Solution Methods for Microeconomic Dynamic Stochastic Optimization Problems

2023-01-29

Christopher D. Carroll[1]

Note: The code associated with this document should work (though the Matlab code may be out of date), but has been superceded by the set of tools available in the Econ-ARK toolkit, more specifically the HARK Framework. The SMM estimation code at the end has specifically been superceded by the SolvingMicroDSOPs REMARK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* **?** and Cagetti (**?**). 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 **?**, 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.

Contents

1 Introduction

Calculating the mathematically optimal amount to save is remarkably difficult. Under well-founded assumptions about the nature of risk (and attitudes toward risk), the problem cannot be solved analytically; computational solutions are the only option. To avoid having to solve this hard problem, past generations of economists showed impressive ingenuity in reformulating the question. Budding graduate students are still taught a host of tricks whose purpose is partly to avoid the resort to numerical solutions: Quadratic or Constant Absolute Risk Aversion utility, perfect markets, perfect insurance, perfect foresight, the "timeless" perspective, the restriction of uncertainty to very special kinds,[1] and more.

The motivation for these reformulations is to exchange an intractable general problem for a tractable specific alternative. Unfortunately, the burgeoning literature on numerical solutions has shown that the features that yield tractability also profoundly change the solution. These tricks are excuses to solve a problem that has defined away the central difficulty: Understanding the proper role of uncertainty (and other complexities like constraints) in optimal intertemporal choice.

The temptation to use such tricks (and the tolerance for them in leading academic journals) is palpably lessening, thanks to advances in mathematical analysis, increasing computing power, and the growing capabilities of numerical computation software. Together, such tools permit today's laptop computers to solve problems that required supercomputers a decade ago (and, before that, could not be solved at all).

These points are not unique to the consumption/saving problem; the same propositions apply to almost any question that involves both intertemporal choice and uncertainty, including many aspects of the behavior of firms and governments.

Given the ubiquity of such problems, one might expect that the use of numerical methods for solving dynamic optimization problems would by now be nearly as common as the use of econometric methods in empirical work.

Of course, we remain far from that equilibrium. The most plausible explanation for the gap is that barriers to the use of numerical methods have remained forbiddingly high.

These lecture notes provide a gentle introduction to a particular set of solution tools and show how they can be used to solve some canonical problems in consumption choice and portfolio allocation. Specifically, the notes describe and solve optimization problems for a consumer facing uninsurable idiosyncratic risk to nonfinancial income (e.g., labor or transfer income),[2] 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

[1]E.g., lognormally distributed rate-of-return risk – but no labor income risk – under CRRA utility (the **?**-**?** model).

[2]Expenditure shocks (such as for medical needs, or to repair a broken automobile) are usually treated in a manner similar to labor income shocks. See **?** and **?** 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.

is described and solved, an extension with portfolio choice between a riskless and a risky asset is also solved. Finally, a simple example is presented of 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 (**?**) and Cagetti (**?**)).

2 The Problem

We are interested in the behavior a consumer whose goal in period t is to maximize expected discounted utility from consumption over the remainder of a lifetime that ends in period T:

$$\max\ \mathbb{E}_t\left[\sum_{n_{\boldsymbol{\theta}}=0}^{T-t}\beta^{n_{\boldsymbol{\theta}}}\mathrm{u}(\mathbf{c}_{t+n})\right], \tag{1}$$

and whose circumstances evolve according to the transition equations[3]

$$\begin{aligned}
\mathbf{a}_t &= \mathbf{m}_t - \mathbf{c}_t \\
\mathbf{b}_{t+1} &= \mathbf{a}_t\mathsf{R}_{t+1} \\
\mathbf{y}_{t+1} &= \mathbf{p}_{t+1}\boldsymbol{\theta}_{t+1} \\
\mathbf{m}_{t+1} &= \mathbf{b}_{t+1} + \mathbf{y}_{t+1}
\end{aligned} \tag{2}$$

where the variables are

$\beta-$	pure time discount factor
$\mathbf{a}_t-$	assets after all actions have been accomplished in period t
$\mathbf{b}_{t+1}-$	'bank balances' (nonhuman wealth) at the beginning of $t+1$
$\mathbf{c}_t-$	consumption in period t
$\mathbf{m}_t-$	'market resources' available for consumption ('cash-on-hand')
$\mathbf{p}_{t+1}-$	'permanent labor income' in period $t+1$
$\mathsf{R}_{t+1}-$	interest factor $(1+\mathsf{r}_{t+1})$ from period t to $t+1$
$\mathbf{y}_{t+1}-$	noncapital income in period $t+1$.

For now, we will assume that the exogenous variables evolve as follows:

$\mathsf{R}_t = \mathsf{R}\ \forall\ t$	- constant interest factor $= 1+\mathsf{r}$
$\mathbf{p}_{t+1} = \boldsymbol{\Phi}_{t+1}\mathbf{p}_t$	- permanent labor income dynamics
$\log\ \boldsymbol{\theta}_{t+n} \sim \mathcal{N}(-\sigma_{\boldsymbol{\theta}}^2/2, \sigma_{\boldsymbol{\theta}}^2)$	- lognormal transitory shocks $\forall\ n>0$.

Using the fact about lognormally distributed variables `ELogNorm`[4] that if $\log\boldsymbol{\varphi}\ \sim$

[3]The usual analysis of dynamic programming problems combines the equations below into a single expression; here, they are disarticulated to highlight the important point that several distinct processes (intertemporal choice, stochastic shocks, intertemporal returns, income growth) are involved in the transition from one period to the next.

[4]This fact is referred to as `ELogNorm` in the handout `MathFactsList`, in the references as **?**; further citation to facts in that handout will be referenced simply by the name used in the handout for the fact in question, e.g. `LogELogNorm` is the name of the fact that implies that $\log\mathbb{E}[\boldsymbol{\theta}]=0$.

$\mathcal{N}(\varphi, \sigma_\varphi^2)$ then $\log \mathbb{E}[\boldsymbol{\varphi}] = \varphi + \sigma_\varphi^2/2$, assumption the assumption about the distribution of shocks guarantees that $\log \mathbb{E}[\boldsymbol{\theta}] = 0$ which means that $\mathbb{E}[\boldsymbol{\theta}]$=1 (the mean value of the transitory shock is 1).

Equation (3) indicates that we are allowing for a predictable average profile of income growth over the lifetime $\{\boldsymbol{\Phi}\}_0^T$ (allowing, for example, for typical career wage paths).[5]

Finally, the utility function is of the Constant Relative Risk Aversion (CRRA), form, $\mathrm{u}(\bullet) = \bullet^{1-\rho}/(1-\rho)$.

As is well known, this problem can be rewritten in recursive (Bellman equation) form

$$\mathbf{v}_t(\mathbf{m}_t, \mathbf{p}_t) = \max_{\mathbf{c}_t} \ \mathrm{u}(\mathbf{c}_t) + \mathbb{E}_t[\beta \mathbf{v}_{t+1}(\mathbf{m}_{t+1}, \mathbf{p}_{t+1})] \tag{3}$$

subject to the Dynamic Budget Constraint (DBC) (2) given above, where $\mathbf{v}_t$ measures total expected discounted utility from behaving optimally now and henceforth.

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 possible). In the consumption problem 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, implying that

$$\mathbf{v}_T(\mathbf{m}_T, \mathbf{p}_T) = \frac{\mathbf{m}_T^{1-\rho}}{1-\rho}. \tag{4}$$

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 $\mathrm{v}_T(m_T) = \mathrm{u}(m_T)$.[6] For our CRRA utility function, $\mathrm{u}(xy) = x^{1-\rho}\mathrm{u}(y)$, so equation (4) can be rewritten as

$$\mathbf{v}_T(\mathbf{m}_T, \mathbf{p}_T) = \mathbf{p}_T^{1-\rho} \frac{m_T^{1-\rho}}{1-\rho} = \mathbf{p}_{T-1}^{1-\rho} \boldsymbol{\Phi}_T^{1-\rho} \frac{m_T^{1-\rho}}{1-\rho} = \mathbf{p}_{T-1}^{1-\rho} \boldsymbol{\Phi}_T^{1-\rho} \mathrm{v}_T(m_T).$$

[5]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 **?**.

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

Now define a new optimization problem:

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

The accumulation equation is the normalized version of the transition equation for $\mathbf{m}_{t+1}$.[7] 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 (5) which has only a single state variable m_t, we can obtain the levels of the value function, consumption, and all other variables of interest simply by multiplying the results by the appropriate function of $\mathbf{p}_t$, e.g. $\mathbf{c}_t(\mathbf{m}_t, \mathbf{p}_t) = \mathbf{p}_t \mathrm{c}_t(\mathbf{m}_t/\mathbf{p}_t)$ or $\mathbf{v}_t(\mathbf{m}_t, \mathbf{p}_t) = \mathbf{p}_t^{1-\rho} \mathrm{v}_t(m_t)$. 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, **?** shows how a bank's optimization problem can be normalized by the level of the bank's productivity.

4 The Usual Theory, and A Bit More Notation

The first order condition for (5) with respect to c_t is

$$\begin{aligned} \mathrm{u}'(c_t) &= \mathbb{E}_t[\beta \mathcal{R}_{t+1} \boldsymbol{\Phi}_{t+1}^{1-\rho} \mathrm{v}'_{t+1}(m_{t+1})] \\ &= \mathbb{E}_t[\beta \mathsf{R} \quad \boldsymbol{\Phi}_{t+1}^{-\rho} \mathrm{v}'_{t+1}(m_{t+1})] \end{aligned}$$

and because the `Envelope` theorem tells us that

$$\mathrm{v}'_t(m_t) = \mathbb{E}_t[\beta \mathsf{R} \boldsymbol{\Phi}_{t+1}^{-\rho} \mathrm{v}'_{t+1}(m_{t+1})] \tag{5}$$

we can substitute the LHS of (5) for the RHS of (5) to get

$$\mathrm{u}'(c_t) = \mathrm{v}'_t(m_t) \tag{6}$$

and rolling this equation forward one period yields

$$\mathrm{u}'(c_{t+1}) = \mathrm{v}'_{t+1}(a_t \mathcal{R}_{t+1} + \boldsymbol{\theta}_{t+1}) \tag{7}$$

[7] Derivation:

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

while substituting the LHS in equation (5) gives us the Euler equation for consumption

$$\mathrm{u}'(c_t) = \mathbb{E}_t[\beta \mathsf{R} \mathbf{\Phi}_{t+1}^{-\rho} \mathrm{u}'(c_{t+1})]. \tag{8}$$

Now note that in equation (7) neither m_t nor c_t has any *direct* effect on v_{t+1} - only the difference between them (i.e. unconsumed market resources or 'assets' a_t) matters. It is therefore possible (and will turn out to be convenient) to define a function[8]

$$\mathfrak{v}_t(a_t) = \mathbb{E}_t[\beta \mathbf{\Phi}_{t+1}^{1-\rho} \mathrm{v}_{t+1}(\mathcal{R}_{t+1} a_t + \boldsymbol{\theta}_{t+1})] \tag{9}$$

that returns the expected $t+1$ value associated with ending period t with any given amount of assets. This definition implies that

$$\mathfrak{v}_t'(a_t) = \mathbb{E}_t[\beta \mathsf{R} \mathbf{\Phi}_{t+1}^{-\rho} \mathrm{v}_{t+1}'(\mathcal{R}_{t+1} a_t + \boldsymbol{\theta}_{t+1})] \tag{10}$$

or, substituting from equation (7),

$$\mathfrak{v}_t'(a_t) = \mathbb{E}_t\left[\beta \mathsf{R} \mathbf{\Phi}_{t+1}^{-\rho} \mathrm{u}'\left(\mathrm{c}_{t+1}(\mathcal{R}_{t+1} a_t + \boldsymbol{\theta}_{t+1})\right)\right]. \tag{11}$$

Finally, note for future use that the first order condition (5) can now be rewritten as

$$\mathrm{u}'(c_t) = \mathfrak{v}_t'(m_t - c_t). \tag{12}$$

5 Solving the Next-to-Last Period

The problem in the second-to-last period of life is:

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

and using (1) the fact that $\mathrm{v}_T = \mathrm{u}(c)$; (2) the definition of $\mathrm{u}(c)$; (3) the definition of the expectations operator, and (4) the fact that $\mathbf{\Phi}_T$ is nonstochastic, this becomes

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

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

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, however, there is no general analytical solution to this maximization problem, and so 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, the integral must be calculated numerically, and numerical integration is *very* slow.

[8] The peculiar letter designating our new function is pronounced 'Gothic v'. Letters in this font will be used for end-of-period quantities.

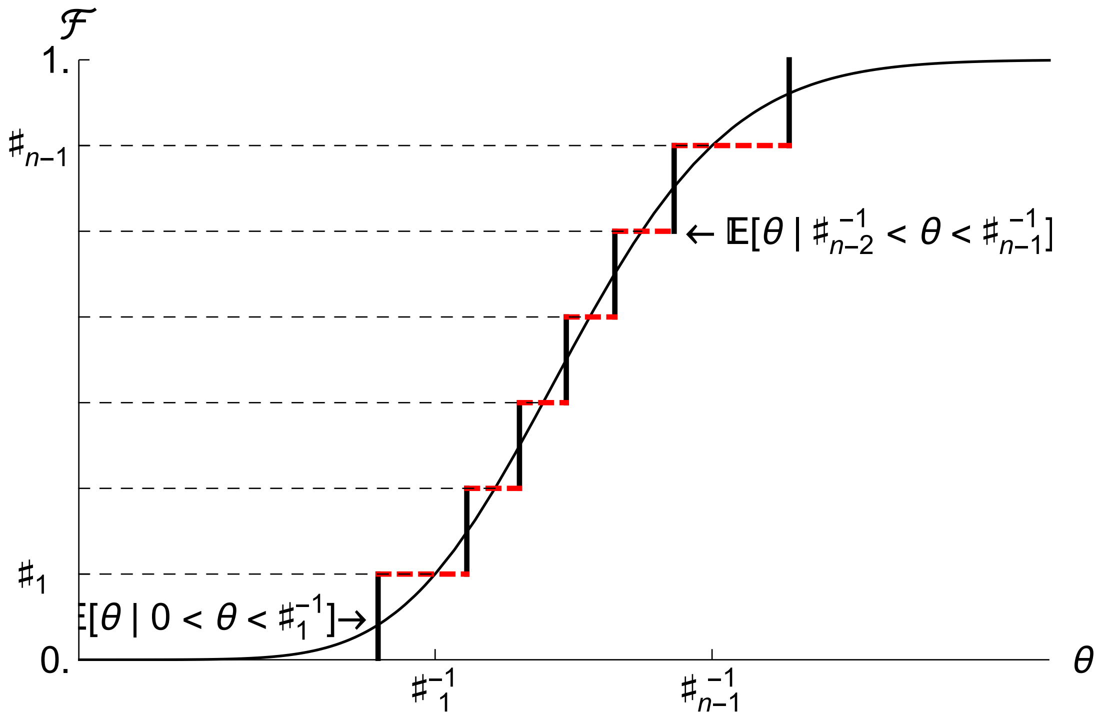

Figure 1 Discrete Approximation to Lognormal Distribution $\mathcal{F}$

5.1 Discretizing the Distribution

Our first time-saving step is therefore to construct a discrete approximation to the lognormal distribution that can be used in place of numerical integration. We calculate an n-point approximation as follows.

Define a set of points from $\sharp_0$ to $\sharp_{n_{\boldsymbol{\theta}}}$ on the $[0, 1]$ interval as the elements of the set $\sharp = \{0, 1/n, 2/n, \ldots, 1\}$.[9] Call the inverse of the $\boldsymbol{\theta}$ distribution $\mathcal{F}^{-1}$, and define the points $\sharp_i^{-1} = \mathcal{F}^{-1}(\sharp_i)$. Then the conditional mean of $\boldsymbol{\theta}$ in each of the intervals numbered 1 to n is:

$$\boldsymbol{\theta}_i \equiv \mathbb{E}[\boldsymbol{\theta} | \sharp_{i-1}^{-1} \leq \boldsymbol{\theta} < \sharp_i^{-1}] = \int_{\sharp_{i-1}^{-1}}^{\sharp_i^{-1}} \vartheta \; d\mathcal{F}\,(\vartheta). \tag{13}$$

The method is illustrated in Figure 1. The solid continuous curve represents the "true" CDF $\mathcal{F}\,(\boldsymbol{\theta})$ for a lognormal distribution such that $\mathbb{E}[\boldsymbol{\theta}] = 1$, $\sigma_{\boldsymbol{\theta}} = 0.1$. The short vertical line segments represent the $n_{\boldsymbol{\theta}}$ equiprobable values of $\boldsymbol{\theta}_i$ which are used to approximate this distribution.[10]

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

[10] More sophisticated approximation methods exist (e.g. Gauss-Hermite quadrature; see **?** 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.

Recalling our definition of $\mathfrak{v}_t(a_t)$, for $t = T-1$

$$\mathfrak{v}_{T-1}(a_{T-1}) = \beta \boldsymbol{\Phi}_T^{1-\rho} \left(\frac{1}{n_{\boldsymbol{\theta}}}\right) \sum_{i=1}^{n_{\boldsymbol{\theta}}} \frac{(\mathcal{R}_T a_{T-1} + \boldsymbol{\theta}_i)^{1-\rho}}{1-\rho} \tag{14}$$

so we can rewrite the maximization problem as

$$\mathrm{v}_{T-1}(m_{T-1}) = \max_{c_{T-1}} \left\{ \frac{c_{T-1}^{1-\rho}}{1-\rho} + \mathfrak{v}_{T-1}(m_{T-1} - c_{T-1}) \right\}. \tag{15}$$

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 (15) in a reasonable amount of time. The *Mathematica* program that solves exactly this problem is called `2period.m`. (The archive also contains parallel Matlab programs, but these notes will dwell on the specifics of the *Mathematica* implementation, which is superior in many respects.)

The first thing `2period.m` does is to read in the file `functions.m` which contains definitions of the consumption and value functions; `functions.m` also defines the function `SolveAnotherPeriod` which, given the existence in memory of a solution for period $t+1$, solves for period t.

The next step is to run the programs `setup_params.m`, `setup_grids.m`, `setup_shocks.m`, respectively. `setup_params.m` sets values for the parameter values like the coefficient of relative risk aversion. `setup_shocks.m` calculates the values for the $\boldsymbol{\theta}_i$ defined above (and puts those values, and the (identical) probability associated with each of them, in the vector variables $\boldsymbol{\theta}$`Vals` and $\boldsymbol{\theta}$`Prob`). Finally, `setup_grids.m` constructs a list of potential values of cash-on-hand and saving, then puts them in the vector variables `mVec` = `aVec` = $\{0, 1, 2, 3, 4\}$ respectively. Then `2period.m` runs the program `setup_lastperiod.m` which defines the elements necessary to determine behavior in the last period, in which $\mathrm{c}_T(m) = m$ and $\mathrm{v}_T(m) = \mathrm{u}(m)$.

After all the setup, the only remaining step in `2period.m` is to invoke `SolveAnotherPeriod`, which constructs the solution for period $T-1$ given the presence of the solution for period T (constructed by `setup_lastperiod.m`).

Because we will always be comparing our solution to the perfect foresight solution, we also construct the variables required to characterize the perfect foresight consumption function in periods prior to T. In particular, we construct the list `yExpPDV` (which contains the PDV of expected income – 'expected human wealth'), and `yMinPDV` which contains the minimum possible discounted value of future income at the beginning of period $T-1$ ('minimum human wealth').[11]

The perfect foresight consumption function is also constructed (`setup_PerfectForesightSolution.m`). This program uses the fact that, in *Mathematica*, functions can be saved as objects using the commands `#` and `&`. The `#` denotes the argument of the function, while the `&`, placed at the end of the function, tells *Mathematica* that the function should be saved

[11]This is useful in determining the search range for the optimal level of consumption in the maximization problem.

as an object. In the program, the last period perfect foresight consumption function is saved as an element in the list `cF = {(# - 1 + Last[yExpPDV]) Last[κMin] &}`, where `Last[yExpPDV]` gives the just-constructed PDV of human wealth at the beginning of T (equal to 1, since current income is included in h_T), and `Last[κMin]` gives the perfect foresight marginal propensity to consume (equal to 1, since it is optimal to spend all resources in the last period). Since `#` in the code stands in for what was called m in the model, the discounted total wealth is decomposed into discounted non-human wealth `# - 1` and discounted human wealth `Last[yExpPDV]`. The resulting formula then corresponds to $\bar{c}_T = (m_T - 1 + h_T)\underline{\kappa}_T$, which translates to $\bar{c}_T = m_T$ for $h_T = \underline{\kappa}_T = 1$.

The infinite horizon perfect foresight marginal propensity to save

$$\lambda = (1/\mathsf{R})(\mathsf{R}\beta)^{1/\rho} \tag{16}$$

is also defined because it will be useful in a number of derivations.[12]

The program then constructs behavior for one iteration back from the last period of life by using the function `AddNewPeriodToParamLifeDates`. Using the *Mathematica* command `AppendTo`, various existing lists (which characterized the solution for period T) are redefined to include an additional element representing the relevant formulas in the second to last period of life. For example, `κMin` now has two elements. The second element, given by `1/(1 + Last[λ]/Last[κMin])`, is the perfect foresight marginal propensity to consume in $t = T - 1$.[13]

Next, the program defines a function `𝔳[at_]` (in `functions_stable.m`) which is the exact implementation of (9): It returns the expectation of the value of behaving optimally in period T given any specific amount of assets at the end of $T - 1$, a_{T-1}.

The heart of the program is the next expression (in `functions.m`). This expression loops over the values of the variable `mVec`, solving the maximization problem (given in equation (15)):

$$\max_{\mathtt{c}} \ \ \mathtt{u[c] + \mathfrak{v}[mVec[[i]]-c]} \tag{17}$$

for each of the i values of `mVec` (henceforth let's call these points $m_{T-1,i}$). The maximization routine returns two values: the maximized value, and the value of c which yields that maximized value. When the loop (the `Table` command) is finished, the variable `vAndcList` contains two lists, one with the values $v_{T-1,i}$ and the other with the consumption levels $c_{T-1,i}$ associated with the $m_{T-1,i}$.

5.3 An Interpolated Consumption Function

Now we use the first of the really convenient built-in features of *Mathematica*. Given a set of points on a function (in this case, the consumption function $\mathrm{c}_{T-1}(m)$), *Mathematica* can create an object called an `InterpolatingFunction` which when applied to an input m will yield the value of c that corresponds to a linear interpolation of the value of c from the points in the `InterpolatingFunction` object. We can therefore

[12] Detailed discussion can be found in Carroll (?).

[13] Carroll (?) shows that this is also a recurring formula that extends inductively to earlier periods.

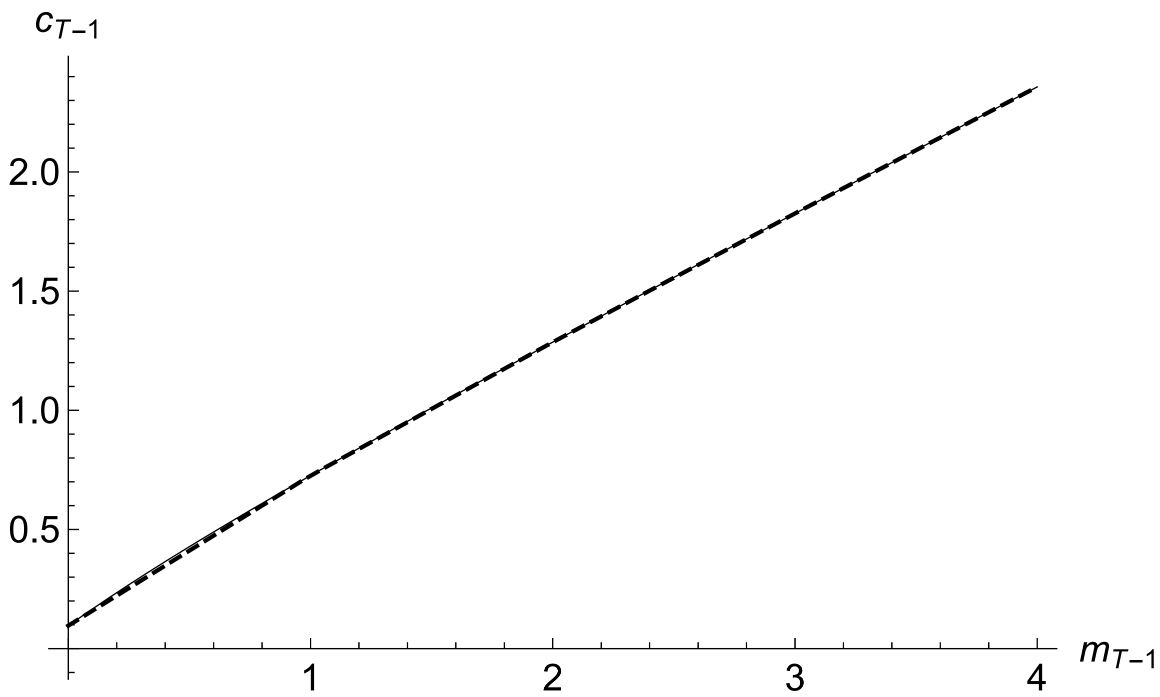

Figure 2 $\mathrm{c}_{T-1}(m_{T-1})$ (solid) versus $\grave{\mathrm{c}}_{T-1}(m_{T-1})$ (dashed)

define an approximation to the consumption function $\grave{\mathrm{c}}_{T-1}(m_{T-1})$ which, when called with an m_{T-1} that is equal to one of the points in `mVec[[i]]` 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[[i]]`, returns the value of c that reflects a linear interpolation between the $c_{T-1,i}$ associated with the two `mVec[[i]]` 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 $\grave{\mathrm{v}}_{T-1}$ `InterpolatingFunctions` that are generated by the program `2PeriodInt.m`. 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).[14]

5.4 Interpolating Expectations

`2period.m` 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:

[14] For some problems, especially ones with discrete choices, value function approximation is unavoidable; nevertheless, even in such problems, the techniques sketched below can be very useful across much of the range over which the problem is defined.

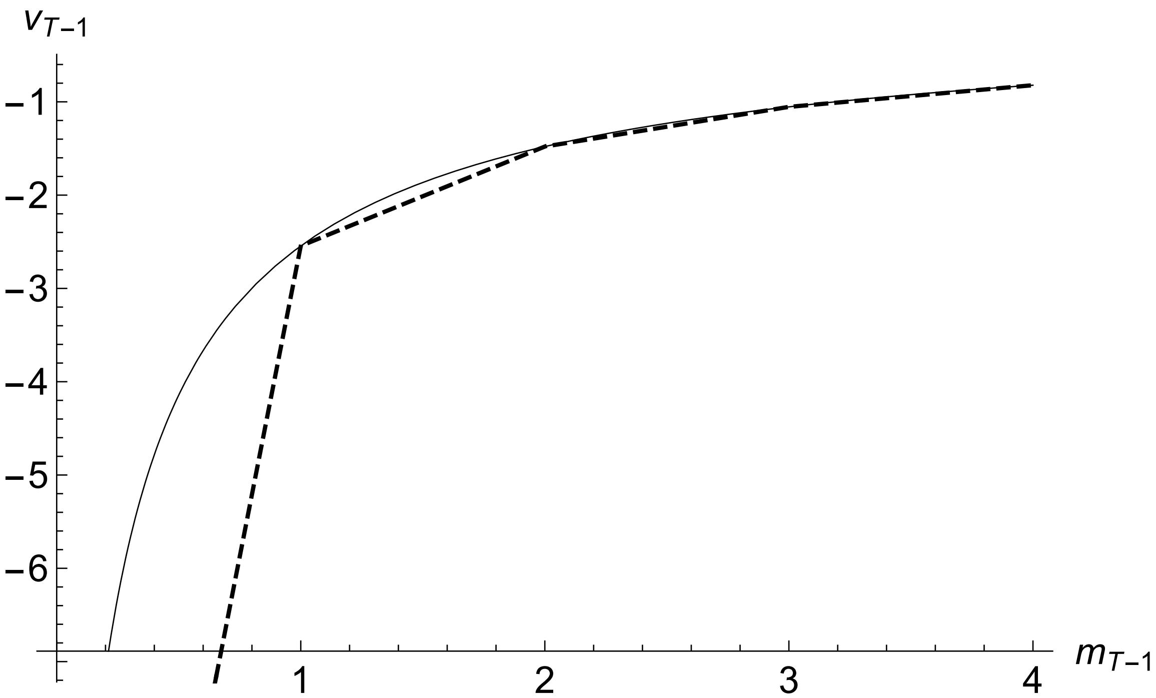

Figure 3 v_{T-1} (solid) versus $\grave{\mathrm{v}}_{T-1}(m_{T-1})$ (dashed)

Since it uses equation (15), 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. But for any given value of a_{T-1}, there is a good chance that the program may end up calculating the corresponding $\mathfrak{v}$ 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.

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 `setup_grids.m`), 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 (9); then construct an `InterpolatingFunction` object $\grave{\mathfrak{v}}_{T-1}(a_{T-1})$ from the list of points on the function captured in the $\vec{a}_{T-1}$ and $\vec{\mathfrak{v}}_{T-1}$ vectors.

Thus, we are now interpolating for the function that reveals the expected value of *ending* the period with a given amount of assets.[15] The program `2periodIntExp.m` solves this problem. Figure 4 compares the true value function to the `InterpolatingFunction` approximation; the functions are of course identical at the gridpoints chosen for a_{T-1} and they appear reasonably close except in the region below $m_{T-1} = 1$.

[15] What we are doing here is closely related to 'the method of parameterized expectations' of ?; the only difference is that our method is essentially a nonparametric version.

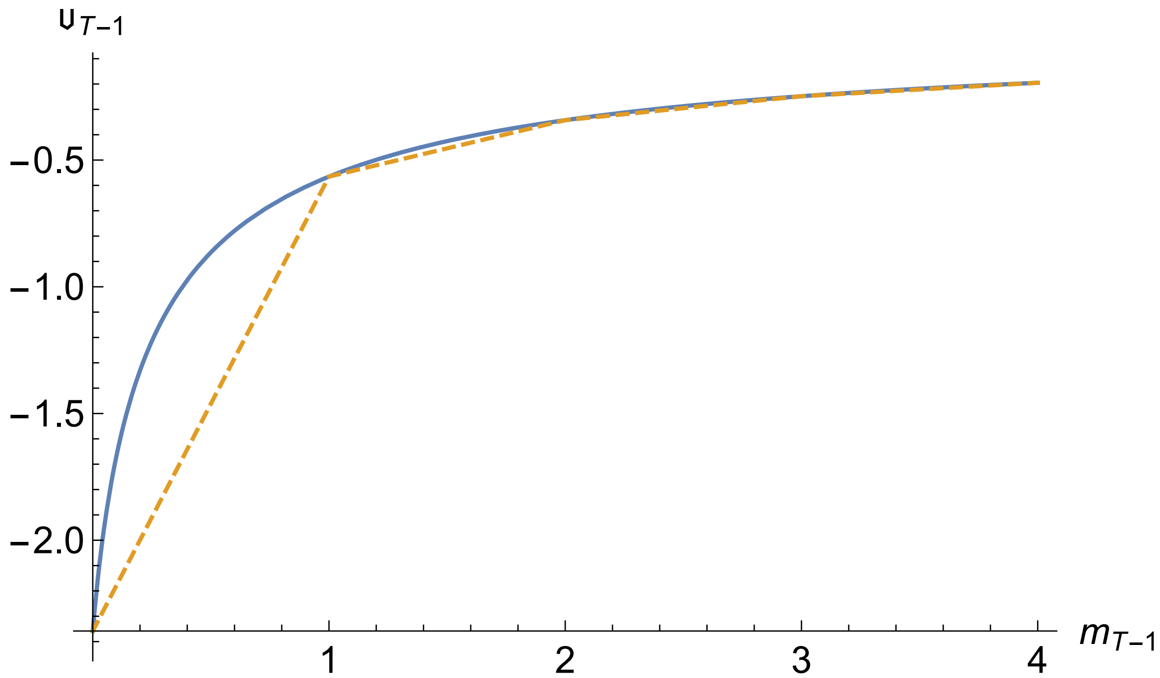

Figure 4 End-Of-Period Value $\mathfrak{v}_{T-1}(a_{T-1})$ (solid) versus $\grave{\mathfrak{v}}_{T-1}(a_{T-1})$ (dashed)

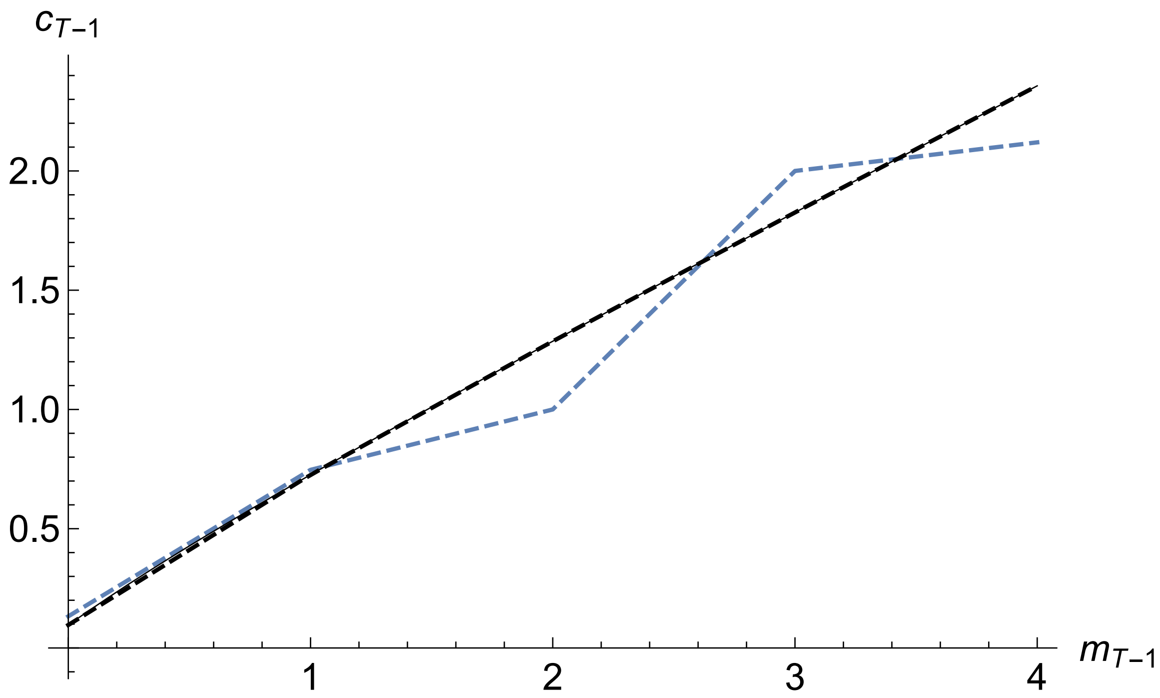

Figure 5 $\mathrm{c}_{T-1}(m_{T-1})$ (solid) versus $\grave{\mathrm{c}}_{T-1}(m_{T-1})$ (dashed)

Nevertheless, the resulting consumption rule obtained when $\grave{\mathfrak{v}}_{T-1}(a_{T-1})$ 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 (**?**) prove 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,

$$\mathbb{E}[-(c - \cancel{c})^2] < -(\mathbb{E}[c] - \cancel{c})^2 \tag{18}$$

where $\cancel{c}$ is the 'bliss point'. 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:

$$\mathbb{E}[-2(c - \cancel{c})] = -2(\mathbb{E}[c] - \cancel{c}). \tag{19}$$

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

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

$$\begin{aligned} \mathrm{u}'(c_{T-1}) &= \beta\, \mathbb{E}_{T-1}[\boldsymbol{\Phi}_T^{-\rho}\mathsf{R}\mathrm{u}'(c_T)] \\ c_{T-1}^{-\rho} &= \mathsf{R}\beta \left(\frac{1}{n_{\boldsymbol{\theta}}}\right) \sum_{i=1}^{n_{\boldsymbol{\theta}}} \boldsymbol{\Phi}_T^{-\rho} \left(\mathsf{R}(m_{T-1} - c_{T-1}) + \boldsymbol{\theta}_i\right)^{-\rho}. \end{aligned} \tag{20}$$

The downward-sloping curve in Figure 6 shows the value of $c_{T-1}^{-\rho}$ for our baseline parameter values for $0 \leq c_{T-1} \leq 4$ (the horizontal axis). The solid upward-sloping curve shows the value of the RHS of (20) 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 (20). The value of c_{T-1} for which the RHS and LHS of (20) 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 (20) under the assumption that $m_{T-1} = 4$, and the intersection of this curve with $\mathrm{u}'(c_{T-1})$ 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.

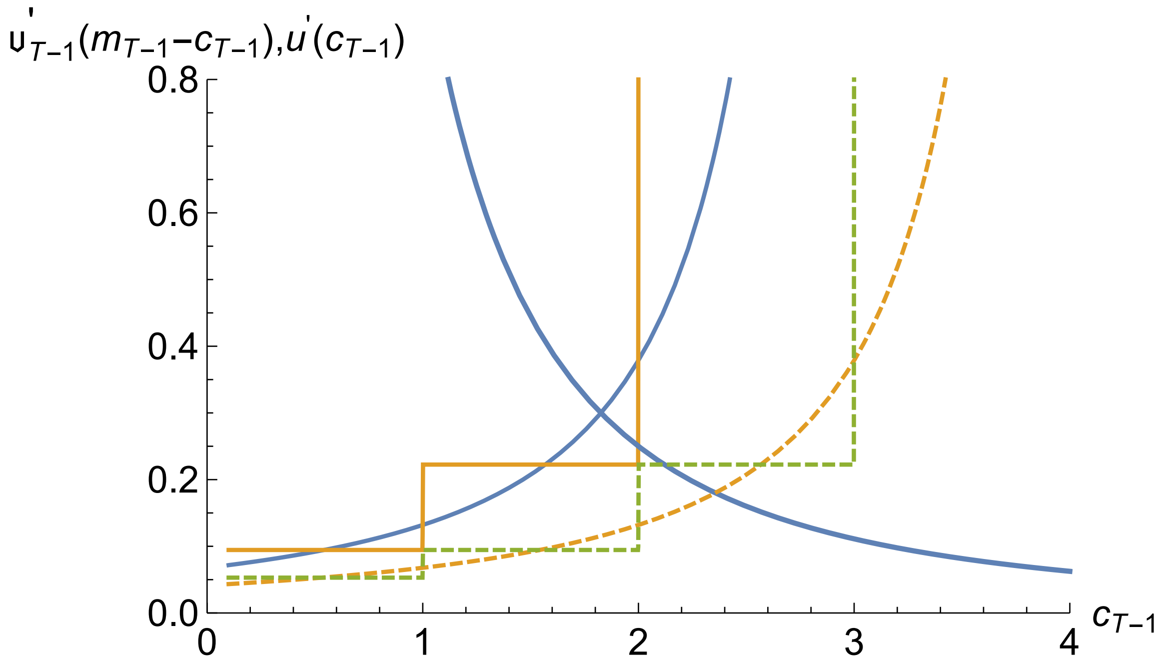

Figure 6 $\mathrm{u}'(c)$ versus $\mathfrak{v}'_{T-1}(3-c), \mathfrak{v}'_{T-1}(4-c), \grave{\mathfrak{v}}'_{T-1}(3-c), \grave{\mathfrak{v}}'_{T-1}(4-c)$

Now consider the derivative of our function $\grave{\mathfrak{v}}_{T-1}(a_{T-1})$. Because we have constructed $\grave{\mathfrak{v}}_{T-1}$ as a linear interpolation, the slope of $\grave{\mathfrak{v}}_{T-1}(a_{T-1})$ 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 $\grave{\mathfrak{v}}_{T-1}(a_{T-1})$ follows a step function.

The solid-line step function in Figure 6 depicts the actual value of $\grave{\mathfrak{v}}'_{T-1}(3-c_{T-1})$. When we attempt to find optimal values of c_{T-1} given m_{T-1} using $\grave{\mathfrak{v}}_{T-1}(a_{T-1})$, the numerical optimization routine will return the c_{T-1} for which $\mathrm{u}'(c_{T-1}) = \grave{\mathfrak{v}}'_{T-1}(m_{T-1}-c_{T-1})$. Thus, for $m_{T-1} = 3$ the program will return the value of c_{T-1} for which the downward-sloping $\mathrm{u}'(c_{T-1})$ curve intersects with the $\grave{\mathfrak{v}}'_{T-1}(3-c_{T-1})$; 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 $\mathrm{u}'(c_{T-1})$ with $\grave{\mathfrak{v}}'_{T-1}(4-c_{T-1})$; 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 $\mathrm{u}'(c_{T-1})$ and $\mathfrak{v}'_{T-1}(m_{T-1}-c_{T-1})$ if our estimate of $\grave{\mathfrak{v}}'_{T-1}$ 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*

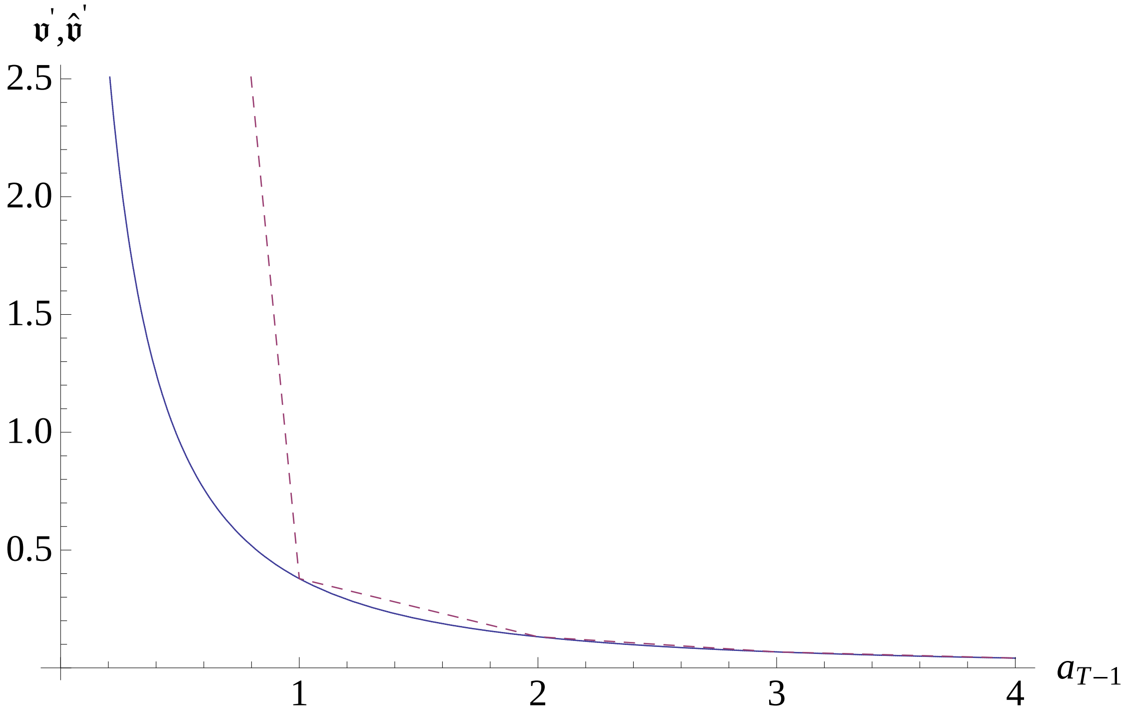

Figure 7 $\mathfrak{v}'_{T-1}(a_{T-1})$ versus $\grave{\mathfrak{v}}'_{T-1}(a_{T-1})$

value of end-of-period assets function $\mathfrak{v}'$. That is, we calculate

$$\mathfrak{v}'_{T-1}(a_{T-1}) = \beta \mathsf{R} \mathbf{\Phi}_T^{-\rho} \left(\frac{1}{n_{\boldsymbol{\theta}}}\right) \sum_{i=1}^{n_{\boldsymbol{\theta}}} (\mathcal{R}_T a_{T-1} + \boldsymbol{\theta}_i)^{-\rho} \tag{21}$$

at the points in `aVec` yielding $\{\{a_{T-1,1}, \mathfrak{v}'_{T-1,1}\}, \{a_{T-1,2}, \mathfrak{v}'_{T-1,2}\} \ldots\}$ and construct $\grave{\mathfrak{v}}'_{T-1}(a_{T-1})$ as the linear interpolating function that fits this set of points.

PlotOPRawVSFOC

The program file `functionsIntExpFOC.m` therefore uses the function `𝔳a[at_]` defined in `functions_stable.m` as the embodiment of equation (21), and constructs the `InterpolatingFunction` 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 increasingly poor job of capturing the nonlinearity of $\mathfrak{v}'_{T-1}(a_{T-1})$ at lower and lower levels of a.

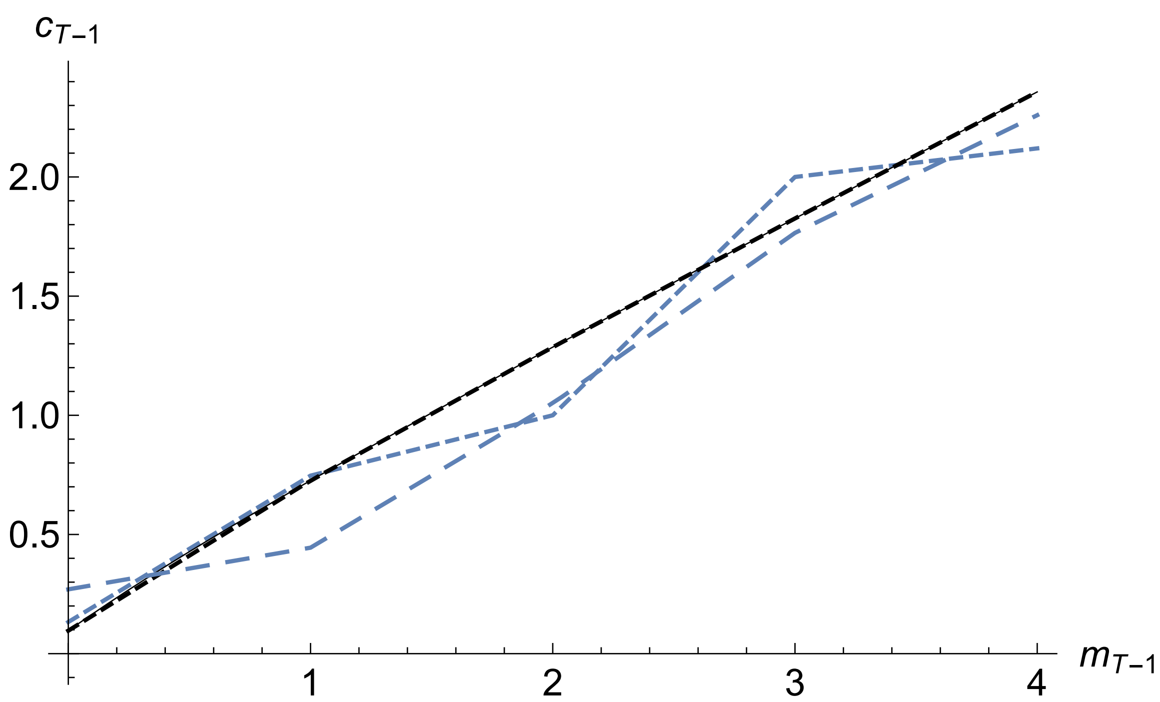

Figure 8 $\mathrm{c}_{T-1}(m_{T-1})$ (solid) Versus Two Methods for Constructing $\grave{\mathrm{c}}_{T-1}(m_{T-1})$

This is where we unveil our next trick. To understand the logic, start by considering the case where $\mathcal{R}_T = \beta = \mathbf{\Phi}_T = 1$ and there is no uncertainty (that is, we know for sure that income next period will be $\boldsymbol{\theta}_T = 1$). The final Euler equation is then:

$$c_{T-1}^{-\rho} = c_T^{-\rho}. \tag{22}$$

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 it as equivalent to a unit of current wealth. Total resources 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 $\mathrm{v}'_{T-1}(m_{T-1}) = \mathrm{u}'(c_{T-1})$ this implies that

$$\mathrm{v}'_{T-1}(m_{T-1}) = \left(\frac{m_{T-1}+1}{2}\right)^{-\rho}. \tag{23}$$

Of course, this is a highly nonlinear function. However, if we raise both sides of (23) to the power $(-1/\rho)$ the result is a linear function:

$$[\mathrm{v}'_{T-1}(m_{T-1})]^{-1/\rho} = \frac{m_{T-1}+1}{2}. \tag{24}$$

This is a specific example of a general phenomenon: A theoretical literature cited in **?** 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 $\mathfrak{v}_{T-1}$, we can define

$$\mathfrak{c}_{T-1}(a_{T-1}) \equiv [\mathfrak{v}'_{T-1}(a_{T-1})]^{-1/\rho} \tag{25}$$

which would be linear in the perfect foresight case. Thus, our procedure is to calculate the values of $\mathfrak{c}_{T-1,i}$ at each of the $a_{T-1,i}$ gridpoints, with the idea that we will construct $\grave{\mathfrak{c}}_{T-1}$ as the interpolating function connecting these points.

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

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

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

$$\grave{\mathfrak{c}}_{T-1}(a_{T-1}) = \grave{\mathfrak{c}}_{T-1}(a_{T-1,1}) + \grave{\mathfrak{c}}^a_{T-1}(a_{T-1,1})(a_{T-1} - a_{T-1,1}) \tag{26}$$

for values of $a_{T-1} < a_{T-1,1}$, where $\grave{\mathfrak{c}}^a_{T-1}(a_{T-1,1})$ is the derivative of the $\grave{\mathfrak{c}}_{T-1}$ 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) $\mathfrak{v}_{T-1}(a_{T-1})$ as a_{T-1} approaches the value $\underline{a}_{T-1} = -\underline{\boldsymbol{\theta}}\mathcal{R}_T^{-1}$. From (21) we have

$$\lim_{a_{T-1}\downarrow \underline{a}_{T-1}} \mathfrak{v}'_{T-1}(a_{T-1}) = \lim_{a_{T-1}\downarrow \underline{a}_{T-1}} \beta \mathsf{R} \boldsymbol{\Phi}_T^{-\rho} \left(\frac{1}{n_{\boldsymbol{\theta}}}\right) \sum_{i=1}^{n_{\boldsymbol{\theta}}} (a_{T-1}\mathcal{R}_T + \boldsymbol{\theta}_i)^{-\rho}. \tag{27}$$