

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

February 2017

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 of 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.

There is an extensive literature on *static* models of discrete/continuous choice: a classic example is Dubin and McFadden (1984). However the focus of our paper is on *dynamic* DC models. A classic example is the 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 costly 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

¹There 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).

²The 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).

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 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 consumption-savings 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. We then show how DC-EGM can be used to solve DC models with taste shocks and investigate its performance as a nested solution method for structural estimation of a DP model of retirement.

Fella (2014) showed how the 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 smooths out some of the kinks in the value functions thereby simplifying the numerical solution of the model. 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

³In 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).

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 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 microeconomic 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); Ejr n s and J rgensen (2016); 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. Thus, 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 where 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.

⁴Contrary 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.

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 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 (2016) 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) and Deaton (1991) 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) and Deaton (1991) to allow for a discrete retirement decision. We derive an analytic solution to this problem with logarithmic utility to both illustrate the complexity caused by the addition of a discrete retirement choice and show how DC-EGM computes this solution. While we focus on this simple example for expositional clarity, DC-EGM can be applied to a much more general class of problems

that include taste and income shocks. We will discuss these extensions in section 3 and show how the addition of shocks can 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

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

involving choices of consumption c_t and whether to retire d_t , to maximize lifetime discounted utility. Let $d_t = 0$ denote retirement and $d_t = 1$ continued work, and let $\delta > 0$ be the disutility of work. To simplify the exposition, we assume retirement is absorbing, i.e. a retiree cannot to return to work.⁵

We solve (1) subject to a sequence of period-specific borrowing constraints, $c_t \leq M_t$, where $M_t = R(M_{t-1} - c_{t-1}) + yd_{t-1}$ is the consumer's consumable resources (wealth) at the beginning of period t . There is a fixed, non-stochastic gross interest rate, R , and labor income y for workers. The continuous consumption decision and discrete retirement decision are made at the start of each period, whereas interest earnings and labor income are paid at the end of the period.⁶

Let $V_t(M)$ and $W_t(M)$ be the expected discounted lifetime utility of a worker and a retiree, respectively, in period t of their life. The choice problem of the worker can be expressed recursively through the Bellman equation as

$$V_t(M) = \max \left\{ v_t(M, 0), v_t(M, 1) \right\}, \quad (2)$$

where the *choice-specific value functions* $v_t(M, d)$, $d \in \{0, 1\}$, are given by

$$v_t(M, 0) = \max_{0 \leq c \leq M} \left\{ \log(c) + \beta W_{t+1}(R(M - c)) \right\}, \quad (3)$$

$$v_t(M, 1) = \max_{0 \leq c \leq M} \left\{ \log(c) - \delta + \beta V_{t+1}(R(M - c) + y) \right\}. \quad (4)$$

⁵This allows us to focus primarily on the worker's problem. In the absence of absorbing retirement, the retiree's problem involves a discrete choice (returning to work or staying retired), and can be solved by DC-EGM similarly to the worker's problem.

⁶This timing convention is standard in the literature and it removes the need to include income as a separate state variable when we extend this model to a much wider class of problems with stochastic R and/or y .

The Bellman equation for the retiree's problem, which does not involve a discrete choice, is much simpler and can be written as

$$W_t(M) = \max_{0 \leq c \leq M} \left\{ \log(c) + \beta W_{t+1}(R(M - c)) \right\}. \quad (5)$$

The value function for a retiree $W_t(M)$ has a closed-form solution given by Phelps (1962, p. 742), so we focus on solving the worker's problem, that is, solving for the value function $V_t(M)$ and the optimal consumption rule $c_t(M)$. The fact that the future looks the same from the point of view of the retiree and from the point of view of the worker who decides to retire, can be verified from the fact that the right hand side of (3) is identical to that of (5). Therefore, we have $W_t(M) = v_t(M, 0)$, and the consumption function of the retiree is identical to the *choice-specific consumption function* of the worker who decided to retire, $c_t(M, 0)$ where the second argument denotes the retirement choice.

Note that even if $v_t(M, 0)$ and $v_t(M, 1)$ are concave functions of M , the value function $V_t(M)$ is the maximum of these two concave functions by (2) and will generally not be globally concave (Clausen and Strub, 2013). Further $V_t(M)$ will generally have a *kink point* at the value $M = \bar{M}_t$ where the two choice-specific value functions cross, i.e. $v_t(\bar{M}_t, 1) = v_t(\bar{M}_t, 0)$. We refer to these as *primary kinks* because they constitute *optimal retirement thresholds* for the worker in each period t . The optimal retirement rule is given by $d_t(M) = 1$ if $M < \bar{M}_t$ and $d_t(M) = 0$ if $M \geq \bar{M}_t$.

The worker is indifferent between retiring and working at the primary kink \bar{M}_t and $V_t(M)$ is non-differentiable at this point. However the left and right hand side derivatives, $V_t^-(M)$ and $V_t^+(M)$, exist and satisfy $V_t^-(\bar{M}_t) < V_t^+(\bar{M}_t)$. The discontinuity in the derivative of $V_t(M)$ at \bar{M}_t leads to a discontinuity in the optimal consumption function in the *previous* period $t - 1$ because the Bellman equation for $V_{t-1}(M)$ depends on $V_t(M)$. In turn, this causes a kink in $V_{t-1}(M)$ which we label a *secondary kink* since it is a reflection of the primary kink in $V_t(M)$. Thus, the primary kinks propagate back in time and manifest themselves in an accumulation of secondary kinks in the value functions in earlier periods, resulting in an increasing number of discontinuities in the consumption functions in earlier periods of the life cycle. The discontinuities in consumption rules in period t are caused by the worker's anticipation of landing exactly at the kink points in periods $t + 1, t + 2, \dots, T$ under the optimal consumption policy.

Theorem 1 (Analytical solution to the retirement problem). *Assume that income and disutility*

of work are time-invariant, the discount factor β and the disutility of work δ are not too large, i.e.

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

and instantaneous utility is given by $u(c) = \log(c)$. Then for $\tau \in \{1, \dots, T\}$ the optimal consumption rule in the worker's problem (2)-(4) is given by

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

The segment boundaries are totally ordered with

$$y/R\beta < \overline{M}_{T-\tau}^{l_1} < \dots < \overline{M}_{T-\tau}^{l_{\tau-1}} < \overline{M}_{T-\tau}^{r_{\tau-1}} < \dots < \overline{M}_{T-\tau}^{r_1} < \overline{M}_{T-\tau}, \quad (8)$$

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

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

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

The proof of Theorem 1, in particular the expressions for the kink points $M_{T-\tau}^{l_i}$ and $M_{T-\tau}^{r_i}$, are available in the online appendix. However, we show how this solution is derived when we introduce the DC-EGM algorithm in the next Section.⁷

⁷Note that the assumptions on the parameters β , δ and R are needed to ensure the ordering of the boundaries (8). Modified versions of Theorem 1 hold under weaker conditions, including a version where income and the

Theorem 1 establishes that the optimal consumption rule of the worker $c_{T-\tau}(M)$ is piece-wise linear in M , and in period t consists of $2(T-t)+1$ segments. The first segment where $M < y/R\beta$ is the credit constrained region where the agent consumes all available wealth and does not save. The next $T-t-1$ segments are demarcated by the *liquidity constraint kink points* $\overline{M}_t^{l_j}$ which define values of M at which the consumer is liquidity constrained at age $t+j$ but not at any earlier age. The remaining segments are defined by the secondary kinks, $\overline{M}_t^{r_j}$, $j = 1, \dots, T-t-1$ and represent the largest level of saving for which it is optimal to retire at age $t+j$ but not at any earlier age. Finally, \overline{M}_t is the retirement threshold, which denotes the minimum level of wealth which is required to retire. This is also the primary kink point as explained above. The optimal consumption function is discontinuous at points $\overline{M}_t^{r_j}$, and \overline{M}_t , so in total there are $T-t$ downward jumps in the consumption function.

Theorem 1 implies that the value function $V_t(M)$ is piecewise logarithmic with the same kink points, and can be written as $V_t(M) = B_t \log(c_t(M)) + C_t$ for constants (B_t, C_t) that depend on the region that M falls in. $V_t(M)$ has one primary kink at the optimal retirement threshold \overline{M}_t and $T-t-1$ secondary kinks at $\overline{M}_t^{r_j}$, $j = 1, \dots, T-t-1$. In addition there are $T-t$ kinks related to current and future liquidity constraints at $M = y/R\beta$ and $\overline{M}_t^{l_j}$, $j = 1, \dots, T-t-1$. If $R\beta = 1$ the liquidity-related kink points collapse to a single point $M = y/R\beta = y = \overline{M}_t^{l_1} = \dots = \overline{M}_t^{l_{T-t-1}}$, as in the case shown in Figure 3.

2.2 DC-EGM for problems without taste shocks

We are now in a position to introduce a generalization of the EGM algorithm for solving discrete-continuous problems that we call the DC-EGM algorithm. We describe DC-EGM by showing how it can be used to solve for the optimal consumption rule in the last three periods of the worker's problem. As the original EGM constitutes a building block of DC-EGM, we illustrate it as well using the consumption choice problem of the retiree. After explaining the DC-EGM algorithm, we compare its numerical performance and show that DC-EGM can closely approximate the analytic solution in Theorem 1.

DC-EGM is a backward induction algorithm that uses the *inverted Euler equation* to sequentially compute (potentially without root-finding) the choice-specific value functions $v_t(M, d)$ and

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.

the corresponding choice specific consumption functions $c_t(M, d)$ starting at the last period of life, T . Note that in a generic period t of the backward induction, the Bellman equation (4) of a worker who remains working implies the following first order condition for the consumption choice known as the Euler equation

$$\begin{aligned} 0 &= u'(c) - \beta R u'(c_{t+1}(R(M - c) + y)) \\ &= 1/c - \frac{\beta R}{c_{t+1}(R(M - c) + y)}. \end{aligned} \quad (10)$$

Similarly the Bellman equation (3) for the worker who decides to retire implies the following Euler equation

$$\begin{aligned} 0 &= u'(c) - \beta R u'(c_{t+1}^R(R(M - c))) \\ &= 1/c - \frac{\beta R}{c_{t+1}^R(R(M - c))}, \end{aligned} \quad (11)$$

where the second equation shows the case of $u(c) = \log(c)$. Given the period $t + 1$ optimal consumption functions $c_{t+1}(M)$ for workers and $c_{t+1}^R(M)$ for retirees, the solutions to these Euler equations yield the period t *choice-specific consumption functions* of the worker, $c_t(M, d)$. The solution are computed by applying the inverse of the marginal utility to the second component in (10) and (11). When such an inverse function is analytical, specifying an exogeneous grid over end-of-period saving $A = M - c$ facilitates solving for optimal current consumption in closed form without resorting to iterative numerical methods. This is the idea behind the endogenous grid method (EGM) proposed by Carroll (2006) which we build on.

Consider the terminal period T . The optimal consumption rule is to consume all available wealth, and thus is given by $c_T(M, d) = M$. With positive disutility of working, all agents retire (i.e. $\overline{M}_T = 0$) since income is paid at the *end* of the period, so it follows that $d_T(M) = 0$. This solution provides the base for backwards induction.

Now consider a retiree in period $T - 1$. Note that because the Bellman equation of the retiree (5) is identical to that of the worker who decides to retire (3), the Euler equation (11) also characterizes the optimal consumption choices of the retiree, $c_t^R(M)$. In $T - 1$ the closed-form solution of (11) is $c_{T-1}^R(M) = M/(1 + \beta)$. Consider how this solution is computed using the original EGM algorithm by Carroll (2006). EGM uses the Euler equation (11) to construct an *endogenous grid over M*

from an *exogenous grid over savings* $A = M - c$. Let $\vec{A} = \{A_1, \dots, A_J\}$ denote the exogenous grid over savings. Because savings is a *sufficient statistic*, i.e. carries all the information about wealth and consumption in the period, the Euler equation (11) can be solved for c for each point A_j . As mentioned above, when $u'(c)$ is *analytically invertible* the solution is also analytical. For the case $u(c) = \log(c)$, the solution is easily seen to be $c_{T-1}^R(M_{j,T-1}) = A_j/\beta$, where $M_{j,T-1}$ is an element of the endogenous grid \vec{M}_{T-1} implied by the exogenous grid over savings \vec{A} in period $T - 1$. We have $M_{j,T-1} = A_j + c_{T-1}^R(M_{j,T-1}) = A_j(1 + 1/\beta)$, which implies the EGM solution $c_{T-1}^R(M_{j,T-1}) = M_{j,T-1}/(1 + \beta)$.

Thus, at the points of the endogenous grid \vec{M}_t EGM produces an *exact* solution in the sense that the Euler residuals exactly equal zero. Between these points calculation of the optimal consumption requires function approximation, typically linear interpolation. With the latter, EGM produces the exact (linear) solution for the consumption function of a retiree.

Now consider a worker in period $T - 1$. The worker must solve for *two consumption functions* $c_{T-1}(M, 0)$ and $c_{T-1}(M, 1)$ corresponding to the decision to retire or not, respectively. However no special complications are created at this point: we simply apply “standard EGM” to solve for the optimal consumption rule of a worker $c_{T-1}(M, d)$ for each of the discrete decisions d , using the Euler equation (10) or (11) just as we described for the case of a retiree above. Similar to above, with linear interpolation this also results in exact solutions $c_{T-1}(M, 1) = (M + y/R)/(1 + \beta)$ and $c_{T-1}(M, 0) = M/(1 + \beta)$.

To ensure that the credit constraint $c_{T-1} \leq M$ is satisfied in presence of non-capital income y , standard EGM has an additional step.⁸ Namely, from (10) it follows that invoking the EGM algorithm with zero savings, $A_j = 0$, produces an endogenous point $M_{j,T-1} = y/R\beta$. As we show below in Theorem 2, it holds that savings as a function of wealth must be non-decreasing, and therefore for $M \leq y/R\beta$ the savings must remain zero, i.e. $c_{T-1}(M, 1) = M$. To add this additional “credit constrained” segment to the optimal consumption function, it is sufficient to add a point $M_{0,T-1} = 0$ to the endogenous grid \vec{M}_{T-1} , and set corresponding optimal consumption $c_{T-1}(M_{0,T-1}, 1) = 0$. This way, when linear interpolation is used, EGM finds the first two segments of the true solution $c_{T-1}(M, 1)$ given in equation (7) of Theorem 1 *exactly*, including the location of the first kink point.

In summary, when there are discrete choices, DC-EGM invokes the EGM algorithm to calcu-

⁸Note that the credit constraint never binds in the retiree’s problem.

late, via simple linear interpolation as described above, piece-wise linear approximations of the *decision-specific consumption functions* $c_t(M_{j,t}^d, d)$ defined over *decision-specific endogenous grids* $\vec{M}_t^d = \{M_{1,t}^d, \dots, M_{J,t}^d\}$. However what is different about DC-EGM is that we need to compare the choice specific value functions $v_t(M, 0)$ and $v_t(M, 1)$ in order to locate the threshold level of wealth when it becomes optimal to retire, \bar{M}_t . DC-EGM constructs approximations to $v_t(M, 0)$ and $v_t(M, 1)$ over the respective endogenous grids \vec{M}_t^0 and \vec{M}_t^1 alongside the calculation of the optimal consumption functions $c_t(M, 0)$ and $c_t(M, 1)$ by substituting the latter into the Bellman equations (3) and (4). Using the interpolated decision-specific value functions, we then find the optimal retirement threshold (primary kink) \bar{M}_t by finding the point of intersection of the two decision-specific value functions, $v_t(\bar{M}_t, 0) = v_t(\bar{M}_t, 1)$.

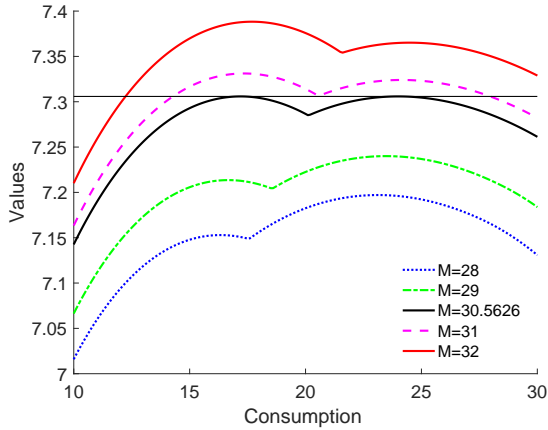
The overall value function for the worker $V_t(M)$ is then computed as an *upper envelope* of the two choice specific value functions $v_t(M, d)$, each defined over the endogenous grid \vec{M}_t^d . Similarly, the overall consumption function of the worker $c_{T-1}(M)$ is combined from choice specific consumption functions $c_{T-1}(M, 1)$ and $c_{T-1}(M, 0)$ depending on whether the level of wealth M is below or above the primary kink point \bar{M}_{T-1} , fully in line with formula (7) of Theorem 1 for $\tau = 1$.

So far DC-EGM seems to be a rather straightforward extension of standard EGM, but at period $T - 2$ we encounter an important additional complication: the *emergence of secondary kinks* due to *multiple local optima for c* in the Bellman equation (4). Recall that $V_t(M)$ is the maximum of decision-specific value functions, and is not globally concave. In particular, $V_{T-1}(M)$ has a *non-concave region* near \bar{M}_{T-1} , where the decision-specific value functions $v_{T-1}(M, 0)$ and $v_{T-1}(M, 1)$ cross. This implies that at time $T - 2$ when we search over c to maximize $\log(c) + \beta V_{T-1}(R(M - c) + y, 1)$ in (4), for some levels of M there will be multiple local optima for c with corresponding multiple solutions to the Euler equations. Thus, DC-EGM must also take care to select the *correct solution to the Euler equation* corresponding to the globally optimal consumption value. This is achieved by the calculation of the *upper envelope* over the overlapping segments of the decision-specific value functions that are produced from different solutions. The *dominated grid points* are then eliminated from the endogenous grid in a way we describe below.

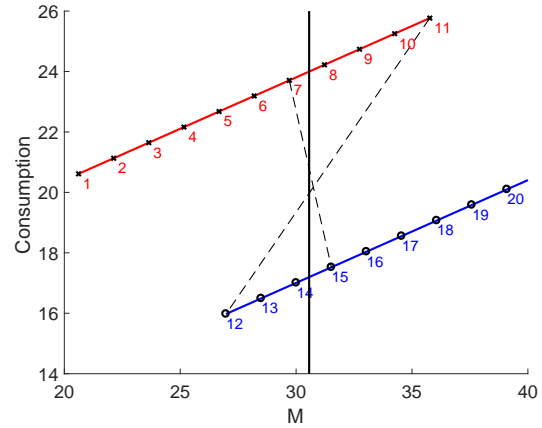
We illustrate this crucial step in the DC-EGM algorithm by showing how suboptimal endogenous grid points are eliminated in the worker's decision specific value function $v_{T-2}(M, 1)$ in

Figure 1: DC-EGM in period $T - 2$ of the retirement problem, where $T = 20$

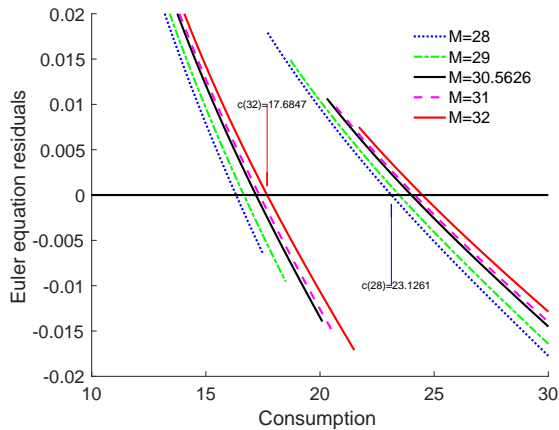
(a) Multiple local optima near secondary kink



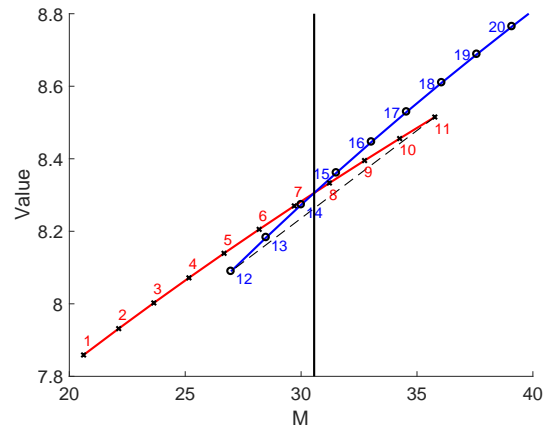
(b) Non-monotonic endogenous grid \vec{M}



(c) Multiple solutions to Euler equations



(d) Construction of the upper envelope



Notes: The plots illustrate how suboptimal endogenous points in the value function of the worker who in period $T - 2$ decides to continue working, $v_{T-2}(M, 1)$, are eliminated in the DC-EGM algorithm. Parameters of the model are $T = 20$, $R = 1$, $\beta = 0.98$, $\delta = 1$, and $y = 20$. Panel (a) plots the maximand of the Bellman equation (4), panel (b) shows the points of the optimal consumption functions computed by the standard EGM, corresponding to the endogenous grid points in panel (d), panel (c) showcases the discontinuous Euler equation by plotting Euler residuals for several values of wealth M , corresponding to panel (a), panel (d) plots the “value correspondence” produced by the standard EGM in the non-concave region of the problem and the location of the kink point found by DC-EGM.

Figure 1.⁹ With the parameter values listed in the Figure 1 notes, in period $T - 1$ the primary kink is $\bar{M}_{T-1} = 30.4382$. In the top left panel (a) we plot the maximand of the Bellman equation (4), $\log(c) + \beta V_{T-1}(R(M - c) + y)$, in period $T - 2$ for different values of M . The kink in $V_{T-1}(M)$ and the non-concave region around it translate directly into a kink and a non-concave region around it in the maximand for various levels of M . Location of the latter kink depends on M , for example, in the lowest dotted line plotted for $M = 28$ the kink at \bar{M}_{T-1} in $V_{T-1}(M)$ induces a kink at $c = 17.5618$. In general we have that the kink in c occurs at $c = M - (\bar{M}_{T-1} - y)/R$, and thus increases monotonically in M . With the kink at $\bar{c} = M - (\bar{M}_{T-1} - y)/R$, the maximand $\log(c) + \beta V_{T-1}(R(M - c) + y)$ has multiple locally optimal values of c on either side of the former. Panel (a) shows that for $M < 30.5626 = \bar{M}_{T-2}^{r_1}$ (notation of Theorem 1) the global optimum is to the right of the kink point, and vice versa for $M > \bar{M}_{T-2}^{r_1}$. At $M = \bar{M}_{T-2}^{r_1}$ the consumer is indifferent between the two locally optimal solutions.

The multiplicity of locally optimal solutions for c in the region near the secondary kink $\bar{M}_{T-2}^{r_1}$ is also reflected in multiple solutions to the corresponding Euler equation as shown in panel (c) of Figure 1. The discontinuity in the Euler residual functions are located at the same kink points.

Panel (b) of Figure 1 shows the implied consumption function and endogenous grid \vec{M}_{T-2}^1 that result from the application of the standard EGM method in period $T - 2$. We label each of the points $(M_{j,T-2}, c_j)$ resulting from the application of EGM to the first 20 exogenous saving grid points $A_1 = 0$, $A_2 = 1$, up to $A_{20} = 19$. The striking result is that EGM produces a *non-monotonic endogenous grid* \vec{M}_{T-2}^1 as is indicated by the dotted line that connects $(M_{11,T-2}, c_{11,T-2}) = (35.76, 25.76)$ to the point $(M_{12,T-2}, c_{12,T-2}) = (26.98, 15.98)$. Evidently, this reflects both a discontinuity and drop in both M and c . Note also that EGM has produced a *consumption correspondence* rather than a *consumption function* because of the two possible consumption values at the endogenous grid points $(M_{12,T-2}, \dots, M_{18,T-2})$. In addition, the jump in this consumption correspondence at $M_{11,T-2}$ going backward to an endogenous grid point $M_{12,T-2} < M_{11,T-2}$ contradicts the theoretical property of the correct solution due to the following theorem.

Theorem 2 (Monotonicity of the Saving Function). *Let $A_t(M, d) = M - c_t(M, d)$ denote the savings function implied by the optimal consumption function $c_t(M, d)$. If $u(c)$ is a concave func-*

⁹In fact, multiple solutions to the Euler equation cause the standard EGM loop to produce a “value correspondence” rather than a value function, while the elimination of suboptimal grid points converts this correspondence back to a proper function.

tion, then for each $t \in \{1, \dots, T\}$ and each discrete choice $d \in \{0, 1\}$ the optimal saving function $A_t(M, d) = M - c_t(M, d)$ is monotone non-decreasing in M .

The proof of Theorem 2 is in Appendix A. It implies that the non-monotonic endogenous grid \vec{M}_{T-2}^1 illustrated in panel (b) of Figure 1 is inconsistent with an optimal solution to the problem. How can this be rectified? In panel (d) of Figure 1 we illustrate the key second step of DC-EGM — refinement of the endogenous grid to discard the suboptimal points produced by the EGM step. This is achieved by constructing the upper envelope over the segments of the discrete choice specific value function correspondence in the region of M where multiple solutions were detected. The detection itself relied on checking for monotonicity of the endogenous grid.

In panel (d) of Figure 1 $M_{12, T-2} < M_{11, T-2}$, and thus two segments are formed (the first indicated with “x” symbols over points $j \in \{1, \dots, 11\}$, and the second indicated with “o” symbols over points $j \in \{11, \dots, 20\}$). The solid vertical line is drawn at the secondary kink point \overline{M}_{T-2}^{r1} (notation of Theorem 1) where the two segments intersect. It follows that to the left of \overline{M}_{T-2}^{r1} the higher first segment contains the global solutions, while to the right of \overline{M}_{T-2}^{r1} it is the second segment that contains global maxima. Therefore, points $j \in \{8, \dots, 14\}$ which correspond to the lower values, are suboptimal and should be discarded. The identified dominated points are eliminated from the endogenous grid \vec{M}_{T-2}^1 , forming the refined endogenous grid $\vec{M}_{T-2}^{*1} = \{M_{1, T-2}, \dots, M_{7, T-2}, M_{15, T-2}, \dots, M_{15, T-2}\}$ that is monotonic. To further increase the accuracy of the solution, the refined grid \vec{M}_{T-2}^{*1} can be supplemented by the kink point \overline{M}_{T-2}^{r1} itself.

With the refined monotonic endogenous grid \vec{M}_{T-2}^{*1} constructed from the upper envelope of the interpolated value functions, we obtain a close approximation to the correct optimal consumption rule $c_{T-2}(M, 1)$ as we can see in panel (b) of Figure 1. That is, for $M \leq \overline{M}_{T-2}^{r1}$ optimal consumption is given by the upper line marked with “x” symbol, while for $M > \overline{M}_{T-2}^{r1}$ optimal consumption is given by the lower line marked with “o” symbols. There is a discontinuous downward jump in consumption from endogenous grid point $M_{7, T-2}$ to the next point $M_{15, T-2}$, unless the kink point \overline{M}_{T-2}^{r1} is also added marking the point of vertical drop in the consumption function. Using this refined monotonic endogenous grid \vec{M}_{T-2}^{*1} , (even with a small number of grid points J) the DC-EGM produces very accurate approximations to the true solutions $V_{T-2}(M)$ and $c_{T-2}(M)$ that capture both the kink in the former and the discontinuity in the latter.

This completes our description of DC-EGM: the described procedure is repeated for all periods

t to solve the retirement problem via backward induction on the Euler and Bellman equations. In the next section we verify that DC-EGM produces accurate approximations to $c_t(M, d)$, $d = 1$, at all periods $t \in \{1, \dots, T\}$. It also generates accurate estimates of the secondary kink points $\overline{M}_t^{r_j}$ that capture discontinuous reductions in consumption that reflects the *anticipated* primary kink \overline{M}_{T-1} under the optimal consumption policy.

2.3 Numerical performance of the DC-EGM

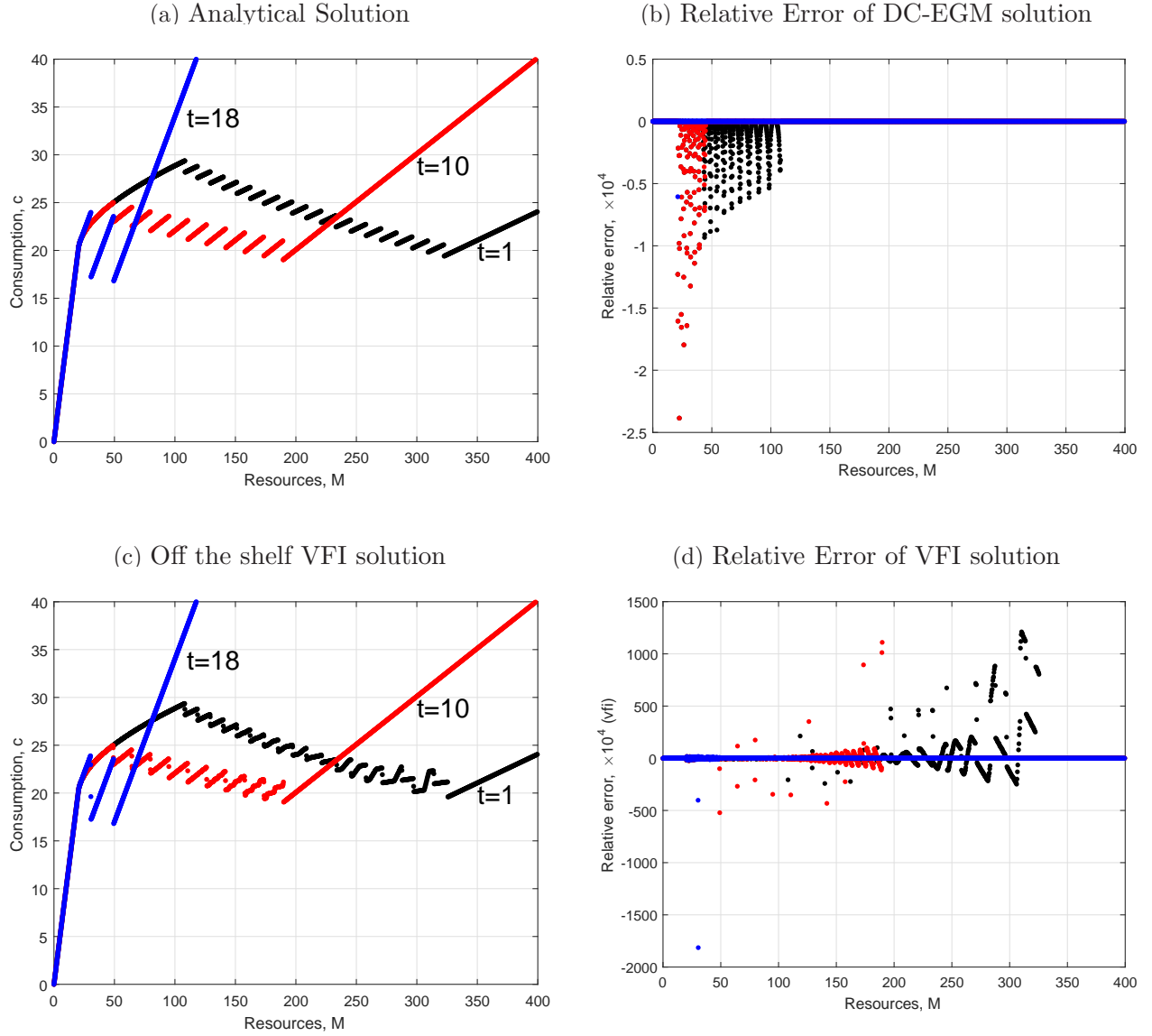
Figure 2 displays the optimal consumption function (7) and compares it to the numerical solution produced by DC-EGM, as well as the numerical solution produced by a naive brute force implementation of value function iteration (VFI). VFI solves the Bellman equations (3) and (4) by backward induction over an *exogenous grid* on M and using numerical optimization to search for optimal consumption as each M grid point.¹⁰ With a sufficient number of grid points, DC-EGM is able to accurately locate all the discontinuities of the analytical consumption rules ($\overline{M}_t^{r_j}$) and the boundary of the credit constrained region $y/R\beta$. Yet, because the kinks points associated with the credit constraint in future periods, $\overline{M}_t^{l_j}$, are not located precisely in the DC-EGM, the right panel of Figure 2 shows small relative errors on the order of 10^{-4} in the intervals $(y/R\beta, \overline{M}_{T-\tau}^{r-1})$ in each period $T - \tau$. Overall, the numerical solution by DC-EGM replicates the analytical solution remarkably well.¹¹

Panels (c)-(d) of Figure 2 show the solution produced by VFI method where the optimal consumption levels are 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

¹⁰Simple linear interpolation of the value function at the exogenous gridpoints over M was used to implement numerical optimization for values of c where implied next period resources $R(M - c) + y$ does not lie on the pre-defined exogenous grid \overline{M} . To enable a fair comparison with DC-EGM, we did use the *analytical expressions* for the value functions and consumption functions in the retiree problem.

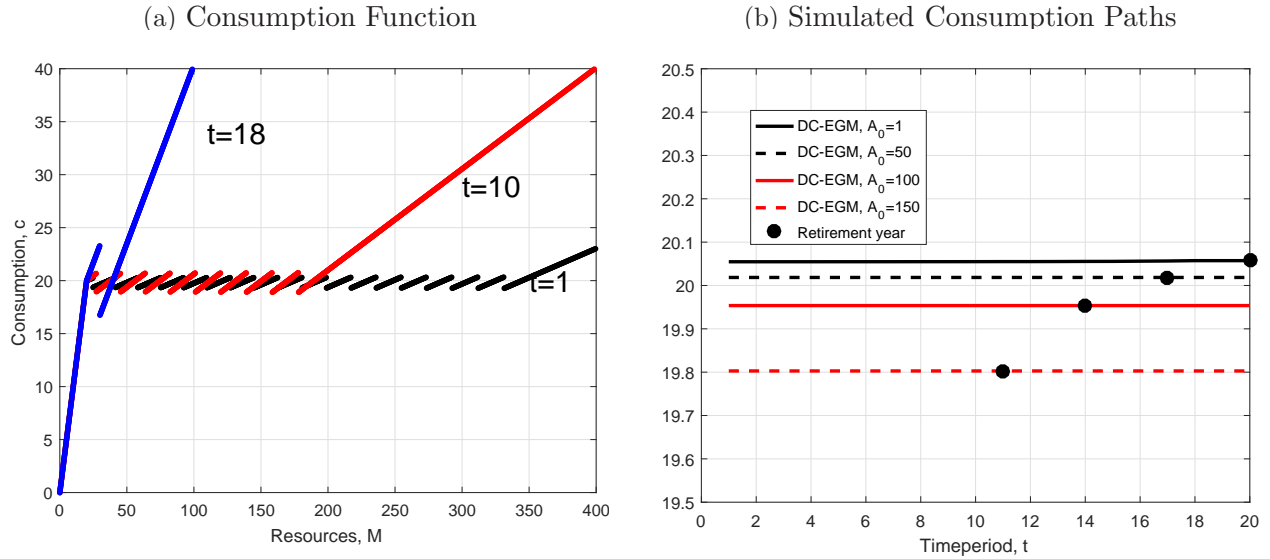
¹¹With 2000 points on the endogenous grid over wealth it took our Matlab/C implementation around 0.17 seconds on a Lenovo ThinkPad laptop with Intel® Core™ i7-4600M CPU @ 2.10 GHz and 8GB RAM to generate the numerical solution by DC-EGM. This is about 20 times faster than VFI which we implemented in 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 established finding in the literature, see e.g. Barillas and Fernández-Villaverde (2007); Jørgensen (2013); Fella (2014); Ameriks, Briggs, Caplin, Shapiro and Tonetti (2015).

Figure 2: 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 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 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 3: Discontinuous Consumption Function and Smooth Consumption Paths



Notes: The plots show optimal consumption functions of the worker in the consumption-savings model 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.

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 2 shows that the VFI solution results in significant approximation errors and is unable to fully capture the numerous discontinuities in the consumption function.

Figure 3 plots the optimal consumption functions and simulated consumption paths under the same assumptions as in Figure 2 except in this case we set $R = 1/\beta = 1.02$. The theoretical prediction for the consumption-savings model without retirement is that, with $R\beta = 1$, simulated consumption paths should be completely 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.

3 DC-EGM for problems with taste shocks

In this section we introduce income and taste shocks into the consumption and retirement model and show how DC-EGM is modified to accommodate these shocks. We show how primary and secondary kinks are smoothed in the presence of the random factors in the model, and explain how the numerical solution is simplified even though the problem remains non-convex in general.

Three effects are at play. First, because the discrete choice policy is expressed in probabilistic terms, the calculation of primary kink points is no more needed. Second, as a result, the process of accumulation of secondary kinks is perturbed: the perturbations caused by the primary kinks that remain throughout the backwards induction process in the deterministic setting, “fade out” in presence of shocks. Third, the calculation of expectations over random income in the problem with taste shocks can be performed with standard numerical algorithms, as apposed to the setting with random income but without taste shocks. We discuss each of these points in detail below.

3.1 Taste and income shocks in the retirement problem

Consider an extension of the model presented in Section 2 where the consumer faces income uncertainty and where choices are affected by discrete 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, $\log \eta_t \sim \mathcal{N}(-\sigma_\eta^2/2, \sigma_\eta^2)$.¹²

Following a vast literature on discrete choice modelling originating with (McFadden, 1973), we assume that random choice specific taste shifters $\sigma_\varepsilon \varepsilon_t(d)$, $d \in \{0, 1\}$, are additively separable, i.i.d. and have an Extreme Value distribution with scale parameter σ_ε . These shocks can be interpreted in the structural sense as the information relevant for the discrete choices that is observed by the agents but not by the econometrician. In this case, the scale parameter σ_ε can be estimated from the data. Alternatively, if the true model is deterministic, σ_ε can be interpreted as a (logit) smoothing parameter that can be chosen to approximate the true model with arbitrary precision,

¹²As 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.

see Theorem 3 below.

Because discrete choice specific taste shocks as well as income shocks only enter the worker problem, the solution of the retiree problem (5) remains the same. As for the worker problem, inclusion of taste shocks require us to rewrite the Bellman equation (2) such that

$$V_t(M, \varepsilon) = \max\{v_t(M, 0) + \sigma_\varepsilon \varepsilon(0), v_t(M, 1) + \sigma_\varepsilon \varepsilon(1)\}, \quad (12)$$

where the value function conditional on the choice to retire $v_t(M, 0)$ is still given by (3). However, the value function conditional on the choice to remain working, $v_t(M, 1)$, is modified to account for the taste and income shocks in the following period, namely

$$v_t(M, 1) = \max_{0 \leq c \leq M} \left\{ \log(c) - \delta + \beta \int EV_{t+1}^{\sigma_\varepsilon}(R(M - c) + y\eta, 1) f(d\eta) \right\}. \quad (13)$$

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

$$\begin{aligned} EV_{t+1}^{\sigma_\varepsilon}(M, 1) &= E \left[\max \left\{ v_{t+1}(M, 0) + \sigma_\varepsilon \varepsilon(0), v_{t+1}(M, 1) + \sigma_\varepsilon \varepsilon(1) \right\} \right] \\ &= \sigma_\varepsilon \log \left(\exp\{v_{t+1}(M, 0)/\sigma_\varepsilon\} + \exp\{v_{t+1}(M, 1)/\sigma_\varepsilon\} \right). \end{aligned} \quad (14)$$

The immediate effect of introducing extreme value taste shocks is the complete elimination of the *primary* kinks, because the location of the indifference point in (12) is now probabilistic from the point of view of the econometrician. Instead of calculating the location of the primary kink \overline{M}_t in the value function of the worker $V(M)$, the discrete choice policy function is now given by the logit choice probabilities $P_t(d|M)$ that arise due to the distributional assumption for the taste shocks.

$$P_t(d|M) = \frac{\exp\{v_t(M, d)/\sigma_\varepsilon\}}{\exp\{v_t(M, 1)/\sigma_\varepsilon\} + \exp\{v_t(M, 0)/\sigma_\varepsilon\}}, \quad d \in \{0, 1\}. \quad (15)$$

Because ε is unobserved it becomes impractical to carry out the calculations in terms of *overall* value function $V(M)$ and *overall* consumption function $c(M)$. Instead, in presence of taste shocks, DC-EGM operates on the *discrete choice specific* value functions $v(M, d)$, consumption functions $c(M, d)$, and computes choice probabilities (15) at each time period t . It is worth noting that the problem is still not globally concave in general, so the upper envelope calculation and the elimination of the suboptimal endogenous grid points is still performed as described in Section 2.2.

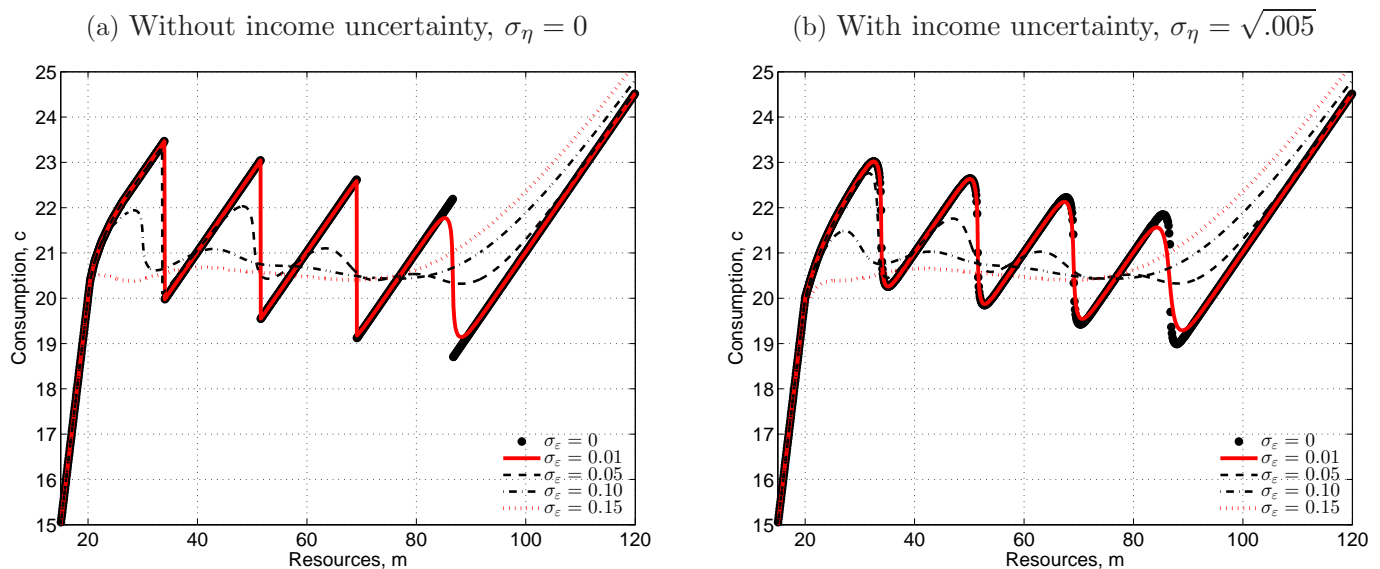
However, when σ_ε is sufficiently large the value function $V_t(M)$ in (12) eventually becomes globally concave. To see this, note that as the variance of the taste shocks increases, the choice-specific value functions are dominated by the noise and the differences between the alternatives become relatively less important. In turn, the components in (12) become similar, and $\lim_{\sigma_\varepsilon \uparrow \infty} EV_t^{\sigma_\varepsilon}(M)/\sigma_\varepsilon = \log(2)$. It follows from (12) then that the value function $v_t(M, 1)$ inherits its globally concave shape from the utility function.

The income shocks¹³ in the model also smooths out the secondary kinks during backwards induction. When the agent cannot perfectly anticipate having next period wealth exactly at the kink point, the secondary kinks are not replicated perfectly in the prior periods. Together these two sources of uncertainty make numerical solution significantly easier by reducing the number of times the secondary envelope routine is called by DC-EGM to refine the non-monotonic endogenous grid, as described in Section 2.2 and Algorithm 2 below.

Figure 4 shows the consumption function $c_t(M, 1)$ for a worker conditional on the choice to continue working for different values of the taste shock scale 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 (which is in line with Theorem 1, not counting the one at the primary kink point \bar{M}_{T-5}). It is evident that taste shocks of larger scale ($\sigma_\varepsilon \geq 0.05$) manage to smooth the function completely — eliminating all four discontinuities in the consumption function. Yet, for $\sigma_\varepsilon = 0.01$ only the last (rightmost) discontinuity is obviously smoothed out. 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 4 that discontinuities in the consumption function can be eliminated with a smaller taste shock ($\sigma_\varepsilon = 0.01$) when additional smoothing through uncertainty

¹³Equivalently, random returns or other stochastic factors in the intertemporal budget constraint.

Figure 4: Optimal Consumption Rules for Worker who remains working, $c_t(M, 1)$.



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 $R = 1$, $\delta = 1$, $\beta = 0.98$, $y = 20$.

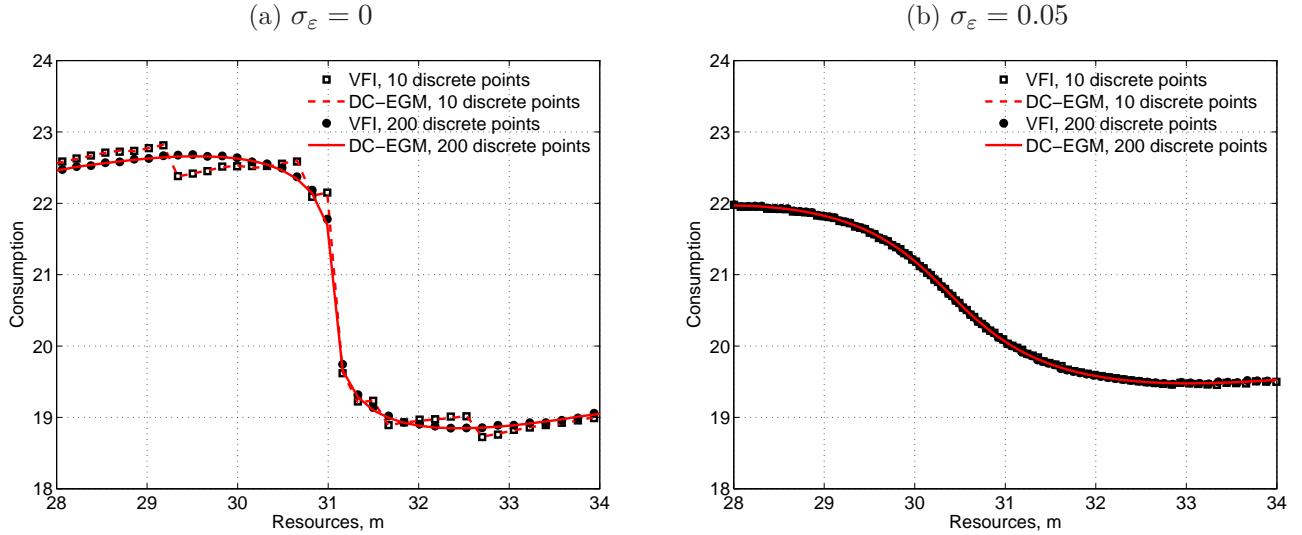
is present in the model.

Because the expected value function in (14) is a smooth function of M around the point where $v_t(M, 1) = v_t(M, 0)$, the maximand in (13) is also smooth in this region. This leads to an additional benefit of the inclusion of taste shocks, namely that standard numerical integration algorithms for the smooth function can be used when calculating the integral in (13). When $\sigma_\varepsilon = 0$, using procedures like Gaussian quadrature typically results in spurious discontinuities as shown in the left panel of Figure 5. This is due to the integrand not being a smooth function, see Appendix B for a detailed discussion. The right panel of Figure 5 illustrates how moderate smoothing ($\sigma_\varepsilon = .05$) significantly reduces this approximation error and removes the artificial kinks.

When the taste shocks ε_t have an interpretation as stochastic noise that is introduced to help solve a difficult DC dynamic program by making it more smooth, σ_ε is the amount of smoothing and has to be chosen and fixed prior to estimation. Theorem 3 shows that the level of σ_ε can always be chosen in such a way that the perturbed model approximates the underlying deterministic model with an arbitrary degree of precision.

Theorem 3 (Extreme Value Homotopy Principle). *In every time period the (expected) value func-*

Figure 5: Artificial Discontinuities in Consumption Functions, $\sigma_\eta^2 = 0.01$, $t = T - 3$.



Notes: The figure 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.

tion of the consumption and retirement problem with extreme value taste shocks $EV_t^{\sigma_\varepsilon}(M_t)$ defined in (14) converges uniformly to the value function of the same problem without taste shocks $V_t(M, 1)$ defined in (2) as the scale of these shocks approaches zero. The following uniform bound holds

$$\forall t : \sup_{M \geq 0} |EV_t^{\sigma_\varepsilon}(M, 1) - V_t(M, 1)| \leq \sigma_\varepsilon \left[\sum_{j=0}^{T-t} \beta^j \right] \log(2). \quad (16)$$

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 3 which holds under weak conditions for arbitrary DC models. Theorem 3 formalizes the results presented graphically in Figure 4, and 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 a smooth “nearby” problem with Extreme value taste shocks. The Extreme value scale parameter σ_ε plays the role of the “homotopy parameter.”

3.2 Extending DC-EGM for Taste Shocks and Income Uncertainty

In presence of taste shocks and income uncertainty, the DC-EGM algorithm remain at large the same, and potentially even more simple because the calculation of the primary kinks \bar{M}_t is replaced by the calculation of choice probabilities using formula (15). The biggest difference is that with taste shocks the DC-EGM operates on the choice specific values, as was mentioned in Section 3.1. We present the pseudo-code of the full algorithm in this section.

If we continue to assume that retirement is an absorbing state, the problem of the retiree remains the same, and we focus again on the workers problem. With taste and income shocks it is given by (12), (3) and (13), with the modified “smoothed” Euler equation taking the form¹⁴

$$\begin{aligned}
 0 &= u'(c) - \beta R \int \left[u'(c_{t+1}(R(M-c) + y\eta, 1)) P_{t+1}(1 | R(M-c) + y\eta) + \right. \\
 &\quad \left. u'(c_{t+1}(R(M-c) + y\eta, 0)) P_{t+1}(0 | R(M-c) + y\eta) \right] f(d\eta) \quad (17) \\
 &= \frac{1}{c} - \beta R \int \left[\frac{P_{t+1}(1 | R(M-c) + y\eta)}{c_{t+1}(R(M-c) + y\eta, 1)} + \frac{P_{t+1}(0 | R(M-c) + y\eta)}{c_{t+1}(R(M-c) + y\eta, 0)} \right] f(d\eta),
 \end{aligned}$$

where $P_t(1|M)$ and $P_t(0|M)$ are the choice probabilities (15), and the last line is written for the case $u(c) = \log(c)$.

Given the next period choice specific consumption functions of the worker $c_{t+1}(M, 0)$ and $c_{t+1}(M, 1)$, we solve (17) via the same backward induction process as described in Section 2.2 for the retirement problem without stochastic shocks. As before, the induction starts at the terminal period T with the easily derived consumption functions $c_T(M, 0) = c_T(M, 1) = M$, choice specific value functions $v_T(M, 0) = u(M) = v_T(M, 1) + \delta$, and the probability to remain working $P_T(1|M) = 1/(1 + \exp(\delta/\sigma_\varepsilon))$.

Then working backward, we choose an exogenous grid over saving $\vec{A} = (A_1, \dots, A_J)$ (which remains fixed throughout the backward induction process here for notational simplicity) and compute optimal consumption $\{c_t(A_1, d), \dots, c_t(A_J, d)\}$ for each point A_j and for each discrete choice d by calculating the inverse marginal utility of the right hand side of the Euler equation (17) and (11) for $d = 1$ and $d = 0$, respectively. Using these consumption functions, we construct the endogenous grid over M as $\vec{M}_t^d = (M_{t,1}^d, \dots, M_{t,J}^d)$ where M_j^d is given by $M_{j,t}^d = c_t(A_j, d) + A_j$, $j \in \{1, \dots, J\}$.

¹⁴See Lemma 1 in the Appendix.

For every d , if the resulting grid points are a monotonically increasing sequence, then no violation of monotonicity of the saving function as per Theorem 2 is indicated, and the DC-EGM method automatically reverts to the standard EGM method of Carroll (2006). However if \vec{M}_t^d is not a monotonically increasing sequence, we apply the same upper envelope procedure as described in Section 2.2 to eliminate the suboptimal elements of \vec{M}_t^d and add a point where the disjoint segments of the value function intersect. The last step amounts to calculating the choice specific value functions $v(M, d)$ alongside the consumption functions, which is achieved by plugging in the computed $c_t(A_1, d)$ into the maximand of the Bellman equation (13) for each point A_j of the exogenous grid on savings.

After the choice specific consumption and value functions $c(M, d)$ and $v(M, d)$ are computed on the *monotonic* endogenous grids \vec{M}^{*dt} , the period t iteration of the DC-EGM algorithm is complete. Unlike in Section 2.2, in presence of taste shocks the calculation of the upper envelope over the choice specific value functions is not performed. Instead, the discrete choice probabilities may be calculated using (15) for any level of wealth M using interpolation of the choice specific values $v(M, 0)$ and $v(M, 1)$.

We continue this procedure from period $T - 1$ backward to period $t = 1$, at which point we have fully approximated the solution to the life cycle retirement problem by DC-EGM. Note that as before, in the formulation with taste shocks none of the steps of the DC-EGM algorithm require iterative root-finding operations here.

Algorithms 1, 2 and 3 provide the pseudo-code for the complete DC-EGM algorithm for the problems with discrete shocks. The full DC-EGM algorithm (Algorithm 1), invokes the EGM step (Algorithm 2) 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 3). By Theorem 3, it can also approximate the solution for the problems without taste shocks if the scale parameter σ_ε is fixed at a sufficiently small value.

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

Algorithm 1 The DC-EGM algorithm

Input: Structural parameters, utility function $u(c)$, number of time periods T , number of grid points J , upper bound on wealth \bar{M} .

```
1: Fix the grid over savings  $\vec{A} = \{A_1, \dots, A_J\}$  such that  $A_1 = 0$  and  $A_j < A_{j+1}$ 
2: for  $t = T, \dots, 1$  do ▷ Backwards induction over time periods
3:   for  $st \in S$  do ▷ For every state (Worker, Retired)
4:     for  $d = \{0, 1\}$  if  $st = \text{Worker}$ , or ▷ For all admissible discrete choices
        $d = 0$  if  $st = \text{Retired}$  do
5:       if  $t = T$  then ▷ Terminal period
6:         Set consumption function  $c_T(M, d) = M$ 
7:         Set policy function  $v_T(M, d) = u(M) + d\delta$ 
8:       else ▷ All periods  $t < T$ 
9:         Call EGM STEP (Algorithm 2)
        Input: next period consumption and value functions  $c_{t+1}(M, d'), v_{t+1}(M, d'), d' \in \{0, 1\}$ 
        Output: consumption and value functions  $c_t(\vec{M}_t^d, d), v_t(\vec{M}_t^d, d)$  over endogenous grid  $\vec{M}_t^d$ 
10:        Call UPPER ENVELOPE (Algorithm 3)
        Input: endogenous grid  $\vec{M}_t^d$ , consumption and value functions  $c_t(\vec{M}_t^d, d), v_t(\vec{M}_t^d, d)$ 
        Output: refined grid  $\vec{M}_t^{*d}$ , consumption and value functions  $c_t(\vec{M}_t^{*d}, d), v_t(\vec{M}_t^{*d}, d)$ 
11:       end if
12:     end for
13:   end for
14: end for
```

Output: The collection of the choice specific consumption and value functions $c_t(M, d)$ and $v_t(M, d)$ defined on the endogenous grids \vec{M}_t^d for both worker and retiree, $d = \{0, 1\}$ and $t = \{1, \dots, T\}$ constitutes the solution of the consumption/savings and retirement model

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.

An important question of how the method handles the situations when the non-convex regions go undetected due to relatively coarse grid \vec{A} 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, $c \leq M$. During the EGM step, the credit constraints are dealt with in exactly same manner as in Carroll (2006), as described in Section 2.2. Let the smallest possible end-of-period

Algorithm 2 The EGM Step, adaptation of the standard EGM algorithm (Carroll, 2006)

Input: Structural parameters, utility function $u(c)$, current period $t < T$, state st , discrete choice d , next period consumption and value functions $c_{t+1}(M, d')$, $v_{t+1}(M, d')$, $d' \in \{0, 1\}$, exogenous grid over savings $\vec{A} = \{A_1, \dots, A_J\}$

- 1: **for** $j = 1, \dots, J$ **do** ▷ Loop over points in \vec{A}
- 2: Calculate next period wealth $M' = A_j R + dy\eta$
- 3: Calculate next period choice probabilities $P_{t+1}(d'|M')$
- 4: Calculate optimal choice specific consumption $c_{t+1}(M', d')$ and $u'(c_{t+1}(M', d'))$
- 5: Calculate choice specific value function $v_{t+1}(M', d')$ in period $t + 1$
- 6: Repeat steps 2 to 5 to compute the right hand side *RHS* of the Euler equation (17) and the expectation of the next period value function *EV* (the last component of the maximand in (13)). Gaussian quadrature or other numerical integration algorithms can be used to calculate the integral over income shocks η . Also, in general case, it may be necessary to integrate over the transition probabilities $P(st'|st, d)$ of the state process st , which is trivial in the retirement model.
- 7: Compute current period optimal consumption $c(A_j, d) = (u')^{-1}(RHS)$
- 8: Compute current period choice specific value function $v(A_j, d) = u(c(A_j, d)) + d\delta + EV$
- 9: Compute the endogenous point $M_{j,t}^d = c(A_j, d) + A_j$
- 10: **end for**
- 11: Add an extra point $M_{0,t}^d = 0$ to \vec{M}_t^d and set $c(0, d) = 0$

Output: Endogenous grid \vec{M}_t^d , consumption and value functions $c_t(\vec{M}_t^d, d)$, $v_t(\vec{M}_t^d, d)$ over it

Algorithm 3 Upper Envelope

Input: Endogenous grid \vec{M}_t^d , choice specific consumption and value functions $c_t(\vec{M}_t^d, d)$, $v_t(\vec{M}_t^d, d)$ calculated on \vec{M}_t^d

- 1: **for** $j = 1, \dots, J - 1$ **do**
- 2: **if** $M_{j,t}^d > M_{j+1,t}^d$ **then** ▷ Detect non-monotonicity in the endogenous grid
- 3: Find $j' > j$ such that $M_{j'+1,t}^d > M_{j',t}^d$
- 4: Define partitions $\vec{N}_1 = \{1, \dots, j\}$, $\vec{N}_2 = \{j, \dots, j'\}$, $\vec{N}_3 = \{j', \dots, J\}$
- 5: Run upper envelope calculation over segments of the “value function correspondence” $v_t(\vec{N}_1, d)$, $v_t(\vec{N}_2, d)$ and $v_t(\vec{N}_3, d)$ computed on the partitions in previous step, as described in Section 2.2 and illustrated in Figure 1, panels (b) and (d).
- 6: Determine a set of suboptimal grid points Q
- 7: Refine the endogenous grid by removing suboptimal points, so that $\vec{M}_t^{*d} = \vec{M}_t^d \setminus Q$
- 8: Remove the corresponding points $c_t(\vec{M}_{Q,t}^d, d)$, $v_t(\vec{M}_{Q,t}^d, d)$
- 9: [Optional] Add the kink point(s) \vec{M} where the uppermost segments in Step 5 intersect
- 10: [Optional] Add the corresponding interpolated values of $c_t(\vec{M}, d)$, $v_t(\vec{M}, d)$
- 11: **end if**
- 12: **end for**

Output: Refined *monotonic* grid \vec{M}_t^{*d} , consumption and value functions $c_t(\vec{M}_t^{*d}, d)$, $v_t(\vec{M}_t^{*d}, d)$ calculated on \vec{M}_t^{*d}

resources $A_1 = 0$ be the first point in the exogenous grid over saving \vec{A} . Assuming that the corresponding point of the endogenous grid $M_t(A_1, d)$ is non-negative¹⁵, it holds that $A(M, d) = 0$ for all $M \leq M_t(A_1, d)$ due to the monotonicity of saving function $A(M, d) = M - c_t(M, d)$ (see Theorem 2). Therefore, the optimal consumption in this region is then given by $c_t(M, d) = M$, and the choice-specific value function is

$$v_t(M, d) = \log(M) - d\delta_t + \beta \int EV_{t+1}(dy\eta)f(d\eta), \quad M \leq M_t(A_1, d). \quad (18)$$

Note that the third component of (18) 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 but not on M . Once this constant is computed, $v_t(M, d)$ essentially has analytical form in the interval $[0, M_t(A_1, d)]$, 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, (18) holds for all d in the interval $0 \leq M \leq \min_d M_t(A_1, d)$. In other words, the choice specific value functions for low wealth have the same shape, which is shifted vertically with d_t -specific coefficients. This implies that the logistic choice probabilities $P_t(d|M)$ 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.

¹⁵It is not hard to show that this holds as long as the per period utility function satisfies the Inada conditions.

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 | $\sigma_\varepsilon \in \{0.01, 0.05\}$ |

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

$$\max_{\{c_t, d_t\}_1^T} \sum_{t=1}^T \beta^t \left(\frac{c_t^{1-\rho} - 1}{1-\rho} - \delta_t d_t \right) \quad (19)$$

where ρ is the CRRA coefficient.

In order to simulate synthetic data from the DGP consistent with the model and the vector 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.¹⁶

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 $s_{i,1}^d = 1$, where we have used $s_{i,t}^d \in \{0, 1\}$ to denote the labor market state, i.e. whether an individual is retired ($s_{i,t}^d = 0$) or working ($s_{i,t}^d = 1$). Each workers initial wealth $M_{i,1}^d$ 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_t^d are calculated. Given the level of resources, discrete-choice specific value functions and choice probabilities are

¹⁶As 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.

computed, and a random draw determines which discrete labor market option d_{it}^d is chosen. After one period lag, the labor force participation decision becomes the labor market state, $s_{i,t+1}^d = d_{it}^d$. The optimal level of consumption, c_{it} , is then computed conditional on d_{it}^d , and the end-of-period wealth is calculated and stored to be used for calculation of resources available in the beginning of period $t + 1$, $M_{i,t+1}^d$. We then add normal additive measurement error with standard deviation $\sigma_\xi = 1$ to get the simulated consumption data, c_{it}^d . This produces simulated panel data $(M_{it}^d, s_{it}^d, d_{it}^d, c_{it}^d)$ for each individual $i \in \{1, \dots, 50,000\}$ in all time periods $t \in \{1, \dots, 44\}$.

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, $\{(M_{it}^d, s_{it}^d, d_{it}^d, c_{it}^d)\}_{i=\{1,\dots,N\}, t=\{1,\dots,T\}}$, 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,

$$c_{it}^d = c_t(M_{it}^d, s_{it}^d, d_{it}^d|\theta) + \xi_{it}, \quad \xi_{it} \sim N(0, \sigma_\xi), \quad \text{i.i.d. } \forall i, t. \quad (20)$$

Let $\xi_{it}^d(\theta) = c_{it}^d - c_t(M_{it}^d, s_{it}^d, d_{it}^d|\theta)$ 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

$$\ell_{it}(\theta, \sigma_\xi) = P(d_{it}^d|M_{it}^d, s_{it}^d, \theta) \frac{\phi(\xi_{it}^d(\theta)/\sigma_\xi)}{\sigma_\xi}, \quad (21)$$

where $\phi(\cdot)$ is the density function of the standard normal distribution. We have ignored the controlled transition probability for the retirement status s_{it}^d , since $P_{tr}(s_{it}^d|s_{i,t-1}^d, d_{i,t-1}^d)$ 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 ($s_{it}^d = 1$) are standard logits

$$P(d_{it}^d | M_{it}^d, s_{it}^d, \theta) = \frac{\exp(v_t(M_{it}^d, s_{it}^d, d_{it}^d | \theta) / \sigma_\varepsilon)}{\sum_{j=0}^1 \exp(v_t(M_{it}^d, s_{it}^d, j | \theta) / \sigma_\varepsilon)} \quad (22)$$

and are computed from the discrete choice specific value functions $v_t(M_{it}^d, s_{it}^d, d_{it}^d | \theta)$ 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 (21) drops out when $s_{it}^d = 0$.

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, $\sigma_\xi^2(\theta) = \frac{1}{NT} \sum_{t=1}^T \xi_{it}^d(\theta)^2$. The concentrated log-likelihood function is, therefore, proportional to

$$\mathcal{L}(\theta) \propto \sum_{i=1}^N \sum_{t=1}^T \left\{ \frac{s_{it}^d}{\sigma_\varepsilon} \left(v_t(M_{it}^d, s_{it}^d, d_{it}^d | \theta) - EV_t(M_{it}^d, s_{it}^d | \theta) \right) - \frac{1}{2} \log \left(\sum_{i=1}^N \sum_{t=1}^T \xi_{it}^d(\theta)^2 \right) \right\}, \quad (23)$$

where $EV_t(M_{it}^d, s_{it}^d | \theta)$ is the the logsum given in (14) evaluated at parameter value θ .¹⁷ The parameter vector $\hat{\theta}$ that maximizes (23) is the ML estimator of the model parameters.

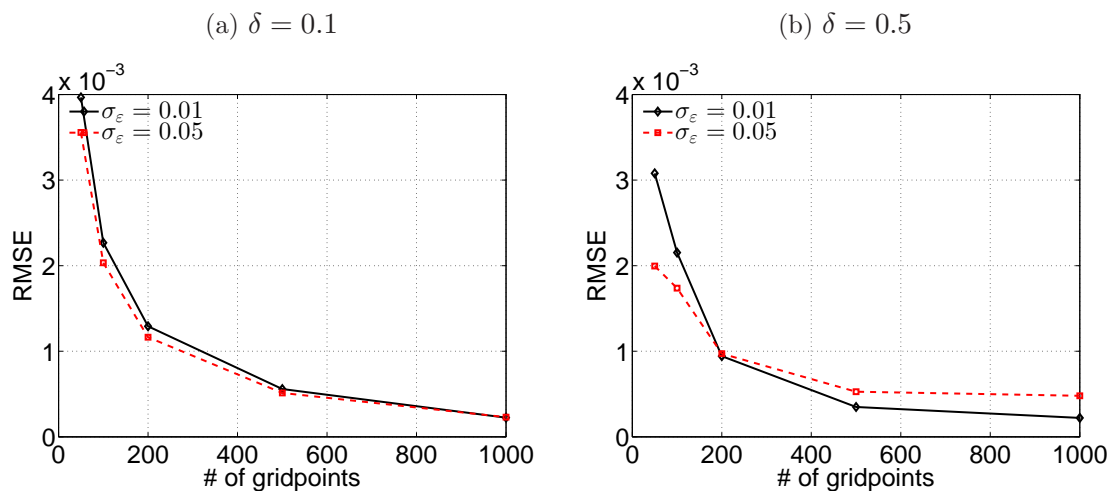
4.3 Taste Shocks as Unobserved State Variables

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 C contains average estimation time for the DC-EGM.

Taste Shocks and Approximation Error. Figure 6 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

¹⁷Following (23), the logsum only has to be evaluated for workers, $s_{it}^d = 1$.

Figure 6: 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.

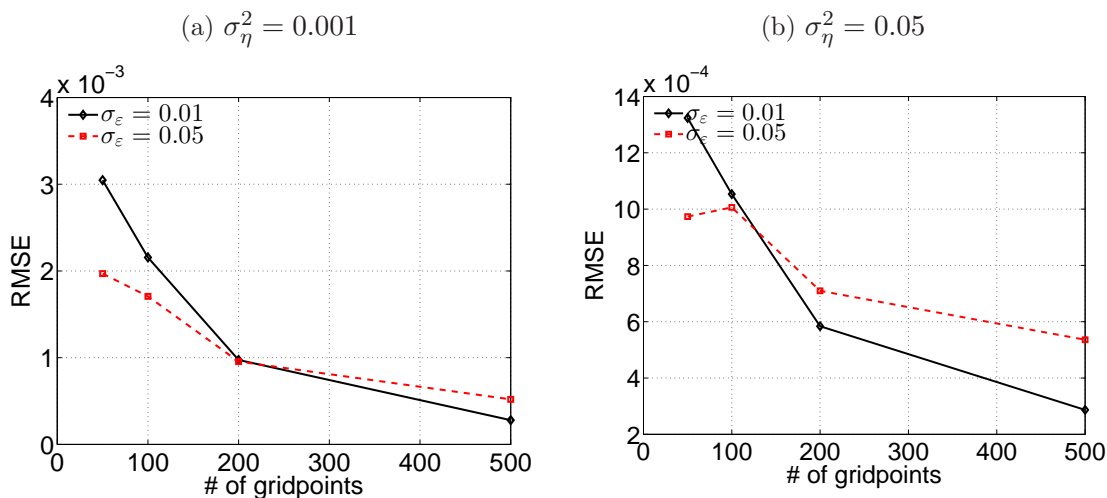
points and rapidly improves as the number of grid points increase from 50 through 1000. Note that standard errors will of course increase with σ_ε due to the increased amount of unexplained variation in the error term and RMSE reflects this too. Bearing this is mind, it is evident that the approximation 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)¹⁸, which is on the order $1.0e^{-4}$ irrespectively of the number of grid points used.

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 7 display RMSE when introducing *income uncertainty*. We report results from two different values of the income variance¹⁹, $\sigma_\eta^2 \in \{0.001, 0.05\}$. The first level, 0.001, does not completely smooth out secondary

¹⁸MCSD results not shown

¹⁹The 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 7: Monte Carlo results: income uncertainty.



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.

kinks while the significantly more uncertain income process with $\sigma_\eta^2 = 0.05$ does (see the right panel of Figure 4).

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 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.

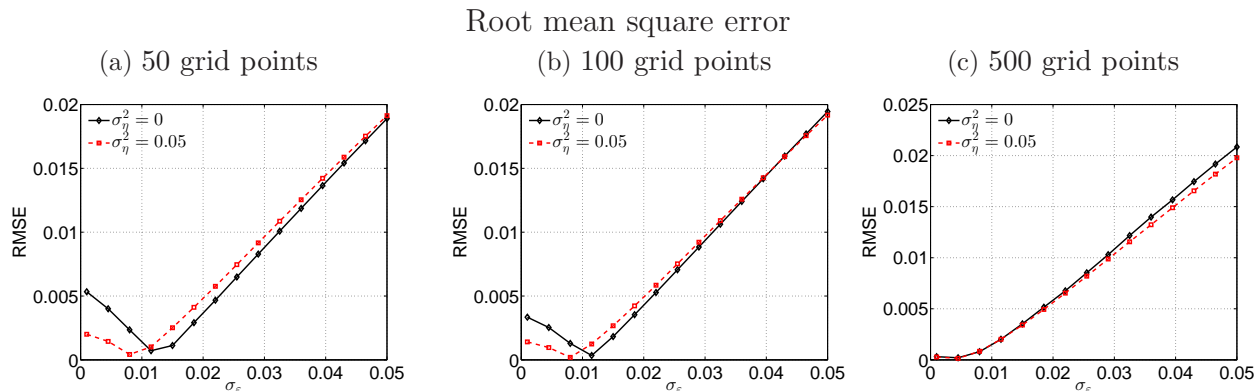
4.4 Taste Shocks as Logit Smoother

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 8 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.

Figure 8: Monte Carlo results: true model without taste shocks (misspecified)



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.

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 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.²⁰

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, 2016); precautionary borrowing and credit card debt (Drue Dahl and Jørgensen, 2015).

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.

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

²⁰DC-EGM algorithm can also be generalized for other specifications including the models with large state space and multidimensional discrete choice. White (2015); Iskhakov (2015); Drue Dahl and Jørgensen (2017) present theoretical foundations for extending endogenous grid methods to multi-dimensional models.

Bellman equation,

$$V_t(M_t, s_t) = \max_{0 \leq c_t \leq M_t, d_t \in D_t} \left[u(c_t, d_t, s_t) + \sigma_\varepsilon \varepsilon_t(d_t) + \beta_t E_t \{ V_{t+1}(M_{t+1}, s_{t+1}) | A_t, d_t \} \right], \quad (24)$$

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 (24) can be written in terms of the deterministic choice-specific value functions $v_t(M_t, s_t | d_t)$ as

$$v_t(M_t, s_t | d_t) = \max_{0 \leq c_t \leq M_t} \left[u(c_t, d_t, s_t) + \beta_t E_t \{ V_{t+1}(M_{t+1}, s_{t+1}) | A_t, d_t \} \right] \quad (25)$$

$$= \max_{0 \leq c_t \leq M_t} \left[u(c_t, d_t, s_t) + \beta_t E_t \left\{ \phi(v_{t+1}(M_{t+1}, s_{t+1} | d_{t+1}), D_{t+1}, \sigma_\varepsilon) | A_t, d_t \right\} \right], \quad (26)$$

where $\phi(x_j, J, \sigma) = \sigma \log \left[\sum_{j \in J} \exp \frac{x_j}{\sigma} \right]$ is the logsum function. The expectation in (26) is now only taken w.r.t. state transitions and idiosyncratic shocks, unlike in (24) and (25).

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 concave²¹ 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 (24) takes the form*

$$u'(c_t, d_t, s_t) = \beta_t R \mathbb{E}_t \left[\sum_{d_{t+1} \in D_{t+1}} u'(c_{t+1}(M_{t+1}, s_{t+1} | d_{t+1}), d_{t+1}, s_{t+1}) P_{t+1}(d_{t+1} | M_{t+1}, s_{t+1}) \right] \quad (27)$$

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

$$P_t(d_t | M_t, s_t) = \exp(v_t(M_t, s_t | d_t) / \sigma_\varepsilon) / \sum_{d \in D_t} \exp(v_t(M_t, s_t | d) / \sigma_\varepsilon). \quad (28)$$

²¹More precisely, a weaker condition is sufficient, namely for every x and arbitrary $\Delta_1 > 0$ and $\Delta_2 > 0$ it must hold that $u(c_t + \Delta_1, d_t, s_t) - u(c_t, d_t, s_t) \geq u(c_t + \Delta_1 + \Delta_2, d_t, s_t) - u(c_t + \Delta_2, d_t, s_t)$, see Theorem 2.

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 (25) given by

$$u'(c_t, d_t, s_t) + \beta_t E \left\{ \frac{\partial V_{t+1}(M_{t+1}, s_{t+1})}{\partial M_{t+1}} \frac{\partial M_{t+1}}{\partial c_t} \right\} = 0 \quad (29)$$

for every value of $d_t \in D_t$. The envelope conditions for (25)

$$\frac{\partial v_t(M_t, s_t|d_t)}{\partial M_t} = \beta_t E \left\{ \frac{\partial V_{t+1}(M_{t+1}, s_{t+1})}{\partial M_{t+1}} \frac{\partial M_{t+1}}{\partial M_t} \right\}, \quad (30)$$

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

$$u'(c_t, d_t, s_t) = \frac{\partial v_t(M_t, s_t|d_t)}{\partial M_t}. \quad (31)$$

The first order condition for (26) is

$$u'(c_t, d_t, s_t) = \beta_t R E_t \left[\sum_{d_{t+1} \in D_{t+1}} \frac{\partial v_{t+1}(M_{t+1}, s_{t+1}|d_{t+1})}{\partial M_{t+1}} P_{t+1}(d_{t+1}|M_{t+1}, s_{t+1}) \right], \quad (32)$$

where choice probabilities $P_{t+1}(d_{t+1}|M_{t+1}, s_{t+1})$ are given by (28). Plugging (31) into (32) completes the proof. \square

The DC-EGM algorithm outlined in Algorithm 1 is readily applicable to the general formulation of the discrete-continuous problem (24), expect for the extra loop that has to be taken over all additional states s_t in Step 3 (Algorithm 1). 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 2.

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 theorem²².

Theorem 2 (Monotonicity 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 2 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 (25) 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

$$f(A, M) = u(M - A, d_t, s_t) + \beta_t E_t \{ V_{t+1}(M_{t+1}(A), s_{t+1}) \} \quad (33)$$

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

²²A similar monotonicity result is also used in Fella (2014).

$f(A', M'') > f(A'', M'')$. Then

$$\begin{aligned}
& f(A', M') - f(A'', M') = \\
& = u(M' - A', d_t, s_t) - u(M' - A'', d_t, s_t) + \\
& + \beta_t [EV_{t+1}(M_{t+1}(A'), s_{t+1}) - EV_{t+1}(M_{t+1}(A''), s_{t+1})] \geq \\
& \geq u(M'' - A', d_t, s_t) - u(M'' - A'', d_t, s_t) + \\
& + \beta_t (EV_{t+1}(M_{t+1}(A'), s_{t+1}) - EV_{t+1}(M_{t+1}(A''), s_{t+1})) = \\
& f(A', M'') - f(A'', M'') > 0.
\end{aligned} \tag{34}$$

For the first inequality we use

$$\begin{aligned}
& u(M' - A', d_t, s_t) - u(M' - A'', d_t, s_t) \geq u(M'' - A', d_t, s_t) - u(M'' - A'', d_t, s_t), \\
& u(M' - A', d_t, s_t) - u(M'' - A', d_t, s_t) \geq u(M' - A'', d_t, s_t) - u(M'' - A'', d_t, s_t), \\
& u(z, d_t, s_t) - u(z - \Delta_M, d_t, s_t) \geq u(z + \Delta_A, d_t, s_t) - u(z + \Delta_A - \Delta_M, d_t, s_t),
\end{aligned} \tag{35}$$

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. \square

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 (17) collapses to

$$\begin{aligned}
u'(c_t(M_t|d_t)) & = \beta \int_0^\infty Ru'(c_{t+1}(M_{t+1}|d_{t+1} = 1)) \cdot \mathbf{1}\{M_{t+1} \leq \bar{M}_{t+1}\} f(d\eta) \\
& + \beta \int_0^\infty Ru'(c_{t+1}(M_{t+1}|d_{t+1} = 0)) \cdot \mathbf{1}\{M_{t+1} > \bar{M}_{t+1}\} f(d\eta).
\end{aligned} \tag{36}$$

where we recall that $M_{t+1} = R(M_t - c_t(M_t|d_t)) + y\eta$. With the change of variables, $q = f(\eta)$, we can write the Euler equation (36) as

$$\begin{aligned}
u'(c_t(M_t|d_t)) & = \beta \int_0^{\bar{q}_t} f^{-1}(q) u'(c_{t+1}(R(M_t - c_t(M_t|d_t)) + yf^{-1}(q), d_{t+1} = 1)) dq \\
& + \beta \int_{\bar{q}_t}^1 f^{-1}(q) u'(c_{t+1}(R(M_t - c_t(M_t|d_t)) + yf^{-1}(q), d_{t+1} = 0)) dq
\end{aligned} \tag{37}$$

where the threshold \bar{q}_t is given by

$$\bar{q}_t = f\left(\frac{\bar{M}_{t+1}}{M_{t+1}}\right). \tag{38}$$

As long as the income shock distribution is not degenerate, the resulting Euler equation (37) 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 $M_{t+1} = \bar{M}_{t+1}$. In turn, this suggests that numerical integration should

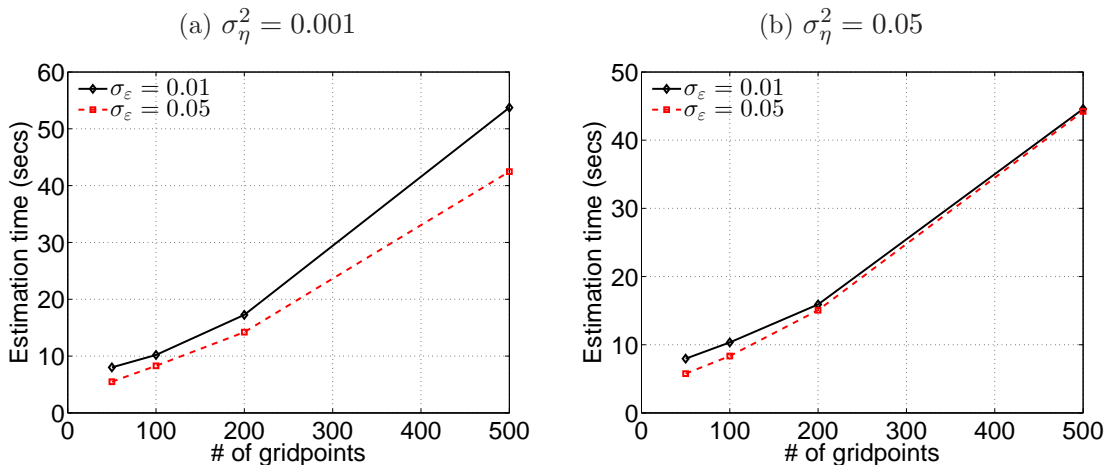
be done twice – once for each case – to ensure that the integral is well-behaved.

In contrast, the naive Euler equation in (36) 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 DC-EGM run times

Figure 9 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 9: Timing: income uncertainty.



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.

D Proof of Extreme Value Homotopy Principle

This appendix proves Theorem 3 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 3 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

$$F(\varepsilon) = \exp\{-\exp\{-\varepsilon\}\}. \quad (39)$$

We have $E\{\varepsilon\} = \gamma$, where $\gamma = 0.577\dots$ 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

$$V_T(s) = \max_{d \in D_T(s)} \max_{c \in C_T(d, s)} u_T(d, c, s). \quad (40)$$

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

$$EV_{T-1}(d, c, s) = \int V_T(s') p_T(\partial s' | d, c, s) \quad (41)$$

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 $t \in \{T - 1, T - 2, \dots, 0\}$ we can define the sequence of functions $\{V_t\}$ recursively using Bellman's equation

$$V_t(s) = \max_{d \in D_t(s)} \max_{c \in C_t(d, s)} \left[u_t(d, c, s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c, s) \right]. \quad (42)$$

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*

$$D = \max_{t \in \{0, 1, \dots, T\}} \sup_{s \in S} |D_t(s)| < \infty \quad (43)$$

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

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

$$EV_t(d, c, s) = \int V_t(s') p_t(\partial s' | d, c, s) \quad (44)$$

is finite and continuous in c .

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

$$c_t(d, s) = \underset{c \in C_t(d, s)}{\operatorname{argmax}} [u_t(d, c, s) + \beta EV_{t+1}(d, c, s)] \quad (45)$$

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

$$\delta_t(s) = \underset{d \in D_t(s)}{\operatorname{argmax}} [u_t(d, c_t(d, s), s) + \beta EV_{t+1}(d, c_t(d, s), s)]. \quad (46)$$

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

$$c_t(s) = c_t(\delta_t(s), s). \quad (47)$$

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

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

$$V_{\sigma, T}(s, \varepsilon) = \max_{d \in D_T(s)} \max_{c \in C_T(d, s)} [u_T(d, c, s) + \sigma \varepsilon(d)]. \quad (48)$$

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.

$$F(\varepsilon_1, \dots, \varepsilon_D) = \prod_{d=1}^D \exp\{-\exp\{-\varepsilon(d)\}\} \quad (49)$$

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

$$\begin{aligned} EV_{\sigma,T}(d, c, s) &= \int_{s'} \int_{\varepsilon'} V_t(s', \varepsilon') F(\partial \varepsilon') p_T(\partial s' | d, c, s) \\ &= \int_{s'} \sigma \log \left(\sum_{d \in D_T(s')} \exp\{u_T(d, c_T(s, d), d)/\sigma\} \right) p_T(\partial s' | d, c, s) \end{aligned} \quad (50)$$

where $c_T(s, d) = \operatorname{argmax}_{c \in C_T(d, s)} u_T(d, c, s)$ 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 (50) 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 2 (Logsum error bounds). *Let $\{v_1, \dots, v_D\}$ be any finite set of D real numbers and let $\sigma > 0$ be a constant. Then we have*

$$0 \leq \sigma \log \left(\sum_{d=1}^D \exp\{v_d/\sigma\} \right) - \max\{v_1, \dots, v_D\} \leq \sigma \log(D). \quad (51)$$

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

$$\log \left(\sum_{d=1}^D \exp\{(v_d - \max\{v_1, \dots, v_D\})/\sigma\} \right) \leq \log \left(\sum_{d=1}^D \exp\{0\} \right) = \log(D). \quad (52)$$

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

$$\log \left(J + \sum_{d=1, d \neq d^*}^D \exp\{(v_d - \max\{v_1, \dots, v_D\})/\sigma\} \right) \geq 0. \quad (53)$$

Combining (52) and (53) with the identity

$$\sigma \log \left(\sum_{d=1}^D \exp\{(v_d - \max\{v_1, \dots, v_D\})/\sigma\} \right) = \sigma \log \left(\sum_{d=1}^D \exp\{v_d/\sigma\} \right) - \max\{v_1, \dots, v_D\} \quad (54)$$

concludes the proof. \square

Lemma 2 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 $\{V_{\sigma,0}, \dots, V_{\sigma,T}\}$ where $V_{\sigma,T}$ is given by equation (48) and the other value functions are given by the Bellman recursion

$$V_{\sigma,t}(s, \varepsilon) = \max_{d \in D_t(s)} \max_{c \in C_t(d,s)} [u_t(d, c, s) + \sigma\varepsilon(d) + \beta EV_{t+1}(d, c, s)] \quad (55)$$

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

$$EV_{\sigma,t+1}(d, c, s) = \sigma \int_{s'} \log \left(\sum_{d' \in D_{t+1}(s')} \exp\{v_{\sigma,t+1}(d', c_{\sigma,t+1}(d', s'), s')/\sigma\} \right) p_{t+1}(\partial s' | d, c, s), \quad (56)$$

where

$$v_{\sigma,t+1}(d, c, s) = u_{t+1}(d, c, s) + \beta EV_{\sigma,t+2}(d, c, s) \quad (57)$$

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

$$c_{\sigma,t+1}(d, s) = \underset{c \in C_{t+1}(d,s)}{\operatorname{argmax}} [v_{\sigma,t+1}(d, c, s)]. \quad (58)$$

Note that we used the Williams-Daly-Zachary Theorem again to obtain the expression for $EV_{t+1}(d, c, s)$ in equation (56) 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 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

$$\delta_{\sigma,t}(s, \varepsilon) = \underset{d \in D_t(s)}{\operatorname{argmax}} [v_{\sigma,t}(d, c_{\sigma,t}(d, s), s) + \sigma\varepsilon(d)], \quad (59)$$

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

$$c_{\sigma,t}(s, \varepsilon) = c_{\sigma,t}(\delta_{\sigma,t}(s, \varepsilon), s). \quad (60)$$

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

$$V_{\sigma,t}(s, \varepsilon) = \max_{d \in D_t(s)} [v_{\sigma,t}(d, c_{\sigma,t}(d, s), s) + \sigma\varepsilon(d)], \quad (61)$$

which follows directly from the Bellman equation (55) and the definition of the v_t function in equation (57). 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

$$\begin{aligned} EV_{\sigma,t}(s) &= \int_{\varepsilon} V_{\sigma,t}(s, \varepsilon) F(\varepsilon) \\ &= \sigma \left(\sum_{d \in D_t(s)} \exp\{v_{\sigma,t}(d, c_{\sigma,t}(d, s), s)/\sigma\} \right). \end{aligned} \quad (62)$$

We are in the position now to state the main result which is a reformulation of Theorem 3 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*

$$\Gamma = \{(V_0, \dots, V_T), (\delta_0, \dots, \delta_T), (c_0, \dots, c_T)\} \quad (63)$$

be the solution to the limiting DP problem without taste shocks given in equations (40), (42), (45) and (46) above. Similarly, let

$$\Gamma_{\sigma} = \{(V_{\sigma,0}, \dots, V_{\sigma,T}), (\delta_{\sigma,0}, \dots, \delta_{\sigma,T}), (c_{\sigma,0}, \dots, c_{\sigma,T})\} \quad (64)$$

be the solution to the the perturbed DP problem with Type I extreme value taste shocks with scale parameter $\sigma > 0$ given in equations (48), (55), (56), (59) and (60). Then as $\sigma \rightarrow 0$ we have

$$\lim_{\sigma \downarrow 0} \Gamma_{\sigma} = \Gamma, \quad (65)$$

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 (62) so that we have uniform bound

$$\forall t \sup_{s \in S} |EV_{\sigma,t}(s) - V_t(s)| \leq \sigma \left[\sum_{j=0}^{T-t} \beta^j \right] \log(D), \quad (66)$$

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

$$\begin{aligned} \lim_{\sigma \downarrow 0} \delta_{\sigma,t}(s, \varepsilon) &= \delta_t(s) \\ \lim_{\sigma \downarrow 0} c_{\sigma,t}(s, \varepsilon) &= c_t(s), \end{aligned} \quad (67)$$

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 (66) by induction using Lemma 2 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 3 (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

$$\forall t, s \quad 0 \leq EV_{\sigma,t}(s) - V_t(s) \leq \sigma \left[\sum_{j=0}^{T-t} \beta^j \right] \log(D). \quad (68)$$

Lemma 3 can be proved by induction using Lemma 2. 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 (40), which can be rewritten in terms of the choice-specific continuous choice rule as

$$V_T(s) = \max_{d \in D_T(s)} [u_T(d, c_T(d, s), s)] \quad (69)$$

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

$$EV_{\sigma,T}(s) = \sigma \log \left(\sum_{d \in D_T(s)} \exp\{u_T(d, c_T(d, s), s)/\sigma\} \right), \quad (70)$$

since it is easy to see that $c_T(d, s) = c_{\sigma,T}(d, s)$ in the final period T . Using Lemma 2, we obtain the bounds

$$0 \leq EV_{\sigma,T}(s) - V_T(s) \leq \sigma \log(D), \quad \forall s \in S, \quad (71)$$

which establishes the base case for our induction proof. Now suppose the inductive hypothesis holds, i.e. the error bounds are given by equation (68) at period $T, T-1, \dots, t+1$. We now want to show that it also holds at period t . We have

$$V_t(s) = \max_{d \in D_t(s)} \left[u_t(d, c_t(d, s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c_t(d, s), s) \right], \quad (72)$$

and

$$EV_{\sigma,t}(s) = \sigma \log \left(\sum_{d \in D_t(s)} \exp \left\{ \frac{1}{\sigma} \left[u_t(d, c_{\sigma,t}(d, s), s) + \beta \int EV_{\sigma,t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d, s), s) \right] \right\} \right). \quad (73)$$

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 (72):

$$\tilde{V}_t(s) = \max_{d \in D_t(s)} \left[u_t(d, c_{\sigma,t}(d, s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d, s), s) \right]. \quad (74)$$

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

$$\tilde{V}_t(s) \leq V_t(s), \quad \forall s \in S. \quad (75)$$

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 (73). We have

$$E\tilde{V}_{\sigma,t}(s) = \sigma \log \left(\sum_{d \in D_t(s)} \exp \left\{ \frac{1}{\sigma} \left[u_t(d, c_{\sigma,t}(d, s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d, s), s) \right] \right\} \right). \quad (76)$$

Note that we can write

$$\begin{aligned} EV_{\sigma,t}(s) &= \sigma \log \left(\sum_{d \in D_t(s)} \exp \left\{ \frac{1}{\sigma} \left[u_t(d, c_{\sigma,t}(d, s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d, s), s) \right. \right. \right. \\ &\quad \left. \left. \left. + \beta \int [EV_{\sigma,t+1}(s') - V_{t+1}(s')] p_{t+1}(\partial s' | s, c_{\sigma,t}(d, s), s) \right] \right\} \right). \end{aligned} \quad (77)$$

By the inductive hypothesis, it follows that

$$\beta \int [EV_{\sigma,t+1}(s') - V_{t+1}(s')] p_{t+1}(\partial s' | d, c_{\sigma,t}(d, s), s) \leq \sigma \beta \left[\sum_{j=0}^{T-t-1} \beta^j \right] \log(D). \quad (78)$$

Thus, it follows from inequality (78) that the following inequality holds

$$EV_{\sigma,t}(s) \leq E\tilde{V}_{\sigma,t}(s) + \sigma \beta \left[\sum_{j=0}^{T-t-1} \beta^j \right] \log(D). \quad (79)$$

From Lemma 2 we have

$$E\tilde{V}_{\sigma,t}(s) - \tilde{V}_t(s) \leq \sigma \log(D). \quad (80)$$

Using inequalities (75) and (80) it follows that

$$0 \leq EV_{\sigma,t}(s) - V_t(s) \leq \sigma \left[\sum_{j=0}^{T-t} \beta^j \right] \log(D), \quad (81)$$

completing the induction step of the argument. It follows by mathematical induction that inequality (68) holds for all $t \in \{0, 1, \dots, T\}$ so Lemma 3 is proved. \square

Note that the bound (68) 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 (68) that (66) 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 $t \in \{0, 1, \dots, T\}$. We will allow for the possibility that there are multiple values of d and c that attain the optimum values in equations (45) and (46) 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 4 (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 (67) 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

$$\begin{aligned}
& u_t(\delta_t(s), c_t(s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | \delta_t(s), c_t(s), s) \\
&= u_t(\delta_t(s), c_t(\delta_t(s), s), s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | \delta_t(s), c_t(\delta_t(s), s), s) \\
&> u_t(d, c, s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c, s) \quad \forall c \neq c_t(s) \in C_t(d, s), d \neq \delta_t(s) \in D_t(s). \quad (82)
\end{aligned}$$

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 (82). We have

$$\begin{aligned}
EV_{\sigma,t}(s) &= \sigma \log \left(\sum_{d \in D_t(s)} \exp \left\{ \frac{1}{\sigma} \left[u_t(d, c_{\sigma,t}(d, s), s) + \beta \int EV_{\sigma,t+1}(s') p_{t+1}(\partial s' | d, c_{\sigma,t}(d, s), s) \right] \right\} \right) \\
&= \int V_{\sigma,t}(s, \varepsilon) F(\varepsilon) \\
&= \int \left[u_t(\delta_{\sigma,t}(s, \varepsilon), c_{\sigma,t}(\delta_{\sigma,t}(s, \varepsilon), s), s) + \sigma \varepsilon (\delta_{\sigma,t}(s, \varepsilon) \right. \\
&\quad \left. + \beta \int EV_{\sigma,t+1}(s') p_{t+1}(\partial s' | \delta_{\sigma,t}(s, \varepsilon), c_{\sigma,t}(\delta_{\sigma,t}(s, \varepsilon), s)) \right] F(\varepsilon) \quad (83)
\end{aligned}$$

By Lemma 3 we have that uniformly for each $t \in \{0, 1, \dots, T\}$ and all $s \in S$

$$\lim_{\sigma \downarrow 0} EV_{\sigma,t}(s) = V_t(s). \quad (84)$$

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

$$\begin{aligned}
\lim_{\sigma_n \downarrow 0} \delta_{\sigma_n,t}(s, \varepsilon) &= d \\
\lim_{\sigma_n \downarrow 0} c_{\sigma_n,t}(\delta_{\sigma_n,t}(s, \varepsilon), s) &= c \quad (85)
\end{aligned}$$

these limits together with the representation of $EV_{\sigma,t}(s)$ in the last equation of (83) implies that

$$V_t(s) = u_t(d, c, s) + \beta \int V_{t+1}(s') p_{t+1}(\partial s' | d, c, s) \quad (86)$$

However because $\delta_t(s)$ and $c_t(s)$ are the unique optimizers of Bellman equation in equation (82) 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 (67) the statement of Lemma 4 hold. \square

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 5 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

$$\begin{aligned} \lim_{\sigma \downarrow 0} \delta_{\sigma,t}(s, \varepsilon) &= d \\ \lim_{\sigma \downarrow 0} c_{\sigma,t}(s, \varepsilon) &= c \end{aligned} \tag{87}$$

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

Lemma 5 (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 4 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 (45) and (46) 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 4, if c is a limit point of $c_{\sigma,t}(s, \varepsilon)$ as $\sigma \rightarrow 0$ we can represent c as

$$c \in \lim_{\sigma \downarrow 0} c_{\sigma,t}(d, s), \tag{88}$$

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 Lemma 4 we can obtain a contradiction, because following the same argument we can show that equation (86) 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 (45) and (46). \square

This concludes the proof of Theorem 2. \square

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.
- DEATON, A. (1991): “Saving and Liquidity Constraints,” *Econometrica*, 59(5), 1221–1248.
- 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 (2017): “A general endogenous grid method for multi-dimensional models with non-convexities and constraints,” *Journal of Economic Dynamics and Control*, 74, 87–107.
- DUBIN, J. A. AND D. L. MCFADDEN (1984): “An Econometric Analysis of Residential Electric Appliance Holdings and Consumption,” *Econometrica*, 52(2), 345–362.
- EJRNÆS, M. AND T. H. JØRGENSEN (2016): “Saving Behavior around Intended and Unintended Childbirths,” Working paper, 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.
- 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,” Working paper, CEPAR.
- 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,” Working paper, 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,” Working paper, Science Po.
- PHELPS, E. (1962): “The Accumulation of Risky Capital: A Sequential Utility Analysis,” *Econometrica*, 30-4, 729–743.

- 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,” Working paper, Johns Hopkins University.