursively.

For a realistic life cycle problem, it would also be necessary at a minimum to calibrate a nonconstant path of expected income growth over the lifetime that matches the empirical profile; allowing for such a calibration is the reason we have included the $\{\mathcal{G}\}_t^T$ vector in our computational specification of the problem.

6.11 Results

The code creates the relevant $\grave{c}_t(m)$ functions for any period in the horizon, at the given values of m. Figure 14 shows $\grave{c}_{T-n}(m)$ for $n = \{20, 15, 10, 5, 1\}$. At least one feature of this figure is encouraging: the consumption functions converge as the horizon extends, something that Carroll (2023b) shows must be true under certain parametric conditions that are satisfied by the baseline parameter values being used here.

7 The Infinite Horizon

{sec:the-infinite-hor

All of the solution methods presented so far have involved period-by-period iteration from an assumed last period of life, as is appropriate for life cycle problems. However, if the parameter values for the problem satisfy certain conditions (detailed in Carroll (2023b)), the consumption rules (and the rest of the problem) will converge to a fixed rule as the horizon (remaining lifetime) gets large, as illustrated in Figure 14. Furthermore, Deaton (1991), Carroll (1992; 1997) and others have argued that the 'buffer-stock' saving behavior that emerges under some further restrictions on parameter values is a good approximation of the behavior of typical consumers over much of the lifetime. Methods for finding the converged functions are therefore of interest, and are dealt with in this section.

Of course, the simplest such method is to solve the problem as specified above for a large number of periods. This is feasible, but there are much faster methods.
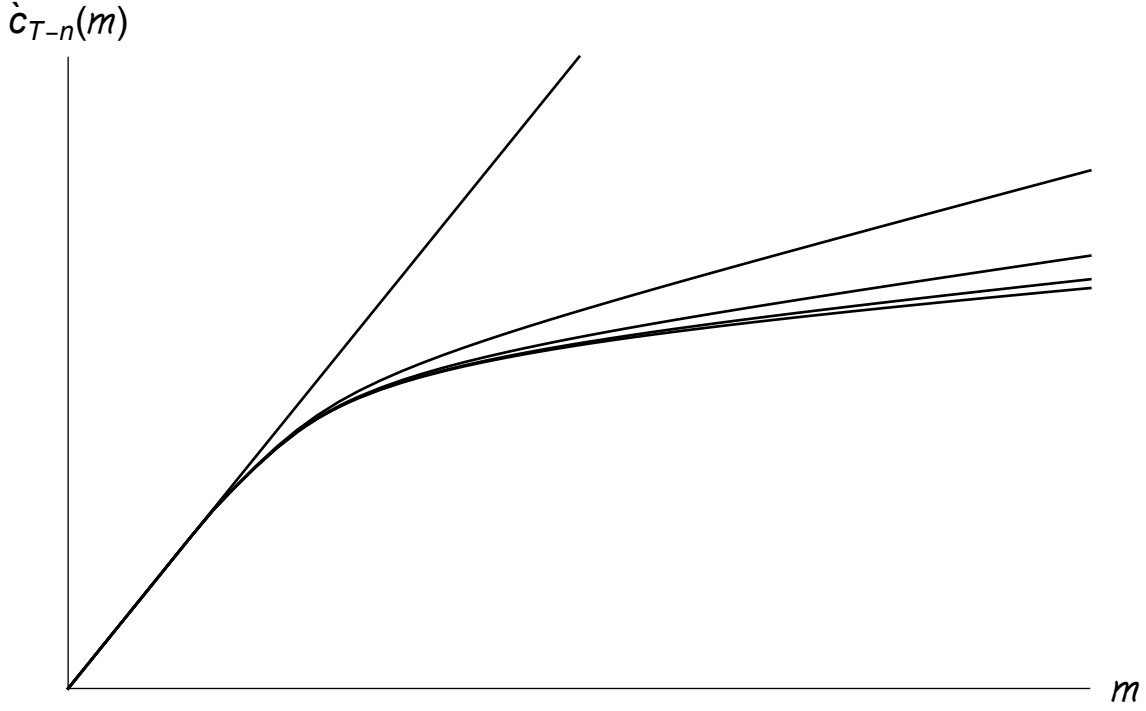
$$\grave{c}_{T-n}(m)$$

Figure 14   Converging $\grave{c}_{T-n}(m)$ Functions as n Increases

7.1 Convergence

In solving an infinite-horizon problem, it is necessary to have some metric that determines when to stop because a solution that is 'good enough' has been found.

A natural metric is defined by the unique 'target' level of wealth that Carroll (2023b) proves will exist in problems of this kind under certain conditions: The $\hat{m}$ such that

$$\mathbb{E}_t[m_{t+1}/m_t] = 1 \text{ if } m_t = \hat{m} \tag{43}$$

where the accent is meant to signify that this is the value that other m's 'point to.'

Given a consumption rule $c(m)$ it is straightforward to find the corresponding $\hat{m}$. So for our problem, a solution is declared to have converged if the following criterion is met: $|\hat{m}_{t+1} - \hat{m}_t| < \epsilon$, where ϵ is a very small number and defines our degree of convergence tolerance.

Similar criteria can obviously be specified for other problems. However, it is always wise to plot successive function differences and to experiment a bit with convergence criteria to verify that the function has converged for all practical purposes.

8 Multiple Control Variables

We now consider how to solve problems with multiple control variables.

8.1 Theory

The new portfolio-share control variable is captured by the archaic Greek character 'stigma'; it represents the share ς of their disposable assets the agent invests in the risky asset (conventionally, the stock market). Designating the return factor for the risky asset as $\mathbf{R}$ and the share of the portfolio invested in $\mathbf{R}$ as ς, the realized portfolio rate of return $\mathfrak{R}$ as a function of the share ς is:

$$\mathfrak{R}(\varsigma) = \mathsf{R} + (\mathbf{R} - \mathsf{R})\varsigma. \tag{44}$$

If we imagine the portfolio share decision as being made simultaneously with the c decision, the traditional way of writing the problem is (substituting the budget constraint):

$$\mathrm{v}_t(m) = \max_{\{c,\varsigma\}} \quad \mathrm{u}(c) + \mathbb{E}[\beta \mathrm{v}_{t+1}((m-c)\mathfrak{R}(\varsigma) + \theta_{t+1})] \tag{45}$$

where we have deliberately omitted the period-designating subscripts for ς and the return factors to highlight the point that, once the consumption and ς decisions have been made, it makes no difference to this equation whether the risky return factor $\mathbf{R}$ is revealed a nanosecond before the end of the current period or a nanosecond after the beginning of the successor period.

8.2 Stages Within a Period

which we will call the 'consumption stage c' and the 'portfolio stage ς.' These could come in either order in the period: We designate the 'portfolio choice first, then consumption' version by $[\varsigma, \mathrm{c}]$ and the 'consumption choice first, then portfolio' as $[\mathrm{c}, \varsigma]$.

In a problem with multiple stages, if we want to refer to a sub-step of a particular stage – say, the Arrival stage of the portfolio stage – we simply add a stage-indicator subscript (in square brackets) to the notation we have been using until now. That is, the Arrival stage of the portfolio problem would be $\mathrm{v}_{\leftarrow[\varsigma]}$.

8.2.1 The (Revised) Consumer's Problem

A slight modification to the consumer's problem specified earlier is necessary to make the stages of the problem completely modular. The difficulty with the earlier formulation is that it assumed that asset returns occurred in the middle step of the consumption problem. Our revised version of the consumption problem takes as its input state the amount of bank balances that have resulted from any prior portfolio decision. The problem is therefore:

$$\mathrm{v}_{[\mathrm{c}]}(m) = \max_c \quad \mathrm{u}(c) + \mathrm{v}_{[\mathrm{c}]_\rightarrow} (\underbrace{m-c}_{a})$$

$$\mathrm{v}_{\leftarrow[\mathrm{c}]}(b) = \mathbb{E}_{\leftarrow[\mathrm{c}]} \left[\mathrm{v}_{[\mathrm{c}]}(\overbrace{b+\theta}^{m}) \right] \tag{46}$$

8.2.2 The Investor's Problem

Consider the standalone problem of an 'investor' whose continuation-value function $v_{[\varsigma]\rightarrow}$ depends on how much wealth $\acute{w}$ they end up after the realization of the stochastic $\mathbf{R}$ return. The expected value that the investor will obtain from any combination of initial w and their optimal choice of the portfolio share ς is the expectation of the continuation-value function over the wealth that results from the portfolio choice:

$$v_{\leftarrow[\varsigma]}(w) = \max_{\varsigma}\ \mathbb{E}_{\leftarrow[\varsigma]}\left[v_{[\varsigma]\rightarrow}\overbrace{(\mathfrak{R}(\varsigma)w)}^{\acute{w}}\right] \tag{47}$$

where we have omitted any period designator like t for the period in which this problem is solved because, with the continuation-value function defined already as $v_{[\varsigma]\rightarrow}(\acute{w})$, the problem is self-contained. The solution to this problem will yield an optimal ς decision rule $\check{\varsigma}(w)$. Finally, we can specify the value of an investor 'arriving' with w as the expected value that will be obtained when the investor invests optimally, generating the *ex ante* optimal stochastic portfolio return factor $\check{\mathfrak{R}}(w) = \mathfrak{R}(\check{\varsigma}(w))$:

$$v_{\leftarrow[\varsigma]}(w) = \mathbb{E}_{\leftarrow}[v_{[\varsigma]\rightarrow}](\overbrace{\check{\mathfrak{R}}(w)})]. \tag{48}$$

The reward for all this notational investment is that it is now clear that *exactly the same code* for solving the portfolio share problem can be used in two distinct problems: a 'beginning-of-period-returns' model and an 'end-of-period-returns' model.

8.2.3 The 'beginning-of-period returns' Problem

The beginning-returns problem effectively just inserts a portfolio choice that happens at a stage immediately before the consumption stage in the optimal consumption problem described in (46), for which we had a beginning-of-stage value function $v_{\leftarrow[c]}(b)$. The agent makes their portfolio share decision within the stage but (obviously) before the risky returns $\mathbf{R}$ for the period have been realized. So the problem's portfolio-choice stage also takes k as its initial state and solves the investor's problem outlined in section 8.2.2 above:

$$v_{\leftarrow[\varsigma]}(k) = \mathbb{E}_{\leftarrow[\varsigma]}[v_{[\varsigma]\rightarrow}(\underbrace{k\check{\mathfrak{R}}}_{b})]$$

$$v_{[\varsigma]\rightarrow}(b) = v_{\leftarrow[c]}(b) \tag{49}$$

Since in this setup bank balances have been determined before the consumption problems starts, we need to rewrite the consumption stage as a function of bank balances that will have resulted from the portfolio investment b, combined with the income shocks θ:

$$v_{\leftarrow[c]}(b) = \max_{c}\ u(c) + \mathbb{E}_{\leftarrow[c]}[v_{[c]\rightarrow}(\underbrace{\overbrace{b + \theta}^{m} - c}_{a})] \tag{50}$$

where, because the consumption stage is the last stage in the period, the continuatibon-value function for the c stage is just the continuation-value function for the period as a whole:

$$v_{[c]\rightarrow}(a) = v_{t\rightarrow}(a) \tag{51}$$

(and recall that $v_{t\rightarrow}(a)$ is exogenously provided as an input to the period's problem via the transition equation assumed earlier: $v_{t\rightarrow}(a) = \beta v_{\leftarrow(t+1)}(a)$).

8.2.4 The 'end-of-period-returns' Problem

If the portfolio share and risky returns are realized at the end of the period, we need to move the portfolio choice stage to immediately before the point at which returns are realized (and after the c choice has been made). The problem is the same as the portfolio problem defined above, except that the input for the investment stage is the assets remaining after the consumption choice: a. So, the portfolio stage of the problem is

$$v_{\leftarrow[\varsigma]}(a) = \mathbb{E}_{\leftarrow[\varsigma]}[v_{[\varsigma]\rightarrow}(\underbrace{a\breve{\mathfrak{R}}}_{k})] \tag{52}$$

where we are designating the post-realization result of the investment as k, and since the ς-stage is the last stage of the problem the end-of-stage k becomes the end-of-period k_t.

The 'state transition' equation between t and $t+1$ is simply $b_{t+1} = k_t$ and the continuation-value function transition is $v_{t\rightarrow}(k) \leftrightarrow \beta v_{\leftarrow(t+1)}(k)$ which reflects the above-mentioned point that there is no substantive difference between the two problems (their $v_{[c]}(m)$ value functions and $c(m)$ functions will be identical).

(Note that we are assuming that there will be only one consumption function in the period, so no stage subscript is necessary to pick out 'the consumption function').

8.2.5 Numerical Solution

we can solve it numerically for the optimal ς at a vector of $\boldsymbol{a}$ (aVec in the code) and then construct an approximated optimal portfolio share function $\breve{\varsigma}(a)$ as the interpolating function among the members of the $\{\boldsymbol{a}, \varsigma\}$ mapping. Having done this, we can now calculate a vector of values and marginal values that correspond to aVec:

$$\begin{aligned} \boldsymbol{v} &= v_{\leftarrow[\varsigma]}(\boldsymbol{a}) \\ \boldsymbol{v}^a &= v^a_{\leftarrow[\varsigma]}(\boldsymbol{a}). \end{aligned} \tag{53}$$ {eq:vShrEnd}

With the $\boldsymbol{v}^a$ approximation described in hand, we can construct our approximation to the consumption function using *exactly the same EGM procedure* that we used in solving the problem *without* a portfolio choice (see (35)):

$$\boldsymbol{c} \equiv (\boldsymbol{v}^a)^{-1/\rho}, \tag{54}$$ {eq:cVecPort}

which, following a procedure identical to that in the EGM subsection 6.8, yields an approximated consumption function $\grave{c}_t(m)$. Thus, again, we can construct the consumption function at nearly zero cost (once we have calculated $\boldsymbol{v}^a$).

8.2.6 The Point

The upshot is that all we need to do is change some of the transition equations and we can use the same solution code (both for the ς-stage and the c-stage) to solve the problem with either assumption (beginning-of-period or end-of-period) about the timing of portfolio choice. There is even an obvious notation for the two problems: $v_{\llcorner t[\varsigma c]}$ can be the period-arrival value function for the version where the portfolio share is chosen at the beginning of the period, and $v_{\llcorner t[c\varsigma]}$ is period-arrival value for the the problem where the share choice is at the end.

What is the benefit of writing effectively the identical problem in two different ways? There are several:

- It demonstrates that, if they are carefully constructed, Bellman problems can be "modular"

  - In a life cycle model one might want to assume that at at some ages agents have a portfolio choice and at other ages they do not. The consumption problem makes no assumption about whether there is a portfolio choice decision (before or after the consumption choice), so there would be zero cost of having an age-varying problem in which you drop in whatever choices are appropriate to the life cycle stage.

- It emphasizes the flexibilty of choice a modeler has to date variables arbitrarily. In the specific example examined here, there is a strong case for preferring the beginning-returns specification because we typically think of productivity or other shocks at date t affecting the agent's state variables before the agent makes that period's choices. It would be awkward and confusing to have a productivity shock dated $t-1$ effectively applying for the problem being solved at t (as in the end-returns specification)

- It may help to identify more efficient solution methods

  - For example, under the traditional formulation in equation (45) it might not occur to a modeler that the endogenous gridpoints solution method can be used, because when portfolio choice and consumption choice are considered simultaneously the EGM method breaks down because the portfolio choice part of the problem is not susceptible to EGM solution. But when the problem is broken into two simpler problems, it becomes clear that EGM can still be applied to the consumption problem even though it cannot be applied to the portfolio choice problem

8.3 Application

In specifying the stochastic process for $\mathbf{R}_{t+1}$, we follow the common practice of assuming that returns are lognormally distributed, $\log \mathbf{R} \sim \mathcal{N}(\phi + \mathsf{r} - \sigma_{\mathbf{r}}^2/2, \sigma_{\mathbf{r}}^2)$ where ϕ is the equity premium over the thin returns r available on the riskless asset.[13]

As with labor income uncertainty, it is necessary to discretize the rate-of-return risk in order to have a problem that is soluble in a reasonable amount of time. We follow the same procedure as for labor income uncertainty, generating a set of $n_{\mathbf{r}}$ equiprobable shocks to the rate of return; in a slight abuse of notation, we will designate the portfolio-weighted return (contingent on the chosen portfolio share in equity, and potentially contingent on any other aspect of the consumer's problem) simply as $\mathfrak{R}_{i,j}$ (where dependence on i is allowed to permit the possibility of nonzero correlation between the return on the risky asset and the θ shock to labor income (for example, in recessions the stock market falls and labor income also declines).

The direct expressions for the derivatives of $\mathrm{v}_{\rightarrow}$ are

$$
\begin{aligned}
\mathrm{v}_{\rightarrow}^a(a_t, \varsigma_t) &= \beta \left(\frac{1}{n_{\mathbf{r}} n_\theta} \right) \sum_{i=1}^{n_\theta} \sum_{j=1}^{n_{\mathbf{r}}} \mathfrak{R}_{i,j} \left(\mathrm{c}_{t+1}(\mathfrak{R}_{i,j} a_t + \theta_i) \right)^{-\rho} \\
\mathrm{v}_{\rightarrow}^\varsigma(a_t, \varsigma_t) &= \beta \left(\frac{1}{n_{\mathbf{r}} n_\theta} \right) \sum_{i=1}^{n_\theta} \sum_{j=1}^{n_{\mathbf{r}}} (\mathbf{R}_{i,j} - \mathsf{R}) \left(\mathrm{c}_{t+1}(\mathfrak{R}_{i,j} a_t + \theta_i) \right)^{-\rho} .
\end{aligned}
\tag{55}
$$

Writing these equations out explicitly makes a problem very apparent: For every different combination of $\{a_t, \varsigma_t\}$ that the routine wishes to consider, it must perform two double-summations of $n_{\mathbf{r}} \times n_\theta$ terms. Once again, there is an inefficiency if it must perform these same calculations many times for the same or nearby values of $\{a_t, \varsigma_t\}$, and again the solution is to construct an approximation to the (inverses of the) derivatives of the $\mathrm{v}_{\rightarrow}$ function.

Details of the construction of the interpolating approximations are given below; assume for the moment that we have the approximations $\grave{\mathrm{v}}_{\rightarrow}^a$ and $\grave{\mathrm{v}}_{\rightarrow}^\varsigma$ in hand and we want to proceed. As noted above in the discussion of (45), nonlinear equation solvers can find the solution to a set of simultaneous equations. Thus we could ask one to solve

$$
\begin{aligned}
c_t^{-\rho} &= \grave{\mathrm{v}}_{t\rightarrow}^a(m_t - c_t, \varsigma_t) \\
0 &= \grave{\mathrm{v}}_{t\rightarrow}^\varsigma(m_t - c_t, \varsigma_t)
\end{aligned}
\tag{56}
$$

simultaneously for c and ς at the set of potential m_t values defined in `mVec`. However, as noted above, multidimensional constrained maximization problems are difficult and sometimes quite slow to solve.

There is a better way. Define the problem

$$
\tilde{\mathrm{v}}_{t\rightarrow}(a_t) = \max_{\varsigma_t} \quad \mathrm{v}_{\rightarrow}(a_t, \varsigma_t)
$$

$$
\text{s.t.}
$$

$$
0 \leq \varsigma_t \leq 1
$$

---

[13]This guarantees that $\mathbb{E}[\mathbf{R}] = \boldsymbol{\varphi}/\mathsf{R}$ is invariant to the choice of σ_ϕ^2; see `LogELogNorm`.

where the tilde over $\tilde{v}(a)$ indicates that this is the v that has been optimized with respect to all of the arguments other than the one still present (a_t). We solve this problem for the set of gridpoints in `aVec` and use the results to construct the interpolating function $\grave{\tilde{v}}_t^a(a_t)$.[14] With this function in hand, we can use the first order condition from the single-control problem

$$c_t^{-\rho} = \grave{\tilde{v}}_t^a(m_t - c_t)$$

to solve for the optimal level of consumption as a function of m_t using the endogenous gridpoints method described above. Thus we have transformed the multidimensional optimization problem into a sequence of two simple optimization problems.

Note the parallel between this trick and the fundamental insight of dynamic programming: Dynamic programming techniques transform a multi-period (or infinite-period) optimization problem into a sequence of two-period optimization problems which are individually much easier to solve; we have done the same thing here, but with multiple dimensions of controls rather than multiple periods.

8.4 Implementation

Following the discussion from section 8.1, to provide a numerical solution to the problem with multiple control variables, we must define expressions that capture the expected marginal value of end-of-period assets with respect to the level of assets and the share invested in risky assets. This is addressed in "Multiple Control Variables."

8.5 Results With Multiple Controls

{subsec:results-with

Figure 15 plots the $t-1$ consumption function generated by the program; qualitatively it does not look much different from the consumption functions generated by the program without portfolio choice.

But Figure 16 which plots the optimal portfolio share as a function of the level of assets, exhibits several interesting features. First, even with a coefficient of relative risk aversion of 6, an equity premium of only 4 percent, and an annual standard deviation in equity returns of 15 percent, the optimal choice is for the agent to invest a proportion 1 (100 percent) of the portfolio in stocks (instead of the safe bank account with riskless return R) is at values of a_t less than about 2. Second, the proportion of the portfolio kept in stocks is *declining* in the level of wealth - i.e., the poor should hold all of their meager assets in stocks, while the rich should be cautious, holding more of their wealth in safe bank deposits and less in stocks. This seemingly bizarre (and highly counterfactual – see Carroll (2002)) prediction reflects the nature of the risks the consumer faces. Those consumers who are poor in measured financial wealth will likely derive a high proportion of future consumption from their labor income. Since by assumption labor income risk is uncorrelated with rate-of-return risk, the covariance between their future consumption

---

[14]A faster solution could be obtained by, for each element in `aVec`, computing $v_{\rightarrow}^{\varsigma}(m_t - c_t, \varsigma)$ of a grid of values of ς, and then using an approximating interpolating function (rather than the full expectation) in the `FindRoot` command. The associated speed improvement is fairly modest, however, so this route was not pursued.
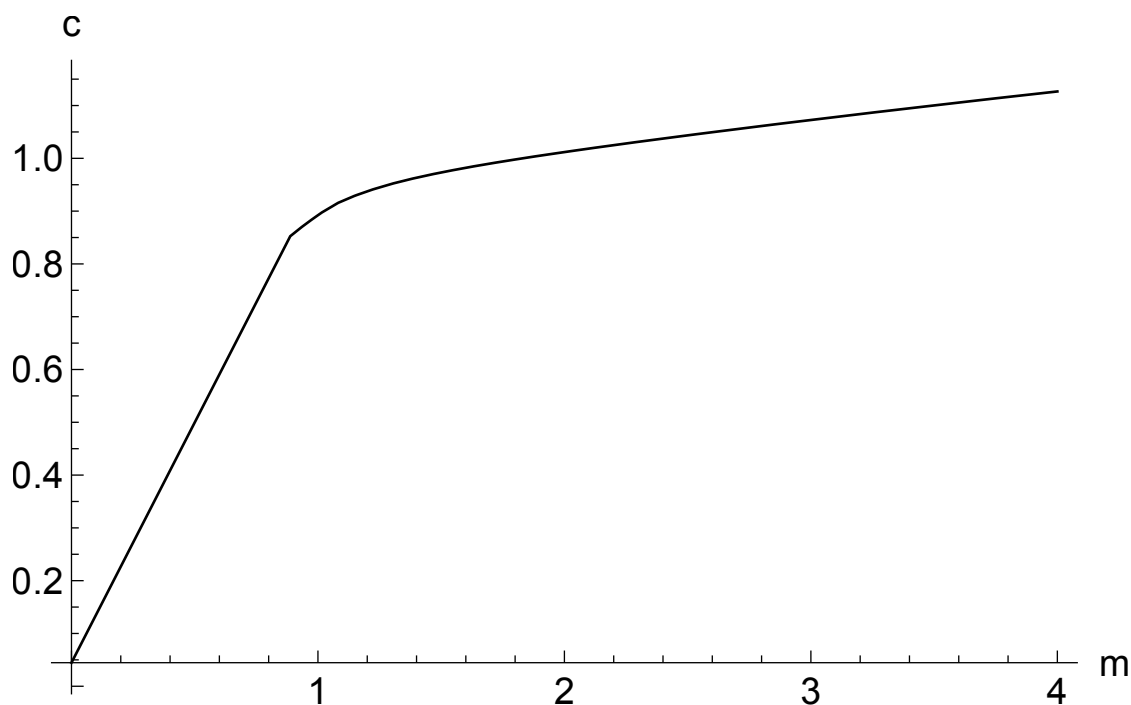
Figure 15 c(m_1) With Portfolio Choice

and future stock returns is relatively low. By contrast, persons with relatively large wealth will be paying for a large proportion of future consumption out of that wealth, and hence if they invest too much of it in stocks their consumption will have a high covariance with stock returns. Consequently, they reduce that correlation by holding some of their wealth in the riskless form.

9 Structural Estimation

This section describes how to use the methods developed above to structurally estimate a life-cycle consumption model, following closely the work of Cagetti (2003).[15] The key idea of structural estimation is to look for the parameter values (for the time preference rate, relative risk aversion, or other parameters) which lead to the best possible match between simulated and empirical moments.

9.1 Life Cycle Model

Realistic calibration of a life cycle model needs to take into account a few things that we omitted from the bare-bones model described above. For example, the whole point of the life cycle model is that life is finite, so we need to include a realistic treatment of

---

[15]Similar structural estimation exercises have been also performed by Palumbo (1999) and Gourinchas and Parker (2002).
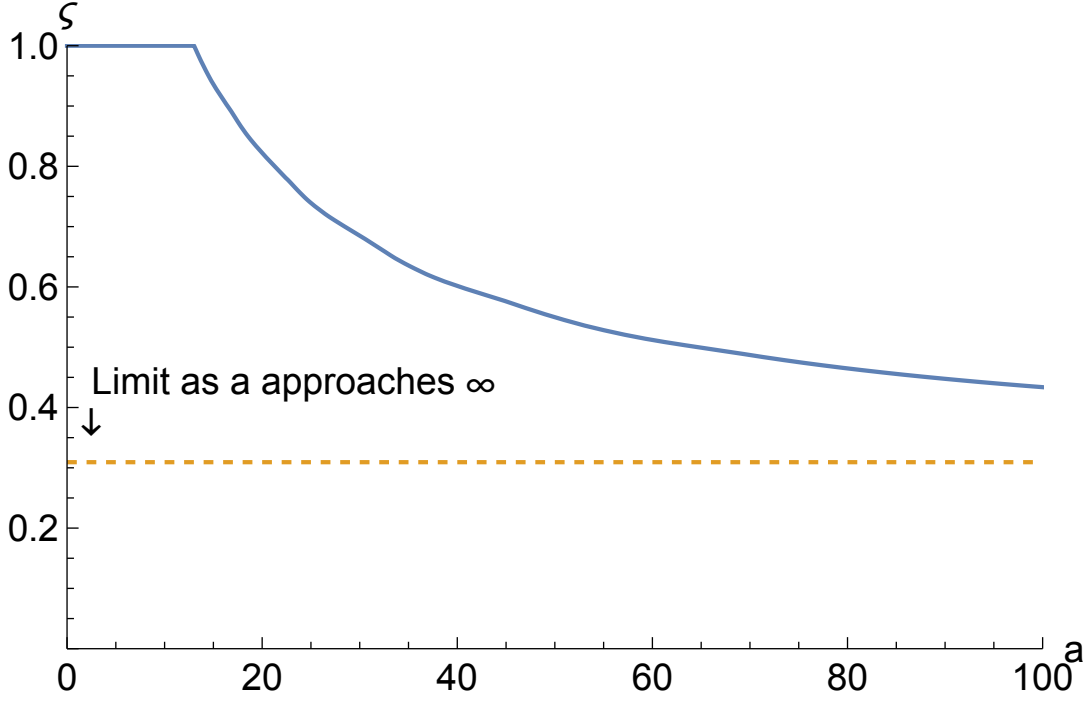
Figure 16 Portfolio Share in Risky Assets in First Period $\varsigma(a)$

{fig:PlotRiskyShare}

life expectancy; this is done easily enough, by assuming that utility accrues only if you live, so effectively the rising mortality rate with age is treated as an extra reason for discounting the future. Similarly, we may want to capture the demographic evolution of the household (e.g., arrival and departure of kids). A common way to handle that, too, is by modifying the discount factor (arrival of a kid might increase the total utility of the household by, say, 0.2, so if the 'pure' rate of time preference were 1.0 the 'household-size-adjusted' discount factor might be 1.2. We therefore modify the model presented above to allow age-varying discount factors that capture both mortality and family-size changes (we just adopt the factors used by Cagetti (2003) directly), with the probability of remaining alive between t and $t + n$ captured by $\mathcal{L}$ and with $\hat{\beta}$ now reflecting all the age-varying discount factor adjustments (mortality, family-size, etc). Using ב (the Hebrew cognate of β) for the 'pure' time preference factor, the value function for the revised problem is

$$v_t(\mathbf{p}_t, \mathbf{m}_t) = \max_{\{c\}_t^T} \quad u(\mathbf{c}_t) + \mathbb{E}_{t \rightarrow} \left[\sum_{n=1}^{T-t} ב^n \mathcal{L}_t^{t+n} \hat{\beta}_t^{t+n} u(\mathbf{c}_{t+n}) \right] \tag{57}$$

{eq:lifecyclemax}

subject to the constraints

$$\mathbf{a}_t = \mathbf{m}_t - \mathbf{c}_t$$
$$\mathbf{p}_{t+1} = \mathcal{G}_{t+1} \mathbf{p}_t \psi_{t+1}$$
$$\mathbf{y}_{t+1} = \mathbf{p}_{t+1} \theta_{t+1}$$
$$\mathbf{m}_{t+1} = \mathsf{R} \mathbf{a}_t + \mathbf{y}_{t+1}$$

where

$$\mathcal{L}_t^{t+n} : \text{probability to } \mathcal{L}\text{ive until age } t+n \text{ given alive at age } t$$

$$\hat{\beta}_t^{t+n} : \text{age-varying discount factor between ages } t \text{ and } t+n$$

$$\psi_t : \text{mean-one shock to permanent income}$$

$$\beth : \text{time-invariant 'pure' discount factor}$$

and all the other variables are defined as in section 2.

Households start life at age $s = 25$ and live with probability 1 until retirement ($s = 65$). Thereafter the survival probability shrinks every year and agents are dead by $s = 91$ as assumed by Cagetti.

Transitory and permanent shocks are distributed as follows:

$$\Xi_s = \begin{cases} 0 & \text{with probability } \wp > 0 \\ \theta_s/\wp & \text{with probability } (1-\wp), \text{ where } \log\theta_s \sim \mathcal{N}(-\sigma_\theta^2/2, \sigma_\theta^2) \end{cases} \quad (58)$$

$$\log\psi_s \sim \mathcal{N}(-\sigma_\psi^2/2, \sigma_\psi^2)$$

where $\wp$ is the probability of unemployment (and unemployment shocks are turned off after retirement).

The parameter values for the shocks are taken from Carroll (1992), $\wp = 0.5/100$, $\sigma_\theta = 0.1$, and $\sigma_\psi = 0.1$.[16] The income growth profile $\mathcal{G}_t$ is from Carroll (1997) and the values of $\mathcal{L}_t$ and $\hat{\beta}_t$ are obtained from Cagetti (2003) (Figure 17).[17] The interest rate is assumed to equal 1.03. The model parameters are included in Table 1.

Table 1 Parameter Values

{table:StrEstParam

σ_θ	0.1	Carroll (1992)
σ_ψ	0.1	Carroll (1992)
$\wp$	0.005	Carroll (1992)
$\mathcal{G}_s$	figure 17	Carroll (1997)
$\hat{\beta}_s, \mathcal{L}_s$	figure 17	Cagetti (2003)
R	1.03	Cagetti (2003)

The structural estimation of the parameters $\beth$ and ρ is carried out using the procedure specified in the following section, which is then implemented in the `StructEstimation.py` file. This file consists of two main components. The first section defines the objects required to execute the structural estimation procedure, while the second section executes the procedure and various optional experiments with

---

[16]Note that $\sigma_\theta = 0.1$ is smaller than the estimate for college graduates estimated in Carroll and Samwick (1997) ($= 0.197 = \sqrt{0.039}$) which is used by Cagetti (2003). The reason for this choice is that Carroll and Samwick (1997) themselves argue that their estimate of σ_θ is almost certainly increased by measurement error.

[17]The income growth profile is the one used by Caroll for operatives. Cagetti computes the time-varying discount factor by educational groups using the methodology proposed by Attanasio et al. (1999) and the survival probabilities from the 1995 Life Tables (National Center for Health Statistics 1998).
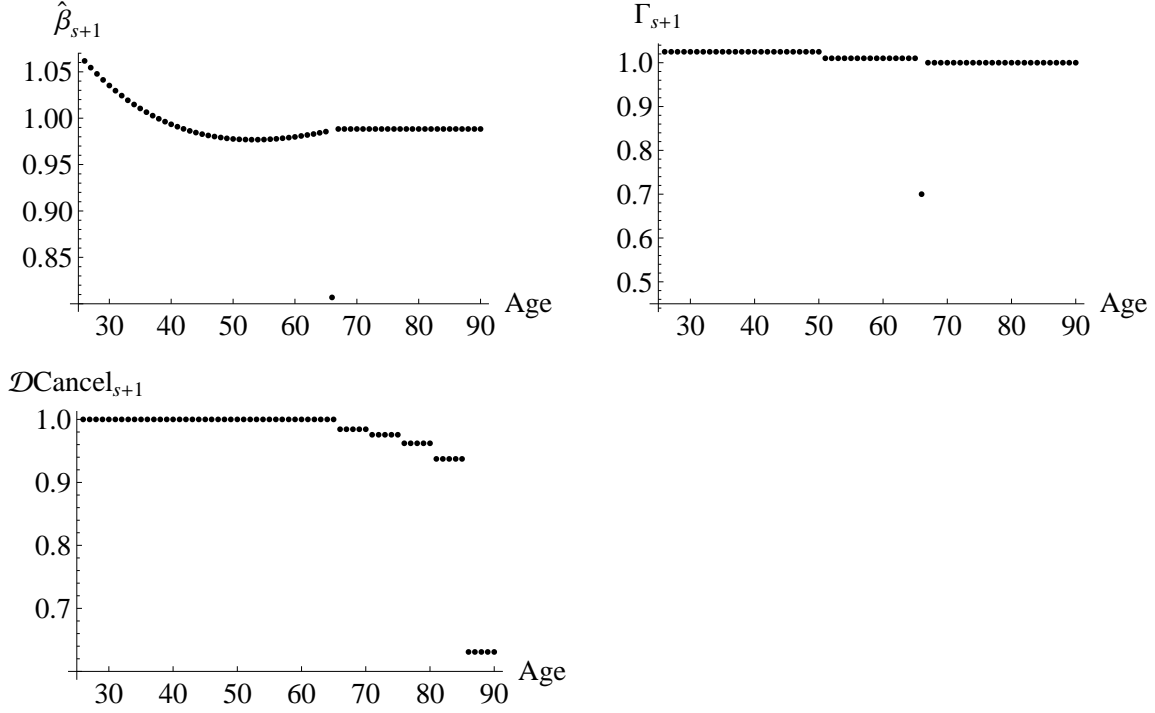
Figure 17 Time Varying Parameters

their corresponding commands. The next section elaborates on the procedure and its accompanying code implementation in greater detail.

9.2 Estimation

When economists say that they are performing "structural estimation" of a model like this, they mean that they have devised a formal procedure for searching for values for the parameters $\beth$ and ρ at which some measure of the model's outcome (like "median wealth by age") is as close as possible to an empirical measure of the same thing. Here, we choose to match the median of the wealth to permanent income ratio across 7 age groups, from age $26-30$ up to $56-60$.[18] The choice of matching the medians rather the means is motivated by the fact that the wealth distribution is much more concentrated at the top than the model is capable of explaining using a single set of parameter values. This means that in practice one must pick some portion of the population who one wants to match well; since the model has little hope of capturing the behavior of Bill Gates, but might conceivably match the behavior of Homer Simpson, we choose to match medians rather than means.

As explained in section 3, it is convenient to work with the normalized version of the

---

[18]Cagetti (2003) matches wealth levels rather than wealth to income ratios. We believe it is more appropriate to match ratios both because the ratios are the state variable in the theory and because empirical moments for ratios of wealth to income are not influenced by the method used to remove the effects of inflation and productivity growth.

model which can be written in Bellman form as:

$$v_t(m_t) = \max_{c_t} \quad u(c_t) + \beth\mathcal{L}_{t+1}\hat{\beta}_{t+1}\,\mathbb{E}_t[(\psi_{t+1}\mathcal{G}_{t+1})^{1-\rho}v_{t+1}(m_{t+1})]$$

$$\text{s.t.}$$

$$a_t = m_t - c_t$$

$$m_{t+1} = a_t \underbrace{\left(\frac{\mathsf{R}}{\psi_{t+1}\mathcal{G}_{t+1}}\right)}_{\equiv\mathcal{R}_{t+1}} + \theta_{t+1}$$

with the first order condition:

$$u^c(c_t) = \beth\mathcal{L}_{t+1}\hat{\beta}_{t+1}\mathsf{R}\,\mathbb{E}_t\left[u^c\left(\psi_{t+1}\mathcal{G}_{t+1}c_{t+1}\left(a_t\mathcal{R}_{t+1}+\theta_{t+1}\right)\right)\right]. \tag{59}$$

{eq:FOCLifeCycle}

The first substantive step in this estimation procedure is to solve for the consumption functions at each age. We need to discretize the shock distribution and solve for the policy functions by backward induction using equation (59) following the procedure in sections 6 and 'Recursion.' The latter routine is slightly complicated by the fact that we are considering a life-cycle model and therefore the growth rate of permanent income, the probability of death, the time-varying discount factor and the distribution of shocks will be different across the years. We thus must ensure that at each backward iteration the right parameter values are used.

Correspondingly, the first part of the `StructEstimation.py` file begins by defining the agent type by inheriting from the baseline agent type `IndShockConsumerType`, with the modification to include time-varying discount factors. Next, an instance of this "life-cycle" consumer is created for the estimation procedure. The number of periods for the life cycle of a given agent is set and, following Cagetti, (2003), we initialize the wealth to income ratio of agents at age 25 by randomly assigning the equal probability values to 0.17, 0.50 and 0.83. In particular, we consider a population of agents at age 25 and follow their consumption and wealth accumulation dynamics as they reach the age of 60, using the appropriate age-specific consumption functions and the age-varying parameters. The simulated medians are obtained by taking the medians of the wealth to income ratio of the 7 age groups.

To complete the creation of the consumer type needed for the simulation, a history of shocks is drawn for each agent across all periods by invoking the `make_shock_history` function. This involves discretizing the shock distribution for as many points as the number of agents we want to simulate and then randomly permuting this shock vector as many times as we need to simulate the model for. In this way, we obtain a time varying shock for each agent. This is much more time efficient than drawing at each time from the shock distribution a shock for each agent, and also ensures a stable distribution of shocks across the simulation periods even for a small number of agents. (Similarly, in order to speed up the process, at each backward iteration we compute the consumption function and other variables as a vector at once.)

With the age-varying consumption functions derived from the life-cycle agent, we can proceed to generate simulated data and compute the corresponding medians. Estimating the model involves comparing these simulated medians with empirical medians,

measuring the model's success by calculating the difference between the two. However, before performing the necessary steps of solving and simulating the model to generate simulated moments, it's important to note a difficulty in producing the target moments using the available data.

Specifically, defining ξ as the set of parameters to be estimated (in the current case $\xi = \{\rho, \beth\}$), we could search for the parameter values which solve

$$\min_{\xi} \sum_{\tau=1}^{7} |\varsigma^{\tau} - \mathbf{s}^{\tau}(\xi)| \tag{60}$$

{eq:naivePowell}

where ς^{τ} and $\mathbf{s}^{\tau}$ are respectively the empirical and simulated medians of the wealth to permanent income ratio for age group τ. A drawback of proceeding in this way is that it treats the empirically estimated medians as though they reflected perfect measurements of the truth. Imagine, however, that one of the age groups happened to have (in the consumer survey) four times as many data observations as another age group; then we would expect the median to be more precisely estimated for the age group with more observations; yet (60) assigns equal importance to a deviation between the model and the data for all age groups.

We can get around this problem (and a variety of others) by instead minimizing a slightly more complex object:

$$\min_{\xi} \sum_{i}^{N} \omega_i |\varsigma_i^{\tau} - \mathbf{s}^{\tau}(\xi)| \tag{61}$$

{eq:StructEstim}

where ω_i is the weight of household i in the entire population,[19] and ς_i^{τ} is the empirical wealth to permanent income ratio of household i whose head belongs to age group τ. ω_i is needed because unequal weight is assigned to each observation in the Survey of Consumer Finances (SCF). The absolute value is used since the formula is based on the fact that the median is the value that minimizes the sum of the absolute deviations from itself.

With this in mind, we turn our attention to the computation of the weighted median wealth target moments for each age cohort using this data from the 2004 Survery of Consumer Finances on household wealth. The objects necessary to accomplish this task are `weighted_median` and `get_targeted_moments`. The actual data are taken from several waves of the SCF and the medians and means for each age category are plotted in figure 18. More details on the SCF data are included in appendix A.

We now turn our attention to the the two key functions in this section of the code file. The first, `simulate_moments`, executes the solving (`solve`) and simulation (`simulation`) steps for the defined life-cycle agent. Subsequently, the function uses the agents' tracked levels of wealth based on their optimal consumption behavior to compute and store the simulated median wealth to income ratio for each age cohort. The second function,

---

[19]The Survey of Consumer Finances includes many more high-wealth households than exist in the population as a whole; therefore if one wants to produce population-representative statistics, one must be careful to weight each observation by the factor that reflects its "true" weight in the population.
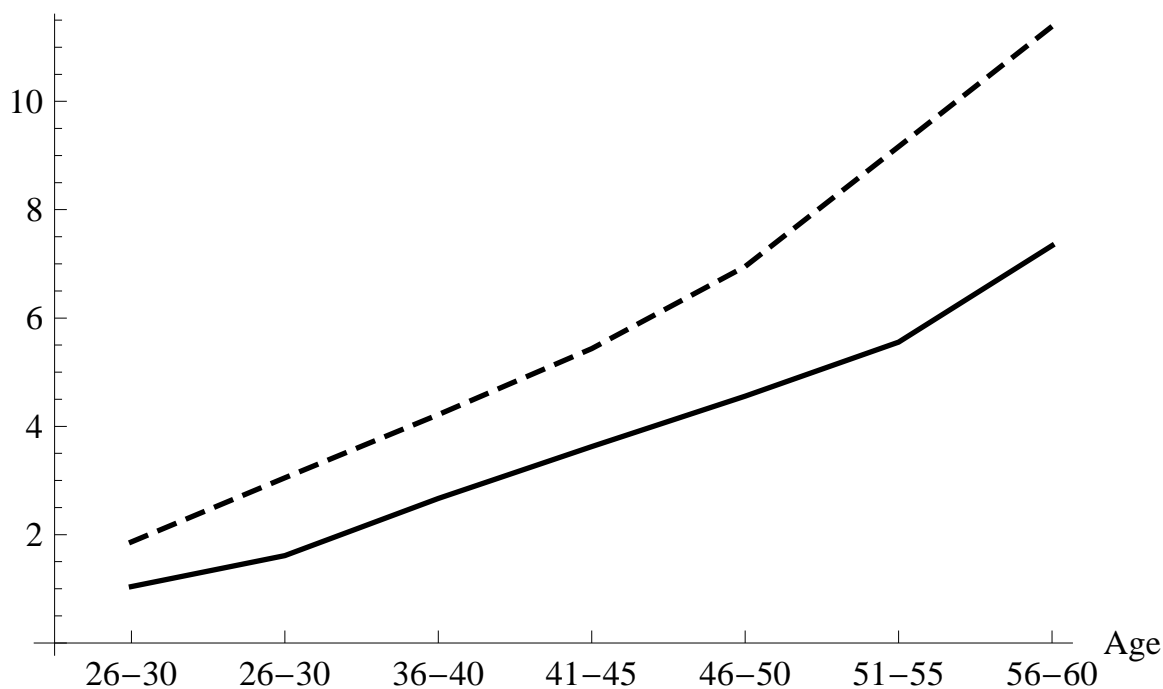
Figure 18 Wealth to Permanent Income Ratios from SCF (means (dashed) and medians (solid))

{fig:MeanMedianSC

smmObjectiveFxn, calls the simulate_moments function to create the objective function described in (61), which is necessary to perform the SMM estimation.

Thus, for a given pair of the parameters to be estimated, the single call to the function smmObjectiveFxn executes the following:

1. solves for the consumption functions for the life-cycle agent

2. simulates the data and computes the simulated medians

3. returns the value of equation (61)

We delegate the task of finding the coefficients that minimize the smmObjectiveFxn function to the minimize_nelder_mead function, which is defined elsewhere and called in the second part of this file. This task can be quite slow and rather problematic if the smmObjectiveFxn function has very flat regions or sharp features. It is thus wise to verify the accuracy of the solution, for example by experimenting with a variety of alternative starting values for the parameter search.

The final object defined in this first part of the StructEstimation.py file is calculateStandardErrorsByBootstrap. As the name suggsts, the purpose of this function is to compute the standard errors by bootstrap.[20] This involves:

1. drawing new shocks for the simulation

---

[20]For a treatment of the advantages of the bootstrap see Horowitz (2001)

2. drawing a random sample (with replacement) of actual data from the SCF

3. obtaining new estimates for ρ and $\beth$

We repeat the above procedure several times (`Bootstrap`) and take the standard deviation for each of the estimated parameters across the various bootstrap iterations.

9.2.1 An Aside to Computing Sensitivity Measures

{subsubsec:sensmea}

A common drawback in commonly used structural estimation procedures is a lack of transparency in its estimates. As Andrews, Gentzkow, and Shapiro (2017) notes, a researcher employing such structural empirical methods may be interested in how alternative assumptions (such as misspecification or measurement bias in the data) would "change the moments of the data that the estimator uses as inputs, and how changes in these moments affect the estimates." The authors provide a measure of sensitivity for given estimator that makes it easy to map the effects of different assumptions on the moments into predictable bias in the estimates for non-linear models.

In the language of Andrews, Gentzkow, and Shapiro (2017), section 9 is aimed at providing an estimator $\xi = \{\rho, \beth\}$ that has some true value ξ_0 by assumption. Under the assumption a_0 of the researcher, the empirical targets computed from the SCF is measured accurately. These moments of the data are precisely what determine our estimate $\hat{\xi}$, which minimizes (61). Under alternative assumptions a, such that a given cohort is mismeasured in the survey, a different estimate is computed. Using the plug-in estimate provided by the authors, we can see quantitatively how our estimate changes under these alternative assumptions a which correspond to mismeasurement in the median wealth to income ratio for a given age cohort.

9.3 Results

The second part of the file `StructEstimation.py` defines a function `main` which produces our ρ and $\beth$ estimates with standard errors using 10,000 simulated agents by setting the positional arguments `estimate_model` and `compute_standard_errors` to true.[21] Results are reported in Table 2.[22]

Table 2 Estimation Results

{tab:EstResults}

ρ	$\beth$
3.69	0.88
(0.047)	(0.002)

---

[21] The procedure is: First we calculate the ρ and $\beth$ estimates as the minimizer of equation (61) using the actual SCF data. Then, we apply the `Bootstrap` function several times to obtain the standard error of our estimates.

[22] Differently from Cagetti (2003) who estimates a different set of parameters for college graduates, high school graduates and high school dropouts graduates, we perform the structural estimation on the full population.

The literature on consumption and saving behavior over the lifecycle in the presenece of labor income uncertainty[23] warns us to be careful in disentangling the effect of time preference and risk aversion when describing the optimal behavior of households in this setting. Since the precautionary saving motive dominates in the early stages of life, the coefficient of relative risk aversion (as well as expected labor income growth) has a larger effect on optimal consumption and saving behavior through their magnitude relative to the interest rate. Over time, life-cycle considerations (such as saving for retirement) become more important and the time preference factor plays a larger role in determining optimal behavior for this cohort.

Using the positional argument `compute_sensitivity`, Figure 19 provides a plot of the plug-in estimate of the sensitivity measure described in 9.2.1. As you can see from the figure the inverse relationship between ρ and $\beth$ over the life-cycle is retained by the sensitivity measure. Specifically, under the alternative assumption that *a particular cohort is mismeasured in the SCF dataset*, we see that the y-axis suggests that our estimate of ρ and $\beth$ change in a predictable way.

Suppose that there are not enough observations of the oldest cohort of households in the sample. Suppose further that the researcher predicts that adding more observations of these households to correct this mismeasurement would correspond to a higher median wealth to income ratio for this cohort. In this case, our estimate of the time preference factor should increase: the behavior of these older households is driven by their time preference, so a higher value of $\beth$ is required to match the affected wealth to income targets under this alternative assumption. Since risk aversion is less important in explaining the behavior of this cohort, a lower value of ρ is required to match the affected empirical moments.

To recap, the sensitivity measure not only matches our intuition about the inverse relationship between ρ and $\beth$ over the life-cycle, but provides a quantitative estimate of what would happen to our estimates of these parameters under the alternative assumption that the data is mismeasured in some way.

By setting the positional argument `make_contour_plot` to true, Figure 20 shows the contour plot of the `smmObjectiveFxn` function and the parameter estimates. The contour plot shows equally spaced isoquants of the `smmObjectiveFxn` function, i.e. the pairs of ρ and $\beth$ which lead to the same deviations between simulated and empirical medians (equivalent values of equation (61)). Interestingly, there is a large rather flat region; or, more formally speaking, there exists a broad set of parameter pairs which leads to similar simulated wealth to income ratios. Intuitively, the flatter and larger is this region, the harder it is for the structural estimation procedure to precisely identify the parameters.

10 Conclusion

Many choices can be made for solving microeconomic dynamic stochastic optimization problems. The set of techniques, and associated code, described in these notes represents

---

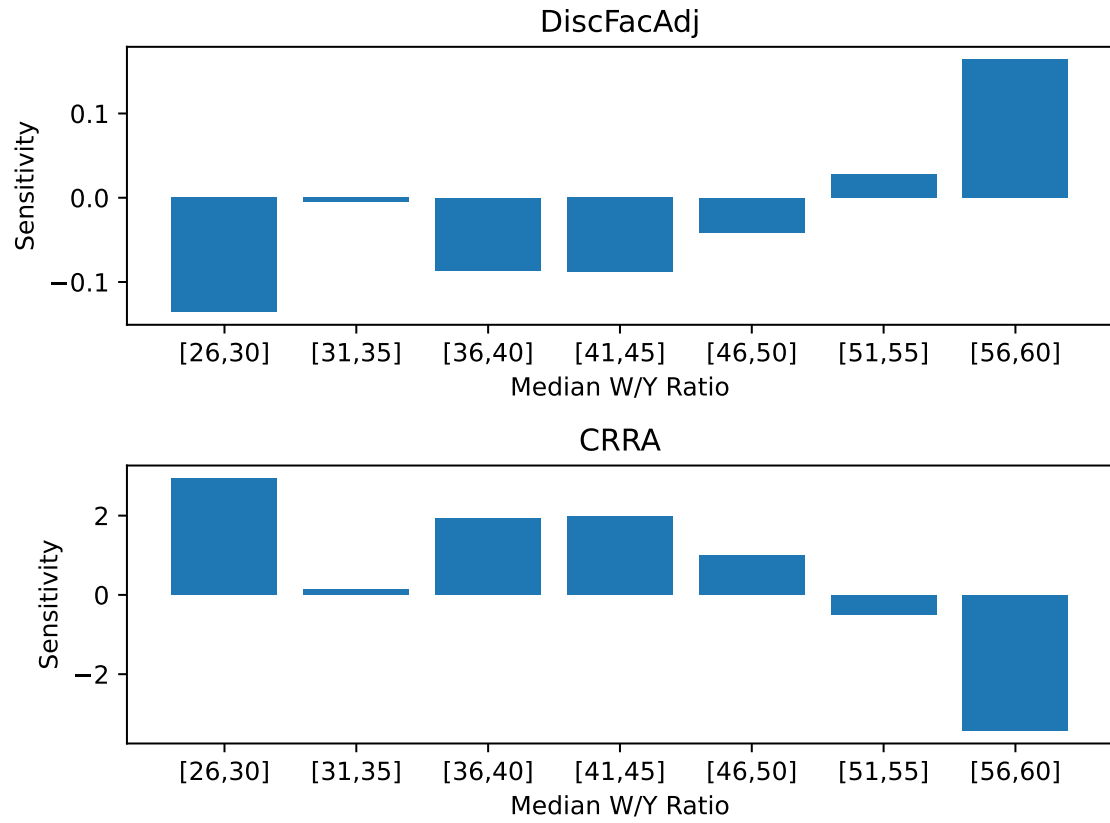[23]For example, see Gourinchas and Parker (2002) for an exposition of this.

Figure 19 Sensitivty of Estimates $\{\rho, \beth\}$ regarding Alternative Mismeasurement Assumptions.
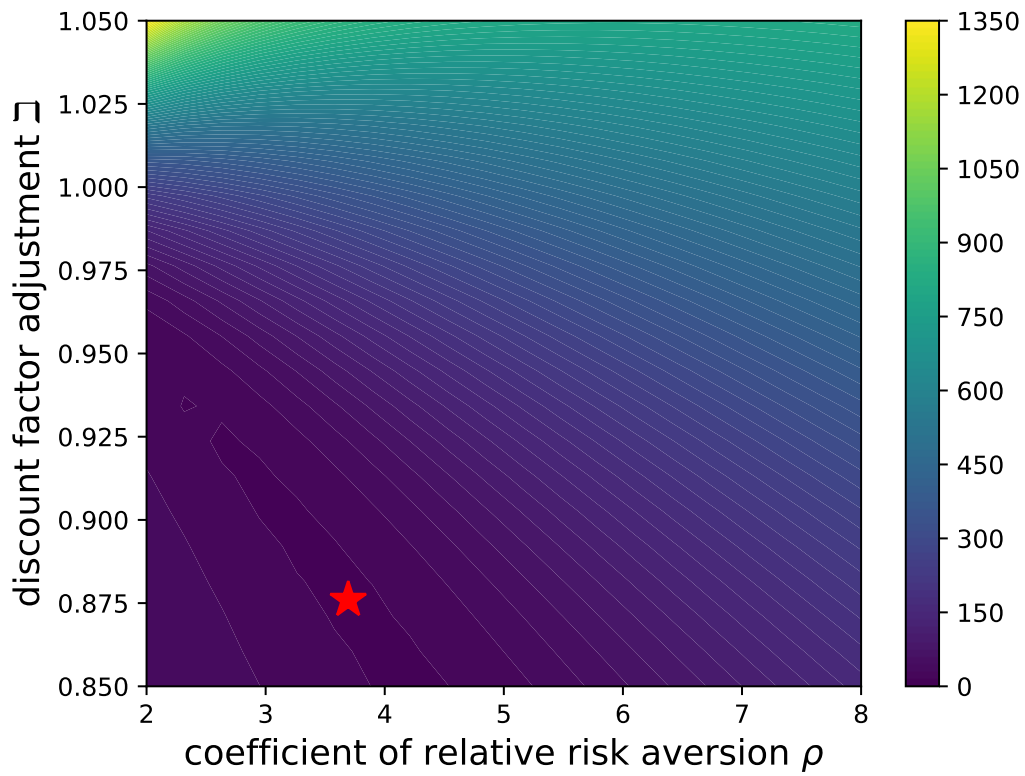
{fig:PlotSensitivity}

Figure 20 Contour Plot (larger values are shown lighter) with $\{\rho, \beth\}$ Estimates (red dot).

{fig:PlotContourMe

an approach that I have found to be powerful, flexible, and efficient, but other problems may require other techniques. For a much broader treatment of many of the issues considered here, see Judd (1998).

Appendices

A SCF Data

Data used in the estimation is constructed using the SCF 1992, 1995, 1998, 2001 and 2004 waves. The definition of wealth is net worth including housing wealth, but excluding pensions and social securities. The data set contains only households whose heads are aged 26-60 and excludes singles, following Cagetti (2003).[24] Furthermore, the data set contains only households whose heads are college graduates. The total sample size is 4,774.

In the waves between 1995 and 2004 of the SCF, levels of *normal* income are reported. The question in the questionnaire is "About what would your income have been if it had been a normal year?" We consider the level of normal income as corresponding to the model's theoretical object P, permanent noncapital income. Levels of normal income are not reported in the 1992 wave. Instead, in this wave there is a variable which reports whether the level of income is normal or not. Regarding the 1992 wave, only observations which report that the level of income is normal are used, and the levels of income of remaining observations in the 1992 wave are interpreted as the levels of permanent income.

Normal income levels in the SCF are before-tax figures. These before-tax permanent income figures must be rescaled so that the median of the rescaled permanent income of each age group matches the median of each age group's income which is assumed in the simulation. This rescaled permanent income is interpreted as after-tax permanent income. Rescaling is crucial since in the estimation empirical profiles are matched with simulated ones which are generated using after-tax permanent income (remember the income process assumed in the main text). Wealth / permanent income ratio is computed by dividing the level of wealth by the level of (after-tax) permanent income, and this ratio is used for the estimation.[25]

---

[24]Cagetti (2003) argues that younger households should be dropped since educational choice is not modeled. Also, he drops singles, since they include a large number of single mothers whose saving behavior is influenced by welfare.

[25]Please refer to the archive code for details of how these after-tax measures of P are constructed.

References

ANDREWS, ISAIAH, MATTHEW GENTZKOW, AND JESSE M SHAPIRO (2017): "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics*, 132(4), 1553–1592.

ATTANASIO, O.P., J. BANKS, C. MEGHIR, AND G. WEBER (1999): "Humps and Bumps in Lifetime Consumption," *Journal of Business and Economic Statistics*, 17(1), 22–35.

CAGETTI, MARCO (2003): "Wealth Accumulation Over the Life Cycle and Precautionary Savings," *Journal of Business and Economic Statistics*, 21(3), 339–353.

CARROLL, CHRISTOPHER D. (1992): "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," *Brookings Papers on Economic Activity*, 1992(2), 61–156, https://www.econ2.jhu.edu/people/ccarroll/BufferStockBPEA.pdf.

——— (1997): "Buffer Stock Saving and the Life Cycle/Permanent Income Hypothesis," *Quarterly Journal of Economics*, CXII(1), 1–56.

——— (2002): "Portfolios of the Rich," in *Household Portfolios: Theory and Evidence*. MIT Press, Cambridge, MA, https://www.econ2.jhu.edu/people/ccarroll/richportfolios.pdf.

——— (2006): "The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems," *Economics Letters*, 91(3), 312–320, https://www.econ2.jhu.edu/people/ccarroll/EndogenousGridpoints.pdf.

——— (2023a): "Solving Microeconomic Dynamic Stochastic Optimization Problems," *Econ-ARK REMARK*.

——— (2023b): "Theoretical Foundations of Buffer Stock Saving," *Revise and Resubmit, Quantitative Economics*.

CARROLL, CHRISTOPHER D., AND MILES S. KIMBALL (1996): "On the Concavity of the Consumption Function," *Econometrica*, 64(4), 981–992, https://www.econ2.jhu.edu/people/ccarroll/concavity.pdf.

CARROLL, CHRISTOPHER D., AND ANDREW A. SAMWICK (1997): "The Nature of Precautionary Wealth," *Journal of Monetary Economics*, 40(1), 41–71.

DEATON, ANGUS S. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59, 1221–1248, https://www.jstor.org/stable/2938366.

——— (1992): *Understanding Consumption*. Oxford University Press, New York.

DEN HAAN, WOUTER J, AND ALBERT MARCET (1990): "Solving the Stochastic Growth Model by Parameterizing Expectations," *Journal of Business and Economic Statistics*, 8(1), 31–34, Available at http://ideas.repec.org/a/bes/jnlbes/v8y1990i1p31-34.html.

GOURINCHAS, PIERRE-OLIVIER, AND JONATHAN PARKER (2002): "Consumption Over the Life Cycle," *Econometrica*, 70(1), 47–89.

HOROWITZ, JOEL L. (2001): "The Bootstrap," in *Handbook of Econometrics*, ed. by James J. Heckman, and Edward Leamer, vol. 5. Elsevier/North Holland.

JUDD, KENNETH L. (1998): *Numerical Methods in Economics*. The MIT Press, Cambridge, Massachusetts.

KOPECKY, KAREN A., AND RICHARD M.H. SUEN (2010): "Finite State Markov-Chain Approximations To Highly Persistent Processes," *Review of Economic Dynamics*, 13(3), 701–714, http://www.karenkopecky.net/RouwenhorstPaper.pdf.

MERTON, ROBERT C. (1969): "Lifetime Portfolio Selection under Uncertainty: The Continuous Time Case," *Review of Economics and Statistics*, 51, 247–257.

PALUMBO, MICHAEL G (1999): "Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle," *Review of Economic Studies*, 66(2), 395–421, Available at http://ideas.repec.org/a/bla/restud/v66y1999i2p395-421.html.

SAMUELSON, PAUL A. (1969): "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Review of Economics and Statistics*, 51, 239–46.

TANAKA, KEN'ICHIRO, AND ALEXIS AKIRA TODA (2013): "Discrete approximations of continuous distributions by maximum entropy," *Economics letters*, 118(3), 445–450.

ZELDES, STEPHEN P. (1989): "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence," *Quarterly Journal of Economics*, 104(2), 275–298.