The Endogenous Grid Method for Discrete-Continuous Dynamic Choice Models with (or without) Taste Shocks †

Fedor Iskhakov
Australian National University and CEPAR UNSW
Thomas H. Jørgensen
University of Copenhagen
John Rust
Georgetown University
Bertel Schjerning
University of Copenhagen

September 2016

Abstract: We present a fast and accurate computational method for solving and estimating a class of dynamic programming models with discrete and continuous choice variables. The solution method we develop for structural estimation extends the endogenous gridpoint method (EGM) to discrete-continuous (DC) problems. Discrete choices can lead to kinks in the value functions and discontinuities in the optimal policy rules, greatly complicating the solution of the model. We show how these problems are ameliorated in the presence of additive choice-specific *IID* extreme value taste shocks which are typically interpreted as "unobserved state variables" in structural econometric applications, or serve as "random noise" to smooth out kinks in the value functions in numerical applications. We present Monte Carlo experiments that demonstrate the reliability and efficiency of the DC-EGM and the associated Maximum Likelihood estimator for structural estimation of a life cycle model of consumption with discrete retirement decisions.

Keywords: Lifecycle model, discrete and continuous choice, retirement choice, endogenous gridpoint method, nested fixed point algorithm, extreme value taste shocks, smoothed max function, structural estimation.

JEL classification: C13, C63, D91

† We acknowledge helpful comments from Chris Carroll and *many other people*, participants at seminars at UNSW, University og Copenhagen, the 2012 conferences of the Society of Economic Dynamics, the Society for Computational Economics, the Initiative for Computational Economics at Zurich (ZICE 2014, 2015). This paper is part of the IRUC research project financed by the Danish Council for Strategic Research (DSF). Iskhakov, Rust and Schjerning gratefully acknowledge this support. Iskhakov gratefully acknowledges the financial support from the Australian Research Council Centre of Excellence in Population Ageing Research (project number CE110001029) and Michael P. Keane's Australian Research Council Laureate Fellowship (project number FL110100247). Jørgensen gratefully acknowledges financial support from the Danish Council for Independent Research in Social Sciences (FSE, grant no. 4091-00040). **Correspondence address:** Research School of Economics, ANU College of Business and Economics, 1018 HW Arndt building, The Australian National University, Canberra, ACT 0200 phone: (+61)261256193, email: `fedor.iskhakov@anu.edu.au`

1 Introduction

This paper develops a fast new solution algorithm for structural estimation of dynamic programming models with discrete and continuous choices. The algorithm we propose extends the Endogenous Grid Method (EGM) by Carroll (2006) to discrete-continuous (DC) models. We refer to it as the DC-EGM algorithm. We embed the DC-EGM algorithm in the inner loop of the nested fixed point (NFXP) algorithm (Rust, 1987), and show that the resulting maximum likelihood estimator produces accurate estimates of the structural parameters at low computational cost.

A classic example of a DC model is a life cycle model with discrete retirement and continuous consumption decisions. While there is a well developed literature on solution and estimation of dynamic discrete choice models, and a separate literature on estimation of life cycle models without discrete choices, there has been far less work on solution and estimation of DC models.[1]

There is good reason why DC models are much less commonly seen in the literature: they are substantially harder to solve. The value functions of models with only continuous choices are typically concave and the optimal policy function can be found from the *Euler equation*. EGM avoids the need to numerically solve the nonlinear Euler equation for the optimal continuous choice at each grid point in the state space. Instead, EGM specifies an exogenous grid over an endogenous quantity, e.g. savings, to analytically calculate the optimal policy rule, e.g., consumption, and endogenously determine the pre-decision state, e.g., beginning-of-period resources.[2] DC-EGM retains the main desirable properties of EGM, namely it avoids the bulk of root-finding operations and handles borrowing constraints in an efficient manner.

Dynamic programs that have only discrete choices are substantially easier to solve, since the optimal decision rule is simply the alternative with highest choice-specific value. However, solving dynamic programming problems that combine continuous and discrete choices is substantially more complicated, since discrete choices introduce kinks and non-concave regions in the value

[1] There are relatively few examples of *structural estimation* or *numerical solution* of DC models. Some prominent examples include the model of optimal non-durable consumption and housing purchases (Carroll and Dunn, 1997), optimal saving and retirement (French and Jones, 2011), and optimal saving, labor supply and fertility (Adda, Dustmann and Stevens, forthcoming).

[2] The EGM is in fact a specific application of what is referred to as "controlling the post-decision state" in operations research and engineering (Bertsekas, Lee, van Roy and Tsitsiklis, 1997). Carroll (2006) introduced the idea in economics by developing the EGM algorithm with the application to the buffer-stock precautionary savings model. Since then the idea became widespread in economics. Further generalizations of EGM include Barillas and Fernández-Villaverde (2007); Hintermaier and Koeniger (2010); Ludwig and Schön (2013); Fella (2014); Iskhakov (2015). Jørgensen (2013) compares the performance of EGM to Mathematical Programming with Equilibrium Constraints (MPEC).

function that lead to discontinuities in the policy function of the continuous choice (consumption). This can lead to situations where the Euler equation has *multiple solutions* for consumption, and hence it is only a *necessary* rather than a sufficient condition for the optimal consumption rule (Clausen and Strub, 2013). This inherent feature of DC problems complicates any method one might consider for solving DC models.

We illustrate how DC-EGM can deal with these inherent complications using a life cycle model with a continuous consumption and binary retirement choice with and without taste shocks. Our example is a simple extension of the classic life cycle model of Phelps (1962) where, in the absence of a retirement decision, the optimal consumption rule could hardly be any simpler — a linear function of resources. However, once the discrete retirement decision is added to the Phelp's problem — in our case allowing a worker with logarithmic utility to also make a binary irreversible retirement decision — the consumption function becomes unexpectedly complex, with multiple discontinuities in the optimal consumption rule. We derive an *analytic solution* for this model, use it to demonstrate the accuracy of the solution obtained numerically by DC-EGM, and then investigate the performance of the Rust's NFXP type nested estimator based on the DC-EGM solution algorithm to estimate the structural parameters of this model.

Fella (2014) showed how EGM could be adapted to solve non-concave problems, including models with discrete and continuous choices. In this paper we focus on discrete choices and show that introducing *IID* Extreme Value Type I choice-specific taste shocks not only facilitates maximum likelihood estimation, but also allows to smooth out some of the kinks in the value functions and thus simplify the numerical solution of DC models. This approach results in *multinomial logit formulas* for the *conditional choice probabilities* for the discrete choices and a closed form expression for the expectation of the value function with respect to these taste shocks.[3]

In econometric applications continuously distributed taste shocks are essential for generating predictions from dynamic programming models that are *statistically non-degenerate*. Such predictions assign a positive (however small) choice probability to *every* alternative, and therefore preclude zero likelihood observations. These shocks are interpreted as *unobserved state variables*, i.e. idiosyncratic shocks observed by agents but not by the econometrician. However, in numerical

[3]In principle, the Extreme Value assumption could be relaxed to allow for other distributions at the cost of numerical approximation of choice probabilities and the conditional expectation of the value function. For example, Bound, Stinebrickner and Waidmann (2010), assume that the discrete choice specific taste shocks are Normal rather than Extreme Value. Yet, we follow the long tradition of discrete choice modeling dating back to (McFadden, 1973) and (Rust, 1987).

or theoretical applications taste shocks can serve as a smoothing device (homotopy perturbation) that facilitates the numerical solution of more advanced DC models that may have excessively many kinks and discontinuities, for example caused by a large number of discrete choices.

The inclusion of Extreme Value Type I taste shocks have a long history in discrete choice modeling dating back to the seminal work by McFadden (1973). This assumption is typically invoked in microeconometric analyses of dynamic discrete choice models where numerical performance boosted by closed form choice probabilities is particularly important, see for example Rust (1994) and the recent survey by Aguirregabiria and Mira (2010). Some recent studies of DC models with Extreme Value taste shocks include Casanova (2010); Ejrnæs and Jørgensen (2015); Iskhakov and Keane (2016); Oswald (2016) and Adda, Dustmann and Stevens (forthcoming).

At first glance, the addition of stochastic shocks would appear to make the problem *harder* to solve, since both the optimal discrete and continuous decision rules will necessarily be functions of these stochastic shocks. However, we show that a variety of stochastic variables in DC models *smooth out* many of the kinks in the value functions and the discontinuities in the optimal consumption rules. In the absence of smoothing, we show that every kink induced by the comparison of the discrete choice specific value functions in any period t propagates backwards in time to all previous periods as a manifestation of the decision maker's anticipation of the future discrete action. The resulting accumulation of kinks during backward induction presents the most significant challenge for the numerical solution of DC models. In presence of taste shocks the decision maker can only anticipate a particular future discrete action to be more or less probable, and thus the primary reason for the accumulation of kinks disappears. Yet, the combination of taste shocks and the stochastic variables in the model is perhaps the most powerful device to prevent the propagation and accumulation of kinks.[4]

In the case when the Extreme Value taste shocks are used as a logit *smoothing device* of an underlying deterministic model of interest, we show that the latter problem can be approximated by the smoothed model to any desirable degree of precision. The scale parameter $\sigma \geq 0$ of the corresponding Extreme Value distribution then serves as a *homotopy* or *smoothing* parameter. When σ is sufficiently large, the non-concave regions near the kinks in the non-smoothed value function disappear and the value functions become globally concave. But even small values of

[4]Contrary to the macro literature that uses stochastic elements such as employment lotteries (Rogerson, 1988; Prescott, 2005; Ljungqvist and Sargent, 2005) to smooth out non-convexities, the taste shock we introduce in DC models in general *do not* fully convexify the problem.

σ smooth out many of the kinks in the value functions and suppress their accumulation in the process of backward induction as noted above. An additional benefit of the taste shocks is that standard integration methods, such as quadrature rules, apply when the expected value function is a smooth function.

We run a series of Monte Carlo simulations to investigate the performance of DC-EGM for structural estimation of the life cycle model with the discrete retirement decision. We find that a maximum likelihood estimator that nests the DC-EGM algorithm performs well. It quickly produces accurate estimates of the structural parameters of the model even when fairly coarse grids over wealth are used. We find the cost of "oversmoothing" to be negligible in the sense that the parameter estimates of a perturbed model with stochastic taste shocks are estimated very accurately even if the true model does not have taste shocks. Thus, even in the case where the addition of taste shocks results in a *misspecification* of the model, the presence of these shocks improves the accuracy of the solution and reduces computation time without increasing the approximation bias significantly. Even when very few grid points are used to solve the model, we find that smoothing the problem improves the root mean square error (RMSE). Particularly, with an appropriate degree of smoothing (σ), we can reduce the number of gridpoints by an order of magnitude without much increase in the RMSE of the parameter estimates.

DC-EGM is applicable to many fields of economics and has been implemented in several recent empirical applications. Ameriks, Briggs, Caplin, Shapiro and Tonetti (2015) study how the need for long term care and bequest motive interact with government-provided support to shape the wealth profile of the elderly. They use an endogenous grid method similar to DC-EGM to solve and estimate the corresponding non-concave model. Iskhakov and Keane (2016) employ DC-EGM to estimate a life-cycle model of discrete labor supply, human capital accumulation and savings for the Australian population. They use the model to evaluate Australia's defined contribution pension scheme with means-tested minimal pension, and quantify the effects of anticipated and unanticipated policy changes. Yao, Fagereng and Natvik (2015) use DC-EGM to analyze how housing and mortgage debt affects consumer's marginal propensity to consume. They estimate a model in which households hold debt, financial assets and illiquid housing and find that a substantial fraction of households are likely to behave in a "hand-to-mouth" fashion despite having significant wealth holdings. Druedahl and Jørgensen (2015) employ a modified version of DC-EGM to analyze the credit card debt puzzle. They solve a model of optimal consumption and debt holdings and show how, for some parameterizations of the model, a large group of consumers find it optimal to simultaneously hold positive gross debt and positive gross assets even though the interest rate on the debt is much higher than the rate on the assets. Ejrnæs and Jørgensen (2015) use DC-EGM to estimate a model of optimal consumption and saving with a fertility choice to analyze the saving behavior around intended and unintended childbirths. They model the fertility process as a discrete choice over effort to conceive a child subject to a biological fecundity constraint and allow for the possibility of unintended child births through imperfect contraceptive control.

In the next section we present a simple extension of the life cycle model of consumption and savings with logarithmic utility studied by Phelps (1962) where we allow for a discrete retirement decision. We derive a closed-form solution to this problem, and discuss its properties. Using this simple model we demonstrate the accuracy of the deterministic version of DC-EGM. We then introduce extreme value taste shocks and show how the implied smoothing affects the value functions and the optimal policy rules. In particular, we show that the error introduced by "extreme value smoothing" is uniformly bounded, and prove that the solution of the smoothed DP problem with taste shocks converges to the solution to the DP problem without taste shocks as scale of the shocks approaches zero. Section 3 presents the full DC-EGM algorithm. In section 4 we show how it is incorporated in the Nested Fixed Point algorithm for maximum likelihood estimation of the structural parameters in the retirement model. We present the results of a series of Monte Carlo experiments in which we explore the performance of the estimator in a variety of settings. We conclude with a short discussion of the range of models that DC-EGM is applicable to and discuss some open issues with this method.

2 An Illustrative Problem: Consumption and Retirement

This section extends the classic life-cycle consumption/savings model of Phelps (1962) to allow for a binary retirement decision. We derive an analytic solution to the simple life cycle problem with logarithmic utility that serves both to illustrate the complexity caused by the addition of a discrete retirement choice and how DC-EGM can be applied. While we focus on this simple illustrative example for expositional clarity, DC-EGM can be applied in a much more general class of problems that we discuss in the conclusion - including the extended version of the retirement model that we use in the Monte Carlo exercise. While we initially illustrate the complexity of the solution without any stochastic elements, we include both taste and income shocks in the simple model and discuss how these additional elements actually simplify the solution of the model using DC-EGM.

2.1 Deterministic model of consumption/savings and retirement

Consider the discrete-continuous (DC) dynamic optimization problem

$$\max_{\{c_t,d_t\}_{t=1}^T} \sum_{t=1}^{T} \beta^t(\log(c_t) - \delta_t d_t) \tag{1}$$

where agents choose consumption c_t and whether to retire to maximize the discounted stream of utilities. Let $d_t = 0$ denote the choice to retire and $d_t = 1$ to continue working, and let $\delta_t > 0$ be the disutility of work at age t. To keep the solution simple, we assume that retirement is absorbing, i.e. once workers retire they are unable to return to work.

Agents solve (1) subject to a sequence of period-specific borrowing constraints, $c_t \leq M_t$ where $M_t = R(M_{t-1} - c_{t-1}) + y_t d_{t-1}$ is the consumer's resources available for consumption in the beginning of period t. We assume a fixed, non-stochastic gross interest rate, R and a deterministic labor income y_t which depends on the previous period's labor supply choice, d_{t-1}. This timing convention is standard in the literature and allows us to avoid a separate state variable when the model is extended in the next sections to allow for wage uncertainty. In turn, consumers choose current period consumption (c_t) simultaneously with labor supply (d_t) before knowing the realization of the wage shock.

Denote $V_t(M_t)$ the maximum expected discounted lifetime utility of a worker, and $W_t(M_t)$ that of a retiree. The choice problem of the worker can be expressed recursively through the Bellman equation

$$V_t(M_t) = \max\{v_t(M_t|d_t = 0), v_t(M_t|d_t = 1)\}, \tag{2}$$

where the *choice-specific value functions* are given as

$$v_t(M_t|d_t = 0) = \max_{0 \leq c_t \leq M_t} \{\log(c_t) + \beta W_{t+1}\big(R(M_t - c_t)\big)\}, \tag{3}$$

$$v_t(M_t|d_t = 1) = \max_{0 \leq c_t \leq M_t} \{\log(c_t) - \delta_t + \beta V_{t+1}\big(R(M_t - c_t) + y_{t+1}\big)\}. \tag{4}$$

The choice problem of the retiree is given by the Bellman equation

$$W_t(M_t) = \max_{0 \le c_t \le M_t} \{\log(c_t) + \beta W_{t+1}\big(R(M_t - c_t)\big)\}. \tag{5}$$

It follows from (3) and (5) that $v_t(M_t|d_t = 0) = W_t(M_t)$. The value function $W_t(M_t)$ is given by Phelps (1962, p. 742) who solves the corresponding optimal consumption problem. In the following we therefore only focus on deriving formulas for $v_t(M_t|d_t = 1)$ and finding optimal consumption rules $c_t(M_t|d_t = 0)$ and $c_t(M_t|d_t = 1)$ for a worker who chooses to retire and to continue working, respectively. It follows that the optimal consumption rule for the retiree is identical to $c_t(M_t|d_t = 0)$.

Note that even if $v_t(M_t, 0)$ and $v_t(M_t, 1)$ are concave functions of M_t, because $V_t(M_t)$ is the maximum of the two, it is generally not concave (Clausen and Strub, 2013). It is not hard to show that V_t will generally have a *kink point* at the value of resources where the two choice-specific value functions cross ($\overline{M}_t$), i.e. where $v_t(\overline{M}_t, 1) = v_t(\overline{M}_t, 0)$. We refer to these points as *primary kinks*.

This kink point at $\overline{M}_t$ is also the *optimal retirement threshold* — the optimal decision for a worker with resources $M_t \le \overline{M}_t$ is to keep working (not to retire) and to use the consumption rule $c_t(M_t|d_t = 1)$, whereas the optimal decision for a worker whose wealth exceeds $\overline{M}_t$ is to retire and to consume $c_t(M_t|d_t = 0)$. The worker is indifferent between retiring and working at the primary kinks ($M = \overline{M}_t$) where the value function is generally non-differentiable. However the left and right hand derivatives do exist and we have $V_t^{-}(\overline{M}_t) < V_t^{+}(\overline{M}_t)$. Through the first order conditions, the discontinuity in the derivative of $V_t(M)$ at $\overline{M}_t$ translates into a discontinuity in the optimal consumption function in the *previous* period $t-1$. In the same time, because the Bellman equation expresses $V_{t-1}(M)$ as a function of $V_t(M)$, the kink point in the latter results in a kink in $V_{t-1}(M)$. In effect, the primary kinks propagate back in time and manifest themselves as discontinuities in the policy functions and additional kinks in the value function. These kinks do not correspond to the points of indifference between the discrete alternatives, but instead appear as reverberations of the primary kinks at the retirement thresholds the consumer expects to encounter in the future. We refer to these as *secondary kinks.*

Let $c_{T-\tau}(M)$ denote the optimal consumption function of the workers in period $t = T - \tau$, i.e. τ periods before the end of the life cycle. Theorem 1 illustrates how complex the solution to Phelps' model becomes once we make the simple extension of allowing a discrete, irreversible retirement choice.

Theorem 1 (Analytical solution to the retirement problem)**.** *Assume that income and disutility of work are time-invariant and the discount factor β and the disutility of work δ are not too large, i.e.*

$$\beta R \leq 1 \text{ and } \delta < (1+\beta)\log(1+\beta). \tag{6}$$

Then $\tau \in \{1, \dots, T\}$ the optimal consumption rule in the workers' problem (2)-(4) *is given by*

$$c_{T-\tau}(M) = \begin{cases} M & \text{if } M \leq y/R\beta, \\ [M + y/R]/(1+\beta) & \text{if } y/R\beta \leq M \leq \overline{M}^{l_1}_{T-\tau}, \\ [M + y(1/R + 1/R^2)]/(1+\beta+\beta^2) & \text{if } \overline{M}^{l_1}_{T-\tau} \leq M \leq \overline{M}^{l_2}_{T-\tau}, \\ \dots & \dots \\ \left[M + y\left(\sum_{i=1}^{\tau-1} R^{-i}\right)\right]\left(\sum_{i=0}^{\tau-1}\beta^i\right)^{-1} & \text{if } \overline{M}^{l_{\tau-2}}_{T-\tau} \leq M \leq \overline{M}^{l_{\tau-1}}_{T-\tau}, \\ \left[M + y\left(\sum_{i=1}^{\tau} R^{-i}\right)\right]\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if } \overline{M}^{l_{\tau-1}}_{T-\tau} \leq M < \overline{M}^{r_{\tau-1}}_{T-\tau}, \\ \left[M + y\left(\sum_{i=1}^{\tau-1} R^{-i}\right)\right]\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if } \overline{M}^{r_{\tau-1}}_{T-\tau} \leq M < \overline{M}^{r_{\tau-2}}_{T-\tau}, \\ \dots & \dots \\ [M + y(1/R + 1/R^2)]\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if } \overline{M}^{r_2}_{T-\tau} \leq M < \overline{M}^{r_1}_{T-\tau}, \\ [M + y/R]\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if } \overline{M}^{r_1}_{T-\tau} \leq M < \overline{M}_{T-\tau}, \\ M\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if } M \geq \overline{M}_{T-\tau}. \end{cases} \tag{7}$$

The segment boundaries are totally ordered with

$$y/R\beta < \overline{M}^{l_1}_{T-\tau} < \cdots < \overline{M}^{l_{\tau-1}}_{T-\tau} < \overline{M}^{r_{\tau-1}}_{T-\tau} < \cdots < \overline{M}^{r_1}_{T-\tau} < \overline{M}_{T-\tau}, \tag{8}$$

and the right-most threshold $\overline{M}_{T-\tau}$ given by

$$\overline{M}_{T-\tau} = \frac{(y/R)e^{-K}}{1-e^{-K}}, \text{ where } K = \delta\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1}, \tag{9}$$

defines the smallest level of wealth sufficient to induce the consumer to retire at age $t = T - \tau$.

The proof of Theorem 1 is given in Appendix C. Note that the assumptions on the parameters β, δ and R are needed to ensure the ordering of the bounderies (8). Modified versions of Theorem 1 hold under weaker conditions, including a version where income and the disutility of work are age-dependent. However, depending on the paths of income and disutility of work some of the intermediate thresholds in Theorem 1 may not exist, or may be equal to each other.

It follows that the optimal consumption rule (7) is piece-wise linear in M, and in period $t = T-\tau$ consists of $2\tau + 1$ segments. The first segment where $M < y/R\beta$ is the credit constrained region. The next $\tau - 1$ segments are connected and bounded by the $\tau - 1$ kink points $\overline{M}^{l_j}_{T-\tau}$ which represents the largest levels of wealth for which the consumer is not liquidity constrained at ages $T-\tau, T-\tau+1, \dots, T-\tau+j-1$, but will be liquidity constrained at age $T-\tau+j$ under the optimal consumption and retirement policy. The remaining segments relate to the retirement choice, namely $\overline{M}^{r_j}_{T-\tau}$, $j = 1, \dots, \tau - 1$ represent the largest level of saving for which it is optimal to retire at age $T - \tau + j$ but not at any earlier age $T-\tau, T-\tau+1, \dots, T-\tau+j-1$. The optimal consumption function is discontinuous at points $\overline{M}^{r_j}_{T-\tau}$, and including the discontinuity at the retirement threshold $\overline{M}_{T-\tau}$ makes altogether τ downward jumps in period $T - \tau$.

Using Theorem 1 it is not hard to show that the value function $V_{T-\tau}(M)$ is piecewise logarithmic with the same kink points, and can be written as

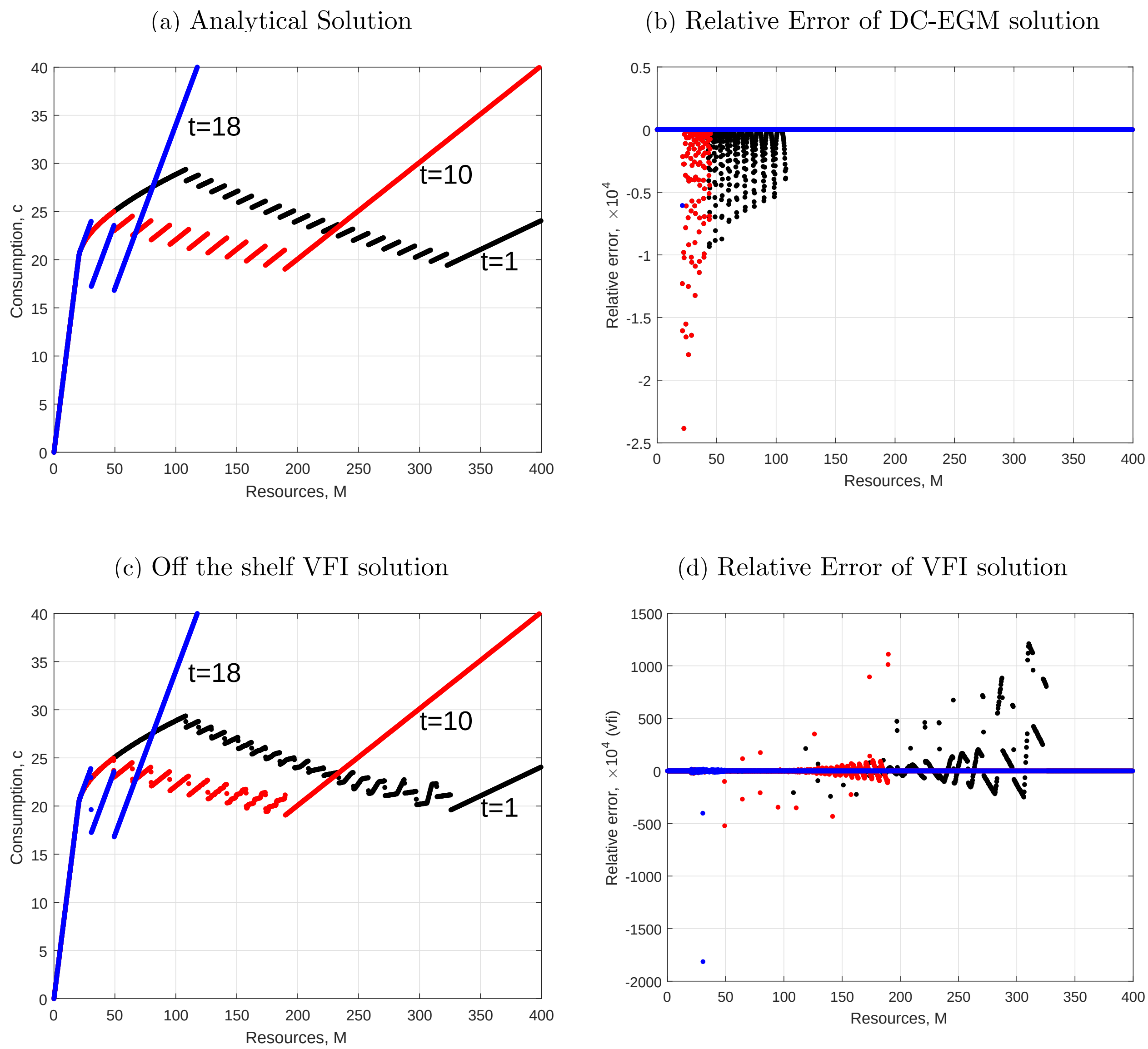
$$V_{T-\tau}(M) = B_{T-\tau} \log(c_{T-\tau}(M)) + C_{T-\tau} \tag{10}$$

for constants $(B_{T-\tau}, C_{T-\tau})$ that depend on the region M falls in. For each $\tau \geq 1$, the value function has one primary kink at the optimal retirement threshold $M = \overline{M}_{T-\tau}$, $\tau - 1$ secondary kinks at $\overline{M}^{r_j}_{T-\tau}$, $j = 1, \dots, \tau - 1$, and τ kinks related to current period and future liquidity constraints at $M = y/R\beta$ and $\overline{M}^{l_j}_{T-\tau}$, $j = 1, \dots, \tau - 1$. If $R\beta = 1$ the liquidity-related kink points collapse to a single point $M = y/R\beta = y = \overline{M}^{l_1}_{T-\tau} = \cdots = \overline{M}^{l_{\tau-1}}_{T-\tau}$.

Figure 1 displays the optimal consumption function (7) and compares it to the numerical solution produced by DC-EGM described below in Section 3, as well as the numerical solution produced by a naive brute force implementation of VFI. With a sufficient number of grid points, DC-EGM is able to accurately locate all the discontinuities of the analytical consumption rules $(\overline{M}^{r_j}_{T-\tau})$ and the boundary of the credit constrained region $y/R\beta$. Yet, because the kinks points $\overline{M}^{l_j}_{T-\tau}$ are not located precisely, the right panel of Figure 1 shows small relative errors on the order of 10^{-4} in the intervals $(y/R\beta, \overline{M}^{\tau-1}_{T-\tau})$ in each period $T - \tau$. Overall, the numerical solution by DC-EGM replicates the analytical solution remarkably well.[5]

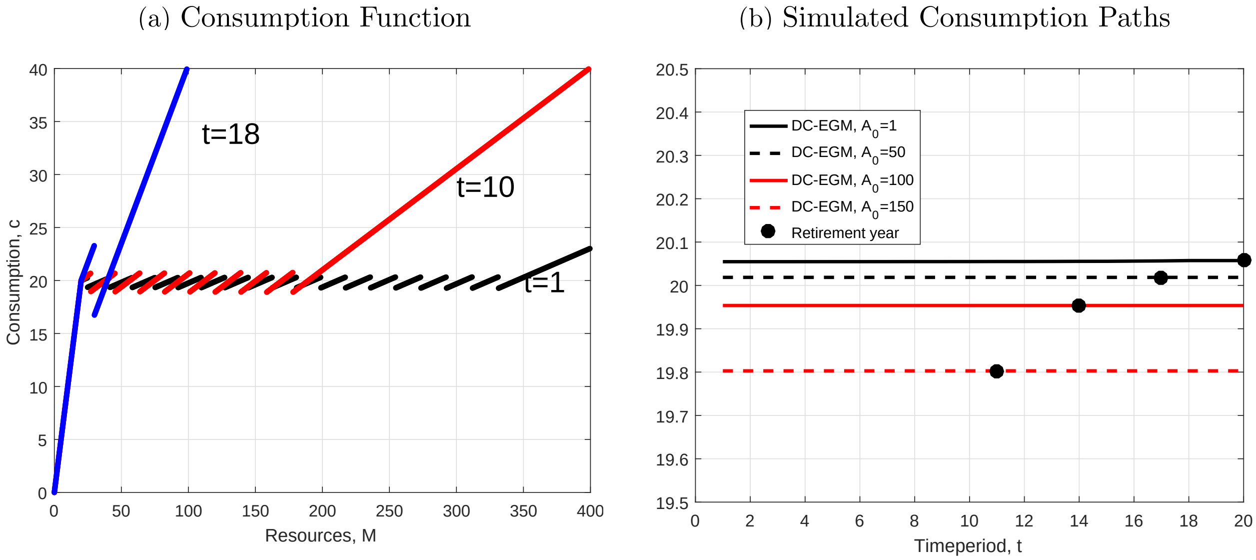
[5]With 2000 points on the endogenous grid over wealth it took our Matlab/C implementation around 0.17 seconds on a Lenovo ThinkPad laptop with Intel® Core™ i7-4600M CPU @ 2.10 GHz and 8GB RAM to generate the numerical solution by DC-EGM. This is about 20 times faster than value function iterations (VFI) which we implemented in Matlab with 500 fixed grid points over wealth. The discretization of consumption is a brute force approach to ensure that global optimum is found. We used 400 equally spaced guesses for each level of wealth. The fact that EGM offers the speedup of one to two orders of magnitude relative to VFI is a well estab-

Figure 1: Optimal Consumption Functions.



Notes: The plots show optimal consumption rules of the worker in the consumption-savings model with $R = 1$, $\beta = 0.98$, $y = 20$, and $T = 20$. Panel (a) illustrates the analytical solution (which is indistinguishable from the the numerical solution produced by DC-EGM), panel (b) illustrates the numerical error from the solution found by DC-EGM. Panel (c) shows the numerical solution found by value function iterations (VFI), and panel (d) shows the associated numerical errors. Both the VFI and DC-EGM solutions were generated using 2000 points in the M-grid. For VFI grid points are equally spaced, the maximum level in the wealth is 600, and 10,000 equally spaced between zero and $M(t)$ points of consumption are used to solve the maximization problem in the Bellman equation.

Figure 2: Discontinuous Consumption Function and Smooth Consumption Paths



Notes: The plots show optimal consumption functions of the worker in the consumption-savings model with with $T = 20$, $d_t = 1$, $y = 20$, $\beta = .98$, and $R = 1/\beta = 1.02$. The left panel illustrates the solution for $t = 1, 10, 18$, while the right panel presents consumption paths simulated over the whole life cycle for several initial levels of wealth. The model was solved by the DC-EGM algorithm.

Panels (c)-(d) of Figure 1 show the solution produced by a traditional value function iterations (VFI) method with the same number of grid points over wealth and optimal consumption levels found by a fine grid search method. This implementation of VFI could admittedly be thought of as too simplistic, with possible improvements in how the grid points are located and spaced, which computational methods are employed to search for optimal consumption in each grid point, etc. Yet, the point we wish to make is that a standard "off the shelf" version of the VFI method may have serious difficulties when solving DC problems due to its failure to adequately capture the secondary kinks in the value function that get "papered over" via naive application of the standard method of linear interpolation of the value functions. The bottom panels of Figure 1 shows that the VFI solution results in significant approximation errors and is unable to fully capture the numerous discontinuities in the consumption function.

Figure 2 plots the optimal consumption functions and simulated consumption paths under the same assumptions as in Figure 1 except in this case we set $R = 1/\beta = 1.02$. The theoretical

lished finding in the literature, see e.g. Barillas and Fernández-Villaverde (2007); Jørgensen (2013); Fella (2014); Ameriks, Briggs, Caplin, Shapiro and Tonetti (2015).

prediction is that, with $R\beta = 1$, simulated consumption paths should be flat, yet the consumption functions shown in the left panel displays numerous discontinuities that accumulate backwards from the final period $T = 20$. Beyond the important economic message that discontinuous consumption functions are not incompatible with consumption smoothing, this also illustrates the remarkable precision of the DC-EGM algorithm. In fact, when we simulate consumption trajectories implied by this incredibly complex solution found numerically, *the simulated consumption profiles are still perfectly flat.*

Before we describe in detail how DC-EGM works, we now illustrate how the incorporation of various types of uncertainty, including Extreme value taste shocks, renders the accumulation of kinks in the value function and discontinuities in the consumption function considerably less severe.

2.2 Adding Taste Shocks and Income Uncertainty

Now consider an extension of the model presented above where the consumer faces income uncertainty and where choices are affected by choice-specific taste shocks. More specifically, assume that income when working is $y_t = y\eta_t$, where η_t is log-normally distributed multiplicative idiosyncratic income shock, $\log \eta_t \sim \mathcal{N}(-\sigma_\eta^2/2, \sigma_\eta^2)$.[6]

The additively separable choice-specific random taste shocks, $\sigma_\varepsilon \varepsilon_t(d_t)$, are i.i.d. Extreme Value type I distributed with scale parameter σ_ε. In this formulation, the extreme value taste shock enters as a structural part of the problem. If the true model does not have taste shocks, σ_ε can be interpreted as a (logit) smoothing parameter, see Theorem 2 below.

The solution of the retiree's problem remains the same, and we focus on the worker's problem. The Bellman equation (2) has to be rewritten to include the taste shocks,

$$V_t(M_t, \varepsilon_t) = \max\{v_t(M_t|d_t = 0) + \sigma_\varepsilon \varepsilon_t(0), v_t(M_t|d_t = 1) + \sigma_\varepsilon \varepsilon_t(1)\}, \tag{11}$$

where the value function conditional on the choice to retire $v_t(M_t|d_t = 0)$ is given by (3). However, the value function conditional on the choice to remain working, $v_t(M_t|d_t = 1)$, is modified to

[6] As mentioned above, we follow the literature in the assumption that idiosyncratic income shocks are realized *after* the labor supply choice is made, which is equivalent to allowing income to be dependent on a lagged choice of labor supply.

account for the taste and income shocks in the following period,

$$v_t(M_t|d_t = 1) = \max_{0\leq c_t\leq M_t} \left\{ \log(c_t) - 1 + \beta \int EV^{\sigma_\varepsilon}_{t+1}(R(M_t - c_t) + y\eta_{t+1}) f(d\eta_{t+1}) \right\}. \tag{12}$$

Because the taste shocks are independent Extreme Value distributed random variables, the expected value function, $EV^{\sigma_\varepsilon}_{t+1}$, is given by the well-known *logsum* formula (McFadden, 1973)

$$\begin{aligned} EV^{\sigma_\varepsilon}_{t+1}(M_{t+1}) &= E\left[\max\left\{v_{t+1}(M_{t+1}|d_{t+1} = 0) + \sigma_\varepsilon\varepsilon(0),\ v_{t+1}(M_{t+1}|d_{t+1} = 1) + \sigma_\varepsilon\varepsilon(1)\right\}\right] \\ &= \sigma_\varepsilon \log\left\{ exp\left[\frac{v_{t+1}(M_{t+1}|d_{t+1} = 0)}{\sigma_\varepsilon}\right] + exp\left[\frac{v_{t+1}(M_{t+1}|d_{t+1} = 1)}{\sigma_\varepsilon}\right]\right\}. \end{aligned} \tag{13}$$

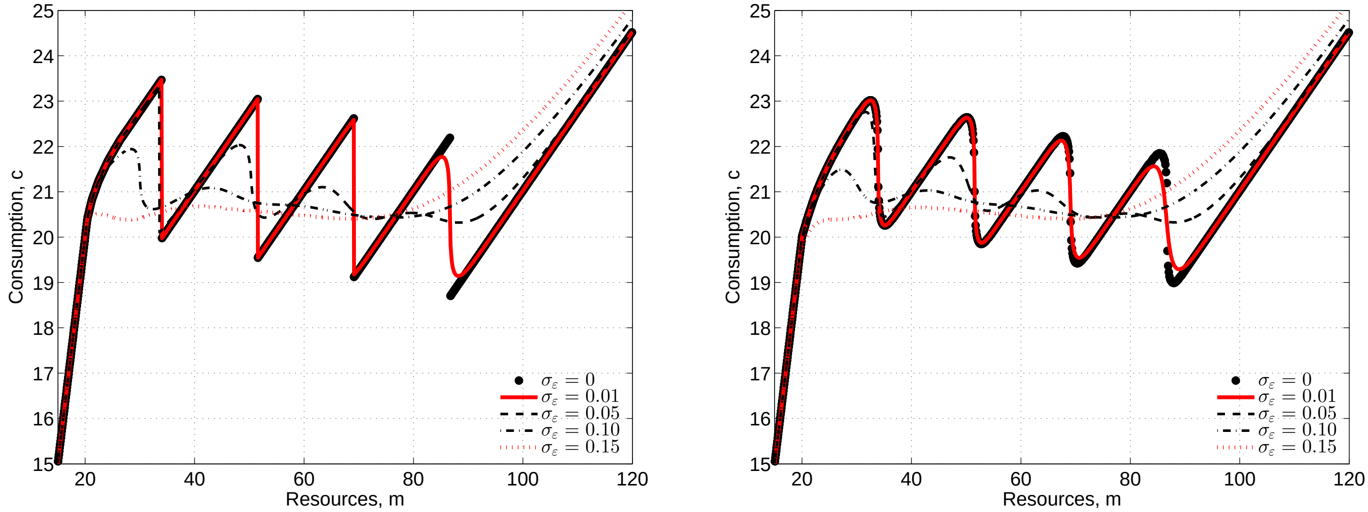
The immediate effect of introducing extreme value taste shocks is the complete elimination of the *primary* kinks due to the smoothing of the logit formula: the expected value function in (13) is a smooth function of M_t around the point where $v_t(M_t|d_t = 1) = v_t(M_t|d_t = 0)$. When σ_ε is sufficiently large the value function $v_t(M_t|d_t = 1)$ eventually becomes globally concave.[7] Even when σ_ε is not large enough to "concavify" the value function completely, by smoothing out the primary kink in period t it still helps to eliminate many of the secondary kinks in the time periods prior to t.

Figure 3 shows the choice specific consumption function $c_t(M_t|d_t = 1)$ for a worker (conditional on the choice to continue working) for different values of smoothing parameter $\sigma_\varepsilon \in \{0, 0.01, 0.05, 0.10, 0.15\}$. The left panel plots the optimal consumption in the absence of income uncertainty ($\sigma_\eta = 0$) while income uncertainty ($\sigma_\eta = \sqrt{0.005}$) is added in the right panel. The plots are drawn for the period $T - 5$, corresponding to 4 discontinuities of the choice specific policy function in line with Theorem 1, without the discontinuity at the retirement threshold $\overline{M}_{T-5}$ in the deterministic model.

[7] To see this, note that as the variance of the taste shocks increases, the choice-specific value functions are dominated by the noise and the differences between the alternatives become relatively less important. In turn, the choice-specific value functions become similar, and $\lim_{\sigma_\varepsilon\downarrow\infty} EV^{\sigma_\varepsilon}_t(M_t)/\sigma_\varepsilon = log(2)$. It follows from (11) then that the value function $v_t(M_t|d_t = 1)$ inherits its globally concave shape from the utility function.

Figure 3: Optimal Consumption Rules for Agent Working Today ($d_{t-1} = 1$).

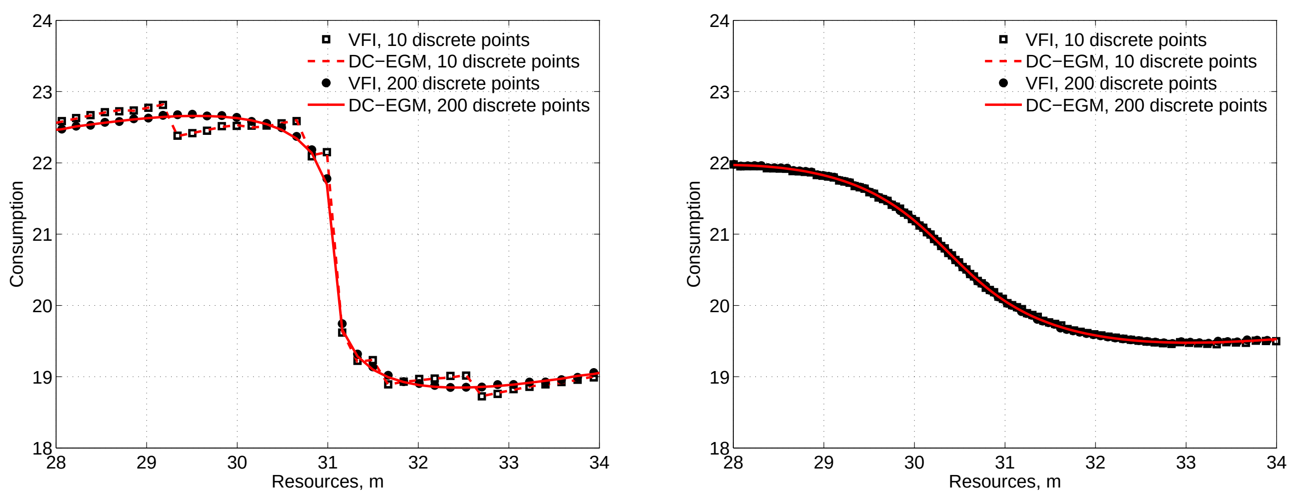
(a) Without income uncertainty, $\sigma_\eta = 0$ (b) With income uncertainty, $\sigma_\eta = \sqrt{.005}$



Notes: The plots show optimal consumption rules of the worker who decides to continue working in the consumption-savings model with retirement in period $t = T - 5$ for a set of taste shock scales σ_ε in the absence of income uncertainty, $\sigma_\eta = 0$, (left panel) and in presence of income uncertainty, $\sigma_\eta = \sqrt{.005}$, (right panel). The rest of the model parameters are $R = 1$, $\beta = 0.98$, $y = 20$.

Figure 4: Artificial Discontinuities in Consumption Functions, $\sigma_\eta^2 = 0.01$, $t = T - 3$.

(a) $\sigma_\varepsilon = 0$ (b) $\sigma_\varepsilon = 0.05$



Notes: Figure 4 illustrates how the number of discrete points used to approximate expectations regarding future income affects the consumption functions from value function iteration (VFI) and DC-EGM. Panel (a) illustrates how using few (10) discrete equiprobable points to approximate expectations produce severe approximation error when there is *no* taste shocks. Panel (b) illustrates how moderate smoothing ($\sigma_\varepsilon = .05$) significantly reduces this approximation error.

It is evident that taste shocks of larger scale ($\sigma_\varepsilon \geq 0.05$) manage to smooth the function completely — eliminating all four discontinuities, and thus, eliminating the non-concavity of the value function in period $T-4$. Yet, for $\sigma_\varepsilon = 0.01$ only the last (rightmost) discontinuity is obviously smoothed out. Thus, even though full "concavification" is not achieved, the presence of extreme value taste shocks makes the consumption function continuous by smoothing out the secondary kinks in the value function.

When the model has other stochastic elements such as wage shocks or random market returns, the accumulation of secondary kinks may be less pronounced due to the additional smoothing. Yet, in the absence of taste shocks, the primary kinks cannot be avoided even if all secondary kinks are eliminated by a sufficiently high degree of uncertainty in the model. It is in this setup which also appears to be mostly used in practical applications, where the introduction of the Extreme Value distributed taste shocks is especially beneficial. The taste shocks *and* other structural shocks together contribute to the reduction of the number of secondary kinks and to the alleviation of the issue of their multiplication and accumulation. It is clear from the right panel of Figure 3 that the non-concavity of the value function can be eliminated with a smaller taste shock ($\sigma_\varepsilon = 0.01$) when additional smoothing through uncertainty is present in the model.

An additional benefit of the inclusion of taste shocks is that numerical integration over the stochastic elements of the model has to be performed on a smooth function $EV_t^{\sigma_\varepsilon}(M_t)$ instead of the kinked value function $V_t(M_t)$. Standard procedures like Gaussian quadrature are readily applicable. When $\sigma_\varepsilon = 0$, performing standard numerical integration typically results in spurious discontinuities as shown in the left panel of Figure 4. This is due to the integrand not being a smooth function, see Appendix B for a detailed discussion. The right panel of Figure 4 illustrates how moderate smoothing ($\sigma_\varepsilon = .05$) significantly reduces this approximation error and removes the artificial kinks.

Thus, the extreme value taste shocks ε_t have a dual interpretation or role in DC models: in structural econometric applications, they can be regarded as unobserved state variables (i.e. variables observed by the consumer but not by the econometrician) that makes their behavior appear probabilistic from the standpoint of a person who does not observe ε_t. ε_t also has an interpretation as stochastic noise that is introduced to help solve a difficult DC dynamic program by smoothing out kinks in the value function similar in some respects to the way stochastic noise is introduced into optimization algorithms to help them find a global optimum of difficult nonlinear programming problems with multiple local optima. In the former case, σ_ε is a scale parameter of taste shocks, and has to be estimated along with other structural parameters. In the latter case, σ_ε is the amount of smoothing and has to be chosen and fixed prior to estimation. Theorem 2 shows that the level of σ_ε can always be chosen in such a way that the perturbed model approximates the true deterministic model with an arbitrary degree of precision. In effect, Theorem 2 formalizes the results presented graphically in Figure 3.

Theorem 2 (Extreme Value Homotopy Principle)**.** *In every time period the (expected) value function of the consumption and retirement problem with extreme value taste shocks $EV_t^{\sigma_\varepsilon}(M_t)$ defined in* (13) *converges uniformly to the value function of the same problem without taste shocks $V_t(M_t)$ defined in* (2) *as the scale of these shocks approaches zero. The following uniform bound holds*

$$\forall t: \quad \sup_{M_t \geq 0} |EV_t^{\sigma_\varepsilon}(M_t) - V_t(M_t)| \leq \sigma_\varepsilon \left[\sum_{j=0}^{T-t} \beta^j \right] \log(2). \tag{14}$$

Consequently, as $\sigma_\varepsilon \downarrow 0$, both continuous and discrete decision rules of the smoothed model with taste shocks converge pointwise to those of the deterministic model.

In Appendix E we prove a more general version of Theorem 2 which holds under very weak conditions for arbitrary DC models with multidimensional continuous or discrete state variables and multiple continuous choice variables. Theorem 2 justifies our claim that the extreme value smoothing can be regarded as a homotopy method for solving the non-smooth limiting problem without taste shocks by solving smooth "nearby" problems with Extreme value taste shocks, and the Extreme value scale parameter plays the role of the "homotopy parameter."

3 The DC-EGM Algorithm

In this section, we describe the generalization of the EGM algorithm for solving discrete-continuous problems that we call the DC-EGM algorithm.

The DC-EGM is a backward induction algorithm that iterates on the Euler equation and sequentially computes the discrete choice specific value functions $v_t(M_t|d_t)$ and the corresponding consumption rules $c_t(M_t|d_t)$ stating at terminal period T. The DC-EGM uses the standard EGM algorithm by Carroll (2006) to find all solutions of the Euler equation *conditional* on the current discrete choice, d_t. We describe this subroutine first.

However, because the problem is generally not convex and the first order conditions are not sufficient, some of the found solutions of the Euler equation do not correspond to the optimal consumption choices. Consequently, the DC-EGM includes a procedure to remove the suboptimal points from the endogenous grids created at the EGM step. We present this subroutine afterwards. Finally, we demonstrate how the DC-EGM efficiently handles credit constraints.

3.1 Finding all solutions to the Euler equation

Because retirement is an absorbing state and retirees only choose consumption, invoking the DC-EGM algorithm is only necessary for solving the workers problem. The consumption/savings problem of the retirees can be solved using the standard EGM method (Carroll, 2006) at very low computational cost. The Euler equation for the worker's problem defined by equations (3), (11) and (12) is given by[8]

$$u'(c_t) \quad = \quad \beta R \mathbb{E}_t \Big[\sum_{j=0,1} u'\big(c_{t+1}(M_{t+1}|d_{t+1}=j)\big) P_{t+1}(d_{t+1}=j|M_{t+1}), \Big] \tag{15}$$

where $P_{t+1}(d_{t+1}|M_{t+1})$ denote conditional choice probabilities over the discrete retirement decision in the *following* period, d_{t+1}. With the assumption of extreme value type I distributed unobserved taste shocks, these choice probabilities have simple logistic form. If there is no taste shocks, $\sigma_\varepsilon = 0$, the choice probabilities reduce to indicator functions.

Conditional on a particular value of the current decision, d_t, we follow the EGM algorithm and form an exogenous *ascending* grid over end-of-period wealth,[9] $\vec{A}_t = \{A^1, \dots, A^G\}$ where $A^j > A^{j-1}$, $\forall j \in \{2, \dots, G\}$ and G is the number of discrete grid points used to approximate the continuous consumption policy function. Because the end-of-period wealth is a *sufficient statistic* for the consumption decision in the current period, the next period resources are given by

$$M_{t+1}(\vec{A}_t) = R\vec{A}_t + d_t y \eta_{t+1}. \tag{16}$$

The utility function in (1) is analytically invertible, therefore the current period consumption for

[8]See Appendix A for derivation.

[9]Referred to as the post-decision state in the operations research literature, Powell (2007).

all discrete labor market choices d_t can be calculated *directly* using the inverted Euler equation

$$c_t(\vec{A}_t|d_t) = (u')^{-1}\Big(\beta \text{RHS}\big(M_{t+1}(\vec{A}_t)\big)\Big), \tag{17}$$

where $\text{RHS}\big(M_{t+1}(\vec{A}_t)\big)$ is the right hand side of (15) evaluated at the points $M_{t+1}(\vec{A}_t)$ using the next period optimal consumption rules $c_{t+1}(M_{t+1}|d_{t+1})$. Finally, combining the current consumption $c_t(\vec{A}_t|d_t)$ found in (17) with the points of $\vec{A}_t$ we get the *endogenous grid* over the current period wealth

$$M_t(\vec{A}_t) = c_t(\vec{A}_t|d_t) + \vec{A}_t. \tag{18}$$

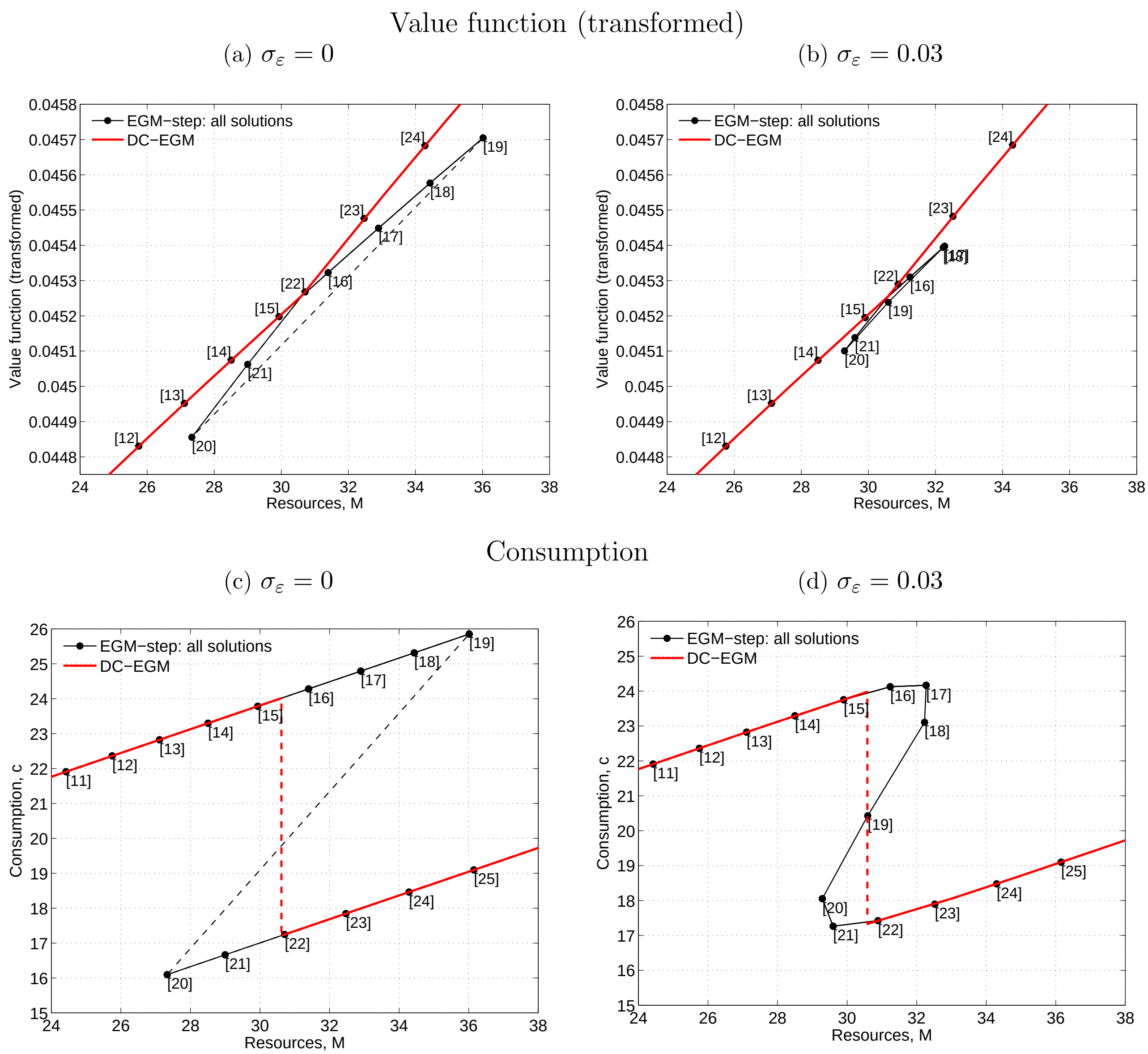
Finally, evaluating the maximand of the equation (12) at the points $c_t(\vec{A}_t|d_t)$, we compute the choice specific value function $v_t\big(M_t(\vec{A}_t)|d_t\big)$.

Algorithm 1 provides a pseudo-code of the described part of the DC-EGM which we call *the EGM step*. The current period discrete choice, d_t, and the next period policy and value functions are inputs to this routine, while the endogenous grid $\vec{M}_t = M_t(\vec{A}_t|d_t)$ and the d_t-specific consumption and value functions, $c_t(\vec{M}_t|d_t) = c_t(\vec{A}_t|d_t)$ and $v_t\big(\vec{M}_t|d_t\big) = v_t\big(M_t(\vec{A}_t)|d_t\big)$ computed on this grid are the outputs.

Figure 5 plots a selection of values of $v_t\big(\vec{M}_t|d_t\big)$ and $c_t(\vec{M}_t|d_t)$ against the endogenous grid $\vec{M}_t$. The points are indexed in the ascending order of the end-of-period wealth forming the grid $\vec{A}_t$. The solid lines approximate the corresponding functions with linear interpolation. It is evident that the interpolated discrete choice specific value function $v_t\big(M|d_t\big)$ is a *correspondence rather than a function* of M because of the existence of the region where multiple values of $v_t\big(M|d_t\big)$ correspond to a single value of M. The same is true for the interpolated discrete choice specific consumption function. The right and the left panels of Figure 5 illustrate the setting with and without the taste shocks respectively. Adding taste shocks with a relatively low variance, $\sigma_\varepsilon = 0.03$, reduces the size of the regions with multiple corresponding values. Dashed lines illustrate discontinuities.

The region where multiple values of $v_t\big(M|d_t\big)$ correspond to a single value of M is the clear evidence of non-concavity of the value function in the following period, and subsequent multiplicity of solutions of the Euler equation. The EGM step approximates *all solutions* to the Euler equation (see Lemma 2 in Appendix A), but because some of these solutions do not correspond to the optimal choices, the value function correspondence has to be cleaned of the suboptimal points to obtain actual value function. We should emphasize, however, that the points produced by the EGM step

Figure 5: Non-concave regions and the elimination of the secondary kinks in DC-EGM.



Notes: The plots illustrate the output from the EGM-step of the DC-EGM algorithm (Algorithm 1) in a non-concave region. The dots are indexed with the index j of the ascending grid over the end-of-period wealth $\vec{A}_t = \{A^1, \dots, A^G\}$ where $A^j > A^{j-1}$, $\forall j \in \{2, \dots, G\}$. The connecting lines show the d_t-specific value functions $v_t(\vec{M}_t|d_t)$ and the consumption function $c_t(\vec{M}_t|d_t)$ linearly interpolated on the endogenous grid $\vec{M}_t$. computed on this grid are the outputs. The left panels illustrate the deterministic case without taste shocks, while in the right panels $\sigma_\varepsilon = 0.03$. The "true" solution, after applying the DC-EGM algorithm is illustrated with a thick solid red line. Dashed lines illustrate discontinuities. The solution is based on $G = 70$ grid points in $\vec{A}_t$, $R = 1$, $\beta = 0.98$, $y = 20$, $\sigma_\eta = 0$.

Algorithm 1 The EGM-step: d_t choice-specific consumption and value functions

1: **Inputs:** Current decision d_t. Choice-specific consumption and value functions $c_{t+1}(\vec{M}_{t+1}|d_{t+1})$ and $v_{t+1}(\vec{M}_{t+1}|d_{t+1})$ associated with the endogenous grid in period $t+1$, $\vec{M}_{t+1}$

2: Let $\vec{\eta} = \{\eta^1, \ldots, \eta^Q\}$ be a vector of quadrature points with associated weights, $\vec{\omega} = \{\omega^1, \ldots, \omega^Q\}$

3: Form an ascending grid over end-of-period wealth, $\vec{A}_t = \{A_t^1, \ldots, A_t^G\}$ where $A_t^j > A_t^{j-1}$, $\forall j \in \{2, \ldots, G\}$

4: **for** $j = 1, \ldots, G$ **do** (Loop over points in $\vec{A}_t$)

5:     **for** $q = 1, \ldots, Q$ **do** (Loop over quadrature points in $\vec{\eta}$)

6:         Compute $M_{t+1}^q(A^j) = RA^j + d_t y \eta_{t+1}^q$

7:         **for** $d_{t+1} = 0, 1$ **do**

8:             Compute $c_{t+1}(M_{t+1}^q(A^j)|d_{t+1})$ by interpolating $c_{t+1}(\vec{M}_{t+1}|d_{t+1})$ at the point $M_{t+1}^q(A^j)$

9:             Compute $v_{t+1}\big(M_{t+1}^q(A^j)|d_{t+1}\big)$ by interpolating $v_{t+1}(\vec{M}_{t+1}|d_{t+1})$ at the point $M_{t+1}^q(A^j)$

10:         **end for**

11:         Compute $\phi_{t+1}\big(M_{t+1}^q(A^j)\big) = \sigma_\varepsilon \log\big(\sum_{j=0,1} \exp(v_{t+1}\big(M_{t+1}^q(A^j)|d_{t+1}=j\big))/\sigma_\varepsilon\big)$

12:         Compute $P_{t+1}(d_{t+1}|M_{t+1}^q(A^j)) = \exp(v_{t+1}\big(M_{t+1}^q(A^j)|d_{t+1}\big)/\sigma_\varepsilon)(\sum_{j=0,1} \exp(v_{t+1}\big(M_{t+1}^q(A^j)|d_{t+1}=j\big))/\sigma_\varepsilon)^{-1}$

13:     **end for**

14:     Compute $\text{RHS}\big(M_{t+1}(A^j)\big) = \beta R \sum_{q=1}^{Q} \sum_{j=1,2} \omega^q \cdot u'\big(c_{t+1}(M_{t+1}^q(A^j)|d_{t+1}=j)\big) \cdot P_{t+1}(d_{t+1}=j|M_{t+1}^q(A^j))$

15:     Compute expected value function $EV_{t+1}\big(M_{t+1}(A^j)\big) = \sum_{q=1}^{Q} \omega^q \cdot \phi_{t+1}\big(M_{t+1}^q(A^j)\big)$

16:     Compute current consumption $c_t(A^j|d_t) = u'^{-1}\Big(\text{RHS}\big(M_{t+1}(A^j)\big)\Big)$

17:     Compute value function $v_t\big(M_t(A^j)|d_t\big) = u(c_t(A^j|d_t)) + \beta EV_{t+1}\big(A^j\big)$

18:     Compute endogenous grid point $M_t(A^j|d_t) = c_t(A^j|d_t) + A_t^j$

19: **end for**

20: Collect the points $M_t(A^j|d_t)$ from the endogenous grid $\vec{M}_t = \{M_t(A^j|d_t), j = 1, \ldots, G\}$ associated with the choice-specific consumption and value functions: $c_t\big(\vec{M}_t|d_t\big) = \{c_t\big(M_t(A^j)|d_t\big), j = 1, \ldots, G\}$, and $v_t\big(\vec{M}_t|d_t\big) = \{v_t\big(M_t(A^j)|d_t\big), j = 1, \ldots, G\}$

21: **Outputs:** $\vec{M}_t$, $c_t(\vec{M}_t|d_t)$ and $v_t\big(\vec{M}_t|d_t\big)$

Notes: The pseudo code is written under the assumption that quadrature rules are used for calculating the expectations, whereas particular implementations can employ other methods for computing the expectation. It is also assumed that interpolation rather than approximation is used in Steps 8 and 9, although the latter is also possible.

necessarily contain the true solutions. This is a notable contrast to the standard solution methods based on an exogenous grid over wealth, which may struggle to find the points of optimality and have to deploy computationally costly global search methods to solve the optimization problem in the Bellman equation.

The next section describes a procedure in DC-EGM algorithm that deals with selecting the true optimal points among the points produced by the EGM step. The true solution found by the full DC-EGM algorithm is illustrated in Figure 5 with a red line for reference.

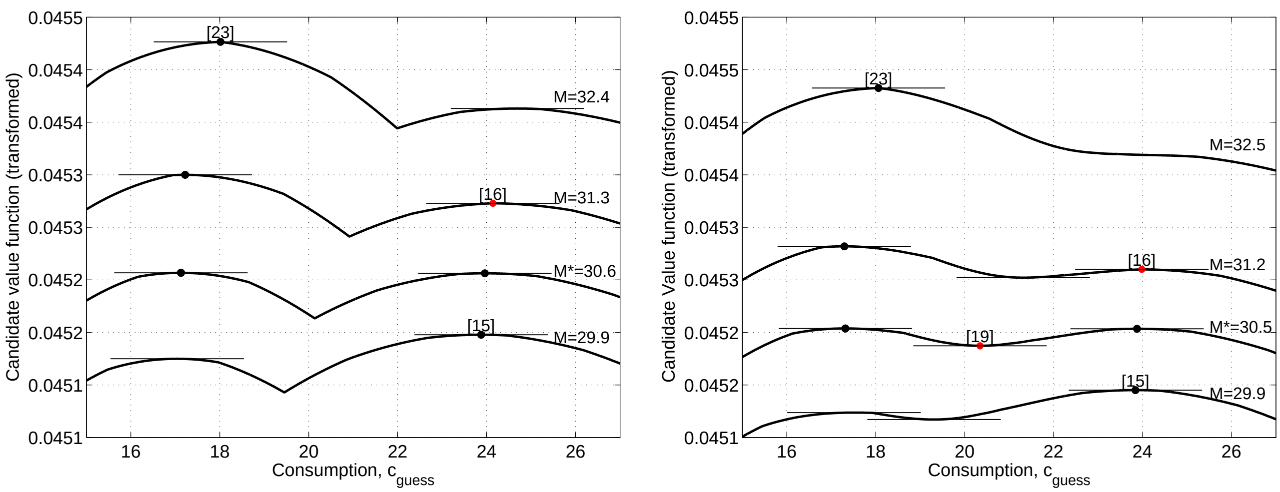
3.2 Calculation of the Upper Envelope

To distinguish between the optimal and suboptimal points produced by the EGM step, the DC-EGM algorithm makes a direct comparison of the values associated with each of the choices. On the plots of the discrete choice specific value function correspondences in Figure 5 (panels a and b), this amounts to computing the *upper envelope* of the correspondence in the regions of M_t where multiple solutions are found.

To provide deeper insight into this process, we plot the maximand of the equation (12) that defines the discrete choice specific value function $v_t(M_t|d_t)$ in Figure 6 as a function of consumption c_{guess} for various values of M_t. The value of $v_t(M_t|d_t)$ is the *global* maximum of the this function. The EGM step (Algorithm 1), however, recovers all critical points where the derivative of the plotted function is zero.[10] The same points in Figures 6 and 5 are indexed with the same indexes for easy comparison.

Figure 6: Local maxima and multiple solutions of the Euler equation.

(a) $\sigma_\varepsilon = 0$ (b) $\sigma_\varepsilon = 0.03$



Notes: The figure plots the maximand of the equation (12), which defines the discrete choice specific value function $v_t(M_t|d_t = 1)$, for the case of $\sigma_\varepsilon = 0$ (panel a) and $\sigma_\varepsilon = 0.03$ (panel b). Horizontal lines indicate the critical points found or approximated by the EGM step of DC-EGM algorithm. The points are indexed with the same indexes as in Figure 5 and the black dots represent global maxima. Model parameters are identical to those of Figure 5.

In the case without taste shocks, $\sigma_\varepsilon = 0$ (panel a), two levels of consumption satisfy the Euler equation (15) in the range $M_t \in [27, 36]$. From Figure 5 we know that points indexed 16 to 21 are suboptimal. Panel (a) in Figure 6 illustrates that the maximand function computed for wealth

[10]More specifically, because the grid $\vec{A}_t$ is finite, for every distinct point of the endogenous grid $\vec{M}_t = M_t(\vec{A}_t)$ it recovers one of the local maxima that corresponds to one of the solutions to the Euler equation. The other local maxima are approximated by interpolation of the value function correspondence between the points of the endogenous grid $\vec{M}_t = M_t(\vec{A}_t)$.

M_t in this range has two local maxima. For example, the 15th point from the EGM step is the global maximum of the maximand computed at $M_t \approx 29.9$, while the 16th point is not the global maximum when resources are $M_t \approx 31.3$.

At some point, the two solutions originating from the two segments of the value function correspondence are *both* optimal. Around $M_t \approx 30.6$ in panel (a) of Figure 6, the decision maker is indifferent between the discrete choices (at the next or some future periods – depending on whether the multiplicity of the solutions was caused by the primary or secondary kink of the next period value function). At this point of indifference, the consumption function is discontinuous, as illustrated with the red dashed line in panel (c) in Figure 5. The intersection point is not necessarily found in the EGM-step outlined above and needs to be additionally computed.[11]

In the smooth case with $\sigma_\varepsilon = 0.03$ the problem of multiplicity of local maxima in the maximand of equation (12) is generally still present, as shown by panel (b) of Figure 6. Correspondingly, there is still a discontinues drop in consumption around M_t around $M_t \in [29, 31]$. In other words, the taste shocks with scale parameter $\sigma_\varepsilon = 0.03$ do not fully "convexify" the value function. Note that in the smooth case there can be three solutions to the Euler equation, only one of which is a global maximum. This configuration is dealt with by the same upper envelope method.

It is clear, that selecting the global maximum among the critical points located by solving the Euler equation during the EGM step amounts to comparing the values of the constructed value function correspondence $v_t(M_t|d_t)$ for each M_t. For comparison, the overlapping segments of $v_t(M_t|d_t)$ may have to be re-interpolated on some common grid, and the upper envelope has to be computed. Algorithm 2 presents the pseudo-code of this calculation. The key insight of the upper envelope algorithm is to use the monotonicity of the end-of-period resources as a function of wealth to detect the regions where multiple values of choice-specific value function $v(M_t|d_t)$ are returned for a single value of M_t (see Step 3 of Algorithm 2). Monotonicity of end-of-period wealth is due to the concavity of the utility function as shown in Theorem 3, see Appendix A. Around every such detected region, the value function correspondence is broken into three segments (Steps 5 to 7), which are then compared point-wise to compute the upper envelope (Step 12). The inferior points are simply dropped from the endogenous grid $\vec{M}_t$, and optionally the approximated kink points at the inserted. Consequently, the consumption and value function correspondences are cleaned up and become *functions*.

[11] In presence of taste shocks, finding the precise indifference points is not essential, but in deterministic settings finding exact intersection points considerably increases the accuracy of the solution.

Algorithm 2 Upper envelope refinement step

1: **Inputs:** Endogenous grid $\vec{M}_t = M_t(\vec{A}_t)$ obtained from the grid over the end-of-period resources $\vec{A}_t = \{A^1, \dots, A^G\}$ where $A^j > A^{j-1}, \forall j \in \{2, \dots, G\}$; saving and value function *correspondences* $c_t(\vec{M}_t|d_t)$ and $v_t\big(\vec{M}_t|d_t\big)$ computed on $\vec{M}_t$
2: **for** $j = 2, \dots, G$ **do** (Loop over the points of endogenous grid)
3:   **if** $M_t(A^j) < M_t(A^{j-1})$ **then** (Criterion for detecting non-concave regions)
4:     Find the first $h \geq j$ such that $M_t(A^h) < M_t(A^{h+1})$
5:     Let $J_1 = \{j' : j' \leq j-1\}$ (Points up to [19] in panel a and [17] in panel b of Figure 5)
6:     Let $J_2 = \{j' : j-1 \leq j' \leq h\}$ (Points [19], [20] in panel a and [17]-[20] in panel b of Figure 5)
7:     Let $J_3 = \{j' : h \leq j'\}$ (Points [20] and up in both panel a and b of Figure 5)
8:     Let $\vec{M}' = \{M_t(A^{j'}) : \min_{i \in J_2} M_t(A^i) = M_t(A^h) \leq M_t(A^{j'}) \leq M_t(A^{j-1}) = \max_{i \in J_2} M_t(A^i)\}$
9:     **for** $i = 1, \dots, |\vec{M}'|$ **do** where $|\vec{M}'|$ is the number of points in $\vec{M}'$
10:       Denote $v_t\big(\vec{M}_t|d_t, J_r\big)$ the segment of $v_t\big(\vec{M}_t|d_t\big)$ computed on the points in the set J_r
11:       Interpolate the segments $v_t\big(\vec{M}_t|d_t, J_r\big)$ at the point $M_t(A^i)$ if $i \notin J^r$, $r = 1, \dots, 3$
12:       **if** $v_t\big(M_t(A^i)|d_t\big) < \max_r v_t\big(M_t(A^i)|d_t, J^r\big)$ **then**
13:         Drop point i from the endogenous grid $\vec{M}_t$
14:       **end if**
15:     **end for**
16:     Find the point $M^\times : v_t\big(M^\times|d_t, J^3\big) = v_t\big(M^\times|d_t, J^1\big)$ [Optional]
17:     Insert $M^\times$ into $\vec{M}_t$ first with associated values $v_t\big(M^\times|d_t, J^3\big)$ and $c_t\big(M^\times|d_t, J^3\big)$ [Optional]
18:     Insert $M^\times$ into $\vec{M}_t$ then with associated values $v_t\big(M^\times|d_t, J^1\big)$ and $c_t\big(M^\times|d_t, J^1\big)$ [Optional]
19:     Set $j = h$
20:   **else**
21:     Keep point j on the endogenous grid $\vec{M}_t$ as is
22:   **end if**
23: **end for**
24: **Outputs:** Refined endogenous grid $\vec{M}_t$, consumption and value *functions* $c_t(\vec{M}_t|d_t)$ and $v_t\big(\vec{M}_t|d_t\big)$

Note: The pseudo code is written using an elementary algorithm for calculation of the upper envelope for a collection of functions defined on their individual grids. More efficient implementations could also be used, see for example (Hershberger, 1989). Inserting the intersection point $M^\times$ into the endogenous grid $\vec{M}_t$ *two times* in step 17 and 18 ensures an accurate representation of the discontinuity in consumption function $c_t(\vec{M}_t|d_t)$. If the optional steps 16-18 are skipped, the secondary kink is smoothed out, but the overall shapes of the consumption and value functions are correct.

While the DC-EGM is similar to the approach proposed in Fella (2014), we explicitly allow for extreme value type I taste shocks to preferences and show how they help with the computational issues specific to the model of discrete-continuous choices. The approach in Fella (2014) does not readily apply to the class of models with taste shocks but should be adjusted along the lines described here. Particularly, the DC-EGM operates with discrete choice specific value functions and optimal consumption rules, and computes integrals of smooth objects. Furthermore, contrary to Fella (2014) who uses instances of increasing marginal utility to detect non-concave regions, DC-EGM uses the value function correspondence. However, both approaches rely on monotonicity of the optimal end-of-period savings function.

Algorithm 3 The DC-EGM algorithm

1: In the terminal period T fix a grid $\vec{M}_T$ over the consumable wealth M_T. On this grid compute consumption rules $c_T(\vec{M}_T|d_T) = \vec{M}_T$ and value functions $v_T(\vec{M}_T|d_T) = (\log(\vec{M}_T) - d_T)$ for every value of discrete choices d_T. This provides the base for backward induction in time
2: **for** $t = T-1, \dots, 1$ **do** (Loop backwards over the time periods)
3: **for** $j = \{0,1\}$ **do** (Loop over the current period discrete choices)
4: Invoke the EGM step (Algorithm 1) with $d_t = j$, $c_{t+1}(\vec{M}_{t+1}|d_{t+1})$ and $v_{t+1}(\vec{M}_{t+1}|d_{t+1})$ as inputs
5: Invoke upper envelope (Algorithm 2) using outputs from Step 4, $\vec{M}_t$, $c_t(\vec{M}_t|d_t)$ and $v_t\big(\vec{M}_t|d_t\big)$ as inputs
6: The endogenous grid $\vec{M}_t$ and consumption and value functions $c_t(\vec{M}_t|d_t)$ and $v_t\big(\vec{M}_t|d_t\big)$ are now computed
7: **end for**
8: **end for**
9: The collection of the choice-specific consumption and value functions $c_t(\vec{M}_t|d_t)$ and $v_t(\vec{M}_t|d_t)$ defined on the endogenous grids $\vec{M}_t$ for $d_t = \{0,1\}$ and $t = \{1, \dots, T\}$ constitutes the solution of the consumption/savings and retirement model

Algorithm 3 presents the pseudo-code of the full DC-EGM algorithm, which invokes the EGM step (Algorithm 1) repeatedly to compute the value function correspondences for all discrete choices, and then finds and removes all suboptimal points on the returned endogenous grids by calling the upper envelope module (Algorithm 2).

An important question of how the method handles the situations when the non-convex regions go undetected due to relatively coarse grid $\vec{A}_t$ is addressed by the Monte Carlo simulations in the next section. We show that even with small number of endogenous grid points the Nested Fixed Point (NFXP) Maximum Likelihood estimator based on the DC-EGM algorithm performs well and is able to identify the structural parameters of the model.

3.3 Credit Constraints

Before turning to the Monte Carlo results, we briefly discuss how DC-EGM handles the credit constraints, $c_t \leq M_t$. During the EGM step, the credit constraints are dealt with in exactly same manner as in Carroll (2006). Let the smallest possible end-of-period resources $A^1 = 0$ be the first point in the grid $\vec{A}_t$. Assuming that the corresponding point of the endogenous grid $M_t(A^1|d_t)$ is positive[12], it holds that $A_t(M|d_t) = 0$ for all $M \leq M_t(A^1|d_t)$ due to the monotonicity of saving function $A_t(M|d_t) = M - c_t(M|d_t)$ (see Theorem 3 in Appendix A). Therefore, the optimal consumption in this region is then given by $c_t(M|d_t) = M$, and the choice-specific value function is

$$v_t(M|d_t) = \log(M) - d_t\delta_t + \beta \int EV_{t+1}(d_t y \eta_{t+1}) f(d\eta_{t+1}), \quad M \leq M_t(A^1|d_t). \tag{19}$$

[12]It is not hard to show that this holds as long as the per period utility function satisfies the Inada conditions.

Note that the third component of (19) is the expected value of having zero savings. It is calculated within the EGM step for the point $A^1 = 0$, and should be saved separately as a constant that depends on d_t but not on M_t. Once this constant is computed, $v_t(M|d_t)$ essentially has analytical form in the interval $[0, M_t(A^1|d_t)]$, and thus can be directly evaluated at any point.

When the per-period utility function is additively separable in consumption and discrete choice like in the retirement model we consider, (19) holds for all $d_t \in D_t$ in the interval $0 \le M \le \min_{d_t \in D_t} M_t(A^1|d_t)$. In other words, the choice specific value functions for low wealth have the same shape (in our case $\log(M)$), which is shifted vertically with d_t-specific coefficients. This implies that the logistic choice probabilities $P_t(d_t|M_t)$ are constant in this interval, and have to only be calculated once.

4 Monte Carlo Results

In this section we investigate the properties of the *approximate* maximum likelihood estimator (MLE) that we obtain using the DC-EGM to approximate the model solution in the inner loop of the Nested Fixed Point algorithm. We specifically focus on role of income uncertainty and taste shocks for the approximation bias induced by a numerical solution with a finite number of grid-points; in particular how approximation bias depends on the number of grid points in smooth as well as non-smooth problems. After a description of the data generating process (DGP), we present the results from a series of Monte Carlo experiments, and show that models used in typical empirical applications are sufficiently smooth to almost eliminate approximation bias using relatively few grid points.

4.1 Data Generation Process

For the Monte Carlo we consider a slightly more general formulation of the consumption/savings and retirement problem defined in (1) with constant relative risk aversion (CRRA) utility

$$\max_{\{c_t, d_t\}_1^T} \sum_{t=1}^{T} \beta^t \left(\frac{c_t^{1-\rho} - 1}{1-\rho} - \delta_t d_t \right) \tag{20}$$

where ρ is the CRRA coefficient.

In order to simulate synthetic data from the DGP consistent with the model and the vector

Table 1: Baseline *true* parameter values.

Description	Value	Description	Value
Time horizon	$T = 44$	Disutility of work	$\delta = 0.5$
Gross interest rate	$R = 1.03$	Discount factor	$\beta = 0.97$
Full time employment income	$y = 1.0$	CRRA coefficient	$\rho = 2.0$
Income variance	$\sigma_\eta = 0$	Taste shocks scale	$\sigma_\varepsilon \in \{0.01, 0.05\}$

of *true* parameter values, we solve the model *very accurately* with 2,000 grid points using the DC-EGM. We refer to this solution as the *true solution* even though this is off course only an accurate finite approximation of the value function.[13]

We consider several specifications of the model in the Monte Carlo experiments below to study various aspects of the performance of the estimator. Once again, we assume that disutility of work is constant over time, i.e. $\delta_t = \delta$. Table 1 presents the parameter values in the baseline specification of the model. Deviations are given explicitly with every Monte Carlo experiment separately. We perform 200 replications for each combination of settings.

For each specification of the model, 50,000 individuals are simulated for $T = 44$ periods. Each individual i is initiated as full-time worker $s^d_{i,1} = 1$, where we have used $s^d_{i,t} \in \{0, 1\}$ to denote the labor market state, i.e. whether an individual is retired ($s^d_{i,t} = 0$) or working ($s^d_{i,t} = 1$). Each workers initial wealth $M^d_{i,1}$ is drawn from a uniform distribution on the interval $[0, 100]$. At the beginning of each time period t, a random log-normal labor market income shock η_t with variance parameter σ_η is drawn if the individual i is working and individual's resources M^d_t are calculated. Given the level of resources, discrete-choice specific value functions and choice probabilities are computed, and a random draw determines which discrete labor market option d^d_{it} is chosen. After one period lag, the labor force participation decision becomes the labor market state, $s^d_{i,t+1} = d^d_{it}$. The optimal level of consumption, c_{it}, is then computed conditional on d^d_{it}, and the end-of-period wealth is calculated and stored to be used for calculation of resources available in the beginning of period $t + 1$, $M^d_{i,t+1}$. We then add normal additive measurement error with standard deviation $\sigma_\xi = 1$ to get the simulated consumption data, c^d_{it}. This produces simulated panel data $(M^d_{it}, s^d_{it}, d^d_{it}, c^d_{it})$ for each individual $i \in \{1, \ldots, 50,000\}$ in all time periods $t \in \{1, \ldots, 44\}$.

[13] As a spot check, we have also compared this solution with the traditional value function iteration approach, where we used a grid search over 1,000 discrete points on the interval $[0, M_t]$ to locate the optimal consumption for each value of wealth. We find that results are essentially identical.

4.2 Maximum Likelihood Estimation

We implement a discrete-continuous version of the Nested Fixed Point (NFXP) Maximum Likelihood estimator devised in Rust (1987, 1988), where we augment the original discrete-choice estimator with a measurement error approach when assessing the likelihood of the observed continuous choices.

Assume that a panel dataset is available, $\{(M^d_{it}, s^d_{it}, d^d_{it}, c^d_{it})\}_{i=\{1,\dots,N\},\, t=\{1,\dots,T\}}$, containing observations on wealth, labor market state, discrete and continuous choices of individuals $i = 1\dots, N$ in time periods $t = 1, \dots, T$. Let $c_t(M_t, s_t, d_t|\theta)$ denote the consumption policy function computed by the DC-EGM for a given vector of model parameters $\theta = (\delta, \beta, \rho, \sigma_\eta, \sigma_\varepsilon)$. We assume that consumption is observed with additive Gaussian measurement error,

$$c^d_{it} = c_t(M^d_{it}, s^d_{it}, d^d_{it}|\theta) + \xi_{it},\ \xi_{it} \sim N(0, \sigma_\xi),\ \text{i.i.d. } \forall i, t. \tag{21}$$

Let $\xi^d_{it}(\theta) = c^d_{it} - c_t(M^d_{it}, s^d_{it}, d^d_{it}|\theta)$ denote the difference between the predicted and the observed consumption. We assume that the measurement error, ξ_{it}, is independent of the taste shocks, $\varepsilon_t(d_t)$, and, thus, the joint likelihood of observation i in period t is given by

$$\ell_{it}(\theta, \sigma_\xi) = P(d^d_{it}|M^d_{it}, s^d_{it}, \theta)\frac{\phi(\xi^d_{it}(\theta)/\sigma_\xi)}{\sigma_\xi}, \tag{22}$$

where $\phi(\cdot)$ is the density function of the standard normal distribution. We have ignored the controlled transition probability for the retirement status s^d_{it}, since $P_{tr}(s^d_{it}|s^d_{i,t-1}, d^d_{i,t-1})$ is always 1 in the data when retirement is absorbing and the labor market state is perfectly controlled by the decision.

The choice probabilities for the workers ($s^d_{it} = 1$) are standard logits

$$P(d^d_{it}|M^d_{it}, s^d_{it}, \theta) = \frac{\exp(v_t(M^d_{it}, s^d_{it}, d^d_{it}|\theta)/\sigma_\varepsilon)}{\sum_{j=0}^{1}\exp(v_t(M^d_{it}, s^d_{it}, j|\theta)/\sigma_\varepsilon)} \tag{23}$$

and are computed from the discrete choice specific value functions $v_t(M^d_{it}, s^d_{it}, d^d_{it}|\theta)$ found by the DC-EGM given a particular value of the parameter vector θ, evaluated at the data. Because retirement is absorbing and thus retirees do not have any discrete choice to make, the first component of individual likelihood contribution (22) drops out when $s^d_{it} = 0$.

The joint log-likelihood function is given by $\tilde{\mathcal{L}}(\theta, \sigma_\xi) = \log\prod_i^N\prod_t^T \ell_{it}(\theta, \sigma_\xi)$ where re-arranging the first order condition with respect to σ_ξ^2 yields the standard ML estimator for the measurement error variance, $\sigma_\xi^2(\theta) = \frac{1}{NT}\sum_{t=1}^{T_i} \xi_{it}^d(\theta)^2$. The concentrated log-likelihood function is, therefore, proportional to

$$\mathcal{L}(\theta) \propto \sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{s_{it}^d}{\sigma_\varepsilon}\Big(v_t(M_{it}^d s_{it}^d, d_{it}^d|\theta) - EV_t(M_{it}^d, s_{it}^d|\theta)\Big) - \frac{1}{2}\log\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\xi_{it}^d(\theta)^2\right)\right\}, \tag{24}$$

where $EV_t(M_{it}^d, s_{it}^d|\theta)$ is the the logsum given in (13) evaluated at parameter value θ.[14] The parameter vector $\hat{\theta}$ that maximizes (24) is the ML estimator of the model parameters.

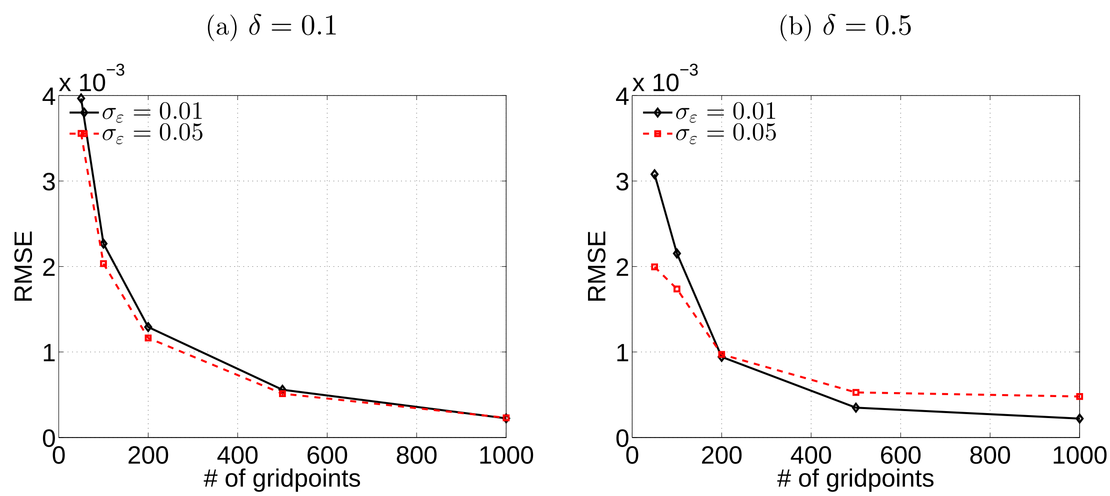
4.3 Taste Shocks as Unobserved State Variables

We are now ready to investigate the effects of smoothing on the accuracy of the ML estimator based on the DC-EGM algorithm. We conduct two Monte Carlo experiments where we vary the degree of smoothing induced by extreme value taste shocks and income uncertainty respectively. Throughout, we focus on estimating the parameter that index disutility of work, δ, while keeping all other fixed at their true values. The Appendix D contains average estimation time for the DC-EGM.

Taste Shocks and Approximation Error. Figure 7 displays the root mean square error (RMSE) of the parameter estimates for the disutility of work, $\hat{\delta}$. Results are shown for varying degree of smoothing, $\sigma_\varepsilon \in \{0.01, 0.05\}$, and different values of the disutility of work parameter, $\delta \in \{0.1, 0.5\}$. With RMSE around $1.0e^{-3}$, the proposed estimator is already accurate with 50 grid points and rapidly improves as the number of grid points increase from 50 through 1000. Note that standard errors will of course increase with σ_ε due to the increased amount of unexplained variation in the error term and RMSE reflects this too. Bearing this is mind, it is evident that the approximation bias decreases as the degree of smoothing increases, i.e., larger values of σ_ε. For higher levels of smoothing, problems with multiplicity of the Euler equation solutions disappear and few grid points are needed to approximate the (smooth) consumption function. This is particularly true when the disutility from work is large ($\delta = .5$) because the non-concave regions are larger in this case. We also calculated the Monte Carlo Standard Deviation (MCSD)[15], which is on the order $1.0e^{-4}$ irrespectively of the number of grid points used.

[14] Following (24), the logsum only has to be evaluated for workers, $s_{it}^d = 1$.
[15] MCSD results not shown

Figure 7: Monte Carlo results: disutility of work.



Notes: The plots illustrate the root mean square error (RMSE) of $\hat{\delta}$. Results are shown for varying degree of smoothing, $\sigma_\varepsilon \in \{0.01, 0.05\}$, and different values of the disutility of work, $\delta \in \{0.1, 0.5\}$. The rest of the parameters are at their baseline levels, see Table 1.

Income Uncertainty. Additional uncertainty about, e.g., future labor market income tend to smooth out *secondary* kinks stemming from multiple solutions to the Euler equations. To illustrate how that additional smoothing affects the proposed estimator, Figure 8 display RMSE when introducing *income uncertainty.* We report results from two different values of the income variance[16], $\sigma_\eta^2 \in \{0.001, 0.05\}$. The first level, 0.001, does not completely smooth out secondary kinks while the significantly more uncertain income process with $\sigma_\eta^2 = 0.05$ does (see the right panel of Figure 3).

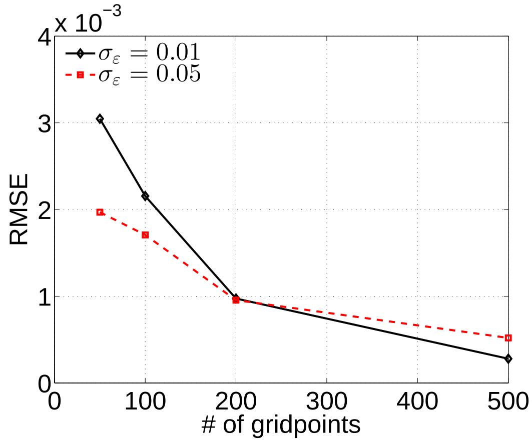
Income uncertainty together with taste shocks smooth the problem to such a degree that the RMSE drops by an order of magnitude when increasing the income variance from .001 to 0.05. Hence, using only few grid points when estimating such a model will result in only minor approximation errors.

As mentioned, standard errors will of course increase with σ_ε due to the increased amount of unexplained variation. The MCSD is quite small and unaffected by the degree of income uncertainty as well as the number of grid points, but increases from 0.00023 to 0.00045 as σ_ε increases from 0.01 to 0.05. This is the main explanation for why RMSE is only smaller for a small

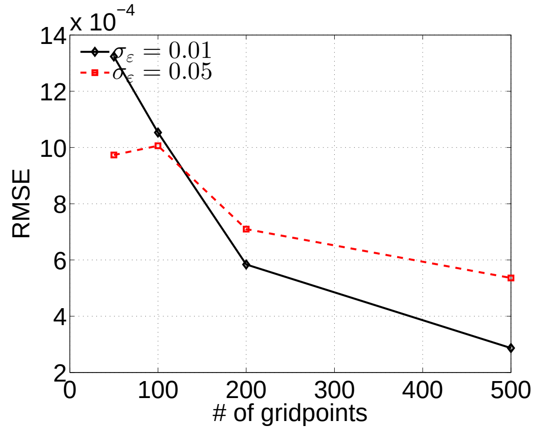
[16]The values of the income variance we use correspond well to the empirical findings, for example in Gourinchas and Parker (2002); Meghir and Pistaferri (2004); Imai and Keane (2004).

Figure 8: Monte Carlo results: income uncertainty.

(a) $\sigma_\eta^2 = 0.001$



(b) $\sigma_\eta^2 = 0.05$



Notes: The plots illustrate the root mean square error (RMSE). Results are shown for varying degree of smoothing, $\sigma_\varepsilon \in \{0.01, 0.05\}$, and different values of the income variance, $\sigma_\eta^2 \in \{0.001, 0.05\}$. The rest of parameters are at their baseline levels, see Table 1.

number of grid points. Sorting out this effect its clear that increasing σ_ε decreases the amount of pure approximation bias - especially when the number of grid points is small. Note that MCSD is very small, in part due to a relatively large sample size, but also because the variance of the iid extreme value error term is extremely small. In most empirical applications, σ_ε would be larger; leading to an even smoother problem than the one we consider here. Hence, with relatively few grid points we can expect to obtain an even smaller approximation bias induced by the finite grid approximation in the DC-EGM.

4.4 Taste Shocks as Logit Smoother

Until now we have assumed that the correct model has unobserved state variables, and thus $\sigma_\varepsilon > 0$ has to be estimated. To investigate how the proposed estimator performs if the data stems from a model in which there are *no* unobserved states, we estimate versions of the model where we *impose* $\sigma_\varepsilon > 0$ and, thus, estimate a misspecified model. This is interesting because if researchers have reasons to believe that the underlying model has no shocks, the inclusion of these shocks acts as a smooth approximation to the true deterministic model. As argued above, solving the smoothed model is much faster since it requires fewer grid points and, thus, is much faster to estimate.

Figure 9 illustrates the RMSE when using 50, 100 and 500 grid points for various levels of smoothing $\sigma_\varepsilon \in [0.001, 0.05]$ while the correct level is $\sigma_\varepsilon = 0$. Intuitively, as the model becomes "more" misspecified (increasing the imposed σ_ε), the RMSE and the MCSD increases. Interestingly, for a given number of discrete grid points, the RMSE is minimized by a $\sigma_\varepsilon > 0$. While large degree of smoothing induces significant approximation bias, the bias is initially falling in σ_ε until some point at which the RMSE increases again. The minimum of the RMSE is attained for lower levels of smoothing if additional stochasticity (i.e. income shocks) is present in the model. This is expected because the income uncertainty smooths the problem and less logit smoothing is needed to obtain the optimal smooth approximation. It is worth noting, however, that the optimal amount of logit smoothing may not be sufficient to completely eliminate the non-convexities in the model. It is therefore essential for the solution method to be able to robustly solve optimization problems with multiple local solutions, the task that DC-EGM performs particularly well.

These results show the potential for great speed gains by smoothing. Using only 50 grid points and imposing $\sigma_\varepsilon = 0.01$ produce a RMSE of around the same level as using 500 grid points and imposing $\sigma_\varepsilon \approx 0$ close to the true model. We can reduce the number of gridpoints by an order of magnitude without increasing the root mean square error significantly simply by choosing the degree of smoothing appropriately. Note, however, that there is naturally a trade-off between lowering the computational cost by increasing smoothing and decreasing the number of grid points and the accuracy of the resulting solution compared to the true solution of the non-smooth model.

5 Discussion and Conclusions

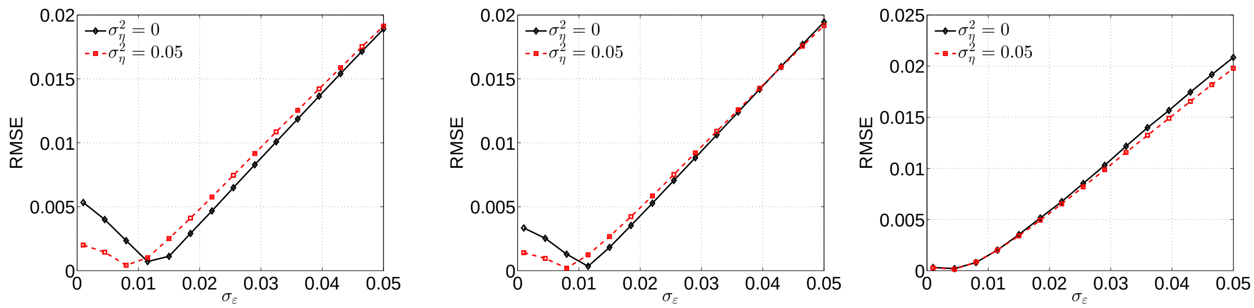
In this paper we have shown how complications from numerous discontinuities in the consumption function to a life cycle model with discrete and continuous choices can be avoided by smoothing the problem and using the DC-EGM algorithm. The proposed algorithm retains all the nice features of the original EGM method, namely that it typically does not require any iterative root-finding operations, and is equally efficient in dealing with borrowing constraints. Moreover, we show that the smoothed model can be successfully estimated by the NFXP estimator based on the DC-EGM algorithm even with small number of grid points, and even when the true DGP is non-smooth.

For expositional clarity, we focused on a simple illustrative example when explaining the details of the DC-EGM algorithm. This also allows us to derive an analytical solution that we can compare to the numerical one. The analytical solution provides economic intuition for why first

Figure 9: Monte Carlo results: true model without taste shocks (misspecified)

Root mean square error

(a) 50 grid points (b) 100 grid points (c) 500 grid points



Notes: The plots illustrate the root mean square error (RMSE) from estimation of a *misspecified* model. The model from which data are simulated is deterministic, $\sigma_\varepsilon = 0$, while the model used to estimate the disutility of work imposes $\sigma_\varepsilon > 0$. Results are shown for varying degree of *imposed* smoothing, $\sigma_\varepsilon \in [0.001, 0.05]$ on the horizontal axes, different levels of income shocks, $\sigma_\eta \in \{0, 0.05\}$, and different number of grid points. The rest of parameters are at their baseline levels, see Table 1.

and second order kinks appear and permits direct evaluation of the precision of the DC-EGM algorithm. Admittedly, the illustrative model of consumption and retirement is very stylized, and the reader may wonder if DC-EGM can be used to solve and estimate larger, more complex and realistic models with more state variables, multiple discrete alternatives, heterogeneous agents, institutional constraints, etc.. The answer is positive. As shown in the Appendix A, the DC-EGM method can be applied to a much more general class of problems as long as the post decision state variable is a sufficient static for the continuous choice in the current period, and the marginal utility function and intra-temporal budget constraint are invertible. When the marginal utility function is *analytically* invertible, DC-EGM also avoids the bulk of costly root-finding operations.[17]

The DC-EGM method has been implemented in several recent empirical applications, where it has proven to be a powerful tool for solving and estimating more complex DC models in various fields: labor supply, human capital accumulation and saving (Iskhakov and Keane, 2016); joint retirement decision of couples (Jørgensen, 2014); consumption, housing purchases and housing debt (Yao, Fagereng and Natvik, 2015); saving decisions and fertility (Ejrnæs and Jørgensen, 2015); precautionary borrowing and credit card debt (Druedahl and Jørgensen, 2015).

[17]DC-EGM algorithm can also be generalized for other specifications including the models with large state space and multidimensional discrete choice. White (2015); Iskhakov (2015); Druedahl and Jørgensen (2016) present theoretical foundations for extending endogenous grid methods to multi-dimensional models.

We have demonstrated in the Monte Carlo experiments that the NFXP maximum likelihood estimator based on the DC-EGM solution algorithm performs very well when decisions are made under uncertainty, e.g. in the presence of extreme valued taste shocks and the existence of income uncertainty. Even when the true model is deterministic, taste shocks can be used as a powerful smoothing device to simplify the solution without much approximation bias due to over-smoothing.

The addition of extreme value taste shocks is not only a convenient smoothing device that simplifies the solution of DC models, it is also an empirically relevant extension required to avoid statistical degeneracy of the model. In empirical applications the variance of these shocks is typically much larger compared to what we have considered here. This makes models smooth enough to almost eliminate approximation bias in parameter estimates even with relatively few grid points. We therefore conclude that DC-EGM is both practical and appears to be a fast and accurate method for use in actual empirical applications.

However from the standpoint of using DC-EGM to find highly accurate solutions to DC problems, while the results we present in this paper are highly encouraging, our conclusions are based on comparing the numerical solution to an analytical solution of a *particular* DC problem. It would be much better to be able to *prove* that DC-EGM provides similar accuracy for an entire *class* of DC problems. Ideally this would be done by deriving bounds on the error between the true decision rule and the approximate decision rule computed by DC-EGM. We conjecture that these bounds, dependent on the number of endogenous gridpoints n used in the DC-EGM algorithm, would converge to zero as $n \to \infty$. We are not aware of any error bounds or convergence proofs even in more straightforward case of concave dynamic problems with only continuous choice that EGM was originally developed for by Carroll (2006). We believe the uniform bounds we derived for the approximation error involved in the use of extreme value smoothing of DC problems may provide one of the tools to derive bounds on the error between the true decision rules and the approximate decision rules calculated by DC-EGM.

A Theoretical foundations of DC-EGM

For the purpose of this Appendix we consider the following more general formulation of the consumption/savings and retirement problem. Let M_t denote consumable wealth that is continuous state variable with particular motion rule described below, and let s_t denote a vector of additional discrete or discretized state variables. Let c_t be the scalar continuous decision (consumption) and d_t be a scalar discrete decision variable with finite set of values that could encode multiple discrete decisions if needed. Consider the dynamic discrete-continuous choice problem given by the Bellman equation,

$$V_t(M_t, s_t) = \max_{0 \le c_t \le M_t, d_t \in D_t} \Big[u(c_t, d_t, s_t) + \sigma_\varepsilon \varepsilon_t(d_t) + \beta_t E_t \big\{ V_{t+1}(M_{t+1}, s_{t+1}) | A_t, d_t \big\} \Big], \qquad (25)$$

where $t = 1, \dots, T-1$, and the last component of the maximand is absent for $t = T$. The choices in the model are restricted by the credit constraint $c_t < M_t$ and feasibility sets D_t. The per period utility includes scaled taste shocks $\sigma_\varepsilon \varepsilon_t(d_t)$, where ε_t is a vector of i.i.d. Extreme Value (Type I) distributed random variables. The dimension of ε_t is equal to the number of alternatives that the discrete choice variable may take, $\varepsilon_t(d_t)$ denotes the component that corresponds to a particular discrete decision. In the general case the discount factor β_t is time-specific to allow for the probability of survival. The expectation is taken over the taste shocks ε_{t+1}, transition probabilities of the state process s_t as well as any serially uncorrelated (or idiosyncratic) shocks that may affect M_{t+1} and s_{t+1}. The expectation is taken conditional on the choices in period t using the *sufficient statistic* $A_t = M_t - c_t$ in place of the continuous (consumption) choice.

Using the well known representation of the expectation of the maximum of Extreme Value distributed random variables, the Bellman equation (25) can be written in terms of the deterministic choice-specific value functions $v_t(M_t, s_t|d_t)$ as

$$\begin{aligned} v_t(M_t, s_t|d_t) &= \max_{0 \le c_t \le M_t} \Big[u(c_t, d_t, s_t) + \beta_t E_t \big\{ V_{t+1}(M_{t+1}, s_{t+1}) | A_t, d_t \big\} \Big] && (26) \\ &= \max_{0 \le c_t \le M_t} \Big[u(c_t, d_t, s_t) + \beta_t E_t \Big\{ \phi\big(v_{t+1}(M_{t+1}, s_{t+1}|d_{t+1}), D_{t+1}, \sigma_\varepsilon \big) | A_t, d_t \Big\} \Big], && (27) \end{aligned}$$

where $\phi(x_j, J, \sigma) = \sigma \log \Big[\sum_{j \in J} \exp \frac{x_j}{\sigma} \Big]$ is the logsum function. The expectation in (27) is now only taken w.r.t. state transitions and idiosyncratic shocks, unlike in (25) and (26).

The crucial assumption for the DC-EGM is that *post decision* state A_t constitutes the sufficient statistic for the continuous choice in period t, i.e. that transition probabilities/densities of the state process (M_t, s_t) depend on A_t rather than M_t or c_t directly. It is also required that A_t as a function of M_t is (analytically) invertible. For our case, assume for concreteness that $A_t = M_t - c_t$, and that $M_{t+1} = RA_t + y(d_t)$, where R is a gross return, and $y(d_t)$ is discrete choice specific income. We also assume that the utility function $u(c_t, d_t, s_t)$ satisfies the following condition.

Assumption 1 (Concave utility)**.** The instantaneous utility $u(c_t, d_t, s_t)$ is concave[18] in c_t and has a monotonic derivative w.r.t. c_t that is (analytically) invertible.

Lemma 1 (Smoothed Euler equation)**.** *The Euler equation for the problem* (25) *takes the form*

$$u'(c_t, d_t, s_t) = \beta_t R \mathbb{E}_t \left[\sum_{d_{t+1} \in D_{t+1}} u'\big(c_{t+1}(M_{t+1}, s_{t+1}|d_{t+1}), d_{t+1}, s_{t+1}\big) P_{t+1}\big(d_{t+1}|M_{t+1}, s_{t+1}\big) \right] \qquad (28)$$

where $u'(c_t, d_t, s_t)$ is the partial derivative of the utility function w.r.t. c_t, $c_{t+1}(M_{t+1}, s_{t+1}|d_{t+1})$ is the choice-specific consumption function in period $t+1$, and $P_{t+1}(d_{t+1}|M_{t+1}, s_{t+1})$ is the conditional

[18]More precisely, a weaker condition is sufficient, namely for every x and arbitrary $\Delta_1 > 0$ and $\Delta_2 > 0$ it must hold that $u(c_t + \Delta_1, d_t, s_t) - u(c_t, d_t, s_t) \ge u(c_t + \Delta_1 + \Delta_2, d_t, s_t) - u(c_t + \Delta_2, d_t, s_t)$, see Theorem 3.

discrete choice probability in period $t+1$, given by

$$P_t(d_t|M_t, s_t) = \exp\big(v_t(M_t, s_t|d_t)/\sigma_\varepsilon\big) \Big/ \sum_{d \in D_t} \exp\big(v_t(M_t, s_t|d)/\sigma_\varepsilon\big). \tag{29}$$

Proof. Discrete choice specific consumption functions $c_t(M_t, s_t|d_t)$ satisfy the the first order conditions for the maximization problems in (26) given by

$$u'(c_t, d_t, s_t) + \beta_t E\left\{\frac{\partial V_{t+1}(M_{t+1}, s_{t+1})}{\partial M_{t+1}} \frac{\partial M_{t+1}}{\partial c_t}\right\} = 0 \tag{30}$$

for every value of $d_t \in D_t$. The envelope conditions for (26)

$$\frac{\partial v_t(M_t, s_t|d_t)}{\partial M_t} = \beta_t E\left\{\frac{\partial V_{t+1}(M_{t+1}, s_{t+1})}{\partial M_{t+1}} \frac{\partial M_{t+1}}{\partial M_t}\right\}, \tag{31}$$

and because $\partial M_{t+1}(d_t) \big/ \partial M_t = R = -\partial M_{t+1}(d_t) \big/ \partial c_t$, it holds for all d_t and $t = 1, \dots, T-1$

$$u'(c_t, d_t, s_t) = \frac{\partial v_t(M_t, s_t|d_t)}{\partial M_t}. \tag{32}$$

The first order condition for (27) is

$$u'(c_t, d_t, s_t) = \beta_t R \mathbb{E}_t\left[\sum_{d_{t+1} \in D_{t+1}} \frac{\partial v_{t+1}(M_{t+1}, s_{t+1}|d_{t+1})}{\partial M_{t+1}} P_{t+1}\big(d_{t+1}|M_{t+1}, s_{t+1}\big)\right], \tag{33}$$

where choice probabilities $P_{t+1}\big(d_{t+1}|M_{t+1}, s_{t+1}\big)$ are given by (29). Plugging (32) into (33) completes the proof. □

The DC-EGM algorithm outlined in Algorithm 3 is readily applicable to the general formulation of the discrete-continuous problem (25), expect for the extra loop that has to be taken over all additional states s_t in Step 3 (Algorithm 3). The expectation over the transition probabilities of the state process is calculated together with the expectation over the other stochastic elements of the model in Algorithm 1.

Lemma 2 (All solutions)**.** *As the auxiliary grid over end-of-period wealth $\vec{A}$ becomes dense on a closed interval $[0, \bar{A}]$ for some upper bound $\bar{A}$, in the sense that the maximum distance between two adjacent points A^j and A^{j+1} approaches zero, the EGM step of DC-EGM algorithm is guaranteed to find all solutions of the Euler equation* (28) *that imply the end-of-period wealth on the interval* $[0, \bar{A}]$.

Proof. Following the Algorithm 1 denote $\textsc{rhs}\big(M_{t+1}(A^j)|d_t\big)$ the right hand side of the Euler equation (28) as a function of the points of the end-of-period wealth grid $\vec{A}$ conditional on discrete choice d_t in period t. The EGM step of the DC-EGM algorithm computes

$$\begin{cases} c_t(A^j|d_t) & = u'^{-1}\Big(\textsc{rhs}\big(M_{t+1}(A^j)\big)\Big), \\ M_t(A^j|d_t) & = u'^{-1}\Big(\textsc{rhs}\big(M_{t+1}(A^j)\big)\Big) + A^j. \end{cases} \tag{34}$$

Both equations in (34) are well defined functions of A^j provided that the utility function $u(\cdot)$ satisfies the Assumption 1. Thus, the system constitutes a well defined *parametric* specification of the curve *composed of the solutions to the Euler equation* $c(M_t, s_t|d_t)$ for all s_t, d_t, where A^j plays the role of a parameter. In the limit as A^j runs through all the values on the interval $[0, \bar{A}]$, all solutions that imply the end-of-period wealth from this interval are found. □

The criteria for selecting the solutions of the Euler equation that correspond to the optimal behavior in the model is based on the monotonicity of the savings function, which is established with the following theorem[19].

Theorem 3 (Monotinicity of savings function)**.** *Denote* $A_t(M_t, s_t|d_t) = M_t - c_t(M_t, s_t|d_t)$ *a discrete choice specific savings function in period* t*. Under the Assumption 1, function* $A_t(M, s_t|d_t)$ *is monotone non-decreasing in* M *for all* t*,*s_t *and* $d_t \in D_t$*.*

Proof. Theorem 3 is an application of Theorem 4 in Milgrom and Shannon (1994) to the current problem. Conditional savings function $A_t(M_t, s_t|d_t)$ is a maximizer in the expression similar to (26) for the discrete choice specific value function $v_t(M_t, s_t|d_t)$. As a function of M and A, the maximand in this expression is given by

$$f(A, M) = u(M - A, d_t, s_t) + \beta_t E_t\big\{V_{t+1}(M_{t+1}(A), s_{t+1})\big\} \tag{35}$$

where $M_{t+1}(A)$ is next period wealth as an *increasing* function of A. It is necessary and sufficient to show that $f(A, M)$ is quasisupermodular in A and satisfies the single crossing property in (A, M). The former is trivial because A is a scalar. For the latter consider $A' > A''$, $M' > M''$ and assume $f(A', M'') > f(A'', M'')$. Then

$$\begin{gathered} f(A', M') - f(A'', M') = \\ = u\left(M' - A', d_t, s_t\right) - u\left(M' - A'', d_t, s_t\right) + \\ +\beta_t\left[EV_{t+1}\left(M_{t+1}(A'), s_{t+1}\right) - EV_{t+1}\left(M_{t+1}(A''), s_{t+1}\right)\right] \geq \\ \geq u\left(M'' - A', d_t, s_t\right) - u\left(M'' - A'', d_t, s_t\right) + \\ +\beta_t\left(EV_{t+1}\left(M_{t+1}(A'), s_{t+1}\right) - EV_{t+1}\left(M_{t+1}(A''), s_{t+1}\right)\right) = \\ f(A', M'') - f(A'', M'') > 0. \end{gathered} \tag{36}$$

For the first inequality we use

$$\begin{gathered} u\left(M' - A', d_t, s_t\right) - u\left(M' - A'', d_t, s_t\right) \geq u\left(M'' - A', d_t, s_t\right) - u\left(M'' - A'', d_t, s_t\right), \\ u\left(M' - A', d_t, s_t\right) - u\left(M'' - A', d_t, s_t\right) \geq u\left(M' - A'', d_t, s_t\right) - u\left(M'' - A'', d_t, s_t\right), \\ u\left(z, d_t, s_t\right) - u\left(z - \Delta_M, d_t, s_t\right) \geq u\left(z + \Delta_A, d_t, s_t\right) - u\left(z + \Delta_A - \Delta_M, d_t, s_t\right), \end{gathered} \tag{37}$$

where $z = M' - A'$, $\Delta_A = A' - A'' > 0$, $\Delta_M = M' - M'' > 0$, and which is due to Assumption 1, i.e. concavity of the utility function. It follows then that $f(A', M') > f(A'', M')$. Similarly, assumption $f(A', M'') \geq f(A'', M'')$ leads to $f(A', M') \geq f(A'', M')$, and thus $f(A, M)$ satisfies the single crossing property, and monotonicity theorem in Milgrom and Shannon (1994) applies. □

[19] A similar monotonicity result is also used in Fella (2014).

B Spurious Discontinuities from Numerical Integration

To illustrate how naive numerical quadrature integration can produce spurious discontinuities in the policy function, we here focus on the illustrative model without smoothing. Particularly, for working households, the smoothed Euler equation in (15) collapses to

$$\begin{aligned} u'\left(c_t\left(M_t|d_t\right)\right) &= \beta\int_0^\infty Ru'(c_{t+1}(M_{t+1}|d_{t+1}=1))\cdot\mathbf{1}\{M_{t+1}\leq\overline{M}_{t+1}\}f(d\eta) \\ &+ \beta\int_0^\infty Ru'(c_{t+1}(M_{t+1}|d_{t+1}=0))\cdot\mathbf{1}\{M_{t+1}>\overline{M}_{t+1}\}f(d\eta). \end{aligned} \tag{38}$$

where we recall that $M_{t+1}=R\left(M_t-c_t\left(M_t|d_t\right)\right)+y\eta$. With the change of variables, $q=f(\eta)$, we can write the Euler equation (38) as

$$\begin{aligned} u'\left(c_t\left(M_t|d_t\right)\right) &= \beta\int_0^{\overline{q}_t} f^{-1}(q)u'\left(c_{t+1}\left(R(M_t-c_t(M_t|d_t))+yf^{-1}(q),d_{t+1}=1\right)\right)dq \\ &+ \beta\int_{\overline{q}_t}^1 f^{-1}(q)u'\left(c_{t+1}\left(R(M_t-c_t(M_t|d_t))+yf^{-1}(q),d_{t+1}=0\right)\right)dq \end{aligned} \tag{39}$$

where the threshold $\overline{q}_t$ is given by

$$\overline{q}_t = f\left(\frac{\overline{M}_{t+1}}{M_{t+1}}\right). \tag{40}$$

As long as the income shock distribution is not degenerate, the resulting Euler equation (39) is continuous and smooth in $c_t(M_t,\mathbb{W})$ through M_{t+1} in spite of the discontinuity in the consumption function $c_{t+1}(M_{t+1},\mathbb{W})$ at $M_{t+1}=\overline{M}_{t+1}$. In turn, this suggests that numerical integration should be done twice – once for each case – to ensure that the integral is well-behaved.

In contrast, the naive Euler equation in (38) is discontinuous in $c_t(M_t,\mathbb{W})$. When using numerical quadrature to evaluate the integral, for a given level of resources, some of the nodes will result in $M_{t+1}\leq\overline{M}_{t+1}$ while others will result in the opposite case. For concreteness, say that 10 nodes are used and the five lowest nodes results in $M_{t+1}\leq\overline{M}_{t+1}$. Say also that for a slightly larger value of current resources perhaps only four nodes satisfy $M_{t+1}\leq\overline{M}_{t+1}$ while now six invokes the alternative. When comparing the solution found in the two (close) values of current period resources, there will be a discontinuous change in the optimal consumption. In the current model, this would result in spurious downward kinks in the consumption function around a secondary kink, as illustrated in the left panel of Figure 3.

C Derivation of the analytical solution to consumption retirement problem

It is straightforward to show using backward induction that the value function for a retiree at age $T-t$ (i.e. t periods before end of life) is a logarithmic function of M[20]

$$v_{T-t}(M|d=0)=\log(M)\left(\sum_{i=0}^t\beta^i\right)+A_t \tag{41}$$

[20] See Phelps (1962); Hakansson (1970).

where

$$A_{T-t} = -\log\left(\sum_{i=0}^{t}\beta^i\right)\left(\sum_{i=0}^{t}\beta^i\right) + \beta[\log(\beta)+\log(R)]\left[\sum_{i=0}^{t-1}\beta^i\left(\sum_{j=0}^{t-1-i}\beta^j\right)\right]. \tag{42}$$

The optimal consumption rule for a retiree is linear in M

$$c_{T-t}(M) = M\left(\sum_{i=0}^{t}\beta^i\right)^{-1}. \tag{43}$$

Recalling that $v_t(M|d=1)$ is the discounted utility of a person of age $T-t$ who decides to work (not retire), we can define the optimal retirement threshold at age t, $\overline{M}_t$ as the value of M that makes the person indifferent between retiring and not retiring at that age

$$v_t(\overline{M}_t|d=0) = v_t(\overline{M}_t|d=1). \tag{44}$$

Since we assume $\delta > 0$ (positive disutility from working), it will be optimal for a person of age t to retire if $M \geq \overline{M}_t$ and work otherwise. We will have a non-convex kink in the value function for working $v_t(M|d=1)$ at the point $\overline{M}_t$ since we have

$$V_t(M) = \max[v_t(M|d=0), v_t(M|d=1)]. \tag{45}$$

As we show below, the two decision-specific value functions are strictly concave and intersect only once at a point $\overline{M}_t$ that we provide an explicit expression for below. We show that $v_t(M|d=1) > v_t(M|d=0)$ for $M < \overline{M}_t$ so it is optimal to work in this region, and $v_t(M|d=1) < v_t(M|d=0)$ for $M > \overline{M}_t$, so it is optimal to retire in this region.

Let $c_t(M|d=0)$ be the optimal consumption of a retiree of age t. This function is given by formula (43) above (with trivial re-indexing). The optimal consumption of a individual who decides not to retire is $c_t(M|d=1)$ given by

$$c_t(M|d=1) = \underset{0\leq c\leq M}{argmax}\,[log(c) - \delta + \beta V_{t+1}(R(M+y_t-c))] \tag{46}$$

The overall optimal consumption rule is then given by

$$c_t(M) = \begin{cases} c_t(M|d=1) & \text{if} \quad M < \overline{M}_t \\ c_t(M|d=0) & \text{if} \quad M \geq \overline{M}_t. \end{cases} \tag{47}$$

It is easy to see that due to the non-convex kink in the value function at $\overline{M}_t$ the optimal consumption function $c_t(M)$ will have a discontinuity at $\overline{M}_t$, and

$$c_t(\overline{M}_t|d=1) > c_t(\overline{M}_t|d=0). \tag{48}$$

This result follows from the condition that

$$V_t^{'-}(\overline{M}_t) < V_t^{'+}(\overline{M}_t). \tag{49}$$

Since there is a kink at $\overline{M}_t$, the derivative $V_t^{'-}(\overline{M}_t)$ must be interpreted as the left hand derivative (derivative from below $\overline{M}_t$), and correspondingly $V_t^{'+}(\overline{M}_t)$ is the right hand derivative of V_t at $M = \overline{M}_t$.

We now establish these results by backward induction, starting at period $T-1$ which is the first period where the consumption/retirement decision is non-trivial (it is easy to see that in the final period of life, it is optimal to retire and consume all remaining savings). For notational simplicity, we drop the time subscripts on income, $y = y_t$, since income is constant here. To derive a formula for the retirement threshold $\overline{M}_{T-1}$ consider the $T-1$ optimization problem

$$c_{T-1}(M|d=1) = \underset{0 \leq c \leq M}{argmax} \left[\log(c) - \delta + \beta \log(R(M-c)+y) \right]. \tag{50}$$

The solution to this is given by

$$c_{T-1}(M|d=1) = \begin{cases} M & \text{if} \quad M < y/R\beta \\ (M + y/R)/(1+\beta) & \text{if} \quad y/R\beta \leq M \leq \overline{M}_{T-1} \end{cases} \tag{51}$$

Note that the worker is liquidity constrained when $M < y/R\beta$ and in this region it is optimal to consume all of her beginning of period savings M and rely on the end of period payment of wage earnings y to finance consumption in her last period of life, T. The value function for the worker at age $T-1$ is

$$v_{T-1}(M|d=1) =$$
$$\begin{cases} \log(M) - \delta + \beta\log(y) & \text{if } M < y/R\beta \\ \log(M + y/R)(1+\beta) - \delta + \beta[\log(\beta) + \log(R)] - \log(1+\beta)(1+\beta) & \text{if } y/R\beta \leq M \leq \overline{M}_{T-1} \end{cases}$$

and the value function for a retiree is given by equation (41). Equating the values of work and retirement and solving for the optimal retirement threshold $\overline{M}_{T-1}$ we have

$$\overline{M}_{T-1} = \frac{(y/R)\exp\{-\delta/(1+\beta)\}}{1-\exp\{-\delta/(1+\beta)\}} \tag{52}$$

provided this is greater than $y/R\beta$ (the threshold below which the consumer is liquidity constrained), otherwise

$$\overline{M}_{T-1} = [y/(R\beta)](1+\beta)^{\frac{(1+\beta)}{\beta}} \exp\{-\delta/\beta\}. \tag{53}$$

However it is easy to see that assumption $\delta < (1+\beta)\log(1+\beta)$ implies that $\overline{M}_{T-1} > y/R\beta$. It is also easy to see that as the disutility of working $\delta \to \infty$ we have $\overline{M}_{T-1} \to 0$, and as $\delta \to 0$, then $\overline{M}_{T-1} \to \infty$, i.e. if there is no disutility of working, the person would never choose to retire.

Note also that at $\overline{M}_{T-1}$ there is a kink in the value function: this is a downward kink (in terms of Clausen and Strub (2013)) as the max of two concave functions $v_{T-1}(M|d=0)$ and $v_{T-1}(M|d=1)$, and this kink in the value function results in a *discontinuity* in the optimal consumption function $c_{T-1}(M)$. There is a drop in consumption equal to $(y/R)/(1+\beta)$ at $\overline{M}_{T-1}$, and with two remaining periods in their life, a retiree has a "marginal propensity to consume" out of wealth equal to $1/(1+\beta)$ the same as a worker. The discontinuous drop in consumption that occurs at the retirement threshold equals the present value of forgone earnings due to retirement, y_{T-1}/R, multiplied by the marginal propensity to consume out of wealth, $1/(1+\beta)$.

To summarize the solution at $T-1$, the optimal retirement threshold is $\overline{M}_{T-1}$ given in equation

(52) and the consumption function is given by

$$c_{T-1}(M) = \begin{cases} M & \text{if} \quad M < y/R\beta \\ (M + y/R)/(1+\beta) & \text{if} \quad y/R\beta \leq M \leq \overline{M}_{T-1} \\ M/(1+\beta) & \text{if} \quad M > \overline{M}_{T-1} \end{cases} \tag{54}$$

and the value function is given by

$$V_{T-1}(M) = \tag{55}$$
$$\begin{cases} \log(M) - \delta + \beta\log(y) & \text{if} \quad M < y/R\beta \\ \log(M + y/R)(1+\beta) - \delta + \beta[\log(\beta) + \log(R)] - \log(1+\beta)(1+\beta) & \text{if} \quad y/R\beta \leq M \leq \overline{M}_{T-1} \\ \log(M)(1+\beta) + \beta[\log(\beta) + \log(R)] - \log(1+\beta)(1+\beta) & \text{if} \quad M > \overline{M}_{T-1}. \end{cases}$$

Now consider going back one more time period in the backward recursion, to $T-2$. We want to illustrate the possibility of *secondary kinks/discontinuities* in the consumption function for a worker $c_{T-2}(M,1)$ caused by the kinks in $V_{T-1}(M)$. Let $\overline{M}_{T-2}$ denote the *primary kink* due to the retirement threshold at $T-2$ and let $\overline{M}^{r_j}_{T-2}$ denote the secondary kinks, where $j = 1, \ldots, N_{T-2}$ and N_{T-t} is the number of secondary kinks t periods before the end of life at age T.

To see how these secondary kinks arise, consider how the $T-2$ consumption function is determined, as the solution to

$$c_{T-2}(M,1) = \underset{0 \leq c \leq M}{argmax} \left[\log(c) - \delta + \beta V_{T-1}(R(M-c) + y)\right]. \tag{56}$$

As shown above $V_{T-1}(M)$ has two kinks: one at $M = y/R\beta$ where the liquidity constraint stops being binding, and the other at $\overline{M}_{T-1}$ where the worker retires. Assume that the initial wealth of the worker at the start of period $T-1$ is low enough so that the worker will be liquidity constrained in period $T-1$. This implies that $R(M-c) + y < y/R\beta$. Then substituting the liquidity-constrained formula for $V_{T-1}(M)$ from (55) into the period $T-2$ optimization (56), we find that optimal consumption is given by $c_{t-2}(M,1) = (M + y/R)/(1+\beta)$. However imposing the liquidity constraint, we must also have $(M+y/R)/(1+\beta) \leq M$ which implies that $M \leq y/R\beta$, and it is easy to verify that for wealth satisfying this constraint, the worker will be liquidity constrained both in period $T-2$ and in period $T-1$ as well.

However for wealth above $y/R\beta$ the worker is no longer liquidity constrained in period $T-2$ but our derivation of the worker's consumption in period $T-2$ is still contingent on the assumption that the worker is liquidity constrained in period $T-1$. This will be true provided that the savings and earnings the worker brings to the start of period $T-1$, $R\beta(M + y/R)/(1+\beta)$, is less than $y/R\beta$, which is equivalent to the inequality $M \leq [y/(R\beta)^2](1 + \beta - R\beta^2)$. It is not hard to show that when $R = 1$ we have $y/\beta < (y/\beta^2)(1 + \beta - \beta^2)$ so the interval for which the consumer will consume $(M+y)/(1+\beta)$ is non-empty when $R = 1$. For $R > 1$ the inequality $y/(R\beta) < [y/(R\beta)^2](1+\beta-R\beta^2)$ is equivalent to $R\beta < 1$, so under this assumption this interval will also exist, otherwise the interval is empty and the consumer goes from consuming $c_{T-2}(M,1) = M$ to consuming an amount we derive below.

In the next region, wealth is sufficiently high in period $T-2$ so the consumer is not liquidity constrained at $T-2$ and the saving and earning will keep the consumer out of the liquidity constrained region at $T-1$, but the worker's wealth is not high enough to retire at $T-1$. The relevant expression for $V_{T-1}(M)$ in this case is given by the middle expression in equation (55). This implies an optimal consumption level equal to $c_{T-2}(M,1) = (M+y(1/R+1/R^2))/(1+\beta+\beta^2)$.

For even larger there will come a point where the consumer can save enough in period $T-2$ to retire in period $T-1$, i.e. savings will exceed the $\overline{M}_{T-1}$ threshold. Thus, there is some wealth level $\overline{M}^r_{T-2}$ at which the the relevant expression for the worker's period $T-1$ value function $V_{T-1}(M)$ is given by the last, retirement, formula in (55). The optimal consumption in this region is $c_{T-2}(M,1) = (M + y/R)/(1 + \beta + \beta^2)$. It is important to carefully check values of c such that savings, $M + y - c$ is in the "convex region" of $V_{T-1}(M)$ around the $T-1$ retirement threshold $\overline{M}_{T-1}$. In this region there will be *two local optima* for c, one involving the higher consumption $(M + y(1/R + 1/R^2))/(1 + \beta + \beta^2)$ and the other involving the lower consumption $(M + y/R)/(1 + \beta + \beta^2)$ that enables the worker to retire at $T-1$.

These two solutions are reflected in the two possible solutions to the first order condition for optimal consumption given by

$$0 = \frac{1}{c} - \begin{cases} (\beta+\beta^2)/(M - c + y(1/R + 1/R^2)) & \text{if } R(M-c)+y < \overline{M}_{T-1} \\ (\beta+\beta^2)/(M - c + y/R) & \text{if } R(M-c)+y \geq \overline{M}_{T-1} \end{cases} \tag{57}$$

For $M < \overline{M}^r_{T-2}$ the global optimum will be $c_{T-2}(M,1) = (M + y(1/R + 1/R^2))/(1 + \beta + \beta^2)$ and the consumer will be working in both periods $T-2$ and $T-1$. However for $M > \overline{M}^r_{T-2}$ the consumer will still work at $T-2$ (provided $M < \overline{M}_{T-2}$, the primary kink point at $T-2$, the wealth threshold at which the consumer retires at $T-2$) but will have enough savings to retire at $T-1$. The optimal consumption in this case will be $c_{T-2}(M,1) = (M + y/R)/(1+\beta+\beta^2)$. It is not hard to show that if $M \leq [y/(R\beta)^2](1+\beta-R\beta^2)$, then the quantity $R(M - c_{T-2}(M,1)) + y \leq y/R\beta$, i.e. the consumer will indeed be in the liquidity constrained region $M \leq y/R\beta$ at the start of $T-1$ as we assumed would be the case. We also have that $y/R\beta < [y/(R\beta)^2](1 + \beta - R\beta^2)$ provided that $R\beta \leq 1$, which we assume to be the case. Otherwise this region would be empty and the optimal consumption would be given by $c_{T-2}(M,1) = (M + y(1/R + 1/R^2))/(1+\beta+\beta^2)$ as derived above. We can check that this consumption function, which is also derived under the assumption that the consumer will not be liquidity constrained at period $T-1$, will result in total savings at $T-1$ that satisfies $R(M-c) + y \geq y/R\beta$ (so the consumer is not liquidity constrained at $T-1$) for wealth at $T-2$ at the lower end of this interval (i.e. at $M = y/R\beta$) provided that $R \leq 1/\beta$.

However, at $M = \overline{M}^r_{T-2}$ the consumer will be indifferent between consuming the larger amount $(M + y(1/R + 1/R^2))/(1+\beta+\beta^2)$ knowing they will *not* retire at $T-1$ and consuming the lower amount $(M + y/R)/(1 + \beta + \beta^2)$ and knowing they will retire at $T-1$. We find $\overline{M}^r_{T-2}$ as the solution to the following equation

$$\begin{aligned}
&\log\left((M + y(1/R + 1/R^2))/(1 + \beta + \beta^2)\right) + \\
&\beta V_{T-1}\left((y + R(M - (M + y(1/R + 1/R^2)))/(1 + \beta + \beta^2))\right) \\
&= \log\left((M + y/R)/(1 + \beta + \beta^2)\right) + \beta V_{T-1}\left(y + R(M - (M + y/R))/(1 + \beta + \beta^2)\right).
\end{aligned}$$

Thus, at $M = \overline{M}^r_{T-2}$ the consumer is indifferent between consuming the larger amount $(M + y(1/R + 1/R^2))/(1+\beta+\beta^2)$ or consuming the smaller amount $(M + y/R)/(1 + \beta + \beta^2)$ that provides the additional savings necessary to enable the consumer to retire at $T-1$.

Now we can express period $T-2$ consumption of the worker as the following piece-wise linear function:

$$c_{T-2}(M,1) = \begin{cases} M & \text{if} \quad M < y/R\beta \\ (M+y/R)/(1+\beta) & \text{if} \quad y/R\beta \le M \le [y/(R\beta)^2](1+\beta-R\beta^2) \\ (M+y(1/R+1/R^2))/(1+\beta+\beta^2) & \text{if} \quad [y/(R\beta)^2](1+\beta-R\beta^2) \le M \le \overline{M}^r_{T-2} \\ (M+y/R)/(1+\beta+\beta^2) & \text{if} \quad \overline{M}^r_{T-2} < M < \overline{M}_{T-2}. \end{cases} \tag{58}$$

It is straightforward to verify that $c_{T-2}(M,1)$ has two kinks at $M = [y/(R\beta)^2](1+\beta-R\beta^2)$ and $M = y/R\beta$ followed by a discontinuity at $M = \overline{M}^r_{T-2}$.

To derive the time $T-2$ retirement threshold $\overline{M}_{T-2}$ we solve for the value of M that makes the consumer indifferent between retiring at $T-2$ and working (but with enough wealth so that the person is above the secondary kink $\overline{M}^r_{T-2}$ where their consumption is given by $c_{T-2}(M,1) = (M+y/R)/(1+\beta+\beta^2)$)

$$\log(M)(1+\beta+\beta^2) + A_{T-2} = \log(M+y/R)(1+\beta+\beta^2) - \delta + A_{T-2} \tag{59}$$

where A_{T-2} is defined in equation (42) above. Note that the right hand side of (59) is the value function for a consumer who does not have enough wealth to retire at $T-2$, but since $M > \overline{M}^r_{T-2}$ (the secondary kink point), it follows that the appropriate formula for $V_{T-1}(M)$ will be the one where $M > \overline{M}_{T-1}$ in equation (55) above. The solution to this equation is $\overline{M}_{T-2}$ given by

$$\overline{M}_{T-2} = \frac{(y/R)e^{-K}}{(1-e^{-K})} \tag{60}$$

where K is given by

$$K = \frac{\delta}{(1+\beta+\beta^2)}. \tag{61}$$

Notice that formulas (60) and (52) imply that $\overline{M}_{T-1} < \overline{M}_{T-2}$, i.e. the wealth threshold for retirement decreases as one approaches the end of life, T.

To summarize the solution we found at $T-2$, the optimal retirement threshold $\overline{M}_{T-2}$ is the solution to equation (59), and the optimal consumption function is given by

$$c_{T-2}(M) = \begin{cases} M & \text{if} \quad M < y/R\beta \\ (M+y/R)/(1+\beta) & \text{if} \quad y/R\beta \le M \le [y/(R\beta)^2](1+\beta-R\beta^2) \\ (M+y(1/R+1/R^2))/(1+\beta+\beta^2) & \text{if} \quad [y/(R\beta)^2](1+\beta-R\beta^2) \le M \le \overline{M}^r_{T-2} \\ (M+y/R)/(1+\beta+\beta^2) & \text{if} \quad \overline{M}^r_{T-2} < M \le \overline{M}_{T-2} \\ M/(1+\beta+\beta^2) & \text{if} \quad M > \overline{M}_{T-2}. \end{cases} \tag{62}$$

The optimal consumption function at $T-2$ has two kinks at $M = y/R\beta$ (the level of wealth at which the consumer is no longer liquidity-constrained) and $M = [y/(R\beta)^2](1+\beta-R\beta^2)$, and two discontinuities: one at the secondary kink point $\overline{M}^r_{T-2}$ where consumption drops by $(y/R^2)/(1+\beta+\beta^2)$, and the other at the retirement threshold $\overline{M}_{T-2}$ where consumption drops by another $(y/R)/(1+\beta+\beta^2)$. Note that the secondary kink point $\overline{M}^r_{T-2}$ is precisely the amount of wealth where, while the consumer does not yet retire at $T-2$, *they know they will have enough to retire at* $T-1$. Thus, the drop in consumption at this secondary kink point can be regarded as *saving at* $T-2$ *for their anticipated retirement at time* $T-1$.

The value function at $T-2$ can be expressed this way:

$$V_{T-2}(M) = \begin{cases} \log(c_{T-2}(M)) - \delta + \beta V_{T-1}(R(M - c_{T-2}(M)) + y) & \text{if } \ M < \overline{M}_{T-2} \\ \log(M)(1+\beta+\beta^2) + A_{T-2} & \text{if } \ M \geq \overline{M}_{T-2} \end{cases} \tag{63}$$

Thus, depending on whether the person's wealth at $T-2$ is above or below the secondary kink point $\overline{M}^r_{T-2}$, they will know whether they will have enough (with their $T-2$ earnings y) to retire at $T-1$ or not, and will save/consume accordingly.

Now consider solving the problem at $t = T-3$, three periods before the end of life. The consumption rule will have three kinks including the level of M where the liquidity constraint no longer binds, and three discontinuities, including the retirement threshold $\overline{M}^r_{T-3}$ in period $T-3$. One additional kink in $c_{T-3}(M)$ is added above the end point $[y/(R\beta)^2](1+\beta-R\beta^2)$ of the first linear segment of $c_{T-2}(M)$ and reflects to the liquidity constraint in period $T-2$. The additional discontinuity corresponds to the secondary kink point $\overline{M}^r_{T-2}$.

Note the pattern here: $c_{T-1}(M)$ has one kink and one discontinuity, $c_{T-2}(M)$ has two kinks and two discontinuities, and $c_{T-3}(M)$ will have three kinks and three discontinuities. The important additional point to notice is that c_{T-1}, c_{T-2} and as we show shortly, c_{T-3}, are all *piecewise linear.*

It will be helpful to distinguish the points marking the sequence of connected linear segments of the consumption function due to kinks in the value function arising at the end of the liquidity constrained region $[0, y/R\beta]$ from those at higher levels of wealth that related to retirement decisions — both current retirement and anticipated future retirements. As we noted the will always be an initial linear segment over the interval $[0, y/R\beta]$ where $c_t(M) = M$ for $M \in [0, y/R\beta]$. Thus there will be a kink in the consumption function at $y/R\beta$ related to current period liquidity constraint. We have also shown that for $M > \overline{M}_t$ it will be optimal to retire, so there is a *discontinuity* in $c_t(M)$ at $\overline{M}_t$ which relates to the primary kink in the value function and the decision to retire in the current period.

However at ages $T-t < T-1$ in addition to these two "current period" kinks/discontinuities, there will be a set of kinks and discontinuities related to the future periods, i.e. "future liquidity constraint" kinks $\overline{M}^{l_j}_{T-t}$ and a set of "future retirement threshold" discontinuities $\overline{M}^{r_j}_{T-t}$. These discontinuities correspond to secondary kinks in the same period value function and result from the primary kinks in the value functions of all future periods.

Thus $c_{T-2}(M)$ has one future liquidity constraint kink $\overline{M}^{l_1}_{T-2}$ at $[y/(R\beta)^2](1+\beta-R\beta^2)$ and one future retirement threshold discontinuity at $\overline{M}^{r_1}_{T-2}$. The former represents the level of saving at which the consumer is not liquidity constrained at age $T-2$, but will be liquidity constrained at age $T-1$. The latter is the level of wealth that leads the worker to discontinuously reduce consumption at $T-2$ in order to have enough savings to retire at $T-1$.

In period $T-3$ there will be a total of tree discontinuities in $c_{T-3}(M)$. The last discontinuity occurs at the retirement threshold $\overline{M}_{T-3}$, but there will be two additional discontinuities at the secondary kink points in the value function V_{T-3}. These are denoted $\overline{M}^{r_1}_{T-3}$ and $\overline{M}^{r_2}_{T-3}$. We have the ordering $\overline{M}_{T-3} > \overline{M}^{r_1}_{T-3} > \overline{M}^{r_2}_{T-3}$. The highest secondary kink point $\overline{M}^{r_1}_{T-3}$ is the level of wealth that leads the consumer to save an amount (including current period wage earnings) of $\overline{M}_{T-2}$, which is the retirement threshold at period $T-2$. Thus at wealth levels that just exceed $\overline{M}^{r_1}_{T-3}$ the consumer works in period $T-3$ but discontinuously reduces consumption in order to have enough resources to retire in period $T-2$. At wealth levels that are just below $\overline{M}^{r_2}_{T-3}$, the consumer works in both periods $T-3$ and $T-2$, and retires only in period $T-1$.

The consumption function $c_{T-3}(M)$ will also have two future liquidity constraint kinks

$\overline{M}^{l_1}_{T-3} = [y/(R\beta)^2](1+\beta-R\beta^2)$ and $\overline{M}^{l_2}_{T-3}$ in addition to the current liquidity constraint at $M = y/R\beta$. The first kink will be at the level of saving that is sufficient for the consumer not to be liquidity-constrained at age $T-3$ but not enough to avoid being liquidity constrained at age $T-2$. At $\overline{M}^{l_1}_{T-3}$ the consumer switches from consuming according to the 2nd linear segment of $c_{T-3}(M) = (M+y/R)/(1+\beta)$ to consuming on the third linear segment $c_{T-3}(M) = (M+y(1/R+1/R^2))/(1+\beta+\beta^2)$.

At the second future liquidity constraint kink point $\overline{M}^{l_2}_{T-3}$ the worker has sufficient saving to not be liquidity constrained at both ages $T-3$ and $T-2$, but not enough to avoid being liquidity constrained at age $T-1$. At $\overline{M}^{l_2}_{T-3}$ the worker switches from consuming on the third segment of $c_{T-3}(M) = (M+y(1/R+1/R^2))/(1+\beta+\beta^2)$ to the fourth segment which is the first of the segments created by the retirement threshold kink points $\overline{M}^{r_j}_{T-3}$. Thus for wealth that exceeds $\overline{M}^{l_2}_{T-3}$ consumption switches to $c_{T-3}(M) = (M+y(1/R+1/R^2+1/R^3))/(1+\beta+\beta^2+\beta^3)$. Then for still higher levels of wealth the worker consumes according to the various piecewise linear segments demarcated by the successive future retirement threshold kink points $\overline{M}^{r_j}_{T-3}$, $j = 2,1$ and finally $\overline{M}_{T-3}$, the retirement threshold at period $T-3$.

Note that the marginal propensity to consume out of wealth is also piecewise linear and monotonically decreasing in M. In the liquidity constrained region the marginal propensity to consume is 1, and in the first of the liquidity constrained consumption segments it is $1/(1+\beta)$, and in the second liquidity constrained segment it is $1/(1+\beta+\beta^2)$. Then in the remaining retirement related consumption segments, the marginal propensity to consume out of wealth is constant and equal to $1/(1+\beta+\beta^2+\beta^3)$.

In summary, the consumption function $c_{T-3}(M)$ is given by

$$c_{T-3}(M) = \begin{cases} M & \text{if} \quad M < y/R\beta \\ (M+y/R)/(1+\beta) & \text{if} \quad y/R\beta \le M \le \overline{M}^{l_1}_{T-3} \\ (M+y(1/R+1/R^2))/(1+\beta+\beta^2) & \text{if} \quad \overline{M}^{l_1}_{T-3} \le M \le \overline{M}^{l_2}_{T-3} \\ (M+y(1/R+1/R^2+1/R^3))/(1+\beta+\beta^2+\beta^3) & \text{if} \quad \overline{M}^{l_2}_{T-3} \le M \le \overline{M}^{r_2}_{T-3} \\ (M+y(1/R+1/R^2))/(1+\beta+\beta^2+\beta^3) & \text{if} \quad \overline{M}^{r_2}_{T-3} \le M < \overline{M}^{r_1}_{T-3} \\ (M+y/R)/(1+\beta+\beta^2+\beta^3) & \text{if} \quad \overline{M}^{r_1}_{T-3} \le M < \overline{M}_{T-3} \\ M/(1+\beta+\beta^2+\beta^3) & \text{if} \quad \overline{M}_{T-3} < M \end{cases} \tag{64}$$

The retirement threshold $\overline{M}_{T-3}$ is given by

$$\overline{M}_{T-3} = \frac{(y/R)e^{-K}}{(1-e^{-K})}, \text{ where } K = \frac{\delta}{(1+\beta+\beta^2+\beta^3)}. \tag{65}$$

We solve for the secondary kinks/discontinuities $\{\overline{M}^{l_i}_{T-3}, \overline{M}^{r_j}_{T-3}\}$, $i = 1,2$ and $j = 1,2$ in the same way as we did for the period $T-2$: we solve for the level of a wealth that makes the consumer indifferent between consuming the higher level of consumption to the "left" of the kink point (more precisely the limit of consumption for wealth approaching the kink point from below) and the lower level of consumption to the "right" of the discontinuity (the limit of consumption for wealth approaching the kink point from above).

Finally, the value function is given by

$$V_{T-3}(M) = \begin{cases} \log(c_{T-3}(M)) - \delta + \beta V_{T-2}(R(M - c_{T-3}(M)) + y) & \text{if} \quad M < \overline{M}_{T-3} \\ \log(M)(1+\beta+\beta^2+\beta^3) + A_{T-3} & \text{if} \quad M \ge \overline{M}_{T-3} \end{cases} \tag{66}$$

Due to the monotonicity of the saving function, the fact that $\overline{M}_{T-2} > \overline{M}^r_{T-2}$ implies that $\overline{M}^{l_1}_{T-3} > M^{l_2}_{T-3} > \overline{M}^{r_1}_{T-3} > \overline{M}^{r_2}_{T-3}$. Similarly, it is not hard to show that $\overline{M}_{T-3} > \overline{M}_{T-2}$.

Having solved for the consumption function explicitly by doing backward induction for 3 periods, it is easy to see the general pattern. At t periods before the end of life T, $t \geq 1$, i.e. at period $T-t$, the consumption function $c_{T-t}(M)$ will have a total of t kinks relating to current and future liquidity constraints, namely $y/R\beta$ and $\overline{M}^{l_j}_{T-t}$, $j = 1, \ldots, t-1$; $t-1$ discontinuities relating the the future retirement thresholds denoted $\overline{M}^{r_j}_{T-t}$, $j = 1, \ldots, t-1$, and one discontinuity at the period t retirement threshold $\overline{M}_{T-t}$. Consequently, $c_{T-t}(M)$ will have $2t+1$ linear segments. For every period $T-t$, $t \geq 1$ there will be a kink in the consumption function at $M = y/R\beta$ corresponding to the end of the liquidity constrained region, $[0, y/R\beta]$.

Under the assumptions $R\beta \leq 1$ and $\delta < (1+\beta)\log(1+\beta)$ all the kink/discontinuity points define non-empty intervals such that the following ordering holds

$$y/R\beta < \overline{M}^{l_1}_{T-t} < \overline{M}^{l_2}_{T-t} < \cdots < \overline{M}^{l_{t-1}}_{T-t} < \overline{M}^{r_{t-1}}_{T-t} < \overline{M}^{r_{t-2}}_{T-t} < \cdots < \overline{M}^{r_2}_{T-t} < \overline{M}^{r_1}_{T-t} < \overline{M}_{T-t}. \quad (67)$$

The first of the future liquidity constraint kink points is always at the same value of M,

$$\overline{M}^{l_1}_{T-t} = [y/(R\beta)^2](1+\beta-R\beta^2) \quad \text{for} \;\; t \geq 2. \quad (68)$$

Period $T-t$ retirement threshold $\overline{M}_{T-t}$ is given by

$$\overline{M}_{T-t} = \frac{(y/R)e^{-K}}{(1-e^{-K})}, \text{ where } K = \delta \left(\sum_{i=0}^{t} \beta^i \right)^{-1}. \quad (69)$$

The values of the last $t-2$ future liquidity constraint kink points $\overline{M}^{l_j}_{T-t}$, $j = 2, \ldots, t-1$ and the future retirement threshold discontinuity points $\overline{M}^{r_j}_{T-t}$, $j = 1, \ldots, t-2$ are determined by the values of wealth that make the consumer indifferent between consuming according to the linear segments of the consumption function on either side of each of these kink points as described above.

The value function $V_{T-t}(M)$ can be expressed recursively in terms of the already defined value function $V_{T-t+1}(M)$ one period ahead:

$$V_{T-t}(M) = \begin{cases} \log(c_{T-t}(M)) - \delta + \beta V_{T-t+1}(R(M - c_{T-t}(M)) + y) & \text{if} \quad M < \overline{M}_{T-t} \\ \log(M) \left(\sum_{i=0}^{t} \beta^i\right) + A_{T-t} & \text{if} \quad M \geq \overline{M}_{T-t} \end{cases} \quad (70)$$

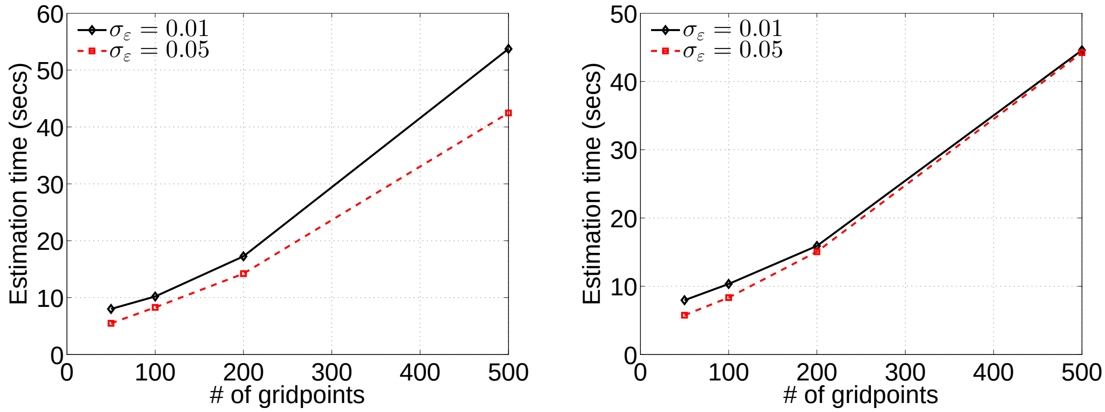
where A_{T-t} was defined in equation (42) above. It is then straightforward to show with the formal mathematical induction argument the general formula (7). $\square$

D DC-EGM run times

Figure 10 illustrates the average estimation time spent to estimate $\hat{\delta}$. Results are shown for varying degree of income uncertainty, $\sigma_\eta \in \{0.001, 0.05\}$, and different values of the disutility of work parameter, $\delta \in \{0.1, 0.5\}$

Figure 10: Timing: income uncertainty.

(a) $\sigma_\eta^2 = 0.001$ (b) $\sigma_\eta^2 = 0.05$



Notes: The plots illustrate the time spent to estimate the model. Results are shown for varying degree of smoothing, $\sigma_\varepsilon \in \{0.01, 0.05\}$, and different values of the income variance, $\sigma_\eta^2 \in \{0.001, 0.05\}$. The rest of parameters are at their baseline levels, see Table 1.

E Proof of Extreme Value Homotopy Principle

This appendix proves Theorem 2 which states that the value function and optimal decision rules in the presence of Type I extreme value distributed taste shocks converge (in an appropriate sense to be defined below) to the value functions and decision rules of a limiting problem without taste shocks. We prove Theorem 2 for a more general class of problems than just the retirement consumption model, and therefore restate it below.

Let ε be a random variable having a standardized Type I extreme value distribution with CDF $F(\varepsilon)$ given by

$$F(\varepsilon) = \exp\{-\exp\{-\varepsilon\}\}. \tag{71}$$

We have $E\{\varepsilon\} = \gamma$, where $\gamma = 0.577\ldots$ is Euler's constant and $\text{var}(\varepsilon) = \pi^2/6$. Then if σ is a positive scaling constant, $\sigma\varepsilon$ will also be a Type I extreme value distribution with expected value $\sigma\gamma$ and variance $\sigma^2\pi^2/6$. In the notation of the illustrative model in the paper, σ corresponds to the scaling parameter of the "perturbed" model σ_ε.

The homotopy convergence result we prove below holds for a considerably more general class of dynamic programming problems than the simple retirement example we analyzed in section 2.1 or even the class defined in Appendix A, where we assumed the continuous choice is a unidimensional variable and we imposed additional assumptions to ensure monotonicity of the savings function. In this appendix we consider a more general class of problems, though we do not strive for maximum possible generality in order to make our proof as straightforward as possible.

Consider a finite horizon DP problem without Type I extreme value taste shocks which we also refer to as the "unperturbed" DP problem. In the last period, T, the agent chooses a vector of k continuous choice variables $c \in C_T(d, s)$, where $C_T(d, s)$ is a compact subset of a R^k and d is one of the discrete choices and s is a potentially multidimensional vector of state variables in some Borel subset S of a finite dimensional Euclidean space. We assume that the discrete choice d is an element of a finite choice set $D_T(s)$. Let $u_T(d, c, s)$ be a utility function that is continuous in c for each s and each $d \in D_T(s)$ and a Borel measurable function of s for each c and d. Then the value function in period T is $V_T(s)$ given by

$$V_T(s) = \max_{d \in D_T(s)} \max_{c \in C_T(d,s)} u_T(d, c, s). \tag{72}$$

Now consider time $T-1$ and let $p_T(s'|s, c, d)$ be a Markov transition probability providing the conditional probability distribution over the state s' at time T given that the state vector at time $T-1$ is s, the discrete choice is d, and the continuous choice is c. Define the conditional expectation of V_T, $EV_{T-1}(d, c, s)$, by

$$EV_{T-1}(d, c, s) = \int V_T(s') p_T(\partial s'|d, c, s) \tag{73}$$

where we use $\partial s'$ to indicate the stochastic next period state variables over which this expectation is taken. In Assumption C below, we assume that this conditional expectation exists and is continuous in c for each $s \in S$ and $d \in D_{T-1}(s)$. Then by backward induction we can define the value function $V_{T-1}(s)$ and, continuing for each $t \in \{T-1, T-2, \ldots, 0\}$ we can define the sequence of functions $\{V_t\}$ recursively using Bellman's equation

$$V_t(s) = \max_{d \in D_t(s)} \max_{c \in C_t(d,s)} \left[u_t(d, c, s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s'|d, c, s) \right]. \tag{74}$$

where $\beta \geq 0$ is the agent's discount factor.

We make the following assumptions on this limiting DP problem without taste shocks that is sufficient to guarantee the existence of a well defined solution.

Assumption B *The choice sets $D_t(s)$ are all finite with a uniformly bounded number of elements D given by*

$$D = \max_{t \in \{0,1,\ldots,T\}} \sup_{s \in S} |D_t(s)| < \infty \tag{75}$$

where $|D_t(s)|$ denotes the number of elements in the finite set $D_t(s)$.

Assumption C *For each $t \in \{0, 1, \ldots, T\}$ and each $s \in S$ and each $d \in D_t(s)$ the function $u_t(d, c, s)$ is continuous in c, and for each $t \in \{1, 2, \ldots, T-1\}$, $s \in S$ and $d \in D_{t-1}(s)$ the function $EV_t(d, c, s)$ is given by*

$$EV_t(d, c, s) = \int V_t(s') p_t(\partial s'|d, c, s) \tag{76}$$

is finite and continuous in c.

Define the *discrete choice-specific continuous choice function* $c_t(d, s)$ by

$$c_t(d, s) = \underset{c \in C_t(d,s)}{argmax} \left[u_t(d, c, s) + \beta EV_{t+1}(d, c, s) \right] \tag{77}$$

and the optimal discrete decision rule $\delta_t(s)$ by

$$\delta_t(s) = \underset{d \in D_t(s)}{argmax} \left[u_t(d, c_t(d, s), s) + \beta EV_{t+1}(d, c_t(d, s), s) \right]. \tag{78}$$

The overall optimal continuous decision rule $c_t(s)$ is then given by

$$c_t(s) = c_t(\delta_t(s), s). \tag{79}$$

The *solution* to the DP problem is given by the collection Γ of the $T+1$ value functions $\{V_0, V_1, \ldots, V_T\}$, the $T+1$ optimal continuous decision rules $\{c_0, c_1, \ldots, c_T\}$ and the $T+1$ optimal discrete decision rules $\{\delta_0, \delta_1, \ldots, \delta_T\}$.

Now we define a family of *perturbed* DP problems index by σ, the scale parameter of the Type I Extreme value distribution. Let ε denote a *vector* of *IID* extreme value random variables with the same dimension as $|D_t(s)|$, the number of elements in the finite choice set $D_t(s)$. Assume the elements of $D_T(s)$ are ordered in some fashion and let $\varepsilon(d)$ be the component of the vector ε corresponding to the choice of alternative $d \in D_t(s)$. We will refer to $\varepsilon(d)$ as the "d^{th} taste shock".

Now consider the last period T. The value function $V_{\sigma,T}(s,\varepsilon)$ is given by

$$V_{\sigma,T}(s,\varepsilon) = \max_{d \in D_T(s)} \max_{c \in C_T(d,s)} \left[u_T(d,c,s) + \sigma\varepsilon(d)\right]. \tag{80}$$

Notice that $V_{\sigma,T}$ is now a function of the vector s and the vector $\varepsilon \in R^{|D_T(s)|}$. If the number of elements of $D_T(s)$ varies with $s \in S$ we can embed the vector ε in R^D where D is the upper bound on the number of discrete choices by Assumption. We can use the convention that if $|D_T(s)| < D$ for some $s \in S$, the function $V_{\sigma,T}(s,\varepsilon)$ depends only on the components of ε corresponding to the feasible choices $d \in D_T(s)$ and not on any components d that are not elements of $D_T(s)$.

The CDF $F(\varepsilon)$ of the vector random variable ε is given by the product of the univariate Type I Extreme value CDFs, i.e.

$$F(\varepsilon_1, \ldots, \varepsilon_D) = \prod_{d=1}^{D} \exp\{-\exp\{-\varepsilon(d)\}\} \tag{81}$$

To compute the expected value of $V_T(s,\varepsilon)$ we apply multivariate integration to get

$$\begin{aligned} EV_{\sigma,T}(d,c,s) &= \int_{s'}\int_{\varepsilon'} V_t(s',\varepsilon')F(\partial\varepsilon')p_T(\partial s'|d,c,s) \\ &= \int_{s'} \sigma \log\left(\sum_{d \in D_T(s')} \exp\{u_T(d,c_T(s,d),d)/\sigma\}\right) p_T(\partial s'|d,c,s) \end{aligned} \tag{82}$$

where $c_T(s,d) = argmax_{c \in C_T(d,s)}\, u_T(d,c,s)$ is the choice-specific continuous choice function. The closed form expression for the expectation over ε' the Type I extreme value random variables is a consequence of a property of extreme value random variables known as *max-stability* i.e. the maximum of a finite collection of Type I extreme value random variables has a (shifted) Type I extreme value distribution. We refer to the log-sum formula inside the integral of the lower equation of (82) as the *smoothed max function.* We now prove a key Lemma that establishes a bound between the usual max function and the smoothed max function.

Lemma 3 (Logsum error bounds)**.** *Let $\{v_1, \ldots, v_D\}$ be any finite set of D real numbers and let $\sigma > 0$ be a constant. Then we have*

$$0 \leq \sigma \log\left(\sum_{d=1}^{D} \exp\{v_d/\sigma\}\right) - \max\{v_1, \ldots, v_D\} \leq \sigma\log(D). \tag{83}$$

Proof. Consider the shifted values $v_d - \max(v_1, \ldots, v_D) \le 0$. It follows that

$$\log\left(\sum_{d=1}^{D} \exp\{(v_d - \max\{v_1, \ldots, v_D\})/\sigma\}\right) \le \log\left(\sum_{d=1}^{D} \exp\{0\}\right) = \log(D). \tag{84}$$

Define $d^* = \arg\max_d(v_d)$ and let $J \ge 1$ denote the number of elements of D for which $v_d = v_{d^*}$. The lower bound is obtained from observing that

$$\log\left(J + \sum_{d=1, d\neq d^*}^{D} \exp\{(v_d - \max\{v_1, \ldots, v_D\})/\sigma\}\right) \ge 0. \tag{85}$$

Combining (84) and (85) with the identity

$$\sigma \log\left(\sum_{d=1}^{D} \exp\{(v_d - \max\{v_1, \ldots, v_D\})/\sigma\}\right) = \sigma \log\left(\sum_{d=1}^{D} \exp\{v_d/\sigma\}\right) - \max\{v_1, \ldots, v_D\} \tag{86}$$

concludes the proof. $\square$

Lemma 3 is the key to all of our subsequent results and the key to Theorem 2 since it shows that the difference between the max function and the smoothed max function is bounded by $\sigma \log(D)$ and this tends to 0 as $\sigma \downarrow 0$. This will imply that the difference between the value functions and decision rules of the unperturbed limiting DP problem and the family of perturbed DP problems with extreme value distributed taste shocks will converge to zero as the scale of the extreme value taste shocks, σ converges to 0.

We can now define the value functions at all time periods for the perturbed problem as the sequence $\{V_{\sigma,0}, \ldots, V_{\sigma,T}\}$ where $V_{\sigma,T}$ is given by equation (80) and the other value functions are given by the Bellman recursion

$$V_{\sigma,t}(s, \varepsilon) = \max_{d \in D_t(s)} \max_{c \in C_t(d,s)} [u_t(d, c, s) + \sigma\varepsilon(d) + \beta EV_{t+1}(d, c, s)] \tag{87}$$

where $EV_{\sigma,t+1}(d, c, s)$ is the conditional expectation of $V_{\sigma,t+1}(s, \varepsilon)$ and is given by

$$EV_{\sigma,t+1}(d, c, s) = \sigma \int_{s'} \log\left(\sum_{d' \in D_{t+1}(s')} \exp\{v_{\sigma,t+1}(d', c_{\sigma,t+1}(d', s'), s')/\sigma\}\right) p_{t+1}(\partial s'|d, c, s), \tag{88}$$

where

$$v_{\sigma,t+1}(d, c, s) = u_{t+1}(d, c, s) + \beta EV_{\sigma,t+2}(d, c, s) \tag{89}$$

and $c_{\sigma,t+1}(d, s)$ is the choice-specific continuous choice rule given by

$$c_{\sigma,t+1}(d, s) = \underset{c \in C_{t+1}(d,s)}{argmax} \left[v_{\sigma,t+1}(d, c, s)\right]. \tag{90}$$

Note that we used the Williams-Daly-Zachary Theorem again to obtain the expression for $EV_{t+1}(d, c, s)$ in equation (88) and we also note that due to the assumption that taste shocks are not only contemporaneously independent across different discrete choices d but also *intertemporally independent* processes, it follows that the value of the ε state vector at time t does not affect the conditional expectation of $V_{\sigma,t+1}$, and hence does not enter the conditional expectation $EV_{t+1}(d,c,s)$. This *conditional independence restriction* on the ε shocks is critical to all results that follow below.

Having defined the set of value functions for the family of perturbed problems we can define the full solution of the perturbed problem as the collection Γ_σ consisting of the value functions $(V_{\sigma,0},\dots,V_{\sigma,T})$, the continuous decision rules $(c_{\sigma,0},\dots,c_{\sigma,T})$ and the the discrete decision rules $(\delta_{\sigma,0},\dots,\delta_{\sigma,T})$. Note that all of these objects depend on both s and ε, which constitute the full vector of state variables in the perturbed problem. In particular, the discrete decision rule $\delta_{\sigma,t}(s,\varepsilon)$ can be defined using the choice-specific continuous choice rule $c_{\sigma,t}(d,s)$ as

$$\delta_{\sigma,t}(s,\varepsilon) = \underset{d\in D_t(s)}{argmax}\left[v_{\sigma,t}(d,c_{\sigma,t}(d,s),s)+\sigma\varepsilon(d)\right], \tag{91}$$

and the unconditional or continuous decision rule can be defined using the choice-specific continuous choice rules by

$$c_{\sigma,t}(s,\varepsilon) = c_{\sigma,t}(\delta_{\sigma,t}(s,\varepsilon),s). \tag{92}$$

To define a notion of convergence of the solution Γ_σ of the family of perturbed DP problems to the solution Γ of the limiting un-perturbed problem, we have to confront the difficulty that the state space for the family of perturbed problems is the set of points of the form (s,ε) for $s\in S$ and $\varepsilon\in R^D$ whereas the state space of the limiting unperturbed problem is just S. We start by noting the following representation for the value functions of the perturbed problem

$$V_{\sigma,t}(s,\varepsilon) = \max_{d\in D_t(s)}\left[v_{\sigma,t}(d,c_{\sigma,t}(d,s),s)+\sigma\varepsilon(d)\right], \tag{93}$$

which follows directly from the Bellman equation (87) and the definition of the v_t function in equation (89). We now compute a *partial expectation* of the value functions $V_{\sigma,t}(s,\varepsilon)$ over the ε holding the s state variable fixed. That is we define the partial expectation $EV_{\sigma,t}(s)$ as the function given by

$$\begin{aligned} EV_{\sigma,t}(s) &= \int_\varepsilon V_{\sigma,t}(s,\varepsilon)F(\varepsilon) \\ &= \sigma\left(\sum_{d\in D_t(s)}\exp\{v_{\sigma,t}(d,c_{\sigma,t}(d,s),s)/\sigma\}\right). \end{aligned} \tag{94}$$

We are in the position now to state the main result which is a reformulation of Theorem 2 for a more general class of DC models than the consumption retirement model in Section 2.

Theorem 2 (Extreme Value Homotopy Principle)**.** *Under assumptions* **B** *and* **C** *above, let*

$$\Gamma = \{(V_0,\dots,V_T),(\delta_0,\dots,\delta_T),(c_0,\dots,c_T)\} \tag{95}$$

be the solution to the limiting DP problem without taste shocks given in equations (72), (74), (77) and (78) above. Similarly, let

$$\Gamma_\sigma = \{(V_{\sigma,0},\dots,V_{\sigma,T}),(\delta_{\sigma,0},\dots,\delta_{\sigma,T}),(c_{\sigma,0},\dots,c_{\sigma,T})\} \tag{96}$$

be the solution to the the perturbed DP problem with Type I extreme value taste shocks with scale *parameter $\sigma > 0$ given in equations (80), (87), (88), (91) and (92). Then as $\sigma \to 0$ we have*

$$\lim_{\sigma \downarrow 0} \Gamma_\sigma = \Gamma, \tag{97}$$

where the convergence of value functions is defined in terms of the partial expectations of the value functions for the perturbed problems with taste shocks, $EV_{\sigma,t}(s)$ given in equation (94) so that we have uniform bound

$$\forall t \;\; \sup_{s \in S} |EV_{\sigma,t}(s) - V_t(s)| \leq \sigma \left[\sum_{j=0}^{T-t} \beta^j \right] \log(D), \tag{98}$$

and the decision rules converge pointwise for all (s, ε), $s \in S$ and $\varepsilon \in R^D$, i.e.

$$\begin{aligned} \lim_{\sigma \downarrow 0} \delta_{\sigma,t}(s, \varepsilon) &= \delta_t(s) \\ \lim_{\sigma \downarrow 0} c_{\sigma,t}(s, \varepsilon) &= c_t(s), \end{aligned} \tag{99}$$

assuming that the decision rules of the limiting problem $\delta_t(s), c_t(s)$ are singletons, otherwise the limits are elements of the sets $(\delta_t(s), c_t(s))$.

Proof. We prove Theorem 2 in three steps. First, we prove (98) by induction using Lemma 3 and showing that the bounds are independent of s. Second, we prove convergence of decision rules assuming that the limiting problem Γ has unique solution. Third, we extend the latter result to non-singleton solution sets.

Lemma 4 (DP error bounds)**.** *Let $V_t(s)$ be the value function for the unperturbed DP problem and let $EV_{\sigma,t}(s)$ be the partial expectation of the value function $V_{\sigma,t}(s, \varepsilon)$ to the perturbed DP problem. Then we have*

$$\forall t, s \;\; 0 \leq EV_{\sigma,t}(s) - V_t(s) \leq \sigma \left[\sum_{j=0}^{T-t} \beta^j \right] \log(D). \tag{100}$$

Lemma 4 can be proved by induction using Lemma 3. We work out the first several steps of the inductive argument, starting at period T. In period T $V_T(s)$ is given by equation (72), which can be rewritten in terms of the choice-specific continuous choice rule as

$$V_T(s) = \max_{d \in D_T(s)} [u_T(d, c_T(d, s), s)] \tag{101}$$

and similarly, we have $EV_{\sigma,T}(s)$ is given by

$$EV_{\sigma,T}(s) = \sigma \log \left(\sum_{d \in D_T(s)} \exp\{u_T(d, c_T(d, s), s)/\sigma\} \right), \tag{102}$$

since it is easy to see that $c_T(d, s) = c_{\sigma,T}(d, s)$ in the final period T. Using Lemma 3, we obtain the bounds

$$0 \leq EV_{\sigma,T}(s) - V_T(s) \leq \sigma \log(D), \; \forall s \in S, \tag{103}$$

which establishes the base case for our induction proof. Now suppose the inductive hypothesis holds, i.e. the error bounds are given by equation (100) at period $T, T-1, \ldots, t+1$. We now want to show that it also holds at period t. We have

$$V_t(s) = \max_{d \in D_t(s)} \left[u_t(d, c_t(d,s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c_t(d,s), s) \right], \tag{104}$$

and

$$EV_{\sigma,t}(s) = \sigma \log \left(\sum_{d \in D_t(s)} \exp \left\{ \frac{1}{\sigma} \left[u_t(d, c_{\sigma,t}(d,s), s) + \beta \int EV_{\sigma,t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d,s), s) \right] \right\} \right). \tag{105}$$

Note that $c_{\sigma,t}(d,s)$ is the choice-specific continuous decision rule for the perturbed problem. Define a function $\tilde{V}_t(s)$ by substituting $c_{\sigma,t}(d,s)$ for $c_t(d,s)$ in equation (104):

$$\tilde{V}_t(s) = \max_{d \in D_t(s)} \left[u_t(d, c_{\sigma,t}(d,s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d,s), s) \right]. \tag{106}$$

Since $c_{\sigma,t}(d,s)$ is not necessarily an optimal choice-specific consumption for the unperturbed problem, it follows that

$$\tilde{V}_t(s) \le V_t(s), \quad \forall s \in S. \tag{107}$$

Similarly define the function $E\tilde{V}_{\sigma,t}(s)$ by substituting the conditional expectation of V_{t+1} instead of the conditional expectation of $EV_{\sigma,t+1}$ in the formula for $EV_{\sigma,t}(s)$ in equation (105). We have

$$E\tilde{V}_{\sigma,t}(s) = \sigma \log \left(\sum_{d \in D_t(s)} \exp \left\{ \frac{1}{\sigma} \left[u_t(d, c_{\sigma,t}(d,s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d,s), s) \right] \right\} \right). \tag{108}$$

Note that we can write

$$\begin{aligned} EV_{\sigma,t}(s) \;=\; & \sigma \log \left(\sum_{d \in D_t(s)} \exp \left\{ \frac{1}{\sigma} \left[u_t(d, c_{\sigma,t}(d,s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d,s), s) \right.\right.\right. \\ & \left.\left.\left. + \; \beta \int [EV_{\sigma,t+1}(s') - V_{t+1}(s')] p_{t+1}(\partial s' | s, c_{\sigma,t}(d,s), s) \right] \right\} \right). \end{aligned} \tag{109}$$

By the inductive hypothesis, it follows that

$$\beta \int [EV_{\sigma,t+1}(s') - V_{t+1}(s')] p_{t+1}(\partial s' | d, c_{\sigma,t}(d,s), s) \le \sigma \beta \left[\sum_{j=0}^{T-t-1} \beta^j \right] \log(D). \tag{110}$$

Thus, it follows from inequality (110) that the following inequality holds

$$EV_{\sigma,t}(s) \le E\tilde{V}_{\sigma,t}(s) + \sigma \beta \left[\sum_{j=0}^{T-t-1} \beta^j \right] \log(D). \tag{111}$$

From Lemma 3 we have

$$E\tilde{V}_{\sigma,t}(s) - \tilde{V}_t(s) \le \sigma \log(D). \tag{112}$$

Using inequalities (107) and (112) it follows that

$$0 \leq EV_{\sigma,t}(s) - V_t(s) \leq \sigma \left[\sum_{j=0}^{T-t} \beta^j \right] \log(D), \tag{113}$$

completing the induction step of the argument. It follows by mathematical induction that inequality (100) holds for all $t \in \{0, 1, \ldots, T\}$ so Lemma 4 is proved. $\square$

Note that the bound (100) is *uniform* over all states $s \in S$ since the right hand side of inequality does not depend on s. In particular, we do not need to rely on any continuity or boundedness assumptions about $V_t(s)$: this function could potentially be non-smooth or even discontinuous in s and an unbounded function of s, something typical in many economic problems with consumption and saving, including the retirement problem we analyzed in Section 2.

It follows from uniformity of bound (100) that (98) holds.

We turn now to establishing that the decision rules $\delta_{\sigma,t}(s, \varepsilon)$ and $c_{\sigma,t}(s, \varepsilon)$ in the perturbed problem converge the optimal decision rules $\delta_t(s)$ and $c_t(s)$ in the limiting unperturbed DP problem for $t \in \{0, 1, \ldots, T\}$. We will allow for the possibility that there are multiple values of d and c that attain the optimum values in equations (77) and (78) above, so in general we can interpret $c_t(s)$ and $\delta_t(s)$ as *correspondences* (i.e. set-valued functions of s). However the pointwise argument is simplest in the case where there is a unique discrete and continuous decision attaining the optimum so we first present the argument in this case in Lemma 4 below.

Lemma 5 (Policy convergence 1)**.** *Consider a point $s \in S$ for which $\delta_t(s)$ is just a single element $d \in D_t(s)$ and $c_t(s)$ is a single element of the set of feasible continuous choice $C_t(\delta_t(s), s)$ that attains the optimum. Then for (99) holds for any $\varepsilon \in R^D$.*

Since the pair of decisions $(\delta_t(s), c_t(s))$ is the unique optimizer of the Bellman equation in state $s \in S$, we have

$$\begin{aligned}
& u_t(\delta_t(s), c_t(s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | \delta_t(s), c_t(s), s) \\
= \; & u_t(\delta_t(s), c_t(\delta_t(s), s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | \delta_t(s), c_t(\delta_t(s), s), s) \\
> \; & u_t(d, c, s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c, s) \quad \forall c \neq c_t(s) \in C_t(d, s), \; d \neq \delta_t(s) \in D_t(s).
\end{aligned} \tag{114}$$

Let d be any limit point of the sequence $\{\delta_{\sigma,t}(s, \varepsilon)\}$. Since feasibility requires $\delta_{\sigma,t}(s, \varepsilon) \in D_t(s)$ and $D_t(s)$ is a finite set, at least one limit point must exist. Similarly let c be a limit point of the choice-specific continuous decision rule $c_{\delta,t}(\delta_{\sigma,t}(s, \varepsilon), s, \varepsilon)$. This also must have one limit point since feasibility requires that $c_{\sigma,t}(s, \varepsilon) = c_{\sigma,t}(\delta_{\sigma,t}(s, \varepsilon), s, \varepsilon) \in C_t(\delta_t(s, \varepsilon), s)$ where the latter is a compact set due to Assumption C (for any fixed d, however since we are considering a subsequence $\{\delta_{\sigma,t}(s, \varepsilon)\}$ that converges to a fixed point $d \in D_t(s)$ it follows the σ sufficiently small, the sequence of consumptions must be elements of the single compact set $C_t(d, s)$).

Now we show that $d = \delta_t(s)$ and $c = c_t(s)$ since otherwise we would have a contradiction of the strict optimality of the decisions $(\delta_t(s), c_t(s))$ in inequality (114). We have

$$
\begin{aligned}
EV_{\sigma,t}(s) &= \sigma \log \left(\sum_{d \in D_t(s)} \exp \left\{ \frac{1}{\sigma} \left[u_t(d, c_{\sigma,t}(d,s), s) + \beta \int EV_{\sigma,t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d,s), s) \right] \right\} \right) \\
&= \int V_{\sigma,t}(s,\varepsilon) F(\varepsilon) \\
&= \int \Big[u_t(\delta_{\sigma,t}(s,\varepsilon), c_{\sigma,t}(\delta_{\sigma,t}(s,\varepsilon), s), s) + \sigma\varepsilon(\delta_{\sigma,t}(s,\varepsilon) \\
&+ \beta \int EV_{\sigma,t+1}(s') p_{t+1}(\partial s' | \delta_{\sigma,t}(s,\varepsilon), c_{\sigma,t}(\delta_{\sigma,t}(s,\varepsilon), s) \Big] F(\varepsilon) \qquad (115)
\end{aligned}
$$

By Lemma 4 we have that uniformly for each $t \in \{0, 1, \ldots, T\}$ and all $s \in S$

$$\lim_{\sigma \downarrow 0} EV_{\sigma,t}(s) = V_t(s). \qquad (116)$$

However using the fact that for a susequence $\{\sigma_n\}$ converging to zero we have

$$
\begin{aligned}
\lim_{\sigma_n \downarrow 0} \delta_{\sigma_n,t}(s,\varepsilon) &= d \\
\lim_{\sigma_n \downarrow 0} c_{\sigma,t}(\delta_{\sigma,t}(s,\varepsilon), s) &= c \qquad (117)
\end{aligned}
$$

these limits together with the representation of $EV_{\sigma,t}(s)$ in the last equation of (115) implies that

$$V_t(s) = u_t(d,c,s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c, s) \qquad (118)$$

However because $\delta_t(s)$ and $c_t(s)$ are the unique optimizers of Bellman equation in equation (114) above, it follows that $d = \delta_t(s)$ and $c = c_t(s)$. This argument holds for all cluster points of $\{\delta_{\sigma,t}(s,\varepsilon)\}$ and $\{c_{\sigma,t}(s,\varepsilon)\}$ so it follows that for any sequence $\{\sigma_n\}$ with $\lim_n \sigma_n = 0$, the sequences $\{\delta_{\sigma_n,t}(s,\varepsilon)\}$ and $\{c_{\sigma_n,t}(s,\varepsilon)\}$ converge to $\delta_t(s)$ and $c_t(s)$, respectively, proving that the claimed limits in equation (99) the statement of Lemma 5 hold. □

Finally we consider the case where $\delta_t(s)$ and/or $c_t(s)$ are not singletons. We also allow for the optimal decision rules to the perturbed problem, $\delta_{\sigma,t}(s,\varepsilon)$ and $c_{\sigma,t}(s,\varepsilon)$ to be correspondences (corresponding to case where multiple choices attain the optimum in the Bellman equation) the fact that the extreme value taste shocks are continuously distributed over the entire real line implies that for almost all ε $\delta_{\sigma,t}(s,\varepsilon)$ will be a singleton (i.e. there will be a unique discrete choice that maximizes the agent's utility).

We now show in Lemma 6 that even when we allow for nonuniqueness in the optimizing choices of (d,c) in both the perturbed problem and the limiting unperturbed problem, the correspondences $\delta_{\sigma,t}(s,\varepsilon)$ and $c_{\sigma,t}(s,\varepsilon)$ are *upper hemicontinuous,* that is if we have limits given by

$$
\begin{aligned}
\lim_{\sigma \downarrow 0} \delta_{\sigma,t}(s,\varepsilon) &= d \\
\lim_{\sigma \downarrow 0} c_{\sigma,t}(s,\varepsilon) &= c \qquad (119)
\end{aligned}
$$

where we now allow for the possibility that the limits d and c are actual *sets,* upper hemicontinuity requres that $d \subset \delta_t(s)$ and $c \subset c_t(s)$.

Lemma 6 (Policy convergence 2). *Consider a point $s \in S$ where the decision rules $\delta_t(s)$ and $c_t(s)$ are potentially non-unique, i.e. these may be sets of points in $D_t(s)$ and $C_t(\delta_t(s), s)$, respectively. Then the correspondences $\delta_{\sigma,t}(s, \varepsilon)$ and $c_{\sigma,t}(s, \varepsilon)$ are upper hemicontinuous, and for almost all ε $\delta_{\sigma,t}(s, \varepsilon)$ is a singleton, which implies that its limit d is a single element in $\delta_t(s)$.*

The proof is similar to Lemma 5 except that we now allow for the possibility that in the limiting DP model without taste shocks, there may be multiple values of $d \in D_t(s)$ and $c \in C_t(\delta_t(s), s)$ that attain the maximum of the Bellman equation in equations (77) and (78) above. Since the extreme value distribution is continuous the probability that there are any ties in the perturbed DP problem with taste shocks is zero (with respect to the extreme value distribution) and thus for almost all (s, ε) $\delta_{\sigma,t}(s, \varepsilon)$ is a singleton, and thus its limit d is a singleton. Following the reasoning of Lemma 5, if c is a limit point of $c_{\sigma,t}(s, \varepsilon)$ as $\sigma \to 0$ we can represent c as

$$c \in \lim_{\sigma \downarrow 0} c_{\sigma,t}(d, s), \tag{120}$$

that is, c is one of the limit points of the $\{c_{\sigma,t}(s, \varepsilon)\}$. Now suppose that the pair (d, c) is not optimal, i.e. $d \neq \delta_t(s)$ and $c \notin c_t(s)$. Then following the same argument as in Lemma5 we can obtain a contradiction, because following the same argument we can show that equation (118) holds, but if (d, c) are not optimal, this would contradict the fact that $V_t(s)$ attains the maximum over all feasible (d, c) values in equations (77) and (78). □

This concludes the proof of Theorem 2. □

References

Adda, J., C. Dustmann and K. Stevens (forthcoming): "The Career Costs of Children," *Journal of Political Economy*.

Aguirregabiria, V. and P. Mira (2010): "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156(1), 38 – 67, Structural Models of Optimization Behavior in Labor, Aging, and Health.

Ameriks, J., J. S. Briggs, A. Caplin, M. D. Shapiro and C. Tonetti (2015): "Long-Term Care Utility and Late in Life Saving," Working Paper 20973, National Bureau of Economic Research.

Barillas, F. and J. Fernández-Villaverde (2007): "A generalization of the endogenous grid method," *Journal of Economic Dynamics and Control*, 31(8), 2698–2712.

Bertsekas, D. P., Y. Lee, B. van Roy and J. N. Tsitsiklis (1997): "A neuro-dynamic programming approach to retailer inventory management," in *Proceedings of the IEEE Conference on Decision and Control*, vol. 4, pp. 4052–4057.

Bound, J., T. Stinebrickner and T. Waidmann (2010): "Health, economic resources and the work decisions of older men," *Journal of Econometrics*, 156(1), 106–129.

Carroll, C. D. (2006): "The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems," *Economics Letters*, 91(3), 312–320.

Carroll, C. D. and W. Dunn (1997): "Data Sources and Solution Methods for Unemployment expectations, jumping (S, s) triggers, and household balance sheets," in *NBER Macroeconomics Annual*, vol. 12, pp. 165–230. National Bureau of Economic Research, Inc.

Casanova, M. (2010): "Happy Together: A Structural Model of Couples' Joint Retirement Choices," Working paper, Department of Economics, UCLA.

Clausen, A. and C. Strub (2013): "A General and Intuitive Envelope Theorem," Working Paper 248, Edinburgh School of Economics.

Druedahl, J. and C. N. Jørgensen (2015): "Precautionary Borrowing and the Credit Card Debt Puzzle," Unpublished working paper, Department of Economics, University of Copenhagen.

Druedahl, J. and T. H. Jørgensen (2016): "A General Endogenous Grid Method for Multi-Dimensional Models with Non-Convexities and Constraints," Discussion paper 16-09, Department of Economics, University of Copenhagen.

Ejrnæs, M. and T. H. Jørgensen (2015): "Saving Behavior around Intended and Unintended Childbirths," unpublished mimeo, University of Copenhagen.

Fella, G. (2014): "A generalized endogenous grid method for non-smooth and non-concave problems," *Review of Economic Dynamics*, 17(2), 329–344.

French, E. and J. B. Jones (2011): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79(3), 69–732.

Gourinchas, P.-O. and J. A. Parker (2002): "Consumption over the life cycle," *Econometrica*, 70(1), 47–89.

Hakansson, N. H. (1970): "Optimal Investment and Consumption Strategies Under Risk for a Class of Utility Functions," *Econometrica*, 38-5, 587–607.

Hershberger, J. (1989): "Finding the upper envelope of n line segments in O(n log n) time," *Information Processing Letters*, 33(4), 169–174.

Hintermaier, T. and W. Koeniger (2010): "The method of endogenous gridpoints with occasionally binding constraints among endogenous variables," *Journal of Economic Dynamics and Control*, 34(10), 2074–2088.

Imai, S. and M. P. Keane (2004): "Intertemporal labor supply and human capital accumulation," *International Economic Review*, 45(2), 601–641.

Iskhakov, F. (2015): "Multidimensional endogenous gridpoint method: Solving triangular dynamic stochastic optimization problems without root-finding operations," *Economics Letters*, 135, 72 – 76.

Iskhakov, F. and M. Keane (2016): "An Analysis of the Australian Social Security System using a Life-Cycle Model of Labor Supply with Asset Accumulation and Human Capital," unpublished mimeo, University of New South Wales.

Jørgensen, T. H. (2013): "Structural estimation of continuous choice models: Evaluating the EGM and MPEC," *Economics Letters*, 119(3), 287–290.

——— (2014): "Leisure Complementarities in Retirement," mimeo, University of Copenhagen.

Ljungqvist, L. and T. Sargent (2005): "Lotteries for Consumers versus Lotteries for Firms," in *Frontiers in Applied General Equilibrium Modeling: in Honor of Herbert Scarf*, ed. by T. Kehoe, T. Srinivasan and J. Whalley, pp. 95–118. Cambridge University Press.

Ludwig, A. and M. Schön (2013): "Endogenous Grids in Higher Dimensions: Delaunay Interpolation and Hybrid Methods," Working Paper 65, University of Cologne.

McFadden, D. (1973): "Conditional logit analysis of qualitative choice behavior," in *Frontiers in econometrics*, ed. by P. Zarembka. Academic Press, New York.

Meghir, C. and L. Pistaferri (2004): "Income variance dynamics and heterogeneity," *Econometrica*, 72(1), 1–32.

Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157–80.

Oswald, F. (2016): "Regional Shocks, Migration and Homeownership," Unpublished working paper, Science Po.

Phelps, E. (1962): "The Accumulation of Risky Capital: A Sequential Utility Analysis," *Econometrica*, 30-4, 729–743.

Powell, W. B. (2007): *Approximate Dynamic Programming: Solving the curses of dimensionality*, vol. 703. John Wiley & Sons.

Prescott, E. (2005): "Nonconvexities in Quantitative General Equilibrium Studies of Business Cycles," in *Frontiers in Applied General Equilibrium Modeling: in Honor of Herbert Scarf*, ed. by T. Kehoe, T. Srinivasan and J. Whalley, pp. 95–118. Cambridge University Press.

Rogerson, R. (1988): "Indivisible Labor, Lotteries, and Equilibrium," *Journal of Monetary Economics*, 21, 3–16.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

——— (1988): "Maximum Likelihood Estimation of Discrete Control Processes," *SIAM Journal on Control and Optimization*, 26(5), 1006–1023.

Rust, J. (1994): "Structural estimation of Markov decision processes," in *Handbook of Econometrics, Vol. 4*, ed. by D. M. R. Engle, vol. chap. 51, pp. 3081–3143. Elsevier Science B.V., Amsterdam.

White, M. N. (2015): "The Method of Endogenous Gridpoints in Theory and Practice," *Journal of Economic Dynamics & Control*, 60, 26–41.

Yao, J., A. Fagereng and G. Natvik (2015): "Housing, Debt and the Marginal Propensity to Consume," Unpublished working paper, Johns Hopkins.