Solution Methods for Microeconomic Dynamic Stochastic Optimization Problems

2024-03-27

Christopher D. Carroll[1]

Note: The Jupyter Notebook 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 the HARK Framework. The MSM results at the end has specifically have been superseded by the SolvingMicroDSOPs REMARK.

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 *Mathematica*, Matlab, and Python software that solves the problems described in the text.

Keywords Dynamic Stochastic Optimization, Method of Simulated Moments, Structural Estimation

JEL codes E21, F41

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, http://www.econ2.jhu.edu/people/ccarroll/, ccarroll@jhu.edu, Phone: (410) 516-7602

The notes were originally written for my Advanced Topics in Macroeconomic Theory class at Johns Hopkins University; instructors elsewhere are welcome to use them for teaching purposes. Relative to earlier drafts, this version incorporates several improvements related to new results in the paper "Theoretical Foundations of Buffer Stock Saving" (especially tools for approximating the consumption and value functions). Like the last major draft, it also builds on material in "The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems" published in *Economics Letters*, available at http://www.econ2.jhu.edu/people/ccarroll/EndogenousArchive.zip, and by including sample code for a method of simulated moments estimation of the life cycle model *a la* Gourinchas and Parker (2002) and Cagetti (2003). Background derivations, notation, and related subjects are treated in my class notes for first year macro, available at http://www.econ2.jhu.edu/people/ccarroll/public/lecturenotes/consumption. I am grateful to several generations of graduate students in helping me to refine these notes, to Marc Chan for help in updating the text and software to be consistent with Carroll (2006), to Kiichi Tokuoka for drafting the section on structural estimation, to Damiano Sandri for exceptionally insightful help in revising and updating the method of simulated moments estimation section, and to Weifeng Wu and Metin Uyanik for revising to be consistent with the 'method of moderation' and other improvements. All errors are my own. This document can be cited as Carroll (2023a) in the references.

Contents

1 Introduction 3

2 The Problem 3

3 Normalization 4

4 The Usual Theory, and A Bit More Notation 6
- 4.1 Notation for Steps . . . 6
- 4.2 The Usual Theory . . . 6
- 4.3 Implementation in Python . . . 8

5 Solving the Next-to-Last Period 8
- 5.1 Discretizing the Distribution . . . 8
- 5.2 The Approximate Consumption and Value Functions . . . 10
- 5.3 An Interpolated Consumption Function . . . 11
- 5.4 Interpolating Expectations . . . 11
- 5.5 Value Function versus First Order Condition . . . 14
- 5.6 Transformation . . . 16
- 5.7 The 'Natural' Borrowing Constraint and the a_{T-1} Lower Bound . . . 19
- 5.8 The Method of Endogenous Gridpoints . . . 21
- 5.9 Improving the a Grid . . . 22
- 5.10 The Method of Moderation . . . 23
- 5.11 Approximating the Slope Too . . . 28
- 5.12 Value . . . 30
- 5.13 Refinement: A Tighter Upper Bound . . . 33
- 5.14 Extension: A Stochastic Interest Factor . . . 34
- 5.15 Imposing 'Artificial' Borrowing Constraints . . . 36

6 Recursion 38
- 6.1 Theory . . . 38
- 6.2 Program Structure . . . 39
- 6.3 Results . . . 39

7 Multiple Control Variables 39
- 7.1 Theory . . . 40
- 7.2 Application . . . 41
- 7.3 Implementation . . . 43
- 7.4 Results . . . 43

8 The-Infinite-Horizon 45
- 8.1 Convergence . . . 45

9 Structural Estimation 46
9.1 Life Cycle Model 46
9.2 Estimation 48
9.2.1 An Aside to Computing Sensitivity Measures 52
9.3 Results 52

10 Conclusion 55

A Further Details on SCF Data 56

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 the various mathematical and computational techniques that, together, speed the solution by many orders of magnitude compared to "brute force" methods. 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' or MSM; sometimes called 'simulated method of moments' or SMM) 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, and more) into a single step in which the agent makes an optimal choice taking account of all the elements above. For the detailed analysis here, we will be careful to disarticulate everything that happens in the problem explicitly into separate steps 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 account **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 $\mathbf{b}$ and income $\mathbf{y}$ 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)):

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

which are available to be spent on consumption $\mathbf{c}_t$.

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

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

For now, we will assume that 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

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

subject to the Dynamic Budget Constraint (DBC) implicitly defined by equations (1)-(3) and 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). Here, 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 referred to simply as 'permanent income.')

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

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

[2]This equation assumes that there are no shocks to permanent income. 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).

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 equation (8) can be rewritten as

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

Now define a new optimization problem:

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

where the last equation is the normalized version of the transition equation for $\mathbf{m}_{t+1}$.[4] Then it is easy to see that for $t = T-1$,

$$\mathbf{v}_{T-1}(\mathbf{m}_{T-1}, \mathbf{p}_{T-1}) = \mathbf{p}_{T-1}^{1-\rho}\mathrm{v}_{T-1}(m_{T-1})$$

and so on back to all earlier periods. Hence, if we solve the problem (**??**) which (when optimal consumption is chosen) has only a single state variable m_t, we can obtain the levels of the value function, consumption, and all other variables from the corresponding permanent-income-normalized solution objects by multiplying each by $\mathbf{p}_t$, e.g. $\mathbf{c}_t(\mathbf{m}_t, \mathbf{p}_t) = \mathbf{p}_t\mathrm{c}_t(\mathbf{m}_t/\mathbf{p}_t)$ (or, for the value function, $\mathbf{v}_t(\mathbf{m}_t, \mathbf{p}_t) = \mathbf{p}_t^{1-\rho}\mathrm{v}_t(m_t)$). We have thus reduced the problem from two continuous state variables to one (and thereby enormously simplified its solution).

For some problems it will not be obvious that there is an appropriate 'normalizing' variable, but many problems can be normalized if sufficient thought is given. For example, Valencia (2006) shows how a bank's optimization problem can be normalized by the level of the bank's productivity.

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

[4] Derivation:

$$\begin{aligned}\mathbf{m}_{t+1}/\mathbf{p}_{t+1} &= (\mathbf{m}_t - \mathbf{c}_t)\mathsf{R}/\mathbf{p}_{t+1} + \mathbf{y}_{t+1}/\mathbf{p}_{t+1} \\ m_{t+1} &= \left(\frac{\mathbf{m}_t}{\mathbf{p}_t} - \frac{\mathbf{c}_t}{\mathbf{p}_t}\right)\mathsf{R}\frac{\mathbf{p}_t}{\mathbf{p}_{t+1}} + \frac{\mathbf{y}_{t+1}}{\mathbf{p}_{t+1}} \\ &= \underbrace{(m_t - c_t)}_{a_t}(\mathsf{R}/\mathcal{G}_{t+1}) + \theta_{t+1}.\end{aligned}$$

4 The Usual Theory, and A Bit More Notation

4.1 Notation for Steps

Since we have dissected the problem into individual steps, if we are to discuss each of the steps separately we need a notation to signal which step we are discussing.

Generically, we want to think of the Bellman problem itself as having three steps:

1. **Arrival**: Incoming state variables (e.g., k_t) 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, predictable income growth $\mathcal{G}$ have been realized (so that, e.g., m_t's value is known) and the agent solves the optimization problem

3. **Continuation**: After all decisions have been made, it is possible to calculate the consequences of the decision, taking as given the 'outgoing' state variables (e.g., a) – sometimes called 'post-state' variables.

When we want to refer to a specific step within period t we will do so by parenthesizing the (t) and preceding it by an indicator character:

Step	Indicator	Example	Explanation
Arrival	$-$	$\mathrm{v}_{-(t)}(k_t)$	value entering t
Decision	$\sim$	$\mathrm{v}_{\sim(t)}(m_t)$	value of t-decision
Continuation	$+$	$\mathrm{v}_{+(t)}(a_t)$	value of outgoing state

This allows us to capture the fact that the value of the consumer's problem can be computed at any of the three steps.

Note that there is no need to use these subscripts for the model's variables; while a variable like $a_t = m_t - c_t$ takes on its value in the transition from the Decision to the Continuation step, it have only one unique value over the course of the period and therefore a notation like $a_{-(t)}$ would be useless because the variable does not have a value until the continuation step is reached. Each variable in our problem has a unique value defined at some point during the period, so there is no ambiguity in referring to them with normal notation like a_t.

4.2 The Usual Theory

Using this new notation, the first order condition for (**??**) with respect to c_t is

$$\begin{aligned} \mathrm{u}^c(c_t) &= \mathbb{E}_t[\beta \mathcal{R}_{t+1} \mathcal{G}_{t+1}^{1-\rho} \mathrm{v}^m_{\sim(t+1)}(m_{t+1})] \\ &= \mathbb{E}_t[\beta \mathsf{R} \quad \mathcal{G}_{t+1}^{-\rho} \mathrm{v}^m_{\sim(t+1)}(m_{t+1})] \end{aligned} \tag{11}$$

and because the `Envelope` theorem tells us that

$$\mathrm{v}^m_{\sim(t)}(m_t) = \mathbb{E}_t[\beta \mathsf{R} \mathcal{G}_{t+1}^{-\rho} \mathrm{v}^m_{\sim(t+1)}(m_{t+1})] \tag{12}$$

we can substitute the LHS of (12) for the RHS of (11) to get

$$\mathrm{u}^c(c_t) = \mathrm{v}^m_{\sim(t)}(m_t) \tag{13}$$

and rolling this equation forward one period yields

$$\mathrm{u}^c(c_{t+1}) = \mathrm{v}^m_{\sim(t+1)}(a_t\mathcal{R}_{t+1} + \theta_{t+1}) \tag{14}$$

while substituting the LHS in equation (11) gives us the Euler equation for consumption:

$$\mathrm{u}^c(c_t) = \mathbb{E}_t[\beta\mathsf{R}\mathcal{G}^{-\rho}_{t+1}\mathrm{u}^c(c_{t+1})]. \tag{15}$$

Although our focus so far has been on the consumer's consumption problem once m is known, for constructing the computational solution it will be useful to distinguish a sequence of three moves (we use the word 'moves' to capture the fact that we are not really thinking of these as occurring at different moments in time – only that we are putting the things that happen within a given moment into an ordered sequence). The first step captures calculations that need to be performed before the shocks are realized, the middle step is after shocks have been realized but before the consumption decision has been made (this corresponds to the timing of v in the treatment above), and the final step captures the situation *after* the consumption decision has been made.

We need to define notation for these three moves within a period. We will use $-$ as a marker for the step before shocks have been realized, $\sim$ to indicate the move in which the choice is made, and $+$ as the indicator for the situation once the choice has been made.

$$\mathrm{v}_{-(t)}(k_t) = \mathbb{E}_{-(t)}[\mathrm{v}_{\sim(t)}(\overbrace{k_t\mathcal{R}_t + \theta_t}^{=m_t})] \tag{16}$$

$$\mathrm{v}_{\sim(t)}(m_t) = \mathrm{u}(\mathrm{c}_t(m_t)) + \mathrm{v}_{+(t)}(a_t) \tag{17}$$

and

$$\mathrm{v}_{+(t)}(a_t) = \beta\mathrm{v}_{-(t+1)}(\underbrace{k_{t+1}}_{=a_t \text{ from (10)}}) \tag{18}$$

where the last line illustrates the notation for addressing the beginning step of a future period.

Putting all this together, we can write the 'continuation value' function from the perspective of the end of the period t as

$$\mathrm{v}_{+(t)}(a_t) = \mathbb{E}_t[\beta\mathcal{G}^{1-\rho}_{t+1}\mathrm{v}_{\sim(t+1)}(\mathcal{R}_{t+1}a_t + \theta_{t+1})] \tag{19}$$

because it returns the expected $t+1$ value associated with ending period t with any given amount of assets. Differentiating with respect to a, we get

$$\mathrm{v}^a_{+(t)}(a_t) = \mathbb{E}_t[\beta\mathsf{R}\mathcal{G}^{-\rho}_{t+1}\mathrm{v}^m_{\sim(t+1)}(\mathcal{R}_{t+1}a_t + \theta_{t+1})] \tag{20}$$

or, substituting from equation (14),

$$\mathrm{v}^a_{+(t)}(a_t) = \mathbb{E}_t\left[\beta\mathsf{R}\mathcal{G}^{-\rho}_{t+1}\mathrm{u}^c\left(\mathrm{c}_{t+1}(\mathcal{R}_{t+1}a_t + \theta_{t+1})\right)\right]. \tag{21}$$

Finally, note for future use that the first order condition (11) can now be rewritten compactly as

$$\mathrm{u}^c(c_t) = \mathrm{v}^a_{+(t)}(m_t - c_t). \tag{22}$$

4.3 Implementation in Python

The code implementing the tasks outlined each of the sections to come is available in the `SolvingMicroDSOPs` jupyter notebook, written in Python. The notebook imports various modules, including the standard `numpy` and `scipy` modules used for numerical methods in Python, as well as some user-defined modules designed to provide numerical solutions to the consumer's problem from the previous section. Before delving into the computational exercise, it is essential to touch on the practicality of these custom modules.

5 Solving the Next-to-Last Period

Our peculiar notation described above is useful not only in designating the step of a value function; it is also useful in clarifying the perspective from which an expectation is taken. So, using this notation, the problem in the second-to-last period of life is:

$$\mathrm{v}_{T-1}(m_{T-1}) = \max_{c_{T-1}} \ \mathrm{u}(c_{T-1}) + \beta\, \mathbb{E}_{+(T-1)} \left[\mathcal{G}_T^{1-\rho} \mathrm{v}_{+(T)}((m_{T-1} - c_{T-1})\mathcal{R}_T + \theta_T) \right],$$

where, to review, $\mathbb{E}_{+(T-1)}$ indicates that the expectation is taken as of the end of period $T-1$.

Using (1) the fact that $\mathrm{v}_T = \mathrm{u}(c_T)$; (2) the definition of $\mathrm{u}(c_T)$; (3) the definition of the expectations operator; and (4) the fact that $\mathcal{G}_T$ is nonstochastic, this becomes:

$$\mathrm{v}_{T-1}(m_{T-1}) = \max_{c_{T-1}} \ \frac{c_{T-1}^{1-\rho}}{1-\rho} + \beta \mathcal{G}_T^{1-\rho} \int_0^\infty \frac{((m_{T-1} - c_{T-1})\mathcal{R}_T + \vartheta)^{1-\rho}}{1-\rho} d\mathcal{F}(\vartheta)$$

where $\mathcal{F}$ is the cumulative distribution function for θ_T.

In principle, the maximization implicitly defines a function $\mathrm{c}_{T-1}(m_{T-1})$ that yields optimal consumption in period $T-1$ for any given level of resources m_{T-1}. Unfortunately, there is no general analytical solution to this maximization problem; for any given m_{T-1} we must use numerical computational tools to find the c_{T-1} that maximizes the expression. This is excruciatingly slow because for every potential c_{T-1} 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!).

5.1 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 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:

$$\begin{aligned}\mathbb{E}[g(\theta)] &= \int_{\underline{\theta}}^{\overline{\theta}} (\vartheta) d\mathcal{F}(\vartheta) \\ &\approx \sum_{\theta=1}^{n} w_i g(\theta_i)\end{aligned}$$

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:

$$\theta_i \equiv \mathbb{E}[\theta | \sharp_{i-1}^{-1} \leq \theta < \sharp_i^{-1}] = \int_{\sharp_{i-1}^{-1}}^{\sharp_i^{-1}} \vartheta \; d\mathcal{F}\; (\vartheta). \tag{23}$$

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

Substituting into our definition of $\mathrm{v}_{+(T)}(a_t)$, for $t = T - 1$

$$\mathrm{v}_{+(T-1)}(a_{T-1}) = \beta \mathcal{G}_T^{1-\rho} \left(\frac{1}{n_\theta}\right) \sum_{i=1}^{n_\theta} \frac{(\mathcal{R}_T a_T + \theta_i)^{1-\rho}}{1-\rho} \tag{24}$$

so we can rewrite the maximization problem that defines the middle stage of the period $\sim$ as

$$\mathrm{v}_{\sim(T-1)}(m_{T-1}) = \max_{c_{T-1}} \left\{ \frac{c_{T-1}^{1-\rho}}{1-\rho} + \mathrm{v}_{-(T-1)}(m_{T-1} - c_{T-1}) \right\}. \tag{25}$$

In the `SolvingMicroDSOPs` notebook, the section "Discretization of the Income Shock Distribution" provides code which instantiates the `DiscreteApproximation` class defined in the `resources` module. This class creates a discretization of the continuous

[5]These points define intervals that constitute a partition of the domain of $\mathcal{F}$.

[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.

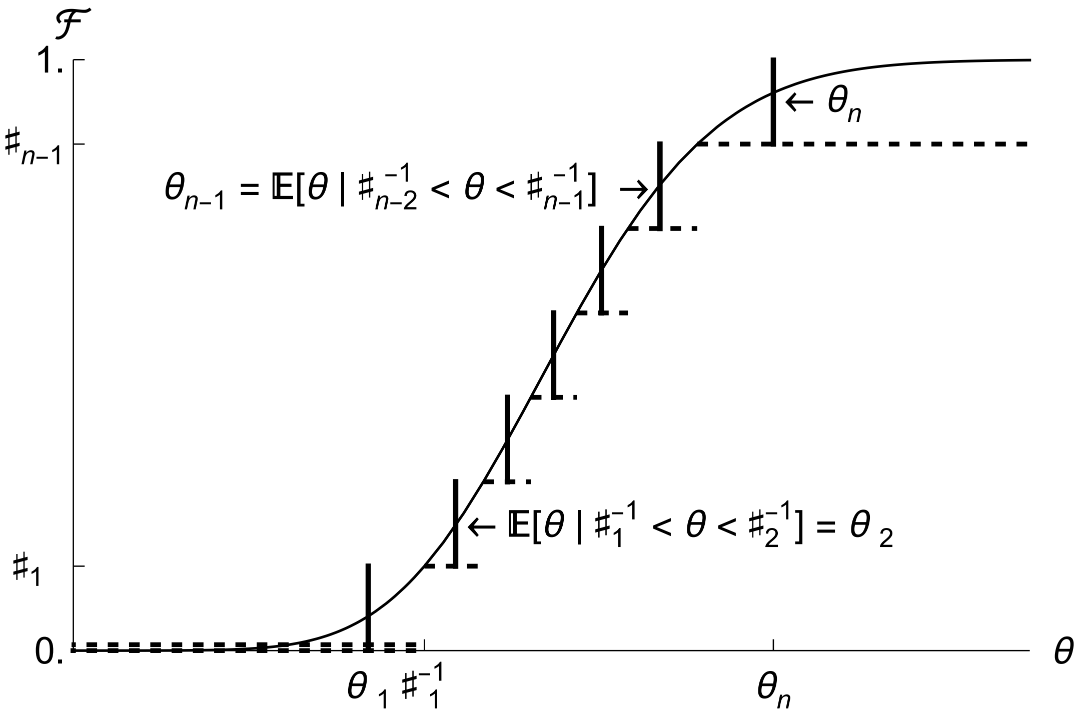

Figure 1 Equiprobable Discrete Approximation to Lognormal Distribution $\mathcal{F}$

log-normal distribution of transitory shocks to income by utilizing seven points, where the mean value is $-.5\sigma^2$, and the standard deviation is $\sigma = .5$.

A close look at the `DiscreteApproximation` class and its subclasses should convince you that the code is simply a computational implementation of the mathematical description of equiprobable discrete approximation in this section. Moreover, the Python code generates a graph of the discretized distribution depicted in 1.

5.2 The Approximate Consumption and Value Functions

Given a particular value of m_{T-1}, a numerical maximization routine can now find the c_{T-1} that maximizes (25) in a reasonable amount of time.

The `SolvingMicroDSOPs` notebook follows a series of steps to achieve this. Initially, parameter values for the coefficient of relative risk aversion (CRRA, ρ), the discount factor (β), the permanent income growth factor ($\mathcal{G}$), and the risk-free interest rate (R) are specified in "Define Parameters, Grids, and the Utility Function."

After defining the utility function, the natural borrowing constraint is defined to be $\underline{a}_{T-1} = -\underline{\theta}\mathcal{R}T^{-1}$, which will be discussed in greater depth in section 5.7. Following the reformulation of the maximization problem, an instance of the `gothic_class` is created using the specifications and the discretized distribution described in the prior lines of code; this is required to provide the numerical solution.

The heart of the program responsible for computing an estimated consumption function begins in "Solving the Model by Value Function Maximization," where a grid characterizing the possible values of market resources (m) is initialized, and for each of the m values, the consumption values c that solve the minimization problem equivalent to (25) are computed using scipy's minimize function.

5.3 An Interpolated Consumption Function

Given a set of points on a function (in this case, the consumption function $\mathrm{c}_{T-1}(m)$), we can create a piecewise linear 'interpolating function' (a 'spline') which when applied to an input m will yield the value of c that corresponds to a linear 'connect-the-dots' interpolation of the value of c from the points, creating a function that aims to provide an approximation of the functions whose points have been sampled.

This is accomplished in "An Interpolated Consumption Function," where the `InterpolatedUnivariateSpline` function is called from the `scipy` module to define an approximation to the consumption function $\grave{\mathrm{c}}_{T-1}(m_{T-1})$. That is, when called with an m_{T-1} that is equal to one of the points in `mVec_int` returns the associated value of $c_{T-1,i}$, and when called with a value of m_{T-1} that is not exactly equal to one of the `mVec_int`, returns the value of c that reflects a linear interpolation between the $c_{T-1,i}$ associated with the two `mVec_int` points nearest to m_{T-1}. Thus if the function is called with $m_{T-1} = 1.75$ and the nearest gridpoints are $m_{j,T-1} = 1$ and $m_{k,T-1} = 2$ then the value of c_{T-1} returned by the function would be $(0.25c_{j,T-1} + 0.75c_{k,T-1})$. We can define a numerical approximation to the value function $\grave{\mathrm{v}}_{T-1}(m_{T-1})$ in an exactly analogous way.

Figures 2 and 3 show plots of the $\grave{\mathrm{c}}_{T-1}$ and $\grave{\mathrm{v}}_{T-1}$ `InterpolatedUnivariateSpline` function calls. While the $\grave{\mathrm{c}}_{T-1}$ function looks very smooth, the fact that the $\grave{\mathrm{v}}_{T-1}$ 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]

5.4 Interpolating Expectations

The `InterpolatedUnivariateSpline` function works well in the sense that it generates a good approximation to the true optimal consumption function. However, there is a clear inefficiency in the program: Since it uses equation (25), for every value of m_{T-1} the program must calculate the utility consequences of various possible choices of c_{T-1} as it searches for the best choice.

For any given value of a_{T-1}, notice that there is a good chance that the program may end up calculating the corresponding $\mathrm{v}_{+(T)}$ many times while maximizing utility from different m_{T-1}'s. For example, it is possible that the program will calculate the value of ending the period with $a_{T-1} = 0$ dozens of times. It would be much more efficient if the program could make that calculation once and then merely recall the value when it is needed again.

[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.

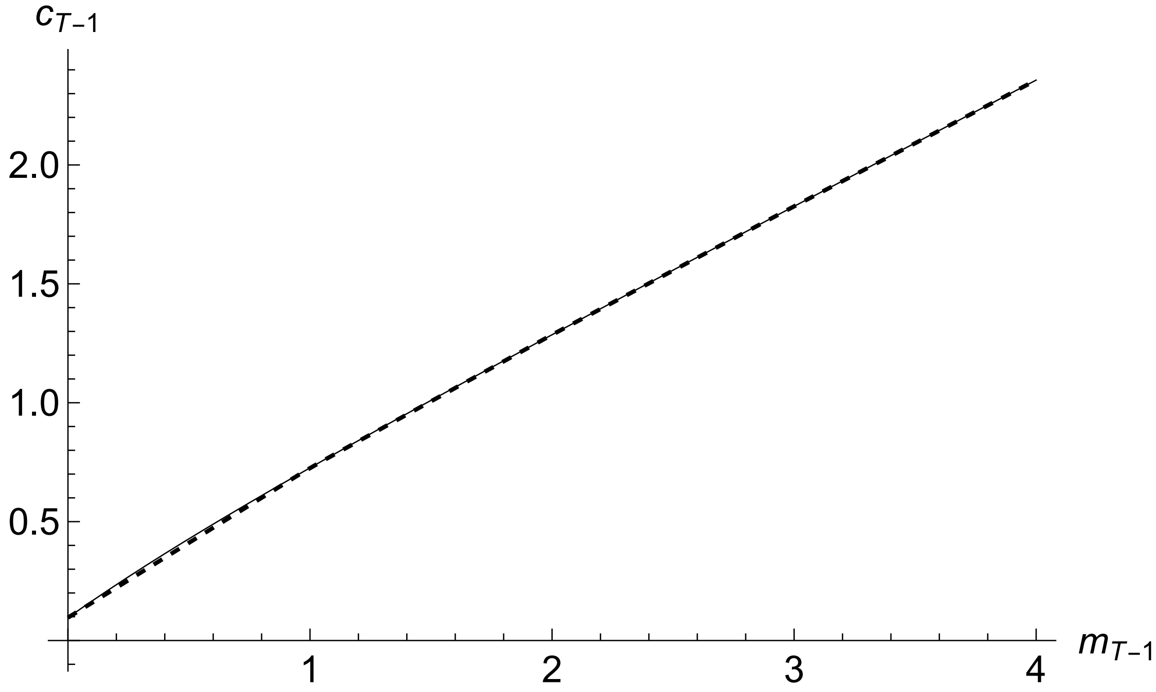

Figure 2 $\mathrm{c}_{T-1}(m_{T-1})$ (solid) versus $\grave{\mathrm{c}}_{T-1}(m_{T-1})$ (dashed)

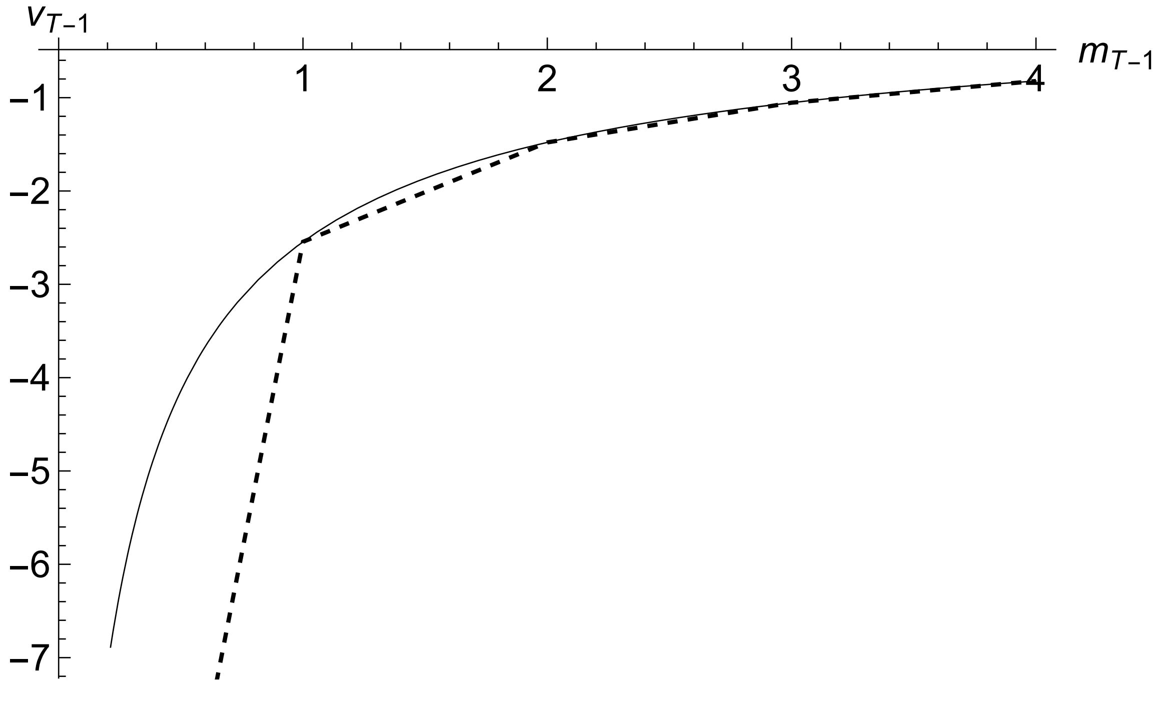

Figure 3 v_{T-1} (solid) versus $\grave{\mathrm{v}}_{T-1}(m_{T-1})$ (dashed)

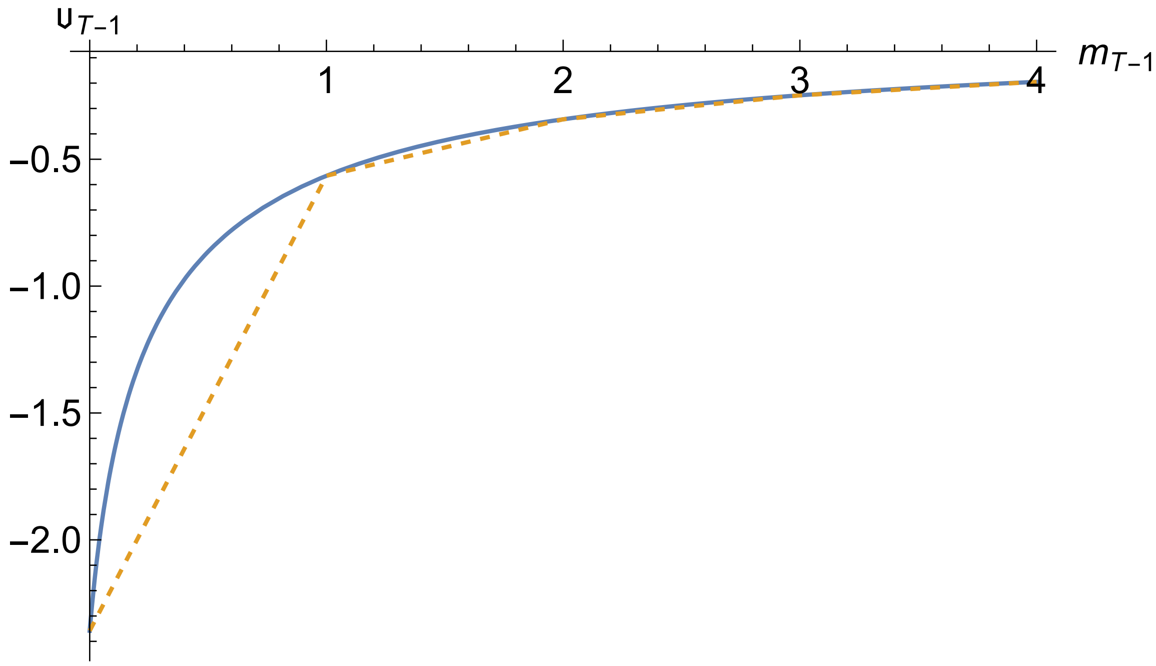

Figure 4 End-Of-Period Value $\mathfrak{v}_{+(T-1)}(a_{T-1})$ (solid) versus $\grave{\mathfrak{v}}_{+(T-1)}(a_{T-1})$ (dashed)

This can be achieved using the same interpolation technique used above to construct a direct numerical approximation to the value function: Define a grid of possible values for saving at time $T-1$, $\vec{a}_{T-1}$ (`aVec` in the code), designating the specific points $a_{T-1,i}$; for each of these values of $a_{T-1,i}$, calculate the vector $\vec{\mathfrak{v}}_{-(T-1)}$ as the collection of points $\mathfrak{v}_{+(T-1),i} = \mathfrak{v}_{+(T-1)}(a_{T-1,i})$ using equation (19); then construct an `InterpolatingUnivariateSpline` object $\grave{\mathfrak{v}}_{+(T-1)}(a_{T-1})$ by passing the lists `aVec` and `gothicvVec` as arguments. These lists contain the points of the $\vec{a}_{T-1}$ and $\vec{\mathfrak{v}}_{-(T-1)}$ vectors, respectively.

As seen in the section "Interpolating Expectations," we are now interpolating for the function that reveals the expected value of *ending* the period with a given amount of assets.[8] The problem is solved in the same block with the remaining following lines of code. Figure 4 compares the true value function to the approximation produced following the interpolation procedure; the functions are of course identical at the gridpoints chosen for a_{T-1} and they appear reasonably close except in the region below $m_{T-1} = 1$.

Nevertheless, the resulting consumption rule obtained when the approximating $\grave{\mathfrak{v}}_{+(T-1)}(a_{T-1})$ is used instead of $\mathfrak{v}_{+(T-1)}(a_{T-1})$ is surprisingly bad, as shown in figure 5. For example, when m_{T-1} goes from 2 to 3, $\grave{c}_{T-1}$ goes from about 1 to about 2, yet when m_{T-1} goes from 3 to 4, $\grave{c}_{T-1}$ goes from about 2 to about 2.05. The function fails even to be strictly concave, which is distressing because Carroll and Kimball (1996) prove

[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.

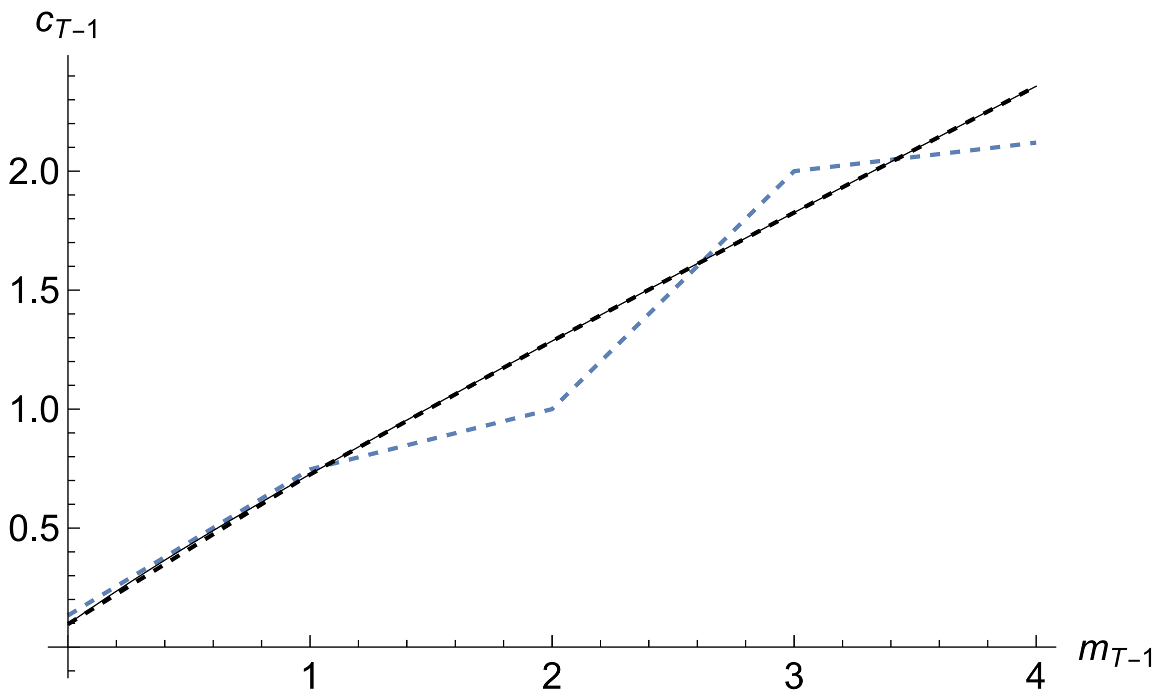

Figure 5 $\mathrm{c}_{T-1}(m_{T-1})$ (solid) versus $\grave{\mathrm{c}}_{T-1}(m_{T-1})$ (dashed)

that the correct consumption function is strictly concave in a wide class of problems that includes this problem.

5.5 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,

$$\mathbb{E}[-(c - \cancel{c})^2] < -(\mathbb{E}[c] - \cancel{c})^2 \tag{26}$$

where $\cancel{c}$ 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:

$$\mathbb{E}[-2(c - \cancel{c})] = -2(\mathbb{E}[c] - \cancel{c}). \tag{27}$$

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.

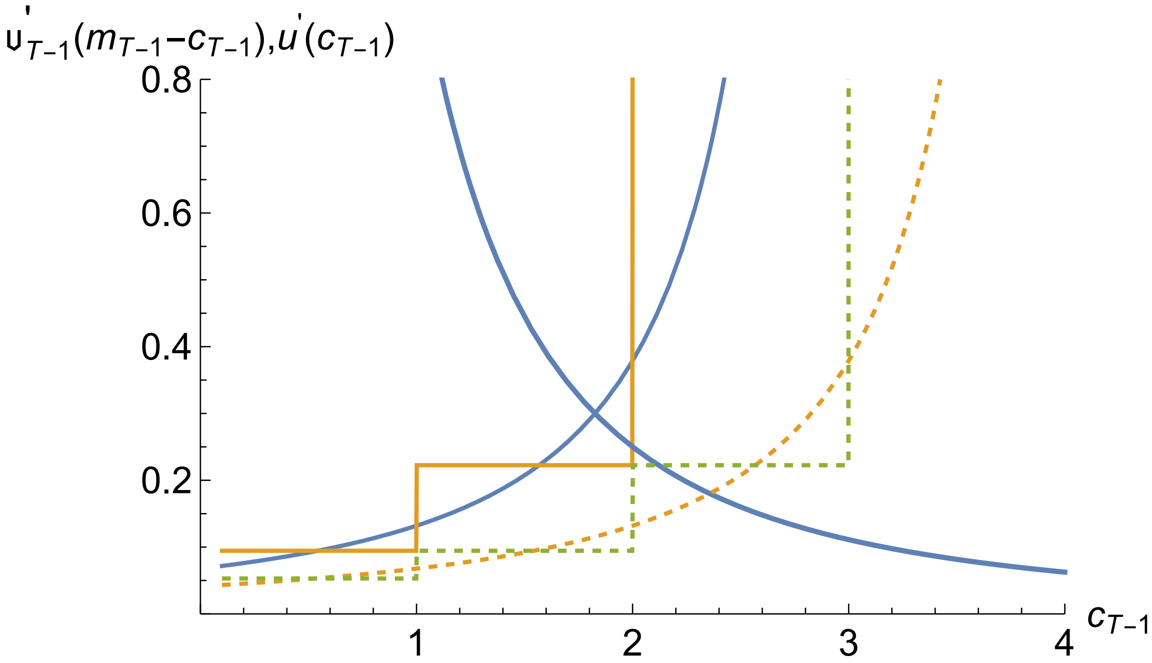

Figure 6 $\mathrm{u}^c(c)$ versus $\mathrm{v}^a_{+(T-1)}(3-c), \mathrm{v}^a_{+(T-1)}(4-c), \grave{\mathrm{v}}^a_{+(T-1)}(3-c), \grave{\mathrm{v}}^a_{+(T-1)}(4-c)$

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

$$\begin{aligned} \mathrm{u}^c(c_{T-1}) &= \beta\, \mathbb{E}_{T-1}[\mathcal{G}_T^{-\rho}\mathsf{R}\mathrm{u}^c(c_T)] \\ c_{T-1}^{-\rho} &= \mathsf{R}\beta\left(\frac{1}{n_\theta}\right)\sum_{i=1}^{n_\theta}\mathcal{G}_T^{-\rho}\left(\mathsf{R}(m_{T-1}-c_{T-1})+\theta_i\right)^{-\rho}. \end{aligned} \tag{28}$$

In the notebook, the "Value Function versus the First Order Condition" section completes the task of finding the values of consumption which satisfy the first order condition in (28) using the brentq function from the `scipy` package. Notice that the use of `u.prime` and `gothic.VP_Tminus1` is possible since they are already defined in the `resources` and `gothic_class` modules.

The downward-sloping curve in Figure 6 shows the value of $c_{T-1}^{-\rho}$ for our baseline parameter values for $0 \le c_{T-1} \le 4$ (the horizontal axis). The solid upward-sloping curve shows the value of the RHS of (28) as a function of c_{T-1} under the assumption that $m_{T-1} = 3$. Constructing this figure is rather time-consuming, because for every value of c_{T-1} plotted we must calculate the RHS of (28). The value of c_{T-1} for which the RHS and LHS of (28) are equal is the optimal level of consumption given that $m_{T-1} = 3$, so the intersection of the downward-sloping and the upward-sloping curves gives the optimal value of c_{T-1}. As we can see, the two curves intersect just below $c_{T-1} = 2$. Similarly, the upward-sloping dashed curve shows the expected value of the RHS of (28) under the assumption that $m_{T-1} = 4$, and the intersection of this curve with $\mathrm{u}^c(c_{T-1})$ yields the optimal level of consumption if $m_{T-1} = 4$. These two curves intersect slightly below $c_{T-1} = 2.5$. Thus, increasing m_{T-1} 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 $\{a_{T-1,i}, a_{i+1,T-1}\}$ 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 $\grave{\mathrm{v}}_{+(T-1)}(a_{T-1})$ follows a step function.

The solid-line step function in Figure 6 depicts the actual value of $\grave{\mathrm{v}}^a_{+(T-1)}(3 - c_{T-1})$. When we attempt to find optimal values of c_{T-1} given m_{T-1} using $\grave{\mathrm{v}}_{+(T-1)}(a_{T-1})$, the numerical optimization routine will return the c_{T-1} for which $\mathrm{u}^c(c_{T-1}) = \grave{\mathrm{v}}^a_{+(T-1)}(m_{T-1} - c_{T-1})$. Thus, for $m_{T-1} = 3$ the program will return the value of c_{T-1} for which the downward-sloping $\mathrm{u}^c(c_{T-1})$ curve intersects with the $\grave{\mathrm{v}}^a_{+(T-1)}(3 - c_{T-1})$; as the diagram shows, this value is exactly equal to 2. Similarly, if we ask the routine to find the optimal c_{T-1} for $m_{T-1} = 4$, it finds the point of intersection of $\mathrm{u}^c(c_{T-1})$ with $\grave{\mathrm{v}}^a_{+(T-1)}(4 - c_{T-1})$; 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 $c_{T-1} = 2$ for both $m_{T-1} = 3$ and $m_{T-1} = 4$.

We would obviously obtain much better estimates of the point of intersection between $\mathrm{u}^c(c_{T-1})$ and $\mathrm{v}^a_{+(T-1)}(m_{T-1} - c_{T-1})$ if our estimate of $\grave{\mathrm{v}}^a_{+(T-1)}$ 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* $\mathrm{v}'_{+(T)}$. That is, we calculate

$$\mathrm{v}^a_{+(T-1)}(a_{T-1}) = \beta \mathsf{R} \mathcal{G}_T^{-\rho} \left(\frac{1}{n_\theta}\right) \sum_{i=1}^{n_\theta} \left(\mathcal{R}_T a_{T-1} + \theta_i\right)^{-\rho} \tag{29}$$

at the points in `aVec` yielding $\{\{a_{T-1,1}, \mathrm{v}^a_{+(T-1),1}\}, \{a_{T-1,2}, \mathrm{v}^a_{+(T-1),2}\} \ldots\}$ and construct $\grave{\mathrm{v}}^a_{+(T-1)}(a_{T-1})$ as the linear interpolating function that fits this set of points.

This is done by making a call to the `InterpolatedUnivariateSpline` function, passing it `aVec` and `vpVec` as arguments. Note that in defining the list of values `vpVec`, we again make use of the predefined `gothic.VP_Tminus1` function. These steps are the embodiment of equation (29), and construct the interpolation of the expected marginal value of end-of-period assets as described above.

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

5.6 Transformation

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

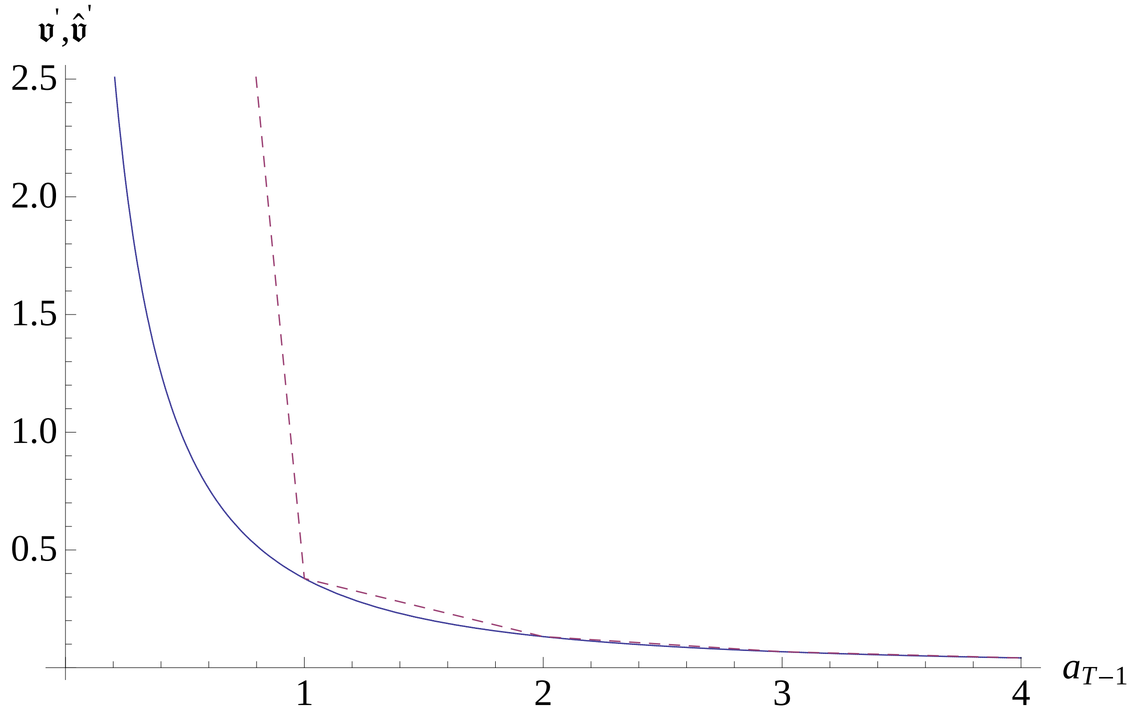

Figure 7 $\mathrm{v}^{a}_{+(T-1)}(a_{T-1})$ versus $\grave{\mathrm{v}}^{a}_{+(T-1)}(a_{T-1})$

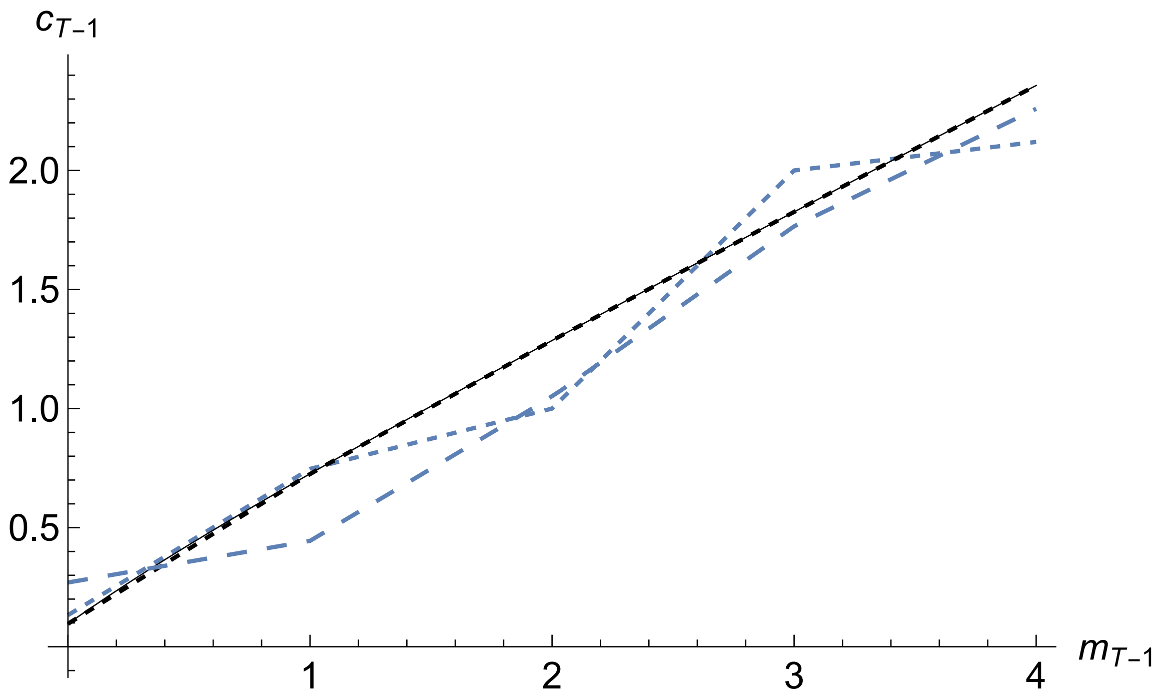

Figure 8 $\mathrm{c}_{T-1}(m_{T-1})$ (solid) Versus Two Methods for Constructing $\grave{\mathrm{c}}_{T-1}(m_{T-1})$

increasingly poor job of capturing the nonlinearity of $\mathrm{v}^{a}_{+(T-1)}(a_{T-1})$ 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 is then:

$$c_{T-1}^{-\rho} = c_T^{-\rho}. \tag{30}$$

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 are comprised of two types: current market resources m_{T-1} and 'human wealth' (the PDV of future income) of $\mathfrak{h}_{T-1} = 1$ (where we use the Gothic font to signify that this is the expectation, as of the END of the period, of the income that will be received in future periods; it does not include current income, which has already been incorporated into m_{T-1}).

The optimal solution is to spend half of total lifetime resources in period $T-1$ and the remainder in period T. Since total resources are known with certainty to be $m_{T-1} + \mathfrak{h}_{T-1} = m_{T-1} + 1$, and since $\mathrm{v}^{m}_{T-1}(m_{T-1}) = \mathrm{u}^{c}(c_{T-1})$ this implies that

$$\mathrm{v}^{m}_{T-1}(m_{T-1}) = \left(\frac{m_{T-1}+1}{2}\right)^{-\rho}. \tag{31}$$

Of course, this is a highly nonlinear function. However, if we raise both sides of (31) to the power $(-1/\rho)$ the result is a linear function:

$$[\mathrm{v}^{m}_{T-1}(m_{T-1})]^{-1/\rho} = \frac{m_{T-1}+1}{2}. \tag{32}$$

This is a specific example of a general phenomenon: A theoretical literature cited 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 this in intuitive terms, the problem we are facing is 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 $\mathrm{v}^{a}_{+(T-1)}(a_{T-1})$, we can define

$$\mathrm{c}_{+(T-1)}(a_{T-1}) \equiv \left(\mathrm{v}^{a}_{+(T-1)}(a_{T-1})\right)^{-1/\rho} \tag{33}$$

which would be linear in the perfect foresight case. Thus, our procedure is to calculate the values of $\mathrm{c}_{+(T-1),i}$ at each of the $a_{T-1,i}$ gridpoints, with the idea that we will construct $\grave{\mathrm{c}}_{+(T-1)}$ as the interpolating function connecting these points (the 'consumed function').

5.7 The 'Natural' Borrowing Constraint and the a_{T-1} Lower Bound

This is the appropriate moment to ask an awkward question we have so far neglected: How should a function like $\grave{\mathrm{c}}_{+(T-1)}$ be evaluated outside the range of points spanned by $\{a_{T-1,1}, ..., a_{T-1,n}\}$ for which we have calculated the corresponding $\mathrm{c}_{+(T-1),i}$ gridpoints used to produce our linearly interpolating approximation $\grave{\mathrm{c}}_{+(T-1)}$ (as described in section 5.3)?

The natural answer would seem to be linear extrapolation; for example, we could use

$$\grave{\mathrm{c}}_{+(T-1)}(a_{T-1}) = \grave{\mathrm{c}}_{+(T-1)}(a_{T-1,1}) + \grave{\mathrm{c}}^{a}_{+(T-1)}(a_{T-1,1})(a_{T-1} - a_{T-1,1}) \tag{34}$$

for values of $a_{T-1} < a_{T-1,1}$, where $\grave{\mathrm{c}}^{a}_{+(T-1)}(a_{T-1,1})$ is the derivative of the $\grave{\mathrm{c}}_{+(T-1)}$ function at the bottommost gridpoint (see below). Unfortunately, this approach will lead us into difficulties. To see why, consider what happens to the true (not approximated) $\mathrm{v}^{a}_{+(T-1)}(a_{T-1})$ as a_{T-1} approaches the value $\underline{a}_{T-1} = -\underline{\theta}\mathcal{R}_T^{-1}$. From (29) we have

$$\lim_{a_{T-1} \downarrow \underline{a}_{T-1}} \mathrm{v}^{a}_{+(T-1)}(a_{T-1}) = \lim_{a_{T-1} \downarrow \underline{a}_{T-1}} \beta \mathsf{R} \mathcal{G}_T^{-\rho} \left(\frac{1}{n_\theta}\right) \sum_{i=1}^{n_\theta} (a_{T-1}\mathcal{R}_T + \theta_i)^{-\rho}. \tag{35}$$

But since $\underline{\theta} = \theta_1$, exactly at $a_{T-1} = \underline{a}_{T-1}$ 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 period 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 (or a negative amount), which yields $-\infty$ utility and ∞ marginal utility (or undefined utility and marginal utility).

These reflections lead us to the conclusion that the consumer faces a 'self-imposed' (or 'natural') borrowing constraint (which results 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 $\underline{a}_{T-1}$).[9] The constraint is 'self-imposed' in the sense that if the utility function were different (say, Constant Absolute Risk Aversion), the consumer would 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 $\mathrm{v}^{a}_{+(T-1)}$ function is approximated by a piecewise linear function like $\grave{\mathrm{v}}^{m}_{+(T-1)}$, because a linear approximation can never reach the correct gridpoint for $\mathrm{v}^{a}_{+(T-1)}(\underline{a}_{T-1}) = \infty$. To see what will happen instead, note first that if we are approximating $\mathrm{v}^{a}_{+(T-1)}$ the smallest value in `aVec` must be greater than $\underline{a}_{T-1}$ (because the expectation for any gridpoint $\leq \underline{a}_{T-1}$ is undefined). Then when the approximating $\mathrm{v}^{a}_{+(T-1)}$ function is evaluated at some value less than

[9] Another term for a constraint of this kind is the 'natural borrowing constraint.'

[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.

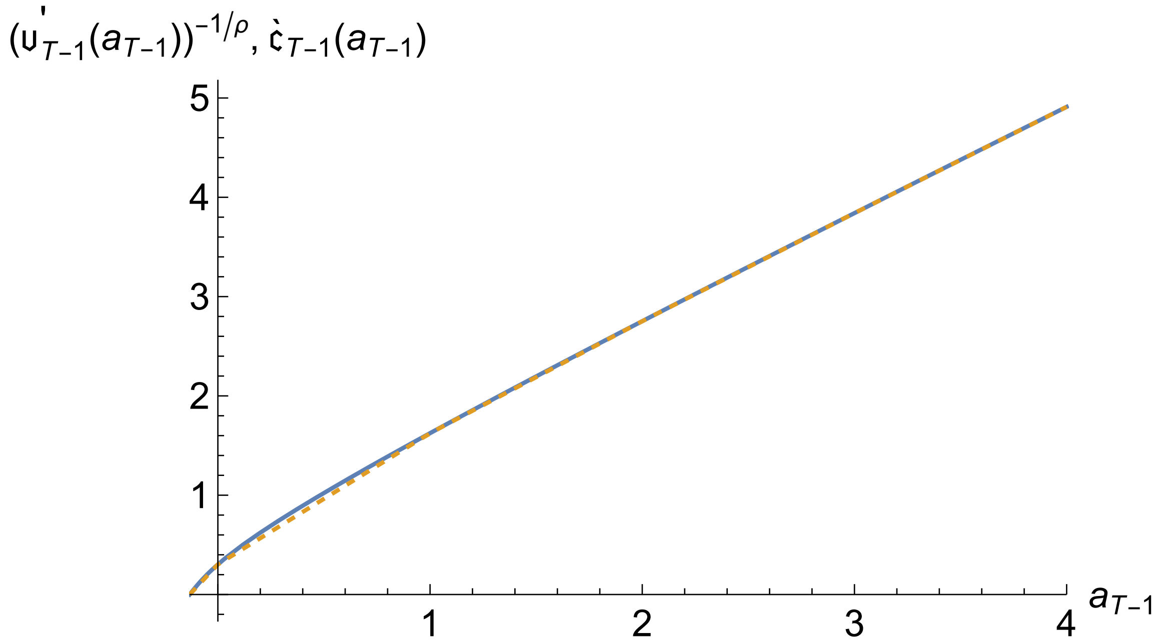

Figure 9 True $\mathrm{c}_{+(T-1)}(a_{T-1})$ vs its approximation $\grave{\mathrm{c}}_{+(T-1)}(a_{T-1})$

the first element in `aVec`, the approximating function will linearly extrapolate the slope that characterized the lowest segment of the piecewise linear approximation (between `aVec[0]` and `aVec[1]`), a procedure that will return a positive finite number, even if the requested a_{T-1} point is below $\underline{a}_{T-1}$. This means that the precautionary saving motive is understated, and by an arbitrarily large amount as the level of assets approaches its true theoretical minimum $\underline{a}_{T-1}$.

The foregoing logic demonstrates that the marginal value of saving approaches infinity as $a_{T-1} \downarrow \underline{a}_{T-1} = -\underline{\theta}\mathcal{R}_T^{-1}$. But this implies that $\lim_{a_{T-1}\downarrow\underline{a}_{T-1}} \mathrm{c}_{+(T-1)}(a_{T-1}) = (\mathrm{v}^a_{+(T-1)}(a_{T-1}))^{-1/\rho} = 0$; that is, as a approaches its minimum possible value, the corresponding amount of c must approach *its* minimum possible value: zero.

The upshot of this discussion is a realization that all we need to do is to augment each of the $\vec{a}_{T-1}$ and $\vec{c}_{T-1}$ vectors with an extra point so that the first element in the list used to produce our interpolation function is $\{a_{T-1,0}, c_{T-1,0}\} = \{\underline{a}_{T-1}, 0.\}$. This is done in section "The Self-Imposed 'Natural' Borrowing Constraint and the a_{T-1} Lower Bound" of the notebook, which can be seen in the defined lists `aVecBot` and `cVec3Bot`.

From there, we plot the lists that have been appended with the natural borrowing contraint and the associated minimal level of consumption. The subseqeunt figure 9 shows the result of this procedure. The solid line calculates the exact numerical value of $\mathrm{c}_{+(T-1)}(a_{T-1})$ while the dashed line is the linear interpolating approximation $\grave{\mathrm{c}}_{+(T-1)}(a_{T-1})$. This figure illustrates the value of the transformation: The true function is close to linear, and so the linear approximation is almost indistinguishable from the true function except at the very lowest values of a_{T-1}.

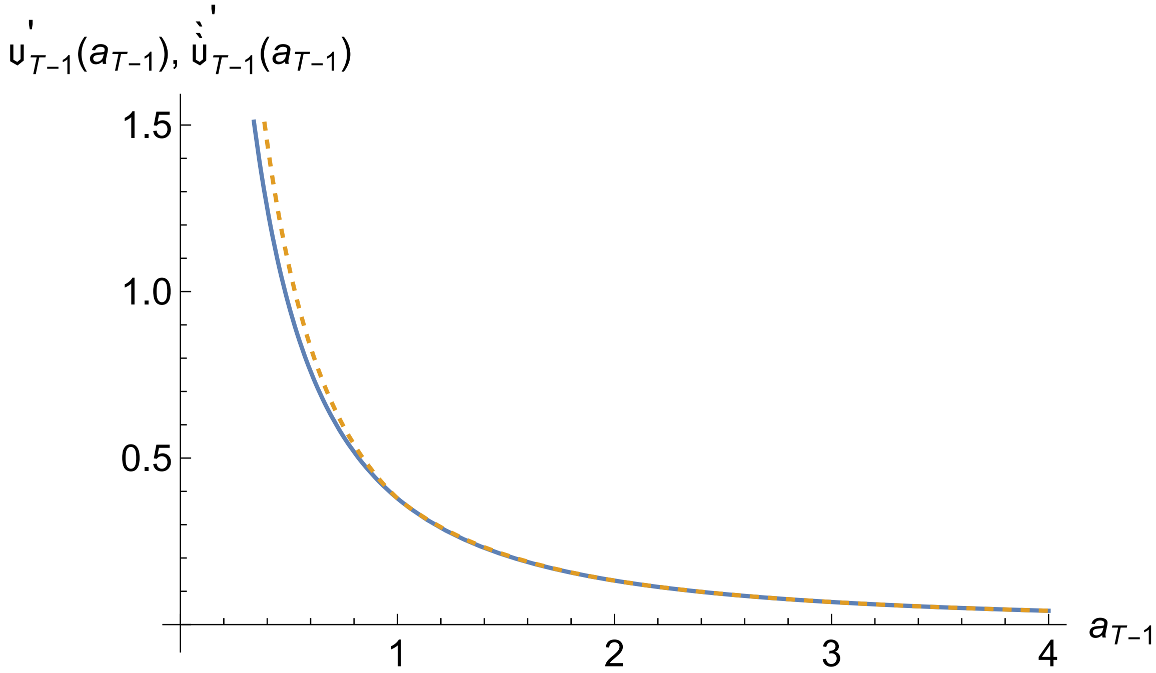

Figure 10 True $\mathrm{v}^a_{+(T-1)}(a_{T-1})$ vs. $\grave{\grave{\mathrm{v}}}^a_{+(T-1)}(a_{T-1})$ Constructed Using $\grave{\mathrm{c}}_{+(T-1)}(a_{T-1})$

Figure 10 similarly shows that when we generate $\grave{\grave{\mathrm{v}}}^a_{+(T-1)}(a_{T-1})$ as $[\grave{\mathrm{c}}_{+(T-1)}(a_{T-1})]^{-\rho}$ (dashed line) we obtain a *much* closer approximation to the true function $\mathrm{v}^a_{+(T-1)}(a_{T-1})$ (solid line) than we did in the previous program which did not do the transformation (Figure 7). (The vertical axis label uses $\mathfrak{v}'$ as an alternative notation for what in these notes we designate as $\mathrm{v}^a_{+(T)}$.

5.8 The Method of Endogenous Gridpoints

Our solution procedure for c_{T-1} still requires us, for each point in $\vec{m}_{T-1}$ (`mVec` in the code), to use a numerical rootfinding algorithm to search for the value of c_{T-1} that solves $\mathrm{u}^c(c_{T-1}) = \mathrm{v}^a_{+(T-1)}(m_{T-1} - c_{T-1})$. Unfortunately, rootfinding is a notoriously computation-intensive (that is, slow!) operation.

Our next trick lets us completely skip the rootfinding step. The method can be understood by noting that any arbitrary value of $a_{T-1,i}$ (greater than its lower bound value $\underline{a}_{T-1}$) will be associated with *some* marginal valuation as of the ++ stage of $T-1$ (that is, at the end of the period), and the further observation that it is trivial to find the value of c that yields the same marginal valuation, using the first order condition,

$$\mathrm{u}^c(c_{T-1,i}) = \mathrm{v}^a_{+(T-1)}(a_{T-1,i}) \tag{36}$$

by using the inverse of the marginal utility function,

$$\begin{aligned} c^{-\rho} &= \mu \\ c &= \mu^{-1/\rho} \end{aligned} \tag{37}$$

which yields the level of consumption that corresponds to marginal utility of μ. Using this to invert both sides of (36), we get

$$\begin{aligned} c_{T-1,i} &= \left(\mathrm{v}^{a}_{+(T-1)}(a_{T-1,i})\right)^{-1/\rho} \\ &\equiv \mathrm{c}_{+(T-1)}(a_{T-1,i}) \\ &\equiv \mathrm{c}_{+(T-1),i}. \end{aligned} \tag{38}$$

where the ++ emphasizes that these are points on the 'consumed' function (that is, the function that reveals how much an optimizing consumer must have consumed in order to end the period with $a_{T-1,i}$.

But with mutually consistent values of $c_{T-1,i}$ and $a_{T-1,i}$ (consistent, in the sense that they are the unique optimal values that correspond to the solution to the problem), we can obtain the $m_{T-1,i}$ that corresponds to both of them from

$$m_{T-1,i} = c_{T-1,i} + a_{T-1,i}. \tag{39}$$

These m_{T-1} gridpoints are "endogenous" in contrast to the usual solution method of specifying some *ex-ante* grid of values of m_{T-1} and then using a rootfinding routine to locate the corresponding optimal c_{T-1}. This routine is performed in "Endogenous Gridpoints." 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 $\vec{\mathrm{c}}_{+(T-1)}$ values in hand, we can generate a set of $m_{T-1,i}$ and $c_{T-1,i}$ pairs that can be interpolated between in order to yield $\grave{\mathrm{c}}(m_{T-1})$ at virtually zero computational cost![11] This is done in the final line of code in this block, and the following code block produces the graph of the interpolated consumption function using this procedure.

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 nonlinearity). Figure 11 plots the actual consumption function c_{T-1} and the approximated consumption function $\grave{\mathrm{c}}_{T-1}$ derived by the method of endogenous grid points. Compared to the approximate consumption functions illustrated in Figure 8, $\grave{\mathrm{c}}_{T-1}$ is quite close to the actual consumption function.

5.9 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 $\grave{\mathrm{c}}_{+(T-1)}$ function tends to be farthest from its true value $\mathrm{c}_{+(T-1)}$ at low values of a.

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

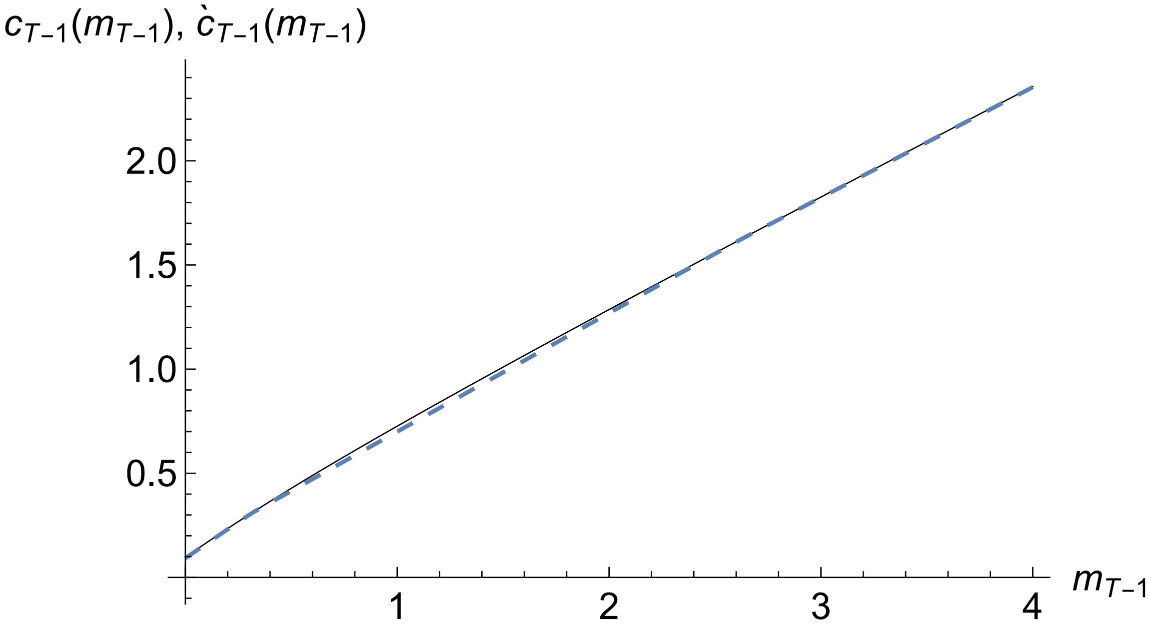

Figure 11 $c_{T-1}(m_{T-1})$ (solid) versus $\grave{c}_{T-1}(m_{T-1})$ (dashed)

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_{T-1} – a method that yields values that are more densely spaced than the uniform grid at low values of a. 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 $\bar{a}_{T-1}$ 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_θ) 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 (captured by the arguments `minval`, `maxval`, and `size`) and returns the new grid incorporating the multi-exponential approach outlined above. Then, a call is made to this function and the improved grid of assets is stored in the object `aVec_eee`. Lastly, the endogenous gridpoint method described in the previous section is performed using this new grid of assets. 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.

5.10 The Method of Moderation

Unfortunately, this endogenous gridpoints solution is not very well-behaved outside the original range of gridpoints targeted by the solution method. (Though other com-

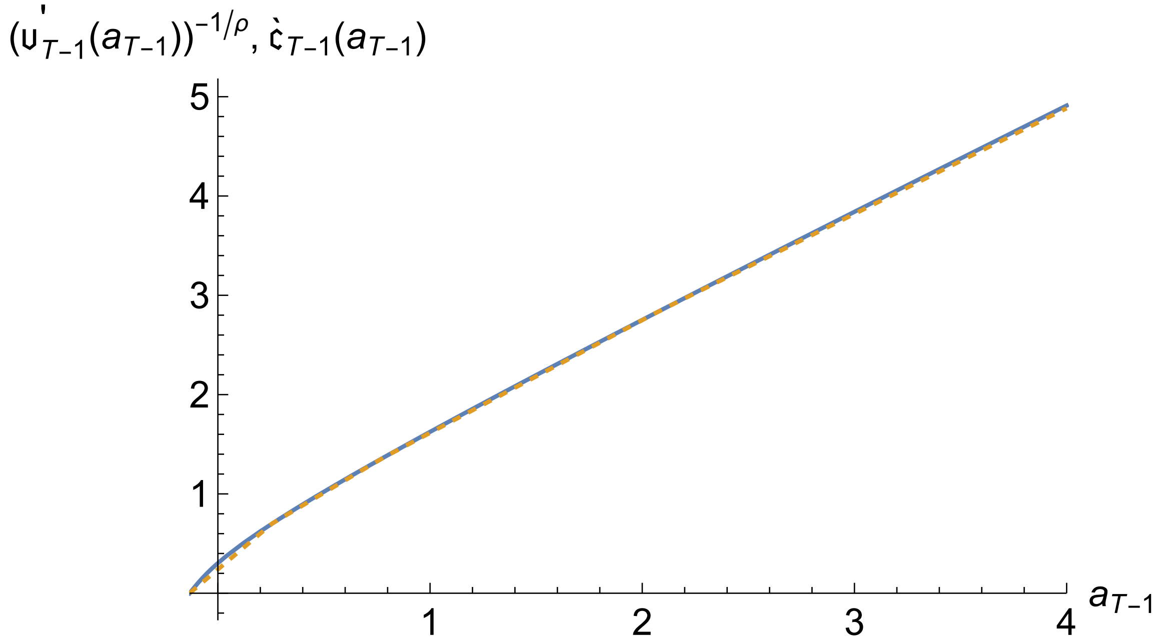

Figure 12 $\mathrm{c}_{+(T-1)}(a_{T-1})$ versus $\grave{\mathrm{c}}_{+(T-1)}(a_{T-1})$, Multi-Exponential `aVec`

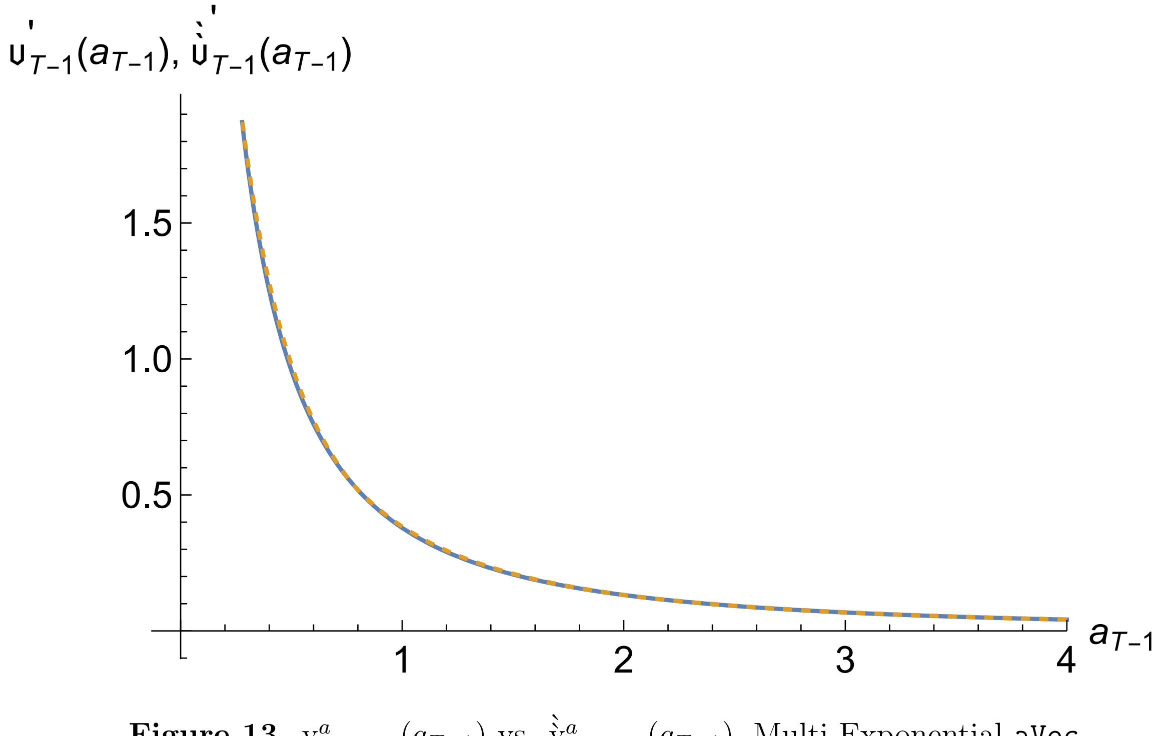

Figure 13 $\mathrm{v}^{a}_{+(T-1)}(a_{T-1})$ vs. $\grave{\grave{\mathrm{v}}}^{a}_{+(T-1)}(a_{T-1})$, Multi-Exponential `aVec`

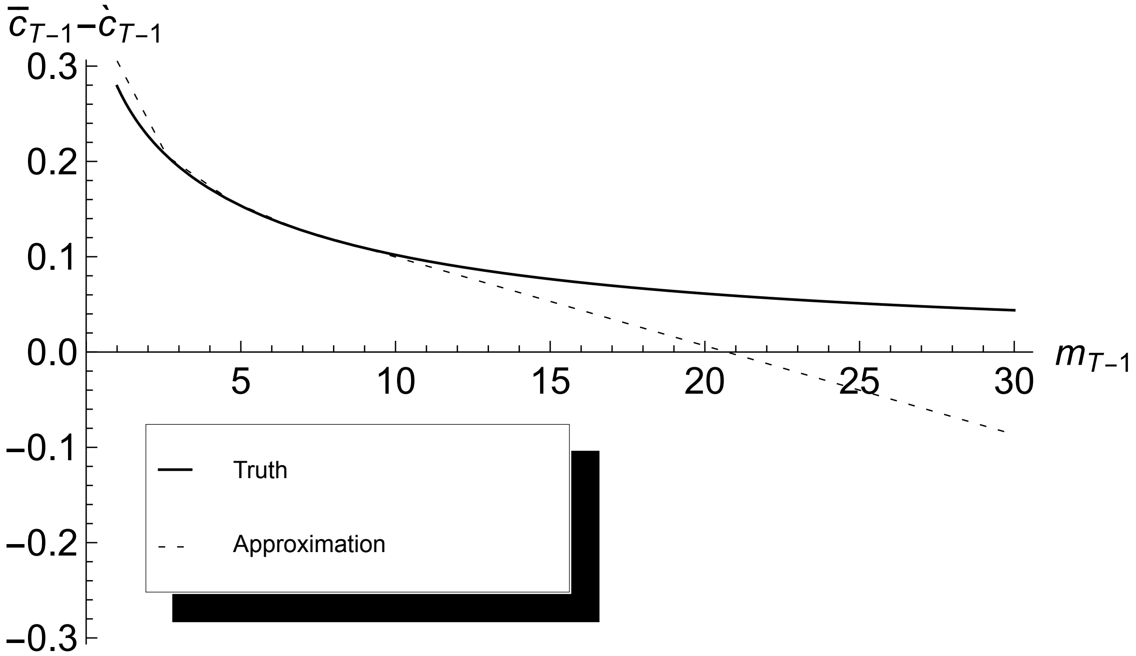

Figure 14 For Large Enough m_{T-1}, Predicted Precautionary Saving is Negative (Oops!)

mon solution methods are no better outside their own predefined ranges). Figure 14 demonstrates the point by plotting the amount of precautionary saving implied by a linear extrapolation of our approximated consumption rule (the consumption of the perfect foresight consumer $\bar{c}_{T-1}$ minus our approximation to optimal consumption under uncertainty, $\grave{c}_{T-1}$). Although theory proves that precautionary saving is always positive, the linearly extrapolated numerical approximation eventually predicts negative precautionary saving (at the point in the figure where the extrapolated locus crosses the horizontal axis).

This error cannot be fixed by extending the upper gridpoint; in the presence of serious uncertainty, the consumption rule will need to be evaluated outside of *any* prespecified grid (because starting from the top gridpoint, a large enough realization of the uncertain variable will push next period's realization of assets above that top; a similar argument applies below the bottom gridpoint). While a judicious extrapolation technique can prevent this problem from being fatal (for example by carefully excluding negative precautionary saving), the problem is often dealt with using inelegant methods whose implications for the accuracy of the solution are difficult to gauge.

As a preliminary to our solution, define $\mathfrak{h}_t$ as end-of-period human wealth (the present discounted value of future labor income) for a perfect foresight version of the problem of a 'risk optimist:' a consumer who believes with perfect confidence that the shocks will always take the value 1, $\theta_{t+n} = \mathbb{E}[\theta] = 1 \ \forall \ n > 0$. The solution to a perfect foresight problem of this kind takes the form[12]

$$\bar{\mathrm{c}}_t(m_t) = (m_t + \mathfrak{h}_t)\underline{\kappa}_t \tag{40}$$

for a constant minimal marginal propensity to consume $\underline{\kappa}_t$ given below.

We similarly define $\underline{\mathfrak{h}}_t$ as 'minimal human wealth,' the present discounted value of labor income if the shocks were to take on their worst possible value in every future period $\theta_{t+n} = \underline{\theta} \ \forall \ n > 0$ (which we define as corresponding to the beliefs of a 'pessimist').

We will call a 'realist' the consumer who correctly perceives the true probabilities of the future risks and optimizes accordingly.

A first useful point is that, for the realist, a lower bound for the level of market resources is $\underline{m}_t = -\underline{\mathfrak{h}}_t$, because if m_t equalled this value then there would be a positive finite chance (however small) of receiving $\theta_{t+n} = \underline{\theta}$ in every future period, which would require the consumer to set c_t to zero in order to guarantee that the intertemporal budget constraint holds (this is the multiperiod generalization of the discussion in section 5.7 about $\underline{a}_{T-1}$). Since consumption of zero yields negative infinite utility, the solution to realist consumer's problem is not well defined for values of $m_t < \underline{m}_t$, and the limiting value of the realist's c_t is zero as $m_t \downarrow \underline{m}_t$.

Given this result, it will be convenient to define 'excess' market resources as the amount by which actual resources exceed the lower bound, and 'excess' human wealth as the amount by which mean expected human wealth exceeds guaranteed minimum human wealth:

$$\begin{aligned} \blacktriangle m_t &= m_t + \overbrace{\underline{\mathfrak{h}}_t}^{=-\underline{m}_t} \\ \blacktriangle \mathfrak{h}_t &= \mathfrak{h}_t - \underline{\mathfrak{h}}_t. \end{aligned}$$

We can now transparently define the optimal consumption rules for the two perfect foresight problems, those of the 'optimist' and the 'pessimist.' The 'pessimist' perceives human wealth to be equal to its minimum feasible value $\underline{\mathfrak{h}}_t$ with certainty, so consumption is given by the perfect foresight solution

$$\begin{aligned} \underline{\mathrm{c}}_t(m_t) &= (m_t + \underline{\mathfrak{h}}_t)\underline{\kappa}_t \\ &= \blacktriangle m_t \underline{\kappa}_t. \end{aligned}$$

The 'optimist,' on the other hand, pretends that there is no uncertainty about future income, and therefore consumes

$$\begin{aligned} \bar{\mathrm{c}}_t(m_t) &= (m_t + \underline{\mathfrak{h}}_t - \underline{\mathfrak{h}}_t + \mathfrak{h}_t)\underline{\kappa}_t \\ &= (\blacktriangle m_t + \blacktriangle \mathfrak{h}_t)\underline{\kappa}_t \\ &= \underline{\mathrm{c}}_t(m_t) + \blacktriangle \mathfrak{h}_t \underline{\kappa}_t. \end{aligned}$$

It seems obvious that the spending of the realist will be strictly greater than that of the pessimist and strictly less than that of the optimist. Figure 15 illustrates the proposition for the consumption rule in period $T-1$.

[12] For a derivation, see Carroll (2023b); $\underline{\kappa}_t$ is defined therein as the MPC of the perfect foresight consumer with horizon $T - t$.

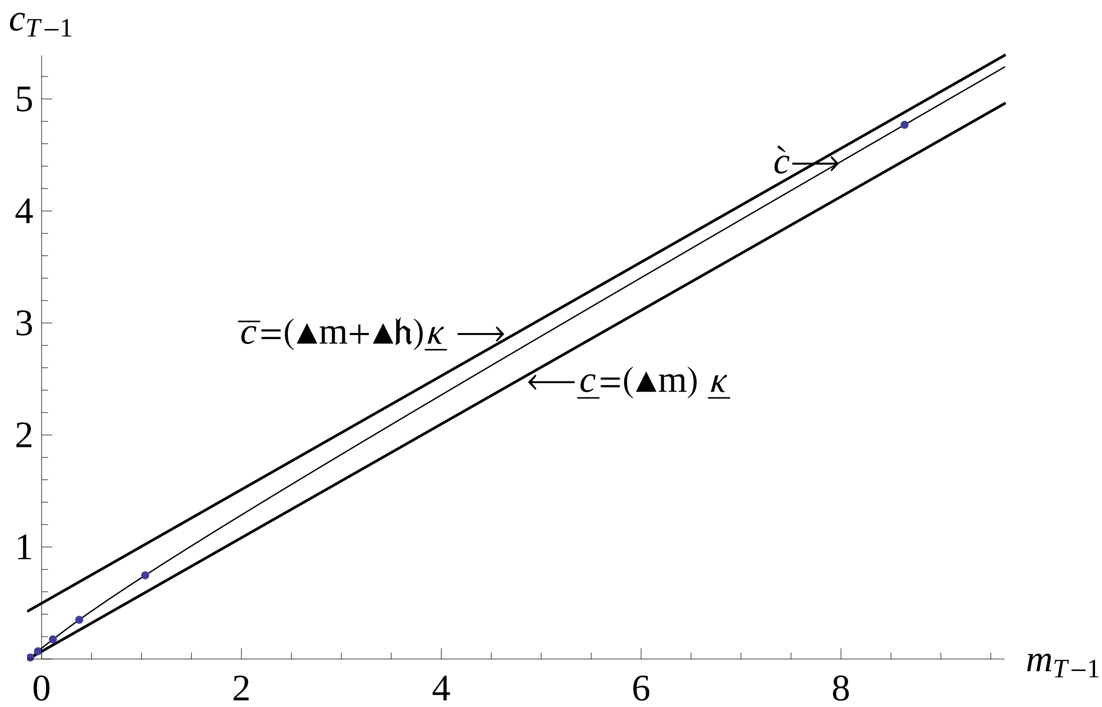

Figure 15 Moderation Illustrated: $\underline{\mathrm{c}}_{T-1} < \grave{\mathrm{c}}_{T-1} < \bar{\mathrm{c}}_{T-1}$

The proof is more difficult than might be imagined, but the necessary work is done in Carroll (2023b) so we will take the proposition as a fact and proceed by manipulating the inequality:

$$\begin{array}{rcl} \blacktriangle m_t \underline{\kappa}_t < & \mathrm{c}_t(\underline{m}_t + \blacktriangle m_t) < (\blacktriangle m_t + \blacktriangle \mathfrak{h}_t)\underline{\kappa}_t & \\ -\blacktriangle m_t \underline{\kappa}_t > & -\mathrm{c}_t(\underline{m}_t + \blacktriangle m_t) & > -(\blacktriangle m_t + \blacktriangle \mathfrak{h}_t)\underline{\kappa}_t \\ \blacktriangle \mathfrak{h}_t \underline{\kappa}_t > & \bar{\mathrm{c}}_t(\underline{m}_t + \blacktriangle m_t) - \mathrm{c}_t(\underline{m}_t + \blacktriangle m_t) & > 0 \\ 1 > & \underbrace{\left(\frac{\bar{\mathrm{c}}_t(\underline{m}_t + \blacktriangle m_t) - \mathrm{c}_t(\underline{m}_t + \blacktriangle m_t)}{\blacktriangle \mathfrak{h}_t \underline{\kappa}_t} \right)}_{\equiv \hat{\varphi}_t} & > 0 \end{array}$$

where the fraction in the middle of the last inequality is the ratio of actual precautionary saving (the numerator is the difference between perfect-foresight consumption and optimal consumption in the presence of uncertainty) to the maximum conceivable amount of precautionary saving (the amount that would be undertaken by the pessimist who consumes nothing out of any future income beyond the perfectly certain component).

Defining $\mu_t = \log \blacktriangle m_t$ (which can range from $-\infty$ to ∞), the object in the middle of the last inequality is

$$\hat{\varphi}_t(\mu_t) \equiv \left(\frac{\bar{\mathrm{c}}_t(\underline{m}_t + e^{\mu_t}) - \mathrm{c}_t(\underline{m}_t + e^{\mu_t})}{\blacktriangle \mathfrak{h}_t \underline{\kappa}_t} \right), \tag{41}$$

and we now define

$$\begin{aligned}\hat{\boldsymbol{\chi}}_t(\mu_t) &= \log\left(\frac{1-\hat{\varrho}_t(\mu_t)}{\hat{\varrho}_t(\mu_t)}\right)\\ &= \log\left(1/\hat{\varrho}_t(\mu_t)-1\right)\end{aligned} \tag{42}$$

which has the virtue that it is linear in the limit as μ_t approaches $+\infty$.

Given $\hat{\boldsymbol{\chi}}$, the consumption function can be recovered from

$$\hat{\mathrm{c}}_t = \bar{\mathrm{c}}_t - \overbrace{\left(\frac{1}{1+\exp(\hat{\boldsymbol{\chi}}_t)}\right)}^{=\hat{\varrho}_t} \blacktriangle\mathfrak{h}_t\underline{\kappa}_t. \tag{43}$$

Thus, the procedure is to calculate $\hat{\boldsymbol{\chi}}_t$ at the points $\vec{\mu}_t$ corresponding to the log of the $\blacktriangle\vec{m}_t$ points defined above, and then using these to construct an interpolating approximation $\grave{\hat{\boldsymbol{\chi}}}_t$ from which we indirectly obtain our approximated consumption rule $\grave{\hat{\mathrm{c}}}_t$ by substituting $\grave{\hat{\boldsymbol{\chi}}}_t$ for $\hat{\boldsymbol{\chi}}$ in equation (43).

Because this method relies upon the fact that the problem is easy to solve if the decision maker has unreasonable views (either in the optimistic or the pessimistic direction), and because the correct solution is always between these immoderate extremes, we call our solution procedure the 'method of moderation.'

Results are shown in Figure 16; a reader with very good eyesight might be able to detect the barest hint of a discrepancy between the Truth and the Approximation at the far righthand edge of the figure – a stark contrast with the calamitous divergence evident in Figure 14.

5.11 Approximating the Slope Too

Until now, we have calculated the level of consumption at various different gridpoints and used linear interpolation (either directly for c_{T-1} or indirectly for, say, $\hat{\boldsymbol{\chi}}_{T-1}$). But the resulting piecewise linear approximations have the unattractive feature that they are not differentiable at the 'kink points' that correspond to the gridpoints where the slope of the function changes discretely.

Carroll (2023b) shows that the true consumption function for this problem is 'smooth:' It exhibits a well-defined unique marginal propensity to consume at every positive value of m. This suggests that we should calculate, not just the level of consumption, but also the marginal propensity to consume (henceforth κ) at each gridpoint, and then find an interpolating approximation that smoothly matches both the level and the slope at those points.

This requires us to differentiate (41) and (42), yielding

$$\begin{aligned}\hat{\varrho}_t^{\mu}(\mu_t) &= (\blacktriangle\mathfrak{h}_t\underline{\kappa}_t)^{-1}e^{\mu_t}\left(\underline{\kappa}_t - \overbrace{\mathrm{c}_t^m(\underline{m}_t+e^{\mu_t})}^{\equiv\boldsymbol{\kappa}_t(m_t)}\right)\\ \hat{\boldsymbol{\chi}}_t^{\mu}(\mu_t) &= \left(\frac{-\hat{\varrho}_t^{\mu}(\mu_t)/\hat{\varrho}_t^2}{1/\hat{\varrho}_t(\mu_t)-1}\right)\end{aligned} \tag{44}$$

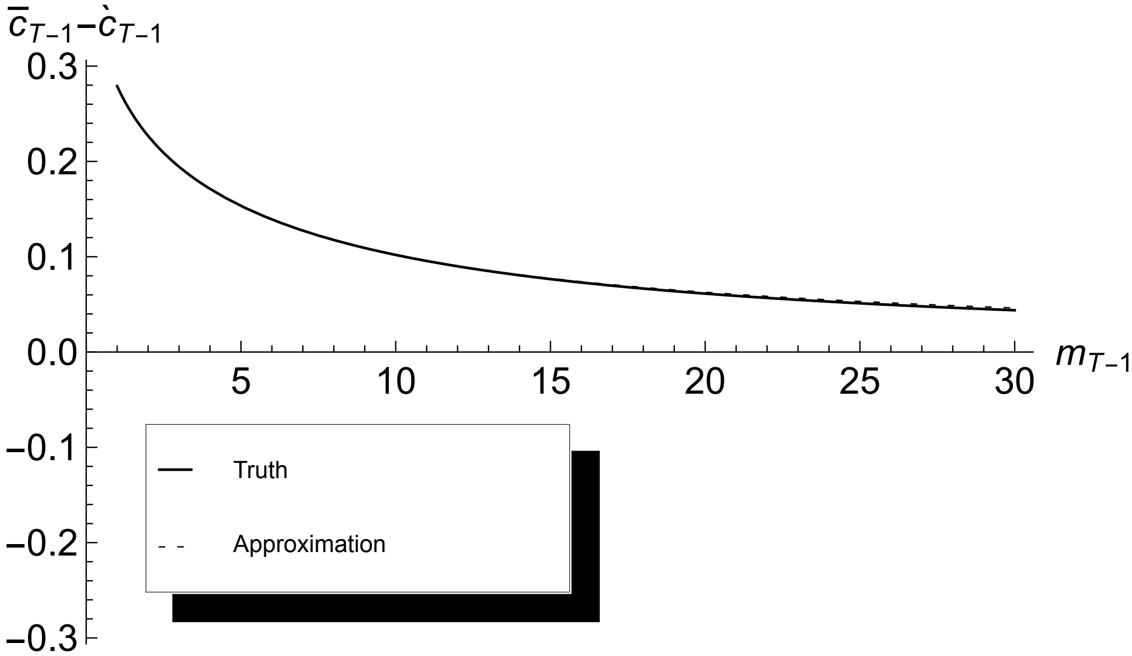

Figure 16 Extrapolated $\grave{\hat{\mathrm{c}}}_{T-1}$ Constructed Using the Method of Moderation

and (dropping arguments) with some algebra these can be combined to yield

$$\hat{\chi}_t^{\mu} = \left(\frac{\underline{\kappa}_t \blacktriangle m_t \blacktriangle \mathfrak{h}_t (\underline{\kappa}_t - \kappa_t)}{(\bar{\mathrm{c}}_t - \mathrm{c}_t)(\bar{\mathrm{c}}_t - \mathrm{c}_t - \underline{\kappa}_t \blacktriangle \mathfrak{h}_t)} \right). \tag{45}$$

To compute the vector of values of (44) corresponding to the points in $\vec{\mu}_t$, we need the marginal propensities to consume (designated κ) at each of the gridpoints, c_t^m (the vector of such values is $\vec{\kappa}_t$). These can be obtained by differentiating the Euler equation (22) (where we define $\mathfrak{m}_t(a) \equiv \mathrm{c}_{+(T)}(a) + a$):

$$\mathrm{u}^c(\mathrm{c}_t) = \mathrm{v}_{+\hat{(\mathrm{T})}}{}^{a}(\mathfrak{m}_t - \mathrm{c}_t) \tag{46}$$

with respect to a, yielding a marginal propensity to *have consumed* $\mathrm{c}_{\bar{t}}^{a}$ at each gridpoint:

$$\begin{aligned} \mathrm{u}''(\mathrm{c}_{+(T)})\mathrm{c}^a_{+(T)} &= \mathrm{v}_{+\hat{(\mathrm{T})}}{}^{a}(\mathfrak{m}_t - \mathrm{c}_t) \\ \mathrm{c}^a_{+(T)} &= \mathrm{v}_{+\hat{(\mathrm{T})}}{}^{a}(\mathfrak{m}_t - \mathrm{c}_t)/\mathrm{u}''(\mathrm{c}_{+(T)}) \end{aligned} \tag{47}$$

and the marginal propensity to consume at the beginning of the period is obtained from the marginal propensity to have consumed by noting that

$$\begin{aligned} \mathfrak{c} &= \mathfrak{m} - a \\ \mathfrak{c}^a + 1 &= \mathfrak{m}^a \end{aligned}$$

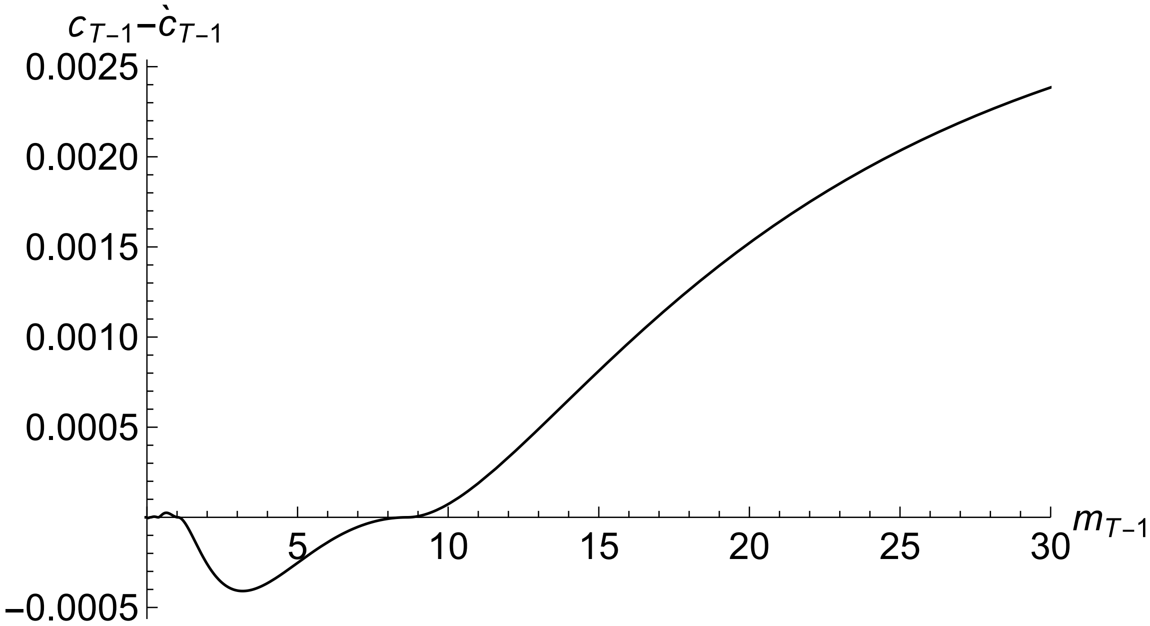

Figure 17 Difference Between True c_{T-1} and $\grave{\hat{\mathrm{c}}}_{T-1}$ Is Minuscule

which, together with the chain rule $\mathfrak{c}^a = \mathrm{c}^m \mathfrak{m}^a$, yields the MPC from

$$\begin{aligned} \mathrm{c}^m(\overbrace{\mathfrak{c}^a + 1}^{=\mathfrak{m}^a}) &= \mathfrak{c}^a \\ \mathrm{c}^m &= \mathfrak{c}^a/(1+\mathfrak{c}^a). \end{aligned} \tag{48}$$

Designating $\grave{\hat{\mathrm{c}}}_{T-1}$ as the approximated consumption rule obtained using an interpolating polynomial approximation to $\hat{\boldsymbol{\chi}}$ that matches both the level and the first derivative at the gridpoints, Figure 17 plots the difference between this latest approximation and the true consumption rule for period $T-1$ up to the same large value (far beyond the largest gridpoint) used in prior figures. Of course, at the gridpoints the approximation will match the true function; but this figure illustrates that the approximation is quite accurate far beyond the last gridpoint (which is the last point at which the difference touches the horizontal axis). (We plot here the difference between the two functions rather than the level plotted in previous figures, because in levels the approximation error would not be detectable even to the most eagle-eyed reader.)

5.12 Value

Often it is useful to know the value function as well as the consumption rule. Fortunately, many of the tricks used when solving for the consumption rule have a direct analogue in approximation of the value function.

Consider the perfect foresight (or "optimist's") problem in period $T-1$:

$$
\begin{aligned}
\bar{\mathrm{v}}_{T-1}(m_{T-1}) &\equiv \mathrm{u}(c_{T-1}) + \beta\mathrm{u}(c_T) \\
&= \mathrm{u}(c_{T-1})\left(1 + \beta((\beta_T\mathsf{R})^{1/\rho})^{1-\rho}\right) \\
&= \mathrm{u}(c_{T-1})\left(1 + \beta(\beta_T\mathsf{R})^{1/\rho-1}\right) \\
&= \mathrm{u}(c_{T-1})\left(1 + (\beta_T\mathsf{R})^{1/\rho}/\mathsf{R}\right) \\
&= \mathrm{u}(c_{T-1})\underbrace{\mathrm{PDV}_t^T(c)/c_{T-1}}_{\equiv \mathbb{C}_t^T}
\end{aligned}
$$

where $\mathbb{C}_t^T = \mathrm{PDV}_t^T(c)$ is the present discounted value of consumption. A similar function can be constructed recursively for earlier periods, yielding the general expression

$$
\begin{aligned}
\bar{\mathrm{v}}_t(m_t) &= \mathrm{u}(\bar{c}_t)\mathbb{C}_t^T \\
&= \mathrm{u}(\bar{c}_t)\underline{\kappa}_t^{-1} \\
&= \mathrm{u}((\blacktriangle m_t + \blacktriangle\mathfrak{h}_t)\underline{\kappa}_t)\underline{\kappa}_t^{-1} \\
&= \mathrm{u}(\blacktriangle m_t + \blacktriangle\mathfrak{h}_t)\underline{\kappa}_t^{1-\rho}\underline{\kappa}_t^{-1} \\
&= \mathrm{u}(\blacktriangle m_t + \blacktriangle\mathfrak{h}_t)\underline{\kappa}_t^{-\rho}
\end{aligned}
\tag{49}
$$

where the second line uses the fact demonstrated in Carroll (2023b) that $\mathbb{C}_t = \kappa_t^{-1}$.

This can be transformed as

$$
\begin{aligned}
\bar{\Lambda}_t &\equiv ((1-\rho)\bar{\mathrm{v}}_t)^{1/(1-\rho)} \\
&= c_t(\mathbb{C}_t^T)^{1/(1-\rho)} \\
&= (\blacktriangle m_t + \blacktriangle\mathfrak{h}_t)\underline{\kappa}_t^{-\rho/(1-\rho)}
\end{aligned}
$$

with derivative

$$
\begin{aligned}
\bar{\Lambda}_t^m &= (\mathbb{C}_t^T)^{1/(1-\rho)}\underline{\kappa}_t, \\
&= \underline{\kappa}_t^{-\rho/(1-\rho)}
\end{aligned}
$$

and since $\mathbb{C}_t^T$ is a constant while the consumption function is linear, $\bar{\Lambda}_t$ will also be linear.

We apply the same transformation to the value function for the problem with uncertainty (the "realist's" problem) and differentiate

$$
\begin{aligned}
\bar{\Lambda}_t &= ((1-\rho)\bar{\mathrm{v}}_t(m_t))^{1/(1-\rho)} \\
\bar{\Lambda}_t^m &= ((1-\rho)\bar{\mathrm{v}}_t(m_t))^{-1+1/(1-\rho)}\,\bar{\mathrm{v}}_t^m(m_t)
\end{aligned}
$$

and an excellent approximation to the value function can be obtained by calculating the values of $\bar{\Lambda}$ at the same gridpoints used by the consumption function approximation, and interpolating among those points.

However, as with the consumption approximation, we can do even better if we realize that the $\bar{\Lambda}$ function for the optimist's problem is an upper bound for the Λ function in the presence of uncertainty, and the value function for the pessimist is a lower bound.

Analogously to (41), define an upper-case

$$\hat{Ϙ}_t(\mu_t) = \left(\frac{\bar{\Lambda}_t(\underline{m}_t + e^{\mu_t}) - \Lambda_t(\underline{m}_t + e^{\mu_t})}{\blacktriangle\mathfrak{h}_t\underline{\kappa}_t(\mathbb{C}_t^T)^{1/(1-\rho)}}\right) \tag{50}$$

with derivative (dropping arguments)

$$\hat{Ϙ}_t^\mu = (\blacktriangle\mathfrak{h}_t\underline{\kappa}_t(\mathbb{C}_t^T)^{1/(1-\rho)})^{-1}e^{\mu_t}\,(\bar{\Lambda}_t^m - \Lambda_t^m) \tag{51}$$

and an upper-case version of the $\boldsymbol{\chi}$ equation in (42):

$$\begin{aligned}\hat{\mathrm{X}}_t(\mu_t) &= \log\left(\frac{1-\hat{Ϙ}_t(\mu_t)}{\hat{Ϙ}_t(\mu_t)}\right) \\ &= \log\left(1/\hat{Ϙ}_t(\mu_t) - 1\right)\end{aligned} \tag{52}$$

with corresponding derivative

$$\hat{\mathrm{X}}_t^\mu = \left(\frac{-\hat{Ϙ}_t^\mu/\hat{Ϙ}_t^2}{1/\hat{Ϙ}_t - 1}\right) \tag{53}$$

and if we approximate these objects then invert them (as above with the $\hat{ϙ}$ and $\hat{\boldsymbol{\chi}}$ functions) we obtain a very high-quality approximation to our inverted value function at the same points for which we have our approximated value function:

$$\hat{\Lambda}_t = \bar{\Lambda}_t - \overbrace{\left(\frac{1}{1+\exp(\hat{\mathrm{X}}_t)}\right)}^{=\hat{Ϙ}_t}\blacktriangle\mathfrak{h}_t\underline{\kappa}_t(\mathbb{C}_t^T)^{1/(1-\rho)} \tag{54}$$

from which we obtain our approximation to the value function and its derivatives as

$$\begin{aligned}\hat{\mathrm{v}}_t &= \mathrm{u}(\hat{\Lambda}_t) \\ \hat{\mathrm{v}}_t^m &= \mathrm{u}^c(\hat{\Lambda}_t)\hat{\Lambda}^m \\ \hat{\mathrm{v}}_t^{mm} &= \mathrm{u}^{cc}(\hat{\Lambda}_t)(\hat{\Lambda}^m)^2 + \mathrm{u}^c(\hat{\Lambda}_t)\hat{\Lambda}^{mm}.\end{aligned} \tag{55}$$

Although a linear interpolation that matches the level of Λ at the gridpoints is simple, a Hermite interpolation that matches both the level and the derivative of the $\bar{\Lambda}_t$ function at the gridpoints has the considerable virtue that the $\bar{\mathrm{v}}_t$ derived from it numerically satisfies the envelope theorem at each of the gridpoints for which the problem has been solved.

If we use the double-derivative calculated above to produce a higher-order Hermite polynomial, our approximation will also match marginal propensity to consume at the gridpoints; this would guarantee that the consumption function generated from the value function would match both the level of consumption and the marginal propensity to consume at the gridpoints; the numerical differences between the newly constructed consumption function and the highly accurate one constructed earlier would be negligible within the grid.

5.13 Refinement: A Tighter Upper Bound

Carroll (2023b) derives an upper limit $\bar{\kappa}_t$ for the MPC as m_t approaches its lower bound. Using this fact plus the strict concavity of the consumption function yields the proposition that

$$\mathrm{c}_t(\underline{m}_t + \blacktriangle m_t) < \bar{\kappa}_t \blacktriangle m_t. \tag{56}$$

The solution method described above does not guarantee that approximated consumption will respect this constraint between gridpoints, and a failure to respect the constraint can occasionally cause computational problems in solving or simulating the model. Here, we describe a method for constructing an approximation that always satisfies the constraint.

Defining $m_t^{\#}$ as the 'cusp' point where the two upper bounds intersect:

$$\begin{aligned}
\left(\blacktriangle m_t^{\#} + \blacktriangle \mathfrak{h}_t\right) \underline{\kappa}_t &= \bar{\kappa}_t \blacktriangle m_t^{\#} \\
\blacktriangle m_t^{\#} &= \frac{\underline{\kappa}_t \blacktriangle \mathfrak{h}_t}{(1 - \underline{\kappa}_t)\bar{\kappa}_t} \\
m_t^{\#} &= \frac{\underline{\kappa}_t \mathfrak{h}_t - \underline{\mathfrak{h}}_t}{(1 - \underline{\kappa}_t)\bar{\kappa}_t},
\end{aligned}$$

we want to construct a consumption function for $m_t \in (\underline{m}_t, m_t^{\#}]$ that respects the tighter upper bound:

$$\begin{array}{rcl}
\blacktriangle m_t \underline{\kappa}_t < & \mathrm{c}_t(\underline{m}_t + \blacktriangle m_t) < \bar{\kappa}_t \blacktriangle m_t & \\
\blacktriangle m_t(\bar{\kappa}_t - \underline{\kappa}_t) > & \bar{\kappa}_t \blacktriangle m_t - \mathrm{c}_t(\underline{m}_t + \blacktriangle m_t) & > 0 \\
1 > & \left(\frac{\bar{\kappa}_t \blacktriangle m_t - \mathrm{c}_t(\underline{m}_t + \blacktriangle m_t)}{\blacktriangle m_t(\bar{\kappa}_t - \underline{\kappa}_t)}\right) & > 0.
\end{array}$$

Again defining $\mu_t = \log \blacktriangle m_t$, the object in the middle of the inequality is

$$\begin{aligned}
\check{\varphi}_t(\mu_t) &\equiv \frac{\bar{\kappa}_t - \mathrm{c}_t(\underline{m}_t + e^{\mu_t})e^{-\mu_t}}{\bar{\kappa}_t - \underline{\kappa}_t} \\
\check{\varphi}_t^{\mu}(\mu_t) &= \frac{\mathrm{c}_t(\underline{m}_t + e^{\mu_t})e^{-\mu_t} - \boldsymbol{\kappa}_t^m(\underline{m}_t + e^{\mu_t})}{\bar{\kappa}_t - \underline{\kappa}_t}.
\end{aligned}$$

As m_t approaches $-\underline{m}_t$, $\check{\varphi}_t(\mu_t)$ converges to zero, while as m_t approaches $+\infty$, $\check{\varphi}_t(\mu_t)$ approaches 1.

As before, we can derive an approximated consumption function; call it $\grave{\check{\mathrm{c}}}_t$. This function will clearly do a better job approximating the consumption function for low values of m_t while the previous approximation will perform better for high values of m_t.

For middling values of m it is not clear which of these functions will perform better. However, an alternative is available which performs well. Define the highest gridpoint below $m_t^{\#}$ as $\bar{\check{m}}_t^{\#}$ and the lowest gridpoint above $m_t^{\#}$ as $\underline{\hat{m}}_t^{\#}$. Then there will be a unique interpolating polynomial that matches the level and slope of the consumption function at these two points. Call this function $\tilde{\mathrm{c}}_t(m)$.

Using indicator functions that are zero everywhere except for specified intervals,

$$\begin{aligned}
\mathbf{1}_{\text{Lo}}(m) &= 1 \text{ if } m \leq \bar{\grave{m}}_t^{\#} \\
\mathbf{1}_{\text{Mid}}(m) &= 1 \text{ if } \qquad \bar{\grave{m}}_t^{\#} < m < \underline{\hat{m}}_t^{\#} \\
\mathbf{1}_{\text{Hi}}(m) &= 1 \text{ if } \qquad\qquad\qquad \underline{\hat{m}}_t^{\#} \leq m
\end{aligned}$$

we can define a well-behaved approximating consumption function

$$\grave{\mathrm{c}}_t = \mathbf{1}_{\text{Lo}}\grave{\grave{\mathrm{c}}}_t + \mathbf{1}_{\text{Mid}}\grave{\grave{\mathrm{c}}}_t + \mathbf{1}_{\text{Hi}}\grave{\grave{\mathrm{c}}}_t. \tag{57}$$

This just says that, for each interval, we use the approximation that is most appropriate. The function is continuous and once-differentiable everywhere, and is therefore well behaved for computational purposes.

We now construct an upper-bound value function implied for a consumer whose spending behavior is consistent with the refined upper-bound consumption rule.

For $m_t \geq m_t^{\#}$, this consumption rule is the same as before, so the constructed upper-bound value function is also the same. However, for values $m_t < m_t^{\#}$ matters are slightly more complicated.

Start with the fact that at the cusp point,

$$\begin{aligned}
\bar{\mathrm{v}}_t(m_t^{\#}) &= \mathrm{u}(\bar{c}_t(m_t^{\#}))\mathbb{C}_t^T \\
&= \mathrm{u}(\blacktriangle m_t^{\#}\bar{\kappa}_t)\mathbb{C}_t^T.
\end{aligned}$$

But for *all* m_t,

$$\bar{\mathrm{v}}_t(m) = \mathrm{u}(\bar{c}_t(m)) + \bar{\mathrm{v}}_{+(\mathrm{T})}(m - \bar{c}_t(m)),$$

and we assume that for the consumer below the cusp point consumption is given by $\bar{\kappa}\blacktriangle m_t$ so for $m_t < m_t^{\#}$

$$\bar{\mathrm{v}}_t(m) = \mathrm{u}(\bar{\kappa}_t\blacktriangle m) + \bar{\mathrm{v}}_{+(\mathrm{T})}((1 - \bar{\kappa}_t)\blacktriangle m),$$

which is easy to compute because $\bar{\mathrm{v}}_{+(\mathrm{T})}(a_t) = \beta\bar{\mathrm{v}}_{t+1}(a_t\mathcal{R} + 1)$ where $\bar{\mathrm{v}}_t$ is as defined above because a consumer who ends the current period with assets exceeding the lower bound will not expect to be constrained next period. (Recall again that we are merely constructing an object that is guaranteed to be an *upper bound* for the value that the 'realist' consumer will experience.) At the gridpoints defined by the solution of the consumption problem can then construct

$$\bar{\Lambda}_t(m) = ((1 - \rho)\bar{\mathrm{v}}_t(m))^{1/(1-\rho)}$$

and its derivatives which yields the appropriate vector for constructing $\check{\mathrm{X}}$ and $\check{\mathrm{\Omega}}$. The rest of the procedure is analogous to that performed for the consumption rule and is thus omitted for brevity.

5.14 Extension: A Stochastic Interest Factor

Thus far we have assumed that the interest factor is constant at R. Extending the previous derivations to allow for a perfectly forecastable time-varying interest factor R_t would be trivial. Allowing for a stochastic interest factor is less trivial.

The easiest case is where the interest factor is i.i.d.,

$$\log \mathbf{R}_{t+n} \sim \mathcal{N}(\mathsf{r} + \phi - \sigma_{\mathbf{r}}^2/2, \sigma_{\mathbf{r}}^2) \ \forall\ n > 0 \tag{58}$$

where ϕ is the risk premium and the $\sigma_{\mathbf{r}}^2/2$ adjustment to the mean log return guarantees that an increase in $\sigma_{\mathbf{r}}^2$ constitutes a mean-preserving spread in the level of the return.

This case is reasonably straightforward because Merton (1969) and Samuelson (1969) showed that for a consumer without labor income (or with perfectly forecastable labor income) the consumption function is linear, with an infinite-horizon MPC[13]

$$\kappa = 1 - \left(\beta\, \mathbb{E}_t[\mathbf{R}_{t+1}^{1-\rho}]\right)^{1/\rho} \tag{59}$$

and in this case the previous analysis applies once we substitute this MPC for the one that characterizes the perfect foresight problem without rate-of-return risk.

The more realistic case where the interest factor has some serial correlation is more complex. We consider the simplest case that captures the main features of empirical interest rate dynamics: An AR(1) process. Thus the specification is

$$\mathbf{r}_{t+1} - \mathbf{r} = (\mathbf{r}_t - \mathbf{r})\gamma + \epsilon_{t+1} \tag{60}$$

where $\mathbf{r}$ is the long-run mean log interest factor, $0 < \gamma < 1$ is the AR(1) serial correlation coefficient, and ϵ_{t+1} is the stochastic shock.

The consumer's problem in this case now has two state variables, m_t and $\mathbf{r}_t$, and is described by

$$\begin{aligned}
\mathrm{v}_t(m_t, \mathbf{r}_t) &= \max_{c_t}\ \mathrm{u}(c_t) + \mathbb{E}_t[\beta_{t+1}\mathcal{G}_{t+1}^{1-\rho}\mathrm{v}_{t+1}(m_{t+1}, \mathbf{r}_{t+1})] \\
&\text{s.t.} \\
a_t &= m_t - c_t \\
\mathbf{r}_{t+1} - \mathbf{r} &= (\mathbf{r}_t - \mathbf{r})\gamma + \epsilon_{t+1} \\
\mathbf{R}_{t+1} &= \exp(\mathbf{r}_{t+1}) \\
m_{t+1} &= \underbrace{(\mathbf{R}_{t+1}/\mathcal{G}_{t+1})}_{\equiv \mathcal{R}_{t+1}} a_t + \theta_{t+1}.
\end{aligned}$$

We approximate the AR(1) process by a Markov transition matrix using standard techniques. The stochastic interest factor is allowed to take on 11 values centered around the steady-state value $\mathbf{r}$. Given this Markov transition matrix, *conditional* on the Markov AR(1) state the consumption functions for the 'optimist' and the 'pessimist' will still be linear, with identical MPC's that are computed numerically. Given these MPC's, the (conditional) realist's consumption function can be computed for each Markov state, and the converged consumption rules constitute the solution contingent on the dynamics of the stochastic interest rate process.

In principle, this refinement should be combined with the previous one; further exposition of this combination is omitted here because no new insights spring from the combination of the two techniques.

[13] See `CRRA-RateRisk` for a derivation.

5.15 Imposing 'Artificial' Borrowing Constraints

Optimization problems often come with additional constraints that must be satisfied. Particularly common is an 'artificial' liquidity constraint that prevents the consumer's net worth from falling below some value, often zero.[14] The problem then becomes

$$\begin{aligned}
\mathrm{v}_{T-1}(m_{T-1}) &= \max_{c_{T-1}} \ \mathrm{u}(c_{T-1}) + \mathbb{E}_{T-1}[\beta \mathcal{G}_T^{1-\rho} \mathrm{v}_{+(T)}(m_T)] \\
&\text{s.t.} \\
a_{T-1} &= m_{T-1} - c_{T-1} \\
m_T &= \mathcal{R}_T a_{T-1} + \theta_T \\
a_{T-1} &\geq 0.
\end{aligned}$$

By definition, the constraint will bind if the unconstrained consumer would choose a level of spending that would violate the constraint. Here, that means that the constraint binds if the c_{T-1} that satisfies the unconstrained FOC

$$c_{T-1}^{-\rho} = \mathrm{v}^a_{+(T-1)}(m_{T-1} - c_{T-1}) \tag{61}$$

is greater than m_{T-1}. Call $\grave{\mathrm{c}}^*_{T-1}$ the approximated function returning the level of c_{T-1} that satisfies (61). Then the approximated constrained optimal consumption function will be

$$\grave{\mathrm{c}}_{T-1}(m_{T-1}) = \min[m_{T-1}, \grave{\mathrm{c}}^*_{T-1}(m_{T-1})]. \tag{62}$$

The introduction of the constraint also introduces a sharp nonlinearity in all of the functions at the point where the constraint begins to bind. As a result, to get solutions that are anywhere close to numerically accurate it is useful to augment the grid of values of the state variable to include the exact value at which the constraint ceases to bind. Fortunately, this is easy to calculate. We know that when the constraint is binding the consumer is saving nothing, which yields marginal value of $\mathrm{v}^a_{+(T-1)}(0)$. Further, when the constraint is binding, $c_{T-1} = m_{T-1}$. Thus, the largest value of consumption for which the constraint is binding will be the point for which the marginal utility of consumption is exactly equal to the (expected, discounted) marginal value of saving 0. We know this because the marginal utility of consumption is a downward-sloping function and so if the consumer were to consume ϵ more, the marginal utility of that extra consumption would be *below* the (discounted, expected) marginal utility of saving, and thus the consumer would engage in positive saving and the constraint would no longer be binding. Thus the level of m_{T-1} at which the constraint stops binding is:[15]

$$\begin{aligned}
\mathrm{u}^c(m_{T-1}) &= \mathrm{v}^a_{+(T-1)}(0) \\
m_{T-1} &= (\mathrm{v}^a_{+(T-1)}(0))^{(-1/\rho)} \\
&= \mathrm{c}_{+(T-1)}(0).
\end{aligned}$$

[14] The word artificial is chosen only because of its clarity in distinguishing this from the case of the 'natural' borrowing constraint examined above; no derogation is intended – constraints of this kind certainly exist in the real world.

[15] The logic here repeats an insight from Deaton (1991).

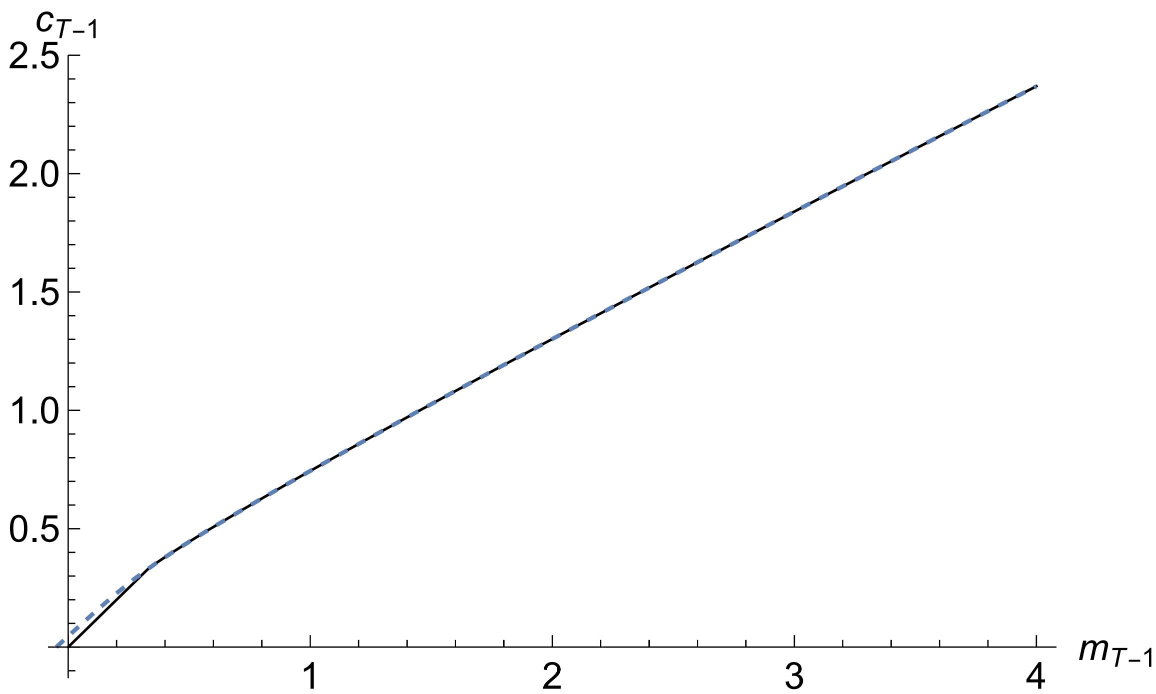

Figure 18 Constrained (solid) and Unconstrained (dashed) Consumption

The constrained problem is solved in section "Artifical Borrowing Constraint" of the notebook, where the variable `constrained` is set to be a boolean type object. If the value of `constrained` is true, then the constraint is binding and their consumption behavior is computed to match (62). The resulting consumption rule is shown in Figure 18. For comparison purposes, the approximate consumption rule from Figure 18 is reproduced here as the solid line; this is accomplished by setting the boolean value of `constrained` to false.

The presence of the liquidity constraint requires three changes to the procedures outlined above:

1. We redefine $\underline{\mathfrak{h}}_t$, which now is the PDV of receiving $\theta_{t+1} = \underline{\theta}$ next period and $\theta_{t+n} = 0 \ \forall \ n > 1$ – that is, the pessimist believes he will receive nothing beyond period $t+1$

2. We augment the end-of-period `aVec` with zero and with a point with a small positive value so that the generated `mVec` will the binding point $m^{\#}$ and a point just above it (so that we can better capture the curvature around that point)

3. We redefine the optimal consumption rule as in equation (62). This ensures that the liquidity-constrained 'realist' will consume more than the redefined 'pessimist,' so that we will have ϱ still between 0 and 1 and the 'method of moderation' will proceed smoothly.

As expected, the liquidity constraint only causes a divergence between the two functions at the point where the optimal unconstrained consumption rule runs into the 45 degree line.

6 Recursion

6.1 Theory

Before we solve for periods earlier than $T-1$, we assume for convenience that in each such period a liquidity constraint exists of the kind discussed above, preventing c from exceeding m. This simplifies things a bit because now we can always consider an `aVec` that starts with zero as its smallest element.

Recall now equations (21) and (22):

$$\begin{aligned} \mathrm{v}^{a}_{+(T)}(a_t) &= \mathbb{E}_t[\beta \mathsf{R} \mathcal{G}_{t+1}^{-\rho} \mathrm{u}^{c}(\mathrm{c}_{t+1}(\mathcal{R}_{t+1} a_t + \theta_{t+1}))] \\ \mathrm{u}^{c}(c_t) &= \mathrm{v}^{a}_{+(T)}(m_t - c_t). \end{aligned}$$

Assuming that the problem has been solved up to period $t+1$ (and thus assuming that we have an approximated $\grave{\mathrm{c}}_{t+1}(m_{t+1})$), our solution method essentially involves using these two equations in succession to work back progressively from period $T-1$ to the beginning of life. Stated generally, the method is as follows. (Here, we use the original, rather than the "refined," method for constructing consumption functions; the generalization of the algorithm below to use the refined method presents no difficulties.)

1. For the grid of values $a_{t,i}$ in `aVec_eee`, numerically calculate the values of $\mathrm{c}_{\bar{t}}(a_{t,i})$ and $\mathrm{c}^{a}_{\bar{t}}(a_{t,i})$,

$$\begin{aligned} \mathrm{c}_{\bar{t},i} &= \left(\mathrm{v}^{a}_{+(T)}(a_{t,i})\right)^{-1/\rho}, \\ &= \left(\beta\, \mathbb{E}_t\left[\mathsf{R} \mathcal{G}_{t+1}^{-\rho} (\grave{\mathrm{c}}_{t+1}(\mathcal{R}_{t+1} a_{t,i} + \theta_{t+1}))^{-\rho}\right]\right)^{-1/\rho}, \\ \mathrm{c}^{a}_{\bar{t},i} &= -(1/\rho)\left(\mathrm{v}^{a}_{+(T)}(a_{t,i})\right)^{-1-1/\rho} \mathrm{v}^{aa}_{+(T)}(a_{t,i}), \end{aligned} \tag{63}$$

   generating vectors of values $\vec{c}_t$ and $\vec{c}^{a}_{\bar{t}}$.

2. Construct a corresponding vector of values of $\vec{m}_t = \vec{c}_t + \vec{a}_t$; similarly construct a corresponding list of MPC's $\vec{\kappa}_t$ using equation (48).

3. Construct a corresponding vector $\vec{\mu}_t$, the levels and first derivatives of $\vec{\varphi}_t$, and the levels and first derivatives of $\vec{\chi}_t$.

4. Construct an interpolating approximation $\grave{\chi}_t$ that smoothly matches both the level and the slope at those points.

5. If we are to approximate the value function, construct a corresponding list of values of $\vec{\mathrm{v}}_t$, the levels and first derivatives of $\vec{\Phi}_t$, and the levels and first derivatives of $\hat{\vec{\mathrm{X}}}_t$; and construct an interpolating approximation function $\hat{\mathrm{X}}_t$ that matches those points.

With $\grave{\chi}_t$ in hand, our approximate consumption function is computed directly from the appropriate substitutions in (43) and related equations. With this consumption rule in hand, we can continue the backwards recursion to period $t-1$ and so on back to the beginning of life.

Note that this loop does not contain an item for constructing $\hat{\mathrm{v}}_t^a(m_t)$. This is because with $\grave{\hat{\mathrm{c}}}_t(m_t)$ in hand, we simply *define* $\hat{\mathrm{v}}_t^m(m_t) = \mathrm{u}^c(\grave{\hat{\mathrm{c}}}_t(m_t))$ so there is no need to construct interpolating approximations - the function arises 'free' (or nearly so) from our constructed $\grave{\hat{\mathrm{c}}}_t(m_t)$ via the usual envelope result (cf. (12)).

6.2 Program Structure

In section "Solve for $c_t(m_t)$ 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 the number of periods specified by `T`, as explained earlier.

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.

6.3 Results

As suggested, the code creates the relevant $\grave{\mathrm{c}}_t(m_t)$ functions for any period in the horizon specified by the variable `T`, at the given values of m. Figure 19 shows $\grave{\mathrm{c}}_{T-n}(m)$ 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 Multiple Control Variables

We now consider how to solve problems with multiple control variables. (To reduce notational complexity, in this section we set $\mathcal{G}_t = 1 \;\forall\; t$.)

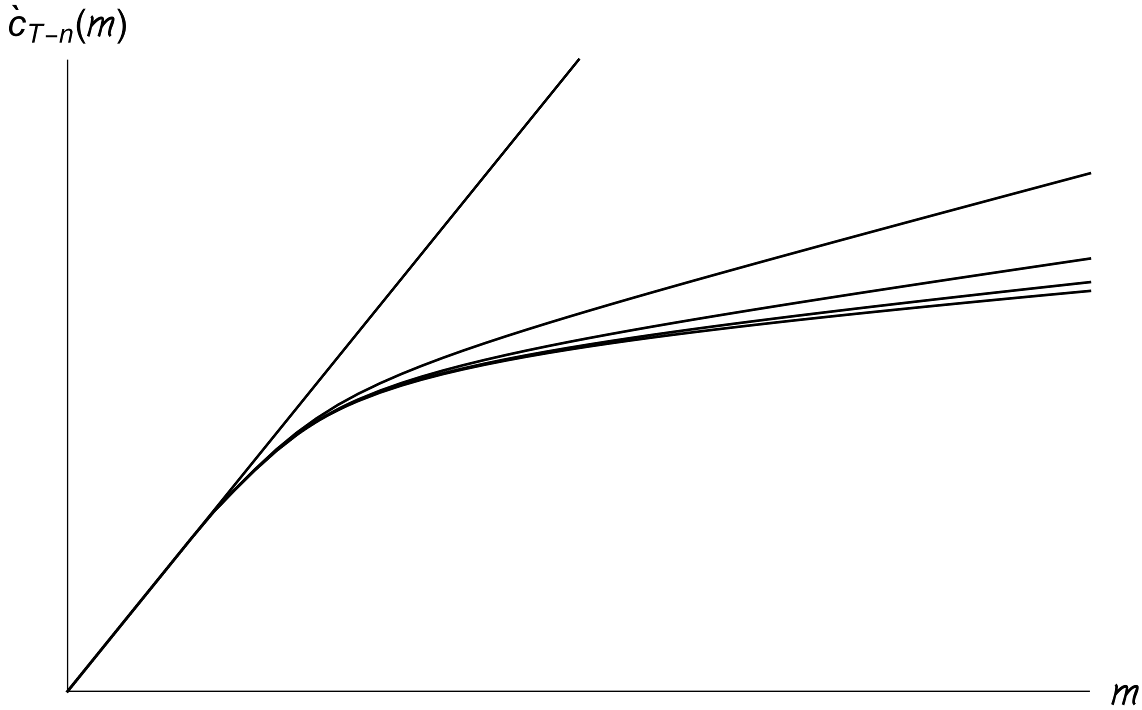

Figure 19 Converging $\grave{c}_{T-n}(m)$ Functions as n Increases

7.1 Theory

The new control variable that the consumer can now choose is the portion of the portfolio to invest in risky assets. Designating the gross return on the risky asset as $\mathbf{R}_{t+1}$, and using ς_t to represent the proportion of the portfolio invested in this asset before the return is realized after the beginning of $t+1$, corresponding to an assumption that the consumer cannot be 'net short' and cannot issue net equity), the overall return on the consumer's portfolio between t and $t+1$ will be:

$$\begin{aligned} \mathfrak{R}_{t+1} &= \mathsf{R}(1-\varsigma_t) + \mathbf{R}_{t+1}\varsigma_t \\ &= \mathsf{R} + (\mathbf{R}_{t+1} - \mathsf{R})\varsigma_t \end{aligned} \tag{64}$$

and the maximization problem is

$$\begin{aligned} \mathrm{v}_t(m_t) = \max_{\{c_t,\varsigma_t\}} \quad & \mathrm{u}(c_t) + \beta\,\mathbb{E}_t[\mathrm{v}_{t+1}(m_{t+1})] \\ \text{s.t.} & \\ \mathfrak{R}_{t+1} &= \mathsf{R} + (\mathbf{R}_{t+1} - \mathsf{R})\varsigma_t \\ m_{t+1} &= (m_t - c_t)\mathfrak{R}_{t+1} + \theta_{t+1} \\ 0 \leq & \varsigma_t \leq 1, \end{aligned}$$

or, more compactly,

$$
\begin{aligned}
\mathrm{v}_t(m_t) = \max_{\{\mathrm{c}_t, \varsigma_t\}} \quad & \mathrm{u}(c_t) + \mathbb{E}_t[\beta \mathrm{v}_{t+1}((m_t - c_t)\mathfrak{R}_{t+1} + \theta_{t+1})] \\
\text{s.t.} & \\
0 \leq \varsigma_t \leq 1. &
\end{aligned}
$$

The first order condition with respect to c_t is almost identical to that in the single-control problem, equation (11), with the only difference being that the nonstochastic interest factor R is now replaced by $\mathfrak{R}_{t+1}$,

$$
\mathrm{u}^c(c_t) = \beta\, \mathbb{E}_t[\mathfrak{R}_{t+1} \mathrm{v}^m_{t+1}(m_{t+1})], \tag{65}
$$

and the Envelope theorem derivation remains the same, yielding the Euler equation for consumption

$$
\mathrm{u}^c(c_t) = \mathbb{E}_t[\beta \mathfrak{R}_{t+1} \mathrm{u}^c(c_{t+1})]. \tag{66}
$$

The first order condition with respect to the risky portfolio share is

$$
\begin{aligned}
0 &= \mathbb{E}_t[\mathrm{v}^m_{\sim(T+1)}(m_{t+1})(\mathbf{R}_{t+1} - \mathsf{R})a_t] \\
&= a_t\, \mathbb{E}_t\left[\mathrm{u}^c\left(\mathrm{c}_{t+1}(m_{t+1})\right)\left(\mathbf{R}_{t+1} - \mathsf{R}\right)\right].
\end{aligned}
$$

As before, it will be useful to define $\mathrm{v}_{+(T)}$ as a function that yields the expected $t+1$ value of ending period t in a given state. However, now that there are two control variables, the expectation must be defined as a function of the chosen values of both of those variables, because expected end-of-period value will depend not just on how much the agent saves, but also on how the saved assets are allocated between the risky and riskless assets. Thus we define

$$
\mathrm{v}_{+(T)}(a_t, \varsigma_t) = \mathbb{E}_t[\beta \mathrm{v}_{t+1}(m_{t+1})]
$$

which has derivatives

$$
\begin{aligned}
\mathrm{v}^a_{+(T)} &= \mathbb{E}_t[\beta \mathfrak{R}_{t+1} \mathrm{v}^m_{t+1}(m_{t+1})] = \mathbb{E}_t[\beta \mathfrak{R}_{t+1} \mathrm{u}^c_{t+1}(\mathrm{c}_{t+1}(m_{t+1}))] \\
\mathrm{v}^\varsigma_{+(T)} &= \mathbb{E}_t[\beta(\mathbf{R}_{t+1} - \mathsf{R})\mathrm{v}^m_{t+1}(m_{t+1})]a_t = \mathbb{E}_t[\beta(\mathbf{R}_{t+1} - \mathsf{R})\mathrm{u}^c_{t+1}(\mathrm{c}_{t+1}(m_{t+1}))]a_t
\end{aligned}
$$

implying that the first order conditions (66) and (67) can be rewritten

$$
\begin{aligned}
\mathrm{u}^c(c_t) &= \mathrm{v}^a_{+(T)}(m_t - c_t, \varsigma_t) \\
0 &= \mathrm{v}^\varsigma_{\bar{t}}(a_t, \varsigma_t).
\end{aligned} \tag{67}
$$

7.2 Application

Our first step is to specify the stochastic process for $\mathbf{R}_{t+1}$. We follow the common practice of assuming that returns are lognormally distributed, $\log \mathbf{R} \sim \mathcal{N}(\phi + \mathsf{r} - \sigma^2_\phi/2, \sigma^2_\phi)$ where ϕ is the equity premium over the returns r available on the riskless asset.[16]

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

[16]This guarantees that $\mathbb{E}[\mathbf{R}] = \boldsymbol{\varphi}$ is invariant to the choice of σ^2_ϕ; see LogELogNorm.

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 shock to labor income (for example, in recessions the stock market falls and labor income also declines).

The direct expressions for the derivatives of $\mathrm{v}_{+(T)}$ are

$$
\begin{aligned}
\mathrm{v}^{a}_{+(T)}(a_t, \varsigma_t) &= \beta\left(\frac{1}{n_{\mathbf{r}} n_\theta}\right) \sum_{i=1}^{n_\theta} \sum_{j=1}^{n_{\mathbf{r}}} \mathfrak{R}_{i,j} \left(\mathrm{c}_{t+1}(\mathfrak{R}_{i,j} a_t + \theta_i)\right)^{-\rho} \\
\mathrm{v}^{\varsigma}_{+(T)}(a_t, \varsigma_t) &= \beta\left(\frac{1}{n_{\mathbf{r}} n_\theta}\right) \sum_{i=1}^{n_\theta} \sum_{j=1}^{n_{\mathbf{r}}} (\mathbf{R}_{i,j} - \mathsf{R}) \left(\mathrm{c}_{t+1}(\mathfrak{R}_{i,j} a_t + \theta_i)\right)^{-\rho}.
\end{aligned}
\tag{68}
$$

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 $n_{\mathbf{r}} \times n$ 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 derivatives of the $\mathrm{v}_{+(T)}$ function.

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

$$
\begin{aligned}
c_t^{-\rho} &= \hat{\mathrm{v}}^{a}_{\bar{t}}(m_t - c_t, \varsigma_t) \\
0 &= \hat{\mathrm{v}}^{\varsigma}_{\bar{t}}(m_t - c_t, \varsigma_t)
\end{aligned}
\tag{69}
$$

simultaneously for c and ς at the set of potential m_t values defined in `mVec`. However, multidimensional constrained maximization problems are difficult and sometimes quite slow to solve. There is a better way. Define the problem

$$
\begin{aligned}
\tilde{\mathrm{v}}_{\bar{t}}(a_t) &= \max_{\varsigma_t} \quad \mathrm{v}_{+(T)}(a_t, \varsigma_t) \\
&\text{s.t.} \\
0 \leq &\varsigma_t \leq 1
\end{aligned}
$$

where the tilde over $\tilde{\mathrm{v}}(a)$ 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 $\grave{\tilde{\mathrm{v}}}^{a}_{t}(a_t)$.[17] With this function in hand, we can use the first order condition from the single-control problem

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

[17]A faster solution could be obtained by, for each element in `aVec`, computing $\mathrm{v}^{\varsigma}_{+(T)}(m_t - c_t, \varsigma)$ 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.

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.

7.3 Implementation

Following the discussion from section 7.1, 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." Inheriting the structure of Python, we establish a new subclass of `gothic_class` called `GothicMC`. This subclass preserves the fundamental structure required to resolve the original problem while adding new methods that capture the previously mentioned points. The essential functions in this new class are found in the final four functions that account for the expected marginal value functions with respect to each of the control variables, both for the terminal period and all earlier periods.

Having the `GothicMC` subclass available, we can proceed with implementing the steps laid out in section 7.2 to solve the problem at hand. Initially, the two distributions that capture the uncertainty faced by consumers in this scenario are discretized. Subsequently, the `GothicMC` class is invoked with the requisite arguments to create an instance that includes the necessary functions to depict the first-order conditions of the consumer's problem. Following that, an improved grid of end-of-period assets is established.

Here is where we can see how the approach described in section 7.2 is reflected in the code. For the terminal period, the optimal share of risky assets is determined for each point in `aVec_eee`, and then the endogenous gridpoints method is employed to compute the optimal consumption level given that the share in the risky asset has been chosen optimally. It's worth noting that this solution takes into account the possibility of a binding artificial borrowing constraint. Lastly, the interpolation process is executed for both the optimal consumption function and the optimal share of the portfolio in risky assets. These values are stored in their respective dictionaries (`mGridPort_life`, `cGridPort_life`, and `varsigmaGrid_life`) and utilized to conduct the recursive process outlined in section 6, thus yielding the numerical solution for all earlier periods.

7.4 Results

Figure 20 plots the first-period consumption function generated by the program; qualitatively it does not look much different from the consumption functions generated by the program without portfolio choice. Figure 21 plots the optimal portfolio share as a

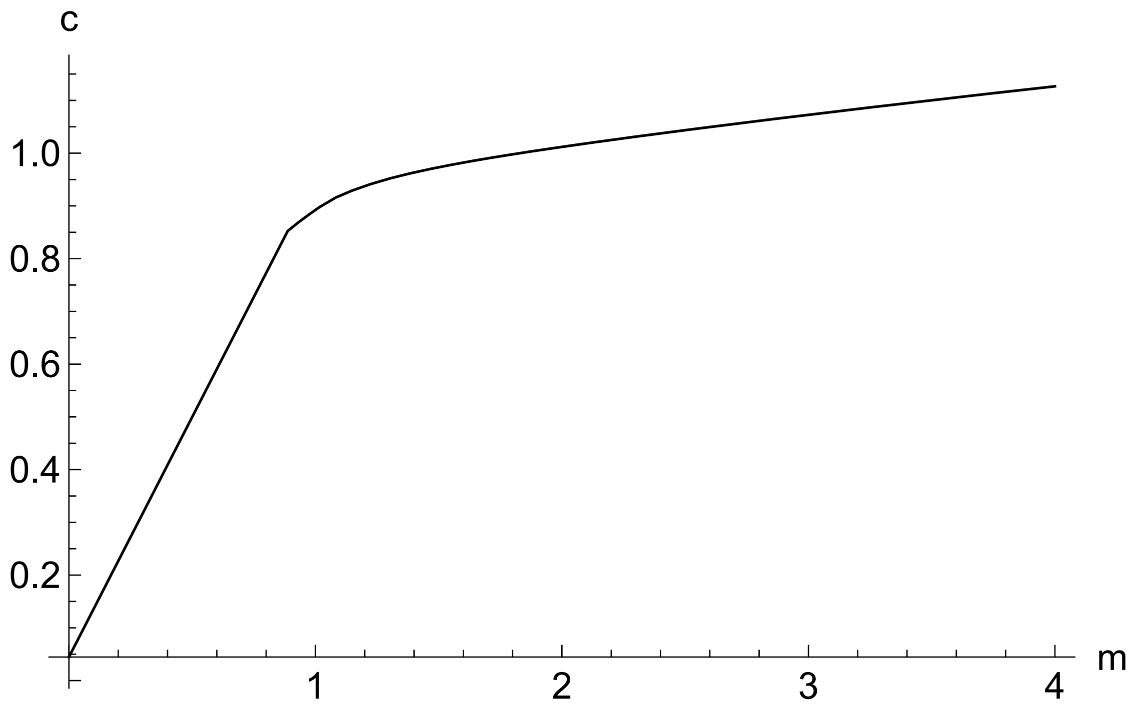

Figure 20 c(m_1) With Portfolio Choice

function of the level of assets. This figure 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) prediction reflects the nature of the risks the consumer faces. Those consumers who are poor in measured financial wealth are likely to 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 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.

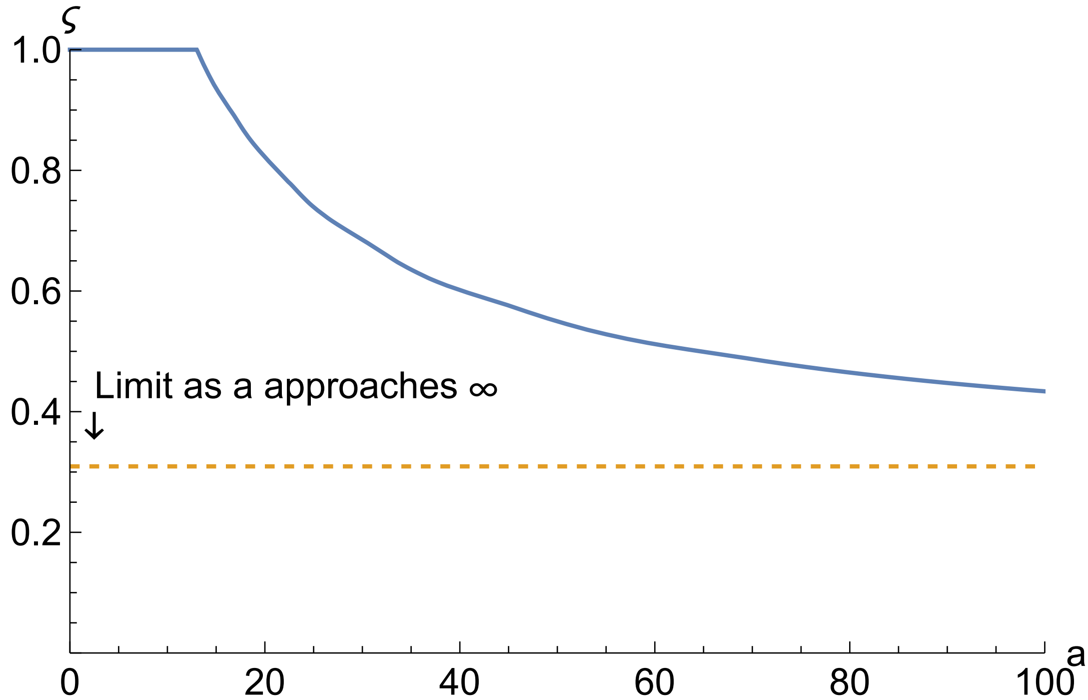

Figure 21 Portfolio Share in Risky Assets in First Period $\varsigma(a)$

8 The-Infinite-Horizon

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 19. 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.

8.1 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

$$\mathbb{E}_t[m_{t+1}/m_t] = 1 \text{ if } m_t = \hat{m} \tag{70}$$

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 measures 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.

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).[18] 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 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 $\beth$ (the Hebrew cognate of β) for the 'pure' time preference factor, the value function for the revised problem is

$$\mathrm{v}_t(\mathbf{p}_t, \mathbf{m}_t) = \max_{\{\mathrm{c}\}_t^T} \quad \mathrm{u}(\mathbf{c}_t) + \mathbb{E}_t \left[\sum_{n=1}^{T-t} \beth^n \mathcal{L}_t^{t+n} \hat{\beta}_t^{t+n} \mathrm{u}(\mathbf{c}_{t+n}) \right] \tag{71}$$

[18]Similar structural estimation exercises have been also performed by Palumbo (1999) and Gourinchas and Parker (2002).

subject to the constraints

$$
\begin{aligned}
\mathbf{a}_t &= \mathbf{m}_t - \mathbf{c}_t \\
\mathbf{p}_{t+1} &= \mathcal{G}_{t+1}\mathbf{p}_t\Psi_{t+1} \\
\mathbf{y}_{t+1} &= \mathbf{p}_{t+1}\theta_{t+1} \\
\mathbf{m}_{t+1} &= \mathsf{R}\mathbf{a}_t + \mathbf{y}_{t+1}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{L}_t^{t+n} &: \text{probability to } \mathcal{L}\text{ive until age } t+n \text{ given alive at age } t \\
\hat{\beta}_t^{t+n} &: \text{age-varying discount factor between ages } t \text{ and } t+n \\
\Psi_t &: \text{mean-one shock to permanent income} \\
\beth &: \text{time-invariant `pure' discount factor}
\end{aligned}
$$

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:

$$
\Xi_s = \begin{cases} 0 & \text{with probability } \wp > 0 \\ \theta_s/\wp & \text{with probability } (1-\wp)\text{, where } \log\theta_s \sim \mathcal{N}(-\sigma_\theta^2/2, \sigma_\theta^2) \end{cases} \tag{72}
$$

$$
\log \boldsymbol{\Psi}_s \sim \mathcal{N}(-\sigma_{\boldsymbol{\Psi}}^2/2, \sigma_{\boldsymbol{\Psi}}^2)
$$

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 $\sigma_{\boldsymbol{\Psi}} = 0.1$.[19] 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 22).[20] The interest rate is assumed to equal 1.03. The model parameters are included in Table 1.

Table 1 Parameter Values

σ_θ	0.1	Carroll (1992)
$\sigma_{\boldsymbol{\Psi}}$	0.1	Carroll (1992)
$\wp$	0.005	Carroll (1992)
$\mathcal{G}_s$	figure 22	Carroll (1997)
$\hat{\beta}_s, \mathcal{L}_s$	figure 22	Cagetti (2003)
R	1.03	Cagetti (2003)

[19] 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.

[20] 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).

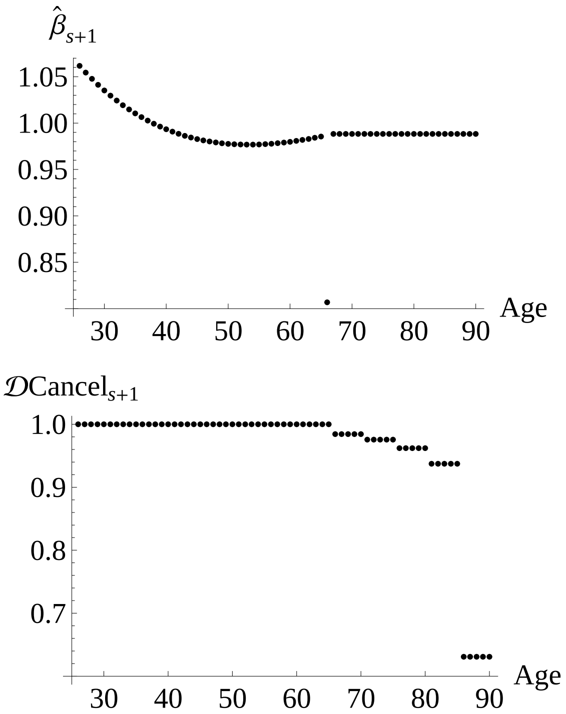

Figure 22 Time Varying Parameters

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 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 $56-60$.[21] 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

[21] 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.

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 model which can be written in Bellman form as:

$$
\begin{aligned}
\mathrm{v}_t(m_t) &= \max_{c_t} \quad \mathrm{u}(c_t) + \daleth \mathcal{L}_{t+1} \hat{\beta}_{t+1} \mathbb{E}_t[(\boldsymbol{\Psi}_{t+1}\mathcal{G}_{t+1})^{1-\rho} \mathrm{v}_{t+1}(m_{t+1})] \\
&\text{s.t.} \\
a_t &= m_t - c_t \\
m_{t+1} &= a_t \underbrace{\left(\frac{\mathsf{R}}{\boldsymbol{\Psi}_{t+1}\mathcal{G}_{t+1}}\right)}_{\equiv \mathcal{R}_{t+1}} + \theta_{t+1}
\end{aligned}
$$

with the first order condition:

$$
\mathrm{u}^c(c_t) = \daleth \mathcal{L}_{t+1} \hat{\beta}_{t+1} \mathsf{R}\, \mathbb{E}_t \left[\mathrm{u}^c \left(\boldsymbol{\Psi}_{t+1}\mathcal{G}_{t+1} \mathrm{c}_{t+1} \left(a_t \mathcal{R}_{t+1} + \theta_{t+1}\right)\right)\right]. \tag{73}
$$

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 (73) following the procedure in sections 5 and 6. 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

$$\min_{\xi} \sum_{\tau=1}^{7} |\varsigma^{\tau} - \mathbf{s}^{\tau}(\xi)| \tag{74}$$

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 (74) 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:

$$\min_{\xi} \sum_{i}^{N} \omega_i \, |\varsigma_i^{\tau} - \mathbf{s}^{\tau}(\xi)| \tag{75}$$

where ω_i is the weight of household i in the entire population,[22] 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 23. 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.

[22]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.

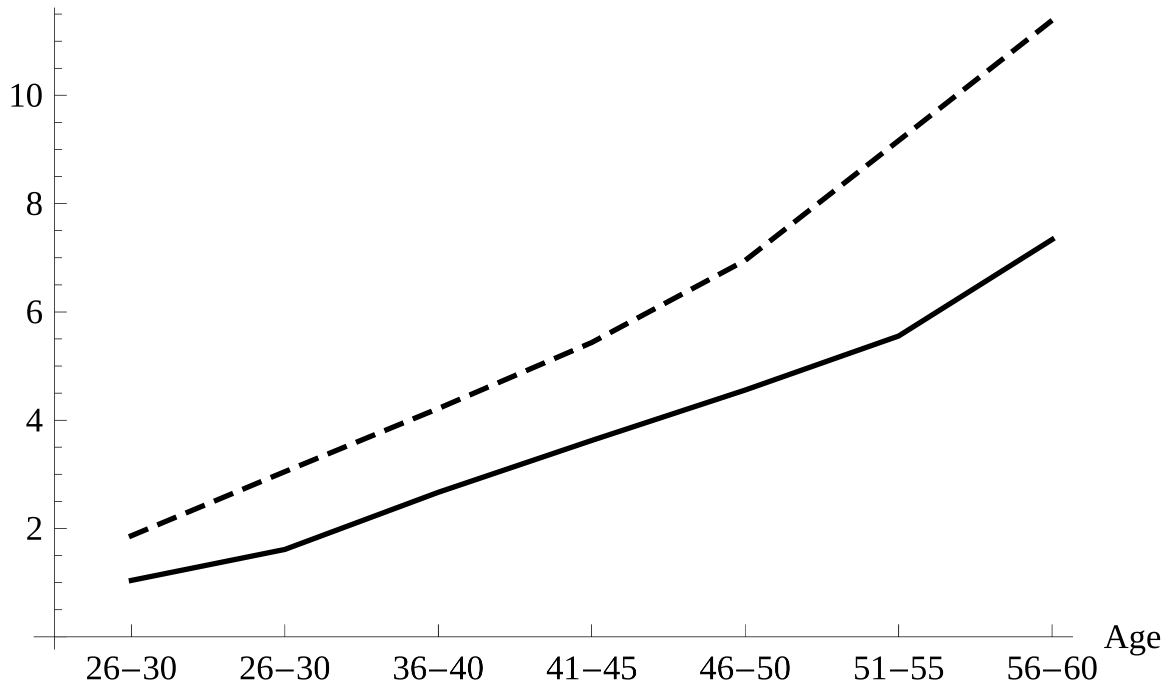

Figure 23 Wealth to Permanent Income Ratios from SCF (means (dashed) and medians (solid))

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, `smmObjectiveFxn`, calls the `simulate_moments` function to create the objective function described in (75), 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 (75)

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 time demanding 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.[23] This involves:

1. drawing new shocks for the simulation
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

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 (75). 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.[24] Results are reported in Table 2.[25]

The literature on consumption and savings behavior over the lifecycle in the presenece of labor income uncertainty [26] warns us to be careful in disentangling the effect of time

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

[24] The procedure is: First we calculate the ρ and $\beth$ estimates as the minimizer of equation (75) using the actual SCF data. Then, we apply the `Bootstrap` function several times to obtain the standard error of our estimates.

[25] 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.

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

Table 2 Estimation Results

ρ	$\beth$
3.69	0.88
(0.047)	(0.002)

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 24 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 25 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 (75)). 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.

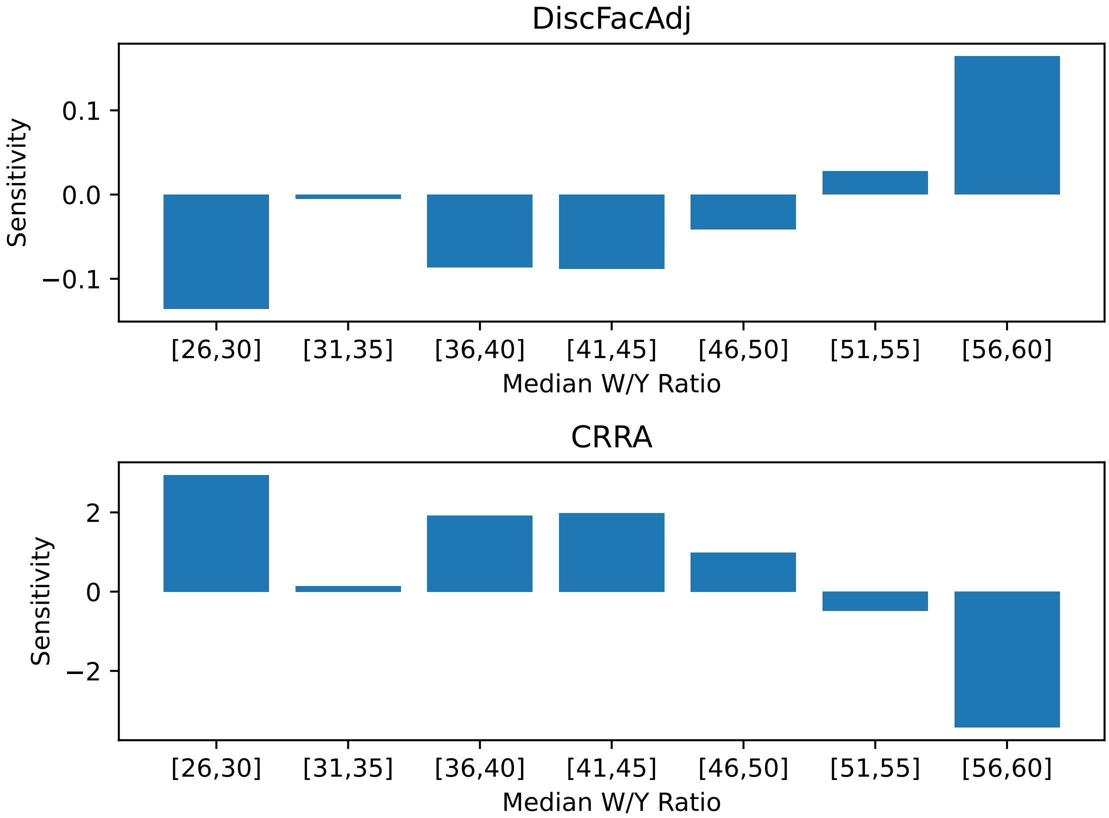

Figure 24 Sensitivty of Estimates $\{\rho, \beth\}$ regarding Alternative Mismeasurement Assumptions.

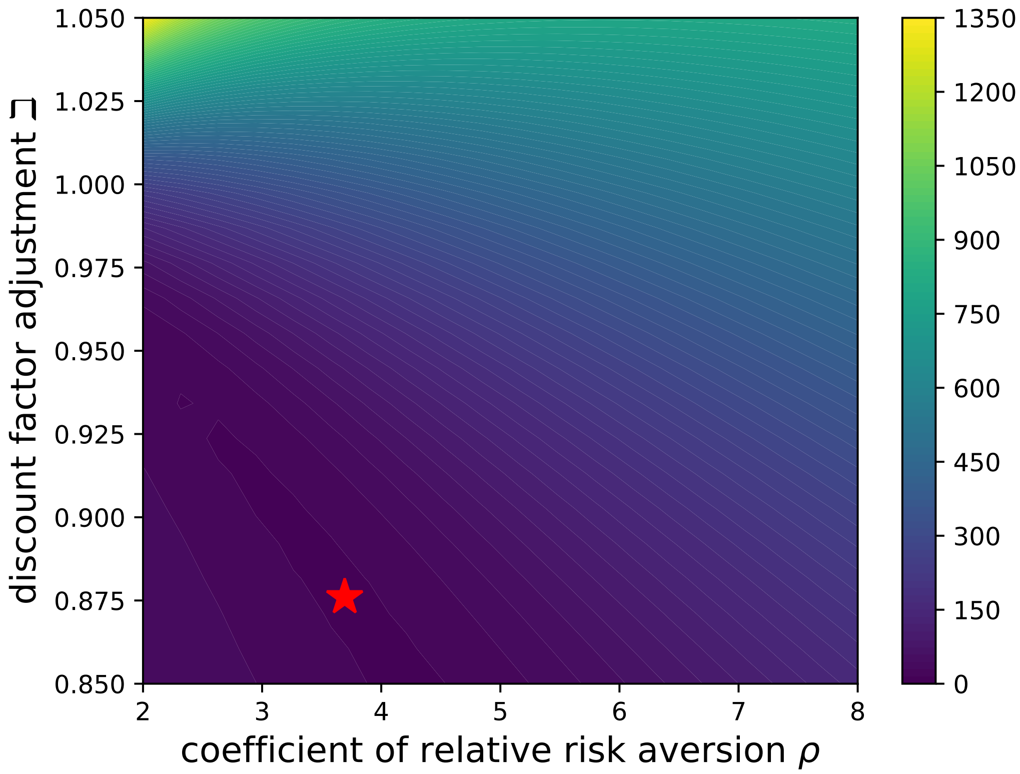

Figure 25 Contour Plot (larger values are shown lighter) with $\{\rho, \beth\}$ Estimates (red dot).

10 Conclusion

There are many choices that can be made for solving microeconomic dynamic stochastic optimization problems. The set of techniques, and associated programs, described in these notes represents 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 Further Details on 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).[27] 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. This 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.[28]

[27] 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.

[28] 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.

——— (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.

Valencia, Fabian (2006): "Banks' Financial Structure and Business Cycles," Ph.D. thesis, Johns Hopkins University.