Estimating Discrete-Continuous Choice Models: The Endogenous Grid Method with Taste Shocks †

Fedor Iskhakov
CEPAR, University New South Wales
Thomas H. Jørgensen
University of Copenhagen
John Rust
Georgetown University
Bertel Schjerning
University of Copenhagen

November 25, 2015

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

JEL classification: C13, C63, D91

† We acknowledge helpful comments from Chris Carroll and *many other people*, participants at seminars at UNSW, University og Copenhagen, the 2012 conferences of the Society of Economic Dynamics, the Society for Computational Economics, the Initiative for Computational Economics at Zurich (ZICE 2014, 2015). This paper is part of the IRUC research project financed by the Danish Council for Strategic Research (DSF). Iskhakov, Rust and Schjerning gratefully acknowledge this support. Iskhakov gratefully acknowledges the financial support from the Australian Research Council Centre of Excellence in Population Ageing Research (project number CE110001029) and Michael P. Keane's Australian Research Council Laureate Fellowship (project number FL110100247). Jørgensen gratefully acknowledges financial support from the Danish Council for Independent Research in Social Sciences (FSE, grant no. 4091-00040). **Correspondence address:** ARC Centre of Excellence in Population Ageing Research (CEPAR), University of New South Wales, Sydney 2052, Australia, phone: (+61)299319202, email: `f.iskhakov@unsw.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 Gridpoint Method (EGM) by Carroll (2006) to discrete-continuous (DC) models. We refer to it as the DC-EGM algorithm. We embed the DC-EGM algorithm in the inner loop of the nested fixed point (NFXP) algorithm (Rust, 1987), and show that the resulting Maximum Likelihood estimator produces accurate estimates of the structural parameters at low computational cost.

A classic example of a DC model is a life cycle model with simultaneous 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*. The EGM avoids the need to numerically solve the nonlinear Euler equation for the optimal policy at each grid point in the state space. Instead, the 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] The DC-EGM retains the main desirable properties of the EGM, namely it avoids all root-finding operations and handles borrowing constraints in an efficient manner.[3]

Dynamic programs that have only discrete choices are substantially easier to solve, since the

[1]There are relatively few examples of *structural estimation* or *numerical solution* of DC models. Some prominent examples include the model of optimal non-durable consumption and housing purchases (Carroll and Dunn, 1997), optimal saving and retirement (French and Jones, 2011), and optimal saving, labor supply and fertility (Adda, Dustmann and Stevens, 2015). These applications approximate the solution by discretizing the continuous choice variables.

[2]The EGM is in fact a specific application of what is referred to as "controlling the post-decision state" in operations research and engineering (Bertsekas, Lee, van Roy and Tsitsiklis, 1997). Carroll (2006) introduced the idea in economics by developing the EGM algorithm with the application to the buffer-stock precautionary savings model. Since then the idea became widespread in economics. Further generalizations of the 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 the EGM to Mathematical Programming with Equilibrium Constraints (MPEC).

[3]The DC-EGM has been implemented in several recent empirical applications, Jørgensen (2014); Yao, Fagereng and Natvik (2015); Ejrnæs and Jørgensen (2015); Druedahl and Jørgensen (2015); Druedahl (2015); Iskhakov and Keane (2015).

optimal decision rule is simply the alternative with highest choice-specific value. However, solving dynamic programming problems that combine continuous and discrete choices is substantially more complicated, since discrete choices introduce kinks and non-concave regions in the value function that lead to discontinuities in the policy function of the continuous choice (consumption). Therefore, the Euler equation for consumption is only *necessary* but not sufficient (Clausen and Strub, 2013). This complication is a feature of the problem itself and complicates the use of any method for solving DC models.

We illustrate this issue by solving and estimating a life cycle problem with continuous consumption and binary retirement decisions. 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 we make what appears to be a slight change to Phelp's problem — allowing a worker with logarithmic utility to also make a binary irreversible retirement decision — the optimal consumption function becomes unexpectedly complex, with multiple discontinuities and non-monotonicities in the optimal consumption rule. We were able to derive an *analytic solution* for the optimal consumption rule of this model, which serves as an illustrative test problem throughout the paper. The complexity in the solution is due to the fact that the value function is a maximum of *choice-specific* value functions, e.g. the maximum of the value of retiring and not-retiring. The max operator creates *kinks* (and therefore non-concave regions) in the value function that cause the Euler equation to have *multiple solutions* posing a significant challenge to *any* numerical method to accurately approximate consumption functions that have multiple discontinuities and non-monotonicities.

Fella (2014) generalizes the EGM to solve non-concave problems, including models with discrete and continuous choices. However, in this paper we show that introducing *IID* Extreme Value Type I choice-specific taste shocks[4] not only facilitates the maximum likelihood estimation, but also ameliorates the difficulties which are inherent to the solution of DC models. This approach results in *conditional choice probabilities* that have the *multinomial logit form* and *closed form* expressions for the conditional expectation of the value function. In econometric applications these extreme value taste shocks are essential for generating predictions from dynamic programming models that

[4]Though there are many ways that stochastic shocks and state variables can be introduced into DC models, we focus on a particularly tractable approach to including multidimensional stochastic taste shocks into discrete choice models that is well known in the econometrics literature (McFadden, 1973; Rust, 1987). In principle, this 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.

are "statistically non-degenerate" — that is, they imply that the probability of choosing any of the alternatives is always positive. 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 these shocks can serve as a smoothing device or "homotopy perturbation" that facilitates the solution of models that are not sufficiently smooth or concave to be solvable by the traditional methods with the desired reliability and accuracy.

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* 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. The combination of taste shocks and the stochastic variables in the model is a powerful device to prevent the propagation and accumulation of kinks.

The Extreme Value distributed taste shocks can either be interpreted as *structural* unobserved state variables or as a logit *smoothing device* of an underlying deterministic model of interest. Let $\sigma \geq 0$ denote the scale parameter of the corresponding Extreme Value distribution. We show that in the latter case σ can be interpreted as a *homotopy* or *smoothing* parameter, that can be chosen in such a way that the deterministic model without taste shocks is approximated by the smoothed model to any desirable degree of precision.

We show that 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 the kinks in the value functions and suppress their accumulation in successive backward induction steps. As noted above, in combination with other structural shocks in the model (i.e. income uncertainty or random returns on savings) taste shocks can completely eliminate the problems that render exact solution of the deterministic model to be cumbersome and unpractical. 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 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.

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) to allow for a discrete retirement decision. We show there is a closed-form solution to this problem even though it is exceptionally complex relative to the original version of the problem without the discrete retirement decision. Using this simple model we illustrate how DC-EGM works and demonstrate its accuracy. We then introduce extreme value taste shocks and show how the implied smoothing affects the consumption and retirement decision rules. Section 3 presents the DC-EGM algorithm. Section 4, shows 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 other problems and models that DC-EGM is applicable to.

2 An Illustrative Problem: Consumption and Retirement

This section extends the classic life-cycle consumption/savings model of Phelps (1962) to allow for a binary retirement decision. We derive an analytic solution to the simplistic life cycle problem with logarithmic utility that serves both to illustrate the complexity caused by the addition of a discrete retirement choice and how the DC-EGM can be applied. While we focus on this simple

illustrative example for expositional clarity, the DC-EGM method can be applied to a much more general class of problems that we discuss in the conclusion - including the extended version of the retirement model that we use in the Monte Carlo exercise. While we initially illustrate the complexity of the solution without any stochastic elements, we include both taste and income shocks to this simple model and discuss how the complexity of the solution reduces.

2.1 Deterministic model of consumption/savings and retirement

Consider the discrete-continuous (DC) dynamic optimization problem

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

where agents choose consumption c_t and whether to retire to maximize the discounted stream of utilities. Let $d_t = 0$ denote the choice to retire and $d_t = 1$ the continue working. We therefore implicitly normalize the disutility of work to 1. To keep the solution as simple as possible, we assume that retirement is absorbing, i.e. once workers retire they are unable to return to work.

Agents solve (1) subject to a sequence of period-specific borrowing constraints, $c_t \leq M_t$ where $M_t = R(M_t - c_t) + y_t d_{t-1}$ is the consumer's resources available for consumption in the beginning of period t. We assume a fixed gross interest rate, R, and a deterministic constant labor income y_t realized in the beginning of period t. Income is a function of the previous period's end-of-period labor market choice, d_{t-1}.

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

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

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

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

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

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

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

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

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

This kink point at $\overline{M}_t$ is also the *optimal retirement threshold* — the optimal decision for a worker whose resources satisfy $M_t \leq \overline{M}_t$ is to keep working (not to retire) and to use the consumption rule $c_t(M_t|d_t = 1)$, whereas the optimal decision for a worker whose wealth exceeds $\overline{M}_t$ is to retire and to consume $c_t(M_t|d_t = 0)$. The optimal consumption rule of the worker $c_t(M_t)$ is then a combination of the two.

At each primary kink point the worker is indifferent between retiring and continuing to work, and the value function is non-differentiable. However the left and right hand derivatives do exist and we have $V_t^-(\overline{M}_t) < V_t^+(\overline{M}_t)$. The discontinuity in the derivative of $V_t(M_t)$ at the primary kink point creates a discontinuity in the optimal consumption function in the *previous* period $t-1$, and a corresponding kink in the value function $V_{t-1}(M_{t-1})$. In effect, the primary kinks propagate back in time. These kinks do not correspond to the points of indifference between the current discrete alternatives, but instead appear as reverberations of the primary kinks in the future. We refer to these points as *secondary kinks.*

Theorem 1 illustrates the complexity in the optimal solution introduced by the seemingly minor change to Phelps' problem of allowing a discrete retirement decision.

Theorem 1 (Analytical solution to the retirement problem)**.** *Let $c_{T-\tau}(M)$ denote the optimal consumption function τ periods before the horizon T in the workers decision problem given by*

(2)-(4) *with constant income* $y_t = y$. *The optimal retirement threshold is* $\overline{M}_{T-\tau}$ *given by*

$$\overline{M}_{T-\tau} = \frac{(y/R)e^{-K}}{1-e^{-K}}, \tag{6}$$

where

$$K = d_{T-\tau}\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1}. \tag{7}$$

The optimal consumption function is given by:

$$c_{T-\tau}(M) = \begin{cases} M & \text{if} \quad M \leq y/R\beta, \\ [M + y/R]/(1+\beta) & \text{if} \quad y/R\beta \leq M \leq \overline{M}_{T-\tau}^{l_1}, \\ [M + y(1/R + 1/R^2)]/(1+\beta+\beta^2) & \text{if} \quad \overline{M}_{T-\tau}^{l_1} \leq M \leq \overline{M}_{T-\tau}^{l_2}, \\ \dots & \dots \quad \dots \\ \left[M + y\left(\sum_{i=1}^{\tau-1} R^{-i}\right)\right]\left(\sum_{i=0}^{\tau-1}\beta^i\right)^{-1} & \text{if} \quad \overline{M}_{T-\tau}^{l_{\tau-2}} \leq M \leq \overline{M}_{T-\tau}^{l_{\tau-1}}, \\ \left(M + \left(\sum_{i=1}^{\tau} R^{-i}\right)\right)\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if} \quad \overline{M}_{T-\tau}^{l_{\tau-1}} \leq M \leq \overline{M}_{T-\tau}^{r_1}, \\ \left[M + \left(\sum_{i=1}^{\tau-1} R^{-i}\right)\right]\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if} \quad \overline{M}_{T-\tau}^{r_1} \leq M \leq \overline{M}_{T-\tau}^{r_2}, \\ \dots & \dots \quad \dots \\ [M + y(1/R + 1/R^2)]\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if} \quad \overline{M}_{T-\tau}^{r_{\tau-2}} \leq M \leq \overline{M}_{T-\tau}^{r_{\tau-1}}, \\ [M + y/R]\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if} \quad \overline{M}_{T-\tau}^{r_{\tau-1}} \leq M \leq \overline{M}_{T-\tau}, \\ M\left(\sum_{i=0}^{\tau}\beta^i\right)^{-1} & \text{if} \quad \overline{M}_{T-\tau} < M. \end{cases} \tag{8}$$

The value function $V_{T-\tau}(M)$ *is piecewise logarithmic, that is, it can be written as* $V_{T-\tau}(M) = B_{T-\tau}\log(c_{T-\tau}(M$ *for constants* $(B_{T-\tau}, C_{T-\tau})$.

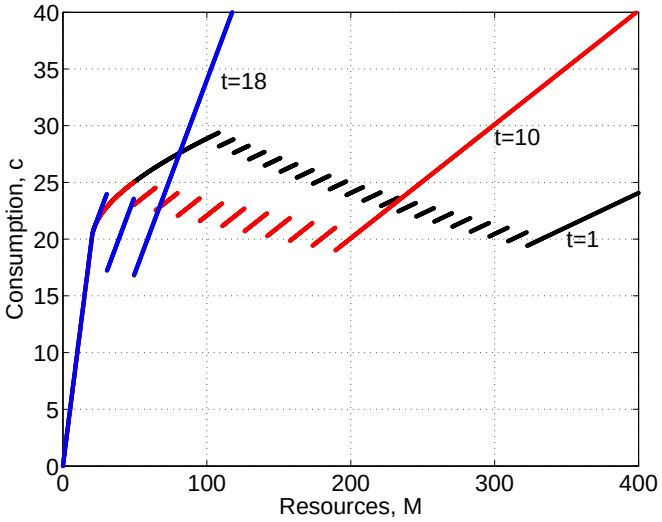
For $\tau \geq 1$, *the value function has one primary kink at the optimal retirement threshold,* $M = \overline{M}_{T-\tau}$. *For* $\tau > 1$, *in addition the value function has* $\tau - 1$ *secondary kinks points associated with future retirement* $\{\overline{M}_{T-\tau}^{r_1}, \dots, \overline{M}_{T-\tau}^{r_{\tau-1}}\}$. *If* $R\beta < 1$, *the value function also has one kink at* $M = y/R$ *associated with binding liquidity constraints in period* $T - \tau$ *and* $\tau - 1$ *kinks* $\{\overline{M}_{T-\tau}^{l_1}, \dots, \overline{M}_{T-\tau}^{l_{\tau-1}}\}$ *associated with the binding liquidity constraints in the future. If* $R\beta = 1$, *we have* $y/R\beta = \overline{M}_{T-\tau}^{l_1} = \cdots = \overline{M}_{T-\tau}^{l_{\tau-1}}$.

Proof. The proof of Theorem 1 is by mathematical induction. The proof is omitted here for space considerations and is available upon request. □

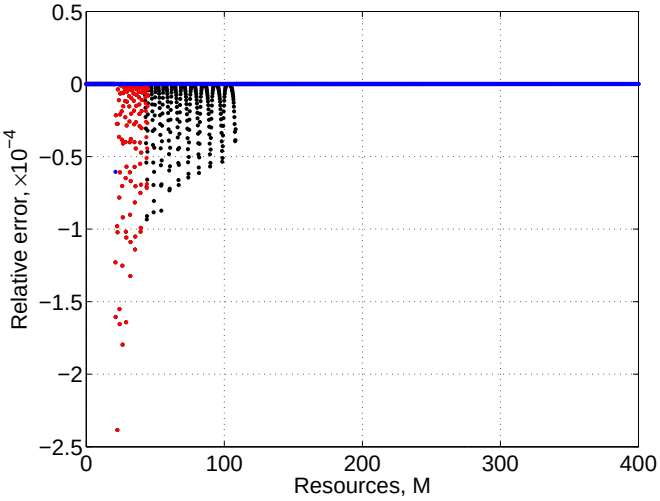
The left panel of Figure 1 illustrates the optimal consumption rule given in Theorem 1 while

Figure 1: Optimal Consumption Functions.

(a) Analytical Solution



(b) Relative Error of DC-EGM solution



Notes: The plots show optimal consumption rules of the worker in the consumption-savings model with $R = 1$, $\beta = 0.98$, $y = 20$, and $T = 20$. The left panel illustrates the analytical solution while the right panel illustrates the numerical error from the solution found by applying the DC-EGM algorithm as discussed in the text.

relative error of the numerical solution found by applying the DC-EGM as described below in Section 3 is illustrated in the right panel.[5] The analytical and numerical solution are practically identical. Only minor differences are found around the piece-wise linear part between the credit constrained region and the discontinuities because the numerical solution applies linear interpolation over this region ignoring the piece-wise nature while the analytical solution explicitly accounts for this feature as described in Theorem 1.

The Figure illustrates that the primary kinks propagates backward in time into *secondary kinks*. For example, $V_{T-2}(M)$ has one secondary kink. This represents a critical level of resources where workers, who do not have enough wealth to retire at $T-2$, can forecast that they will have enough resources at the beginning of period $T-1$ *provided they consume at the lower amount* $c_{T-2}(M) = (M + y/R)/(1 + \beta + \beta^2)$. Instead, the consumer can consume at the higher level $c_{T-2}(M) = (M + y/R + y/R^2)/(1 + \beta + \beta^2)$ but will then not have enough resources at the beginning of period $T-1$ to retire at that time. So the secondary kink $\overline{M}^{r_1}_{T-2}$ represents the

[5] Although the DC-EGM was implemented with 2000 discrete grid-points to approximate the (complex) consumption function, it only took 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 in Figure 1.

critical level of wealth where the worker is indifferent between consuming a higher amount and not retiring next period, and consuming a lower amount and being able to retire next period. The kink in $V_{T-2}(M)$ at $M = \overline{M}^{r_1}_{T-2}$ corresponds to the discontinuity in the consumption function $c_{T-2}(M)$ at this same value of M. The blue line in Figure 1 is the consumption function $c_{T-2}(M) = c_{18}(M)$. We see that it has two discontinuities, one at $M = 30.56 = \overline{M}^{r_1}_{T-2}$ (the *secondary* kink point in V_{T-2}) and another at $M = 49.37 = \overline{M}_{T-2}$ (the *primary* kink point in V_{T-2}).

Before we describe in detail how the DC-EGM works, we turn to a slightly more realistic situation in which consumers face income and taste uncertainty. Uncertainty will tend to smooth the problem and reduce the complexity of the optimal consumption function.

2.2 Adding Taste Shocks and Income Uncertainty

Consider now an extension of the model presented above, where the decision makers receive choice-specific taste shocks and face income uncertainty. More specifically, assume that income when working is $y_t = y\eta_t$, where η_t is log-normally distributed multiplicative idiosyncratic income shock, $\log \eta_t \sim \mathcal{N}(-\sigma_\eta^2/2, \sigma_\eta^2)$.[6]

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

As before, we focus solely on the worker's problem. The Bellman equation (2) has to be rewritten to include the taste shocks,

$$V_t(M_t) = \max\{v_t(M_t|d_t = 0) + \sigma_\varepsilon \varepsilon_t(0), v_t(M_t|d_t = 1) + \sigma_\varepsilon \varepsilon_t(1)\}, \tag{9}$$

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

$$v_t(M_t|d_t = 1) = \max_{0 \le c_t \le M_t} \left\{ \log(c_t) - 1 + \beta \int EV_{t+1}(R(M_t - c_t) + y\eta_{t+1}) f(d\eta_{t+1}) \right\}. \tag{10}$$

[6] 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.

Because the taste shocks are independent Extreme Value distributed random variables, the expected value function, EV_{t+1}, is given by the well-known *logsum* formula

$$\begin{aligned} EV_{t+1}(M_{t+1}) &= E\left[\max\left\{v_{t+1}(M_{t+1}|d_t=0)+\sigma_\varepsilon\varepsilon(0),\ v_{t+1}(M_{t+1}|d_t=1)+\sigma_\varepsilon\varepsilon(1)\right\}\right] \\ &= \sigma_\varepsilon \log\left\{exp\left[\frac{v_{t+1}(M_{t+1}|d_t=0)}{\sigma_\varepsilon}\right]+exp\left[\frac{v_{t+1}(M_{t+1}|d_t=1)}{\sigma_\varepsilon}\right]\right\}. \end{aligned} \tag{11}$$

The immediate effect of introducing the taste shocks is complete elimination of the *primary* kinks due to logit smoothing: the expected value function in (11) is a smooth function of M_t around the point where $v_t(M_t|d_t=1)=v_t(M_t|d_t=0)$. With large enough σ_ε, the value function, $v_t(M_t|d_t=1)$, becomes globally concave.[7] Even when σ_ε is not large enough to "concavify" the value function completely, it may still eliminate at least some of the secondary kinks.

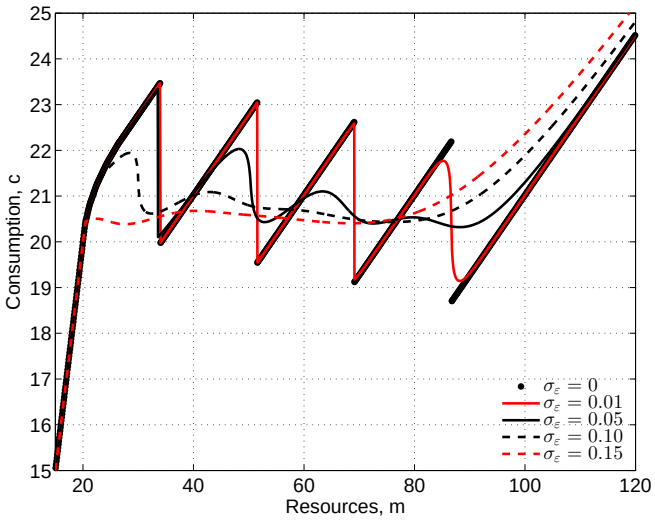
Figure 2 shows the consumption function $c_t(M_t|d_t=1)$ for a worker conditional on the choice to continue working, for different values of smoothing parameter $\sigma_\varepsilon \in \{0, 0.01, 0.05, 0.10, 0.15\}$. The left panel plots the optimal consumption in the absence of income uncertainty ($\sigma_\eta = 0$) while income uncertainty ($\sigma_\eta = \sqrt{0.005}$) is added in the right panel. The plots are drawn for the period $T-5$. Focusing first on the case without income uncertainty, in the absence of taste shocks the choice-specific consumption function has four discontinuities corresponding to the four secondary kinks in the value function V_{T-5} (see Theorem 1). These discontinuities mirror all four (primary and secondary) kinks in the value function in period $t=T-4$.

It is evident that taste shocks of large scale ($\sigma_\varepsilon \geq 0.05$) manage to smooth the function completely, eliminating all four discontinuities (and thus, eliminating the non-concavity of the value function in period $T-4$). Yet, for $\sigma_\varepsilon = 0.01$ only the rightmost discontinuity is distinctively smoothed out. This implies that even when complete "concavitication" is not achieved, the accumulation of the secondary kinks is reduced.

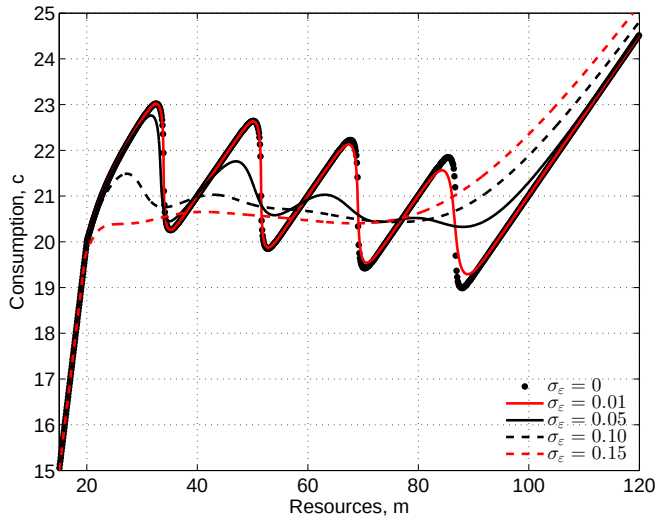
[7] To see this, note that as the variance of the taste shocks increases, the choice-specific value functions are dominated by the noise and the disutility of work becomes relatively less important. In turn, the choice-specific value functions become similar, $\lim_{\sigma_\varepsilon\to\infty} v_t(M_t,1) = v_t(M_t,0)$. As a result, the corresponding value functions will be globally concave.

Figure 2: Optimal Consumption Rules for Agent Working Today ($d_{t-1} = 1$).

(a) Without income uncertainty, $\sigma_\eta = 0$



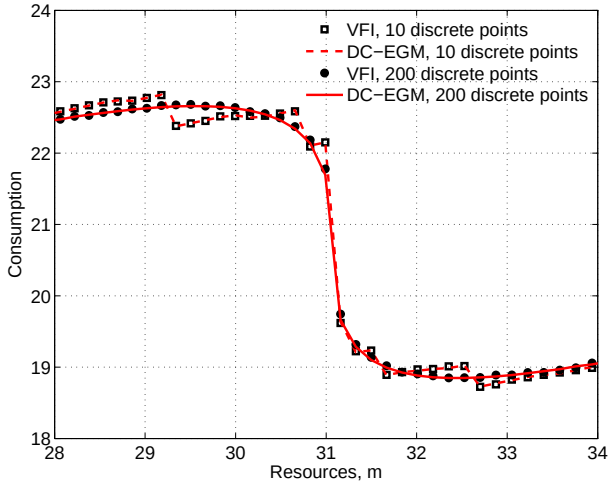
(b) With income uncertainty, $\sigma_\eta = \sqrt{.005}$



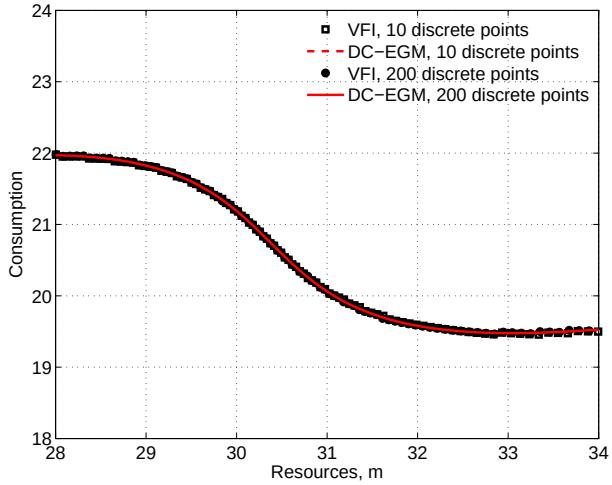
Notes: The plots show optimal consumption rules of the worker who decides to continue working in the consumption-savings model with retirement in period $t = T - 5$ for a set of taste shock scales σ_ε in the absence of income uncertainty, $\sigma_\eta = 0$, (left panel) and in presence of income uncertainty,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

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

(a) $\sigma_\varepsilon = 0$



(b) $\sigma_\varepsilon = 0.05$



Notes: Figure 3 illustrates how the number of discrete points used to approximate expectations regarding future income affects the consumption functions from value function iteration (VFI) and the 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.

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 smoothing of the problem. 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 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 2 that the non-concavity of the value function can be eliminated with a smaller taste shock ($\sigma_\varepsilon = 0.01$) when additional smoothing, through uncertainty, is present in the model.

An additional benefit of the inclusion of taste shocks is that the expected value function EV_t is a smooth function of the continuous state variables. This facilitates standard numerical integration with respect to other stochastic shocks in the model. In particular, standard quadrature rules apply and there is no need to break up the integrals into separate numerical integrals over intervals defined by the kinks in the value function. When $\sigma = 0$, the numerical integrals may not necessarily eliminate the kinks, even if all kinks are eliminated in the true (exact) integral of the value function. This creates the problem that there may be *artificial discontinuities* introduced into the numerically calculated consumption functions when the true consumption functions are continuous in M_t. This problem is illustrated in the left panel of Figure 3. The right illustrates how moderate smoothing ($\sigma_\varepsilon = .05$) significantly reduces this approximation error and knocks out the artificial kinks.

Taste shocks, ε_t, can have a structural interpretation as unobserved state variables, or a smoothing interpretation as a technical device to simplify the problem. In the former interpretation, σ_ε is a scale parameter of taste shocks, and has to be estimated along with other structural parameters. In the latter case, σ_ε is the amount of smoothing and has to be chosen and fixed before the estimation. It can be shown that if the true model does not have taste shocks, the level of σ_ε can always be chosen is such a way, that the perturbed model approximates the true deterministic model with an arbitrary degree of precision. Figure 2 illustrates this idea graphically. As σ_ε approaches zero, the optimal consumption rule approaches the policy function of the original problem in every point.

3 The DC-EGM Algorithm

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

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

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

3.1 Finding all solutions to the Euler equation

Because retirement is an absorbing state and retirees only choose consumption, invoking the DC-EGM algorithm is only necessary for solving the workers problem. The consumption/savings problem of the retirees can be solved using the standard EGM method (Carroll, 2006) at very low computational cost. The Euler equation for the worker's problem is defined by equations (3), (9) and (10) and is given by[8]

$$u'(c_t) \quad = \quad \beta R \mathbb{E}_t \Bigg[\sum_{j=0,1} u'\big(c_{t+1}(M_{t+1}|d_{t+1}=j)\big) P_{t+1}(d_{t+1}=j|M_{t+1}), \Bigg] \tag{12}$$

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

Conditional on a particular value of the current decision, d_t, we follow the EGM algorithm and form an exogenous *ascending* grid over end-of-period wealth,[9] $\vec{A} = \{A^1, \ldots, A^G\}$ where

[8] See Appendix A for derivation.

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

$A^j > A^{j-1}$, $\forall j \in \{2, \dots, G\}$ and G is the number of discrete grid points used to approximate the continuous consumption policy function. Because the end-of-period wealth is a *sufficient statistic* for the consumption decision in the current period, the next period resources are given by

$$M_{t+1}(\vec{A}) = R\vec{A} + d_t y \eta_{t+1}. \tag{13}$$

The utility function in (1) is analytically invertible, therefore the current period consumption can be calculated *directly* using the inverted Euler equation

$$c_t(\vec{A}|d_t) = (u')^{-1}\Big(\beta \text{RHS}\big(M_{t+1}(\vec{A})\big)\Big), \tag{14}$$

where $\text{RHS}\big(M_{t+1}(\vec{A})\big)$ is the right hand side of (12) evaluated at the points $M_{t+1}(\vec{A})$ using the next period optimal consumption rules $c_{t+1}(M_{t+1}|d_{t+1})$. Finally, combining the current consumption $c_t(\vec{A}|d_t)$ found in (14) with the points of $\vec{A}$ we get the *endogenous grid* over the current period wealth

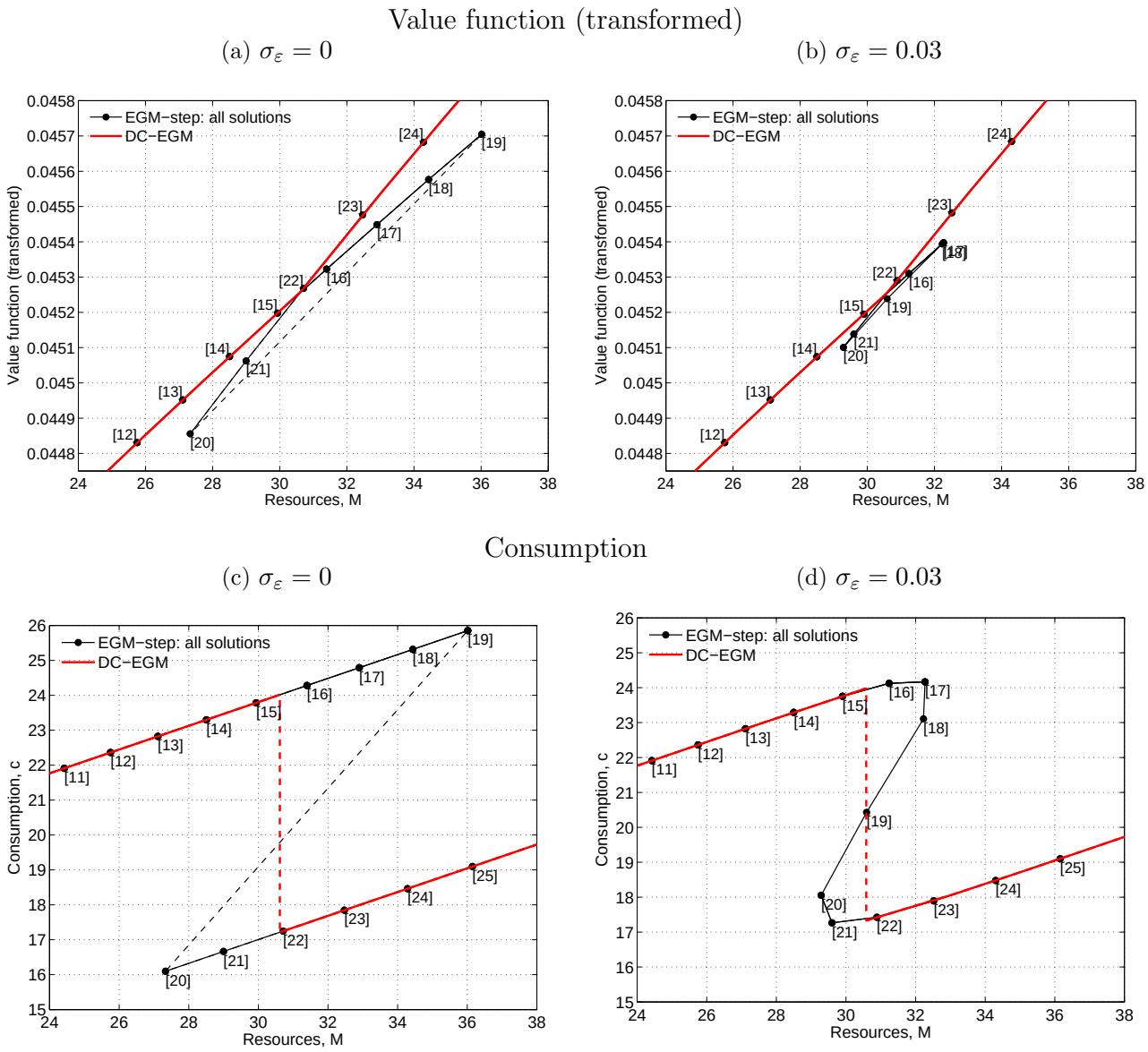
$$M_t(\vec{A}) = c_t(\vec{A}|d_t) + \vec{A}. \tag{15}$$

Further, evaluating the maximand of the equation (10) at the points $c_t(\vec{A}|d_t)$, we compute the choice specific value function $v_t\big(M_t(\vec{A})|d_t\big)$. Algorithm 1 provides a pseudo-code of the described part of the DC-EGM which we call *the EGM step*. The current period discrete choice, d_t, and the next period policy and value functions are inputs to this routine, while the endogenous grid $\vec{M}_t = M_t(\vec{A}|d_t)$ and the d_t-specific consumption and value functions, $c_t(\vec{M}_t|d_t) = c_t(\vec{A}_t|d_t)$ and $v_t\big(\vec{M}_t|d_t\big) = v_t\big(M_t(\vec{A})|d_t\big)$ computed on this grid are the outputs.

Figure 4 plots a selection of values of $v_t\big(\vec{M}_t|d_t\big)$ and $c_t(\vec{M}_t|d_t)$ against the endogenous grid $\vec{M}_t$. The points are indexed in the ascending order of the end-of-period wealth forming the grid $\vec{A}$. The solid lines approximate the corresponding functions with linear interpolation. It is evident that the interpolated discrete choice specific value function $v_t\big(M|d_t\big)$ is a *correspondence rather than a function* of M because of the existence of the region where multiple values of $v_t\big(M|d_t\big)$ correspond to a single value of M. The same is true for the interpolated discrete choice specific consumption function. The right and the left panels of Figure 4 illustrate the setting with and without the taste shocks respectively. Adding taste shocks with a relatively low variance, $\sigma_\varepsilon = 0.03$, reduces the size of the regions with multiple corresponding values. Dashed lines illustrate discontinuities.

The region where multiple values of $v_t\big(M|d_t\big)$ correspond to a single value of M is the clear

Figure 4: Non-concave regions and the elimination of the secondary kinks in DC-EGM.



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

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

1: **Inputs:** Current decision d_t. Choice-specific consumption and value functions $c_{t+1}(\vec{M}_{t+1}|d_{t+1})$ and $v_{t+1}(\vec{M}_{t+1}|d_{t+1})$ associated with the endogenous grid in period $t+1$, $\vec{M}_{t+1}$
2: Let $\vec{\eta} = \{\eta^1, \ldots, \eta^Q\}$ be a vector of quadrature points with associated weights, $\vec{\omega} = \{\omega^1, \ldots, \omega^Q\}$
3: Form an ascending grid over end-of-period wealth, $\vec{A}_t = \{A_t^1, \ldots, A_t^G\}$ where $A_t^j > A_t^{j-1}$, $\forall j \in \{2, \ldots, G\}$
4: **for** $j = 1, \ldots, G$ **do** (Loop over points in $\vec{A}$)
5: **for** $q = 1, \ldots, Q$ **do** (Loop over quadrature points in $\vec{\eta}$)
6: Compute $M_{t+1}^q(A^j) = RA^j + d_t y \eta_{t+1}^q$
7: **for** $d_{t+1} = 0, 1$ **do**
8: Compute $c_{t+1}(M_{t+1}^q(A^j)|d_{t+1})$ by interpolating $c_{t+1}(\vec{M}_{t+1}|d_{t+1})$ at the point $M_{t+1}^q(A^j)$
9: Compute $v_{t+1}\big(M_{t+1}^q(A^j)|d_{t+1}\big)$ by interpolating $v_{t+1}(\vec{M}_{t+1}|d_{t+1})$ at the point $M_{t+1}^q(A^j)$
10: **end for**
11: Compute $\phi_{t+1}\big(M_{t+1}^q(A^j)\big) = \sigma_\varepsilon \log\big(\sum_{j=0,1} \exp(v_{t+1}\big(M_{t+1}^q(A^j)|d_{t+1} = j\big))/\sigma_\varepsilon\big)$
12: Compute $P_{t+1}(d_{t+1}|M_{t+1}^q(A^j)) = \exp(v_{t+1}\big(M_{t+1}^q(A^j)|d_{t+1}\big)/\sigma_\varepsilon)(\sum_{j=0,1} \exp(v_{t+1}\big(M_{t+1}^q(A^j)|d_{t+1} = j\big))/\sigma_\varepsilon)^{-1}$
13: **end for**
14: Compute $\text{RHS}\big(M_{t+1}(A^j)\big) = \beta R \sum_{q=1}^{Q} \sum_{j=1,2} \omega^q \cdot u'\big(c_{t+1}(M_{t+1}^q(A^j)|d_{t+1} = j)\big) \cdot P_{t+1}(d_{t+1} = j|M_{t+1}^q(A^j))$
15: Compute expected value function $EV_{t+1}\big(M_{t+1}(A^j)\big) = \sum_{q=1}^{Q} \omega^q \cdot \phi_{t+1}\big(M_{t+1}^q(A^j)\big)$
16: Compute current consumption $c_t(A^j|d_t) = u'^{-1}\Big(\text{RHS}\big(M_{t+1}(A^j)\big)\Big)$
17: Compute value function $v_t\big(M_t(A^j)|d_t\big) = u(c_t(A^j|d_t)) + \beta EV_{t+1}\big(A^j\big)$
18: Compute endogenous grid $M_t(A^j|d_t) = c_t(A^j|d_t) + A_t^j$
19: **end for**
20: Collect the points $M_t(A^j|d_t)$ to from the endogenous grid $\vec{M}_t = \{M_t(A^j|d_t), j = 1, \ldots, G\}$ associated with the choice-specific consumption and value functions: $c_t\big(\vec{M}_t|d_t\big) = \{c_t\big(M_t(A^j)|d_t\big), j = 1, \ldots, G\}$, and $v_t\big(\vec{M}_t|d_t\big) = \{v_t\big(M_t(A^j)|d_t\big), j = 1, \ldots, G\}$
21: **Outputs:** $\vec{M}_t$, $c_t(\vec{M}_t|d_t)$ and $v_t\big(\vec{M}_t|d_t\big)$

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

evidence of non-concavity of the value function in the following period, and subsequent multiplicity of solutions of the Euler equation. The EGM step approximates *all solutions* to the Euler equation (see Lemma 2 in Appendix A), but because some of these solutions do not correspond to the optimal choices, it produces a *value function correspondence* which has to be cleaned of the suboptimal points to obtain actual value function. We should emphasize, however, that the points produced by the EGM step necessarily contain the true solutions. This is a notable contrast to the standard solution methods based on an exogenous grid over wealth, which may struggle to find the points of optimality and have to deploy computationally costly global search methods to solve the optimization problem in the Bellman equation.

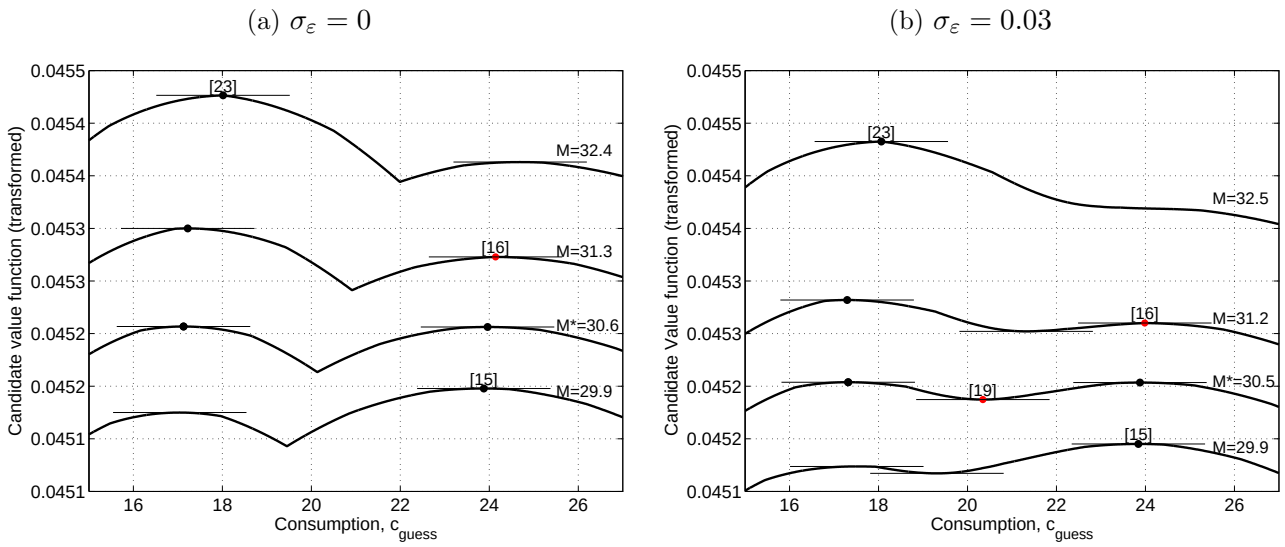
The next section describes a procedure in DC-EGM algorithm that deals with selecting the true optimal points among the points produced by the EGM step. The true solution found by the full DC-EGM algorithm is illustrated in Figure 4 with a red line for reference.

3.2 Calculation of the Upper Envelope

To distinguish between the optimal and suboptimal points produced by the EGM step, the DC-EGM algorithm makes a direct comparison of the values associated with each of the choices. On the plots of the discrete choice specific value function correspondences (panels a and b) in Figure 4, this amounts to computing the *upper envelope* of the correspondence in the regions of M_t where multiple solutions are found.

To provide deeper insight into this process, we plot the maximand of the equation (10) that defines the discrete choice specific value function $v_t(M_t|d_t)$ in Figure 5 as a function of consumption c_{guess} for various values of M_t. The value of $v_t(M_t|d_t)$ is the *global* maximum of the this function. The EGM step (Algorithm 1), however, recovers all critical points where the derivative of the plotted function is zero.[10] The same points in Figures 5 and 4 are indexed with the same indexes for easy comparison.

Figure 5: Local maxima and multiple solutions of the Euler equation.



Notes: The figure plots the maximand of the equation (10), which defines the discrete choice specific value function $v_t(M_t|d_t = 1)$, for the case of $\sigma_\varepsilon = 0$ (panel a) and $\sigma_\varepsilon = 0.03$ (panel b). Horizontal lines indicate the critical points found or approximated by the EGM step of DC-EGM algorithm. The points are indexed with the same indexes as in Figure 4 and the black dots represent global maxima. Model parameters are identical to those of Figure 4.

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

In the case without taste shocks, $\sigma_\varepsilon = 0$ (panel a), two levels of consumption satisfy the Euler equation (12) in the range $M_t \in [27, 36]$. From Figure 4 we know that points indexed 16 to 21 are suboptimal. Panel (a) in Figure 5 illustrates that the maximand function computed for wealth M_t in this range has two local maxima. For example, the 15th point from the EGM step is the global maximum of the maximand computed at $M_t \approx 29.9$, while the 16th point is not the global maximum when resources are $M_t \approx 31.3$.

At some point, the two solutions originating from the two segments of the value function correspondence are *both* optimal. Around $M_t \approx 30.6$ in panel (a) of Figure 5, the decision maker is indifferent between the discrete choices (at the next or some future periods – depending on whether the multiplicity of the solutions was caused by the primary or secondary kink of the next period value function). At this point of indifference, the consumption function is discontinuous, as illustrated with the red dashed line in panel (c) in Figure 4. The intersection point is not necessarily found in the EGM-step outlined above and needs to be additionally computed.[11]

In the smooth case with $\sigma_\varepsilon = 0.03$ the problem of multiplicity of local maxima in the maximand of equation (10) is still present, as shown by panel (b) of Figure 5. Correspondingly, there is still a discontinues drop in consumption around M_t around $M_t \in [29, 31]$. Note that in the smooth case there can be three solutions to the Euler equation, only one of which is a global maximum. This configuration is dealt with by the same upper envelope method.

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

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

Algorithm 2 Upper envelope refinement step

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

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

functions.

Algorithm 3 presents the pseudo-code of the full DC-EGM algorithm. The algorithm invokes the EGM step repeatedly to compute the value function correspondences for all discrete choice, and then finds and removes all suboptimal points on the returned endogenous grids.

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 based on the DC-EGM algorithm performs well and is able to identify the structural parameters of the model.

Algorithm 3 The DC-EGM algorithm

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

3.3 Credit Constraints

Before turning to the Monte Carlo results, we briefly discuss how DC-EGM handles the credit constraints, $c_t \leq M_t$.

During the EGM step, the credit constraints are dealt with in exactly same manner as in the standard EGM by Carroll (2006). Let the smallest possible end-of-period resources $A^1 = 0$ be the first point in the grid $\vec{A}$. Assuming that the corresponding point of the endogenous grid $M_t(A^1|d_t)$ is positive[12], it holds that $A_t(M|d_t) = 0$ for all $M \leq M_t(A^1|d_t)$ due to the monotonicity of saving function $A_t(M|d_t) = M - c_t(M|d_t)$ (see Theorem 2 in Appendix A). Therefore, the optimal consumption in this region is then given by $c_t(M|d_t) = M$, and the choice-specific value function is

$$v_t(M|d_t) = \log(M) - d_t + \beta \int EV_{t+1}(d_t y \eta_{t+1}) f(d\eta_{t+1}), \quad M \leq M_t(A^1|d_t). \tag{16}$$

Note that the third component of (16) is the expected value of having zero savings. It is calculated within the EGM step for the point $A^1 = 0$, and should be saved separately as a constant that depends on d_t but not on M_t. Once this constant is computed, $v_t(M|d_t)$ essentially has analytical form in the interval $[0, M_t(A^1|d_t)]$, and thus can be directly evaluated at any point.

When the per-period utility function is additively separable in consumption and discrete choice like in the retirement model we consider, (16) holds for all $d_t \in D_t$ in the interval $0 \leq M \leq \min_{d_t \in D_t} M_t(A^1|d_t)$. In other words, the choice specific value functions for low wealth have the same shape (in our case $\log(M)$), which is shifted vertically with d_t-specific coefficients.

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

This implies that the logistic choice probabilities $P_t(d_t|M_t)$ are constant in this interval, and have to only be calculated once.

4 Monte Carlo Results

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

4.1 Data Generation Process

For the Monte Carlo we consider slightly more general formulation of the consumption/savings and retirement problem defined in (1) with Constant Relative Risk Aversion (CRRA) utility

$$\max_{\{c_t,d_t\}_1^T} \sum_{t=1}^{T} \beta^t \left(\frac{c_t^{1-\rho} - 1}{1-\rho} - \alpha d_t \right) \tag{17}$$

where ρ is the CRRA coefficient and α ia additional parameter that index the disutility of work.

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 will refer to this solution as the *true solution* even though this is off course only an accurate finite approximation of the value function.[13]

We consider several specifications of the model in the Monte Carlo experiments below to study various aspects of the performance of the estimator. Table 1 presents the parameter values in the baseline specification of the model. Deviations are given explicitly with every Monte Carlo

[13] As a spot check, we have also compared this to 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 optimal consumption for each value of wealth. We find that results are essentially unchanged.

Table 1: Baseline *true* parameter values.

Description	Value	Description	Value
Time horizon	$T = 44$	Disutility of work	$\alpha = 0.5$
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	$\sigma_\varepsilon \in \{0.01, 0.05\}$

separately.

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

4.2 Maximum Likelihood Estimation

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

Assume that a panel dataset is available, $\{(M^d_{it}, s^d_{it}, d^d_{it}, c^d_{it})\}_{i=\{1,\ldots,N\},\, t=\{1,\ldots,T_i\}}$, containing observations on wealth, other states, discrete and continuous choices of individuals $i = 1 \ldots, N$ in time periods $t = 1, \ldots, T_i$. 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 $\theta = (\alpha, \beta, \rho, \sigma_\eta, \sigma_\varepsilon)$. We assume that

consumption is observed with additive Gaussian measurement error,

$$c_{it}^d = c_t(M_{it}^d, s_{it}^d, d_{it}^d|\theta) + \xi_{it},\ \xi_{it} \sim N(0, \sigma_\xi),\ \text{i.i.d. } \forall i, t. \tag{18}$$

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

$$\ell_{it}(\theta, \sigma_\xi) = P(d_{it}^d|M_{it}^d, s_{it}^d, \theta)\frac{\phi(\xi_{it}^d(\theta)/\sigma_\xi)}{\sigma_\xi}, \tag{19}$$

where $\phi(\cdot)$ is the density function of 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 are standard logits

$$P(d_{it}^d|M_{it}^d, s_{it}^d, \theta) = \frac{\exp(v_t(M_{it}^d, s_{it}^d, d_{it}^d|\theta)/\sigma_\varepsilon)}{\sum_{j=0}^1 \exp(v_t(M_{it}^d, s_{it}^d, j|\theta)/\sigma_\varepsilon)} \tag{20}$$

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.

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

$$\mathcal{L}(\theta) \propto \sum_{i=1}^N \sum_{t=1}^{T_i} \left\{ \frac{1}{\sigma_\varepsilon}\Big(v_t(M_{it}^d s_{it}^d, d_{it}^d|\theta) - EV_t(M_{it}^d, s_{it}^d|\theta)\Big) - \frac{1}{2}\log\left(\sum_{i=1}^N \sum_{t=1}^{T_i} \xi_{it}^d(\theta)^2\right) \right\}, \tag{21}$$

where $EV_t(M_{it}^d, s_{it}^d|\theta)$ is the the logsum given in (11) evaluated at parameter value θ. The parameter vector $\hat{\theta}$ that maximizes (21) 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 dis-utility of work, α, while keeping all other fixed at their true values.

Taste Shocks and Approximation Error. Figure 6 displays the root mean square error (RMSE) of the parameter estimates for the dis-utility of work, $\hat{\alpha}$. Results are shown for varying degree of smoothing, $\sigma_\varepsilon \in \{0.01, 0.05\}$, and different values of the disutility of work parameter, $\alpha \in \{0.1, 0.5\}$. With RMSE is around $1.0e^{-3}$, the proposed estimator is already accurate with 50 grid points and rapidly improves as the number of grid points increase from 50 through 1000. Note that standard errors will of course increase with σ_ε due to the increased amount of unexplained variation in the error term and RMSE reflects this too. Bearing this is mind, it is evident that the approximation bias decreases as the degree of smoothing increases, i.e., larger values of σ_ε. For higher levels of smoothing, problems with multiplicity of the Euler equation solutions disappear and few grid points are needed to approximate the (smooth) consumption function. This is particularly true when the dis-utility from work is large ($\alpha = .5$) because the non-concave regions are larger in this case. We also calculated the Monte Carlo Standard Deviation (MCSD)[14], which is on the order $1.0e^{-4}$ irrespectively of the number of grid points used.

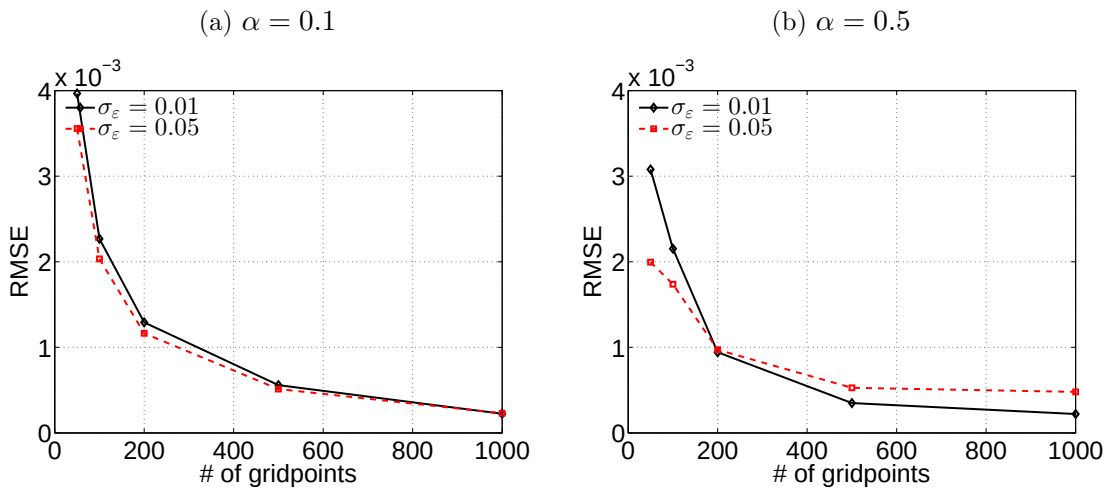
Income Uncertainty. Additional uncertainty about, e.g., future labor market income tend to smooth out *secondary* kinks stemming from multiple solutions to the Euler equations. To illustrate how that additional smoothing affects the proposed estimator, Figure 7 display RMSE when introducing *income uncertainty*. We report results from two different values of the income variance, $\sigma_\eta^2 \in \{0.001, 0.05\}$. The first level, 0.001, does not completely smooth out secondary kinks while the significantly more uncertain income process does (see the right panel of Figure 2).

Income uncertainty together with taste shocks smooth the problem to such a degree that concave regions become few and unimportant and find 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. The MCSD quite is 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

[14]MCSD results not shown

Figure 6: Monte Carlo results: disutility of work.



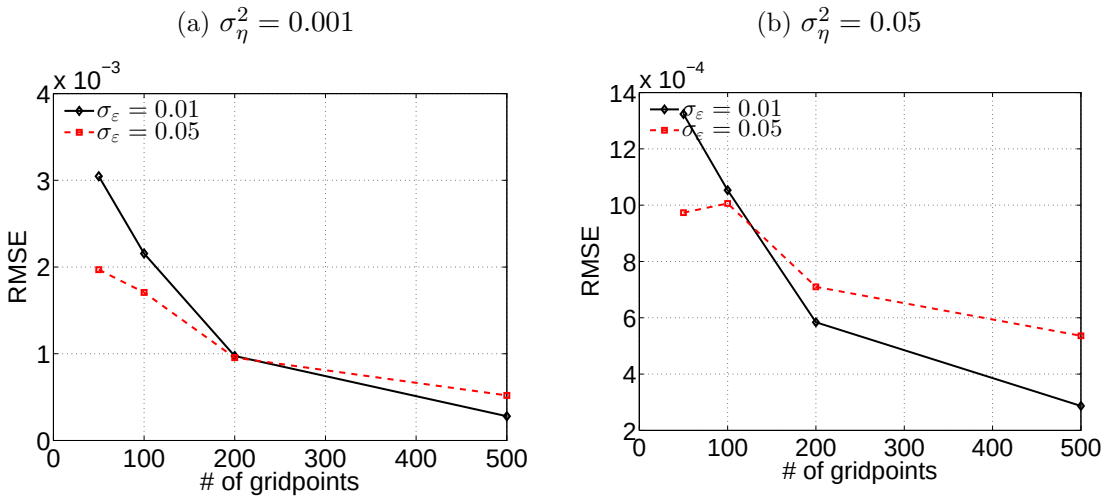
Notes: The plots illustrate the root mean square error (RMSE) of $\hat{\alpha}$. Results are shown for varying degree of smoothing, $\sigma_\varepsilon \in \{0.01, 0.05\}$, and different values of the disutility of work, $\alpha \in \{0.1, 0.5\}$. The rest of the parameters are at their baseline levels,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

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 are 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 *with* unobserved state variables, $\sigma_\varepsilon > 0$, has been estimated. To investigate how the proposed estimator performs if the data used for estimation stems from a model in which there is *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 throughout, solving the smoothed model is much faster since it requires fewer grid points and, thus, is much

Figure 7: Monte Carlo results: income uncertainty.



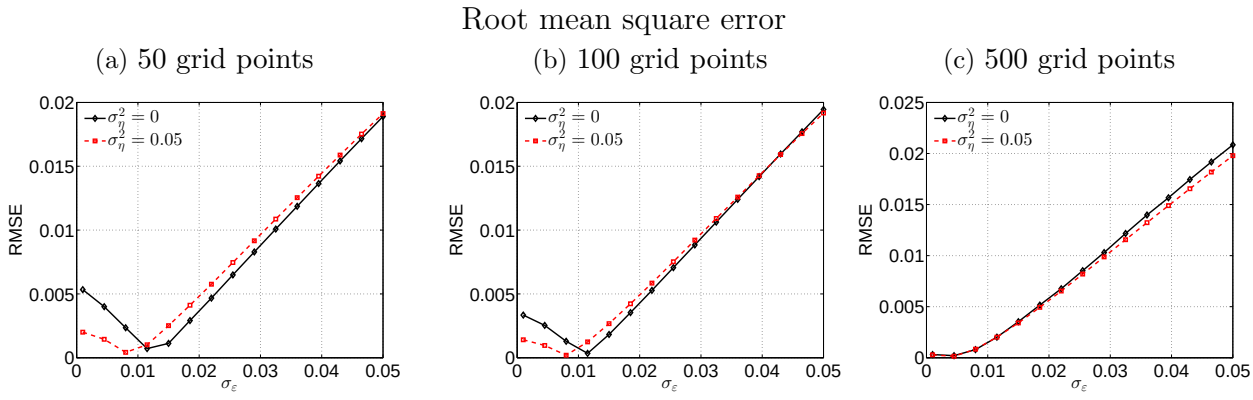
Notes: The plots illustrate the root mean square error (RMSE). Results are shown for varying degree of smoothing, $\sigma_\varepsilon \in \{0.01, 0.05\}$, and different values of the income variance, $\sigma_\eta^2 \in \{0.001, 0.05\}$. The rest of parameters are at their baseline levels, see Table 1.

faster to estimate.

Figure 8 illustrates the RMSE and MC std. 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 MC std. increases. Interestingly, for a given number of discrete grid points, the RMSE is minimized by a $\sigma_\varepsilon > 0$. While very low 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 smoothing (i.e., income shocks) are included in the model. This is natural because the income uncertainty smooths the problem and less logit smoothing is "required" to obtain the optimal smooth approximation.

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.

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 an overcomplicated deterministic solution of 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 does not require any iterative root-finding operations, and is equally efficient in dealing with borrowing constrains.

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, 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 statics for the continuous choice in the current period, and the marginal utility function is

analytically invertible.[15] The algorithms provided in the paper are easily generalized to alternative model specifications including models with large state spaces and multinomial discrete choices.

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

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 empirical 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 adequate for the real life 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 Bellman equation,

$$V_t(M_t, s_t) = \max_{0 \leq c_t \leq M_t, d_t \in D_t} \Big[u(c_t, d_t, s_t) + \sigma_\varepsilon \varepsilon_t(d_t) + \beta_t E_t \big\{ V_{t+1}(M_{t+1}, s_{t+1}) | A_t, d_t \big\} \Big], \qquad (22)$$

[15] See Appendix for a precise formulation of the class of models that are solvable by DC-EGM.

where $t = 1, \ldots, T-1$, 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 distinct options 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 (22) can be written in terms of the deterministic choice-specific value functions $v_t(M_t, s_t|d_t)$ as

$$\begin{aligned} v_t(M_t, s_t|d_t) &= \max_{0 \le c_t \le M_t} \Big[u(c_t, d_t, s_t) + \beta_t E_t \big\{ V_{t+1}(M_{t+1}, s_{t+1}) | A_t, d_t \big\} \Big] && (23) \\ &= \max_{0 \le c_t \le M_t} \Big[u(c_t, d_t, s_t) + \beta_t E_t \Big\{ \phi\big(v_{t+1}(M_{t+1}, s_{t+1}|d_{t+1}), D_{t+1}, \sigma_\varepsilon \big) | A_t, d_t \Big\} \Big], && (24) \end{aligned}$$

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 (24) is now only taken w.r.t. state transitions and idiosyncratic shocks, unlike in (22) and (23).

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

Assumption 1 (Concave utility)**.** The instantaneous utility $u(c_t, d_t, s_t)$ is concave[16] 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* (22) *takes the form*

$$u'(c_t, d_t, s_t) = \beta_t R \mathbb{E}_t \left[\sum_{d_{t+1} \in D_{t+1}} u'\big(c_{t+1}(M_{t+1}, s_{t+1}|d_{t+1}), d_{t+1}, s_{t+1}\big) P_{t+1}\big(d_{t+1}|M_{t+1}, s_{t+1}\big) \right] \quad (25)$$

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

$$P_t(d_t|M_t, s_t) = \exp\big(v_t(M_t, s_t|d_t)/\sigma_\varepsilon\big) \Big/ \sum_{d \in D_t} \exp\big(v_t(M_t, s_t|d)/\sigma_\varepsilon\big). \quad (26)$$

Proof. Discrete choice specific consumption functions $c_t(M_t, s_t|d_t)$ satisfy the the first order con-

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

ditions for the maximization problems in (23) given by

$$u'(c_t, d_t, s_t) + \beta_t E\left\{\frac{\partial V_{t+1}(M_{t+1}, s_{t+1})}{\partial M_{t+1}} \frac{\partial M_{t+1}}{\partial c_t}\right\} = 0 \tag{27}$$

for every value of $d_t \in D_t$. The envelope conditions for (23)

$$\frac{\partial v_t(M_t, s_t|d_t)}{\partial M_t} = \beta_t E\left\{\frac{\partial V_{t+1}(M_{t+1}, s_{t+1})}{\partial M_{t+1}} \frac{\partial M_{t+1}}{\partial M_t}\right\}, \tag{28}$$

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 $t = 1, \ldots, T-1$

$$u'(c_t, d_t, s_t) = \frac{\partial v_t(M_t, s_t|d_t)}{\partial M_t}. \tag{29}$$

The first order condition for (24) is

$$u'(c_t, d_t, s_t) = \beta_t R \mathbb{E}_t\left[\sum_{d_{t+1} \in D_{t+1}} \frac{\partial v_{t+1}(M_{t+1}, s_{t+1}|d_{t+1})}{\partial M_{t+1}} P_{t+1}\big(d_{t+1}|M_{t+1}, s_{t+1}\big)\right], \tag{30}$$

where choice probabilities $P_{t+1}\big(d_{t+1}|M_{t+1}, s_{t+1}\big)$ are given by (26). Plugging (29) into (30) completes the proof. □

The DC-EGM algorithm outlined in Algorithm 3 is readily applicable to the general formulation of the discrete-continuous problem (22), expect for the extra loop that has to be taken over all additional states s_t in Step 3 (Algorithm 3). The expectation over the transition probabilities of the state process is calculated together with the expectation over the other stochastic elements of the model in Algorithm 1.

Lemma 2 (All solutions)**.** *As the auxiliary grid over end-of-period wealth $\vec{A}$ becomes dense on a closed interval $[0, \bar{A}]$ for some upper bound $\bar{A}$, in the sense that the maximum distance between two adjacent points A^j and A^{j+1} approaches zero, the EGM step of DC-EGM algorithm is guaranteed to find all solutions of the Euler equation* (25) *that imply the end-of-period wealth on the interval* $[0, \bar{A}]$.

Proof. Following the Algorithm 1 denote $\textsc{rhs}\big(M_{t+1}(A^j)|d_t\big)$ the right hand side of the Euler equation (25) as a function of the points of the end-of-period wealth grid $\vec{A}$ conditional on discrete choice d_t in period t. The EGM step of the DC-EGM algorithm computes

$$\begin{cases} c_t(A^j|d_t) & = u'^{-1}\Big(\textsc{rhs}\big(M_{t+1}(A^j)\big)\Big), \\ M_t(A^j|d_t) & = u'^{-1}\Big(\textsc{rhs}\big(M_{t+1}(A^j)\big)\Big) + A^j. \end{cases} \tag{31}$$

Both equations in (31) are well defined functions of A^j provided that the utility function $u(\cdot)$ satisfies the Assumption 1. Thus, the system constitutes a well defined *parametric* specification of the curve *composed of the solutions to the Euler equation* $c(M_t, s_t|d_t)$ for all s_t, d_t, where A^j plays the role of a parameter. In the limit as A^j runs through all the values on the interval $[0, \bar{A}]$, all solutions that imply the end-of-period wealth from this interval are found. □

The criteria for selecting the solutions of the Euler equation that correspond to the optimal behavior in the model is based on the monotonicity of the savings function, which is established with the following theorem[17].

Theorem 2 (Monotinicity of savings function)**.** *Denote $A_t(M_t, s_t|d_t) = M_t - c_t(M_t, s_t|d_t)$ a discrete choice specific savings function in period t. Under the Assumption 1, function $A_t(M, s_t|d_t)$ is monotone non-decreasing in M for all t,s_t and $d_t \in D_t$.*

Proof. Theorem 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 (23) for the discrete choice specific value function $v_t(M_t, s_t|d_t)$. As a function of M and A, the maximand in this expression is given by

$$f(A, M) = u(M - A, d_t, s_t) + \beta_t E_t\big\{V_{t+1}(M_{t+1}(A), s_{t+1})\big\} \tag{32}$$

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

$$\begin{gathered} f(A', M') - f(A'', M') = \\ = u\left(M' - A', d_t, s_t\right) - u\left(M' - A'', d_t, s_t\right) + \\ +\beta_t \left[EV_{t+1}\left(M_{t+1}(A'), s_{t+1}\right) - EV_{t+1}\left(M_{t+1}(A''), s_{t+1}\right)\right] \geq \\ \geq u\left(M'' - A', d_t, s_t\right) - u\left(M'' - A'', d_t, s_t\right) + \\ +\beta_t \left(EV_{t+1}\left(M_{t+1}(A'), s_{t+1}\right) - EV_{t+1}\left(M_{t+1}(A''), s_{t+1}\right)\right) = \\ f(A', M'') - f(A'', M'') > 0. \end{gathered} \tag{33}$$

For the first inequality we use

$$\begin{gathered} u\left(M' - A', d_t, s_t\right) - u\left(M' - A'', d_t, s_t\right) \geq u\left(M'' - A', d_t, s_t\right) - u\left(M'' - A'', d_t, s_t\right), \\ u\left(M' - A', d_t, s_t\right) - u\left(M'' - A', d_t, s_t\right) \geq u\left(M' - A'', d_t, s_t\right) - u\left(M'' - A'', d_t, s_t\right), \\ u\left(z, d_t, s_t\right) - u\left(z - \Delta_M, d_t, s_t\right) \geq u\left(z + \Delta_A, d_t, s_t\right) - u\left(z + \Delta_A - \Delta_M, d_t, s_t\right), \end{gathered} \tag{34}$$

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. □

References

Adda, J., C. Dustmann and K. Stevens (2015): "The Career Costs of Children," Working paper.

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

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

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.

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," Discussion paper.

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

Druedahl, J. (2015): "Business Cycle Fluctuations in the Demand for Consumer Durables," Working paper, University of Copenhagen.

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

Ejrnæs, M. and T. H. Jørgensen (2015): "Saving Behavior Around Planned and Unplanned 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.

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

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

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 (2015): "An Analysis of the Australian Social Security System using a Life-Cycle Model of Labor Supply with Asset Accumulation and Human Capital," .

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

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

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.

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

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

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

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.

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