Discussion Papers
Department of Economics
University of Copenhagen

No. 16-09

A General Endogenous Grid Method for Multi-Dimensional Models with Non-Convexities and Constraints

Jeppe Druedahl

Thomas H. Jørgensen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

A General Endogenous Grid Method for Multi-Dimensional Models with Non-Convexities and Constraints*

Jeppe Druedahl†

Thomas H. Jørgensen‡

September 1, 2016

Abstract

The endogenous grid method (EGM) significantly speeds up the solution of stochastic dynamic programming problems by simplifying or completely eliminating root-finding. We propose a general and parsimonious EGM extended to handle 1) multiple continuous states and choices, 2) multiple occasionally binding constraints, and 3) non-convexities such as discrete choices. Our method enjoys the speed gains of the original one-dimensional EGM, while avoiding expensive interpolation on multi-dimensional irregular endogenous grids. We explicitly define a broad class of models for which our solution method is applicable, and illustrate its speed and accuracy using a consumption-saving model with both liquid assets and illiquid pension assets and a discrete retirement choice. (JEL: C13, C63, D91)

Keywords: Endogenous grid method, post-decision states, stochastic dynamic programming, continuous and discrete choices, occasionally binding constraints.

*We thank Anders Munk-Nielsen, Fedor Iskhakov, John Rust, Bertel Schjerning, Matthew White, and seminar participants at the Center for Computational Economics at UCHP, for fruitful discussions. We thank the managing editor and three anonymous referees for their helpful comments. Financial support from the Danish Council for Independent Research in Social Sciences is gratefully acknowledged (FSE, grant no. 4091-00040 and 5052-00086B).

†Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark. Centre for Computational Economics (CCE). E-mail: jeppe.druedahl@econ.ku.dk. Website: http://econ.ku.dk/druedahl.

‡Department of Economics, University of Copenhagen. Centre for Computational Economics (CCE). E-mail: thomas.h.jorgensen@econ.ku.dk. Webpage: www.tjeconomics.com.

1 Introduction

The real-world decision problems households and firms face are often not as well-behaved as assumed in economic models. Important interactions between different choices imply that they are best studied collectively. Accounting for occasionally binding constraints and the discrete nature of many choices, such multi-dimensional dynamic models can be very hard to solve. This has forced, and continue to force, researchers on a quest for faster solution methods in order to be able to solve and analyze behavior from increasingly more complex dynamic economic models.

We contribute to this literature by proposing a parsimonious solution method building on the *endogenous grid point method* (EGM) of Carroll (2006). Our method significantly reduces computational time compared to standard methods for a broad class of stochastic dynamic programming models with multiple continuous states and choices, multiple occasionally binding constraints, and non-convexities such as discrete choices.

The central challenge when using the EGM to solve models with non-convexities is to determine which solutions to the FOCs are globally optimal. This problem arises as non-convexities typically imply that the FOCs are only *necessary*, but not *sufficient*.[1] We are the first to provide an *upper envelope algorithm* for multi-dimensional models solving this task. The previous upper envelope algorithms in Fella (2014) and Iskhakov, Jørgensen, Rust and Schjerning (2015) for one-dimensional models rely on monotonicity assumptions, which have no counterpart in multi-dimensional models.

We are furthermore the first to show how to easily transform the irregular state grids implied by the multi-dimensional EGM into regular, e.g. rectilinear, grids. Interpolation on irregular grids have previously been shown to be the key bottleneck for the performance of multi-dimensional EGM (without non-convexities). Ludwig and Schön (2014) show that using Delaunay-triangulations and so-called visibility walks can result in the EGM being slower than time iterations in a two-dimensional model without non-convexities. White (2015) shows how assuming a specific form of monotonicity can be used to avoid the triangulation step, but his method still requires expensive visibility walks, and cannot handle non-convexities.

Finally, we are the first to show how multiple occasionally binding constraints can be handled parsimoniously within an EGM without any prior knowledge on where in the state space which constraints are binding. Previously, Hintermaier and Koeniger (2010) have shown how to handle two occasionally binding constraints, but only in a hybrid form of EGM with a time iteration step and utilizing specific properties of their model's Kuhn-

[1] Proving that the value function is differentiable at optimal interior choices, and that the FOCs are thus *necessary* at all, is in itself a non-trivial task for models with non-convexities; recent theoretical results from Clausen and Strub (2013), however, turn out to be very helpful in this regard.

Tucker multipliers.[2] The general framework for a multi-dimensional EGM presented in White (2015) does not allow for constraints at all. As our solution method generalizes the various previous generalizations of EGM we, for brevity, refer to it as the G^{2}EGM.[3]

To illustrate the potential of our method, we solve an illustrative model of consumption/saving with both liquid assets and illiquid pension assets and a discrete retirement choice. For a given level of precision measured in terms of the value function, our G^{2}EGM is about *20 times faster* than a highly optimized implementation of the standard work horse method of value function iteration (VFI). The main benefit of our method compared to VFI is that we avoid searching for the optimal choices using an expensive constrained multi-dimensional *global* search algorithm. The speed-up of our G^{2}EGM is even larger if we alternatively measure precision in terms of the average errors in the Euler-equation for consumption. Further extending our illustrative model with a labor supply decision on the intensive margin and human capital accumulation (i.e a model with three continuous states and choices), we show that the speed-gain is roughly the same. As our proposed solution method can be applied in many fields of economics, it consequently makes it possible to estimate richer life cycle models than previously using full-solution estimators, and thus perform policy analysis based on more realistic models. For example, models that would take almost three weeks to estimate using VFI can be estimated within a day using our method.

We also explicitly define a broad model class in terms of *necessary* and *sufficient* conditions on model fundamentals where our method is applicable. Hereby researchers can check whether a particular model of interest is solvable using our method.[4] In broad terms the three central conditions of our model class are: *i)* there must exist a "low" dimensional vector of post-decision state variables which collectively is a *sufficient* statistic for the continuous states and choices[5], *ii)* the FOCs must be at least *necessary*, *iii)* the model must imply *injectivity* in a specific sense. Examples of model types which could be solved using our method are models of consumption and human capital accumulation (Imai and Keane, 2004), models of retirement and health with a consumption floor (French and Jones, 2011), models of life insurance and consumption over the life cycle

[2] Barillas and Fernández-Villaverde (2007) also develop a hybrid form of EGM with a value function iteration step.

[3] In a broader context our paper is also related to the growing literature on solving high-dimensional dynamic economic models. See Maliar and Maliar (2014) for a review, and Den Haan, Judd and Juillard (2011) for a comparison of competing methods. The use of post-decision states, which is central for the EGM, is also widespread in the engineering literature (see Powell (2011) and Bertsekas (2012)), but to the best of our knowledge they focus exclusively on some form of value function iteration. Hull (2015) discusses how these insights can be used when solving dynamic economic models.

[4] MATLAB and C++ code for the illustrative models are also available from the authors web pages.

[5] In our two-dimensional illustrative model, for example, the post-decision (end-of-period) levels of liquid assets and illiquid pension assets contain all relevant information for forecasting future (beginning-of-period) asset levels.

(Hong and Ríos-Rull, 2012), models of liquid and illiquid assets with fixed (i.e. non-convex) transaction costs and Epstein-Zin-Wiel preferences (Kaplan and Violante, 2014; Berger and Vavra, 2015), and models of human capital accumulation, savings and career and fertility choices (Adda, Dustmann and Stevens, forthcoming).

The paper proceeds in two more or less self-contained parts: The intention is that readers only interested in the overall idea can stop reading after section 4. Specifically, in the following section 2, we describe an illustrative model of retirement and saving in both liquid assets and illiquid pension assets. In section 3 we describe how the illustrative model can solved using our G^{2}EGM. In section 4 we report speed and accuracy comparisons with VFI. From section 5 we instead focus on the general case and introduce a broad model class in terms of necessary and sufficient conditions where our G^{2}EGM is applicable. In section 6 we describe in detail how our G^{2}EGM can be used to solve all models in this class. Section 7 concludes the paper with final remarks.

2 Illustrative Model

In this section, we formulate a consumption-saving model with both liquid and illiquid assets and an absorbing discrete retirement choice, which we later use to illustrate our proposed solution method. We assume that the saved liquid assets can always be accessed while the saved illiquid pension assets can only be accessed at retirement. The problem when retired and working thus differ.

Retired households. In retirement, households solve a standard consumption-savings problem. The resources available for consumption in period t is denoted m_t, such that post-decision (or end-of-period) assets a_t after consumption c_t is given by

$$a_t = m_t - c_t$$

Next period resources are given by

$$m_{t+1} \quad = \quad R_a a_t + y_{t+1}$$

where R_a is the gross rate of return, and $y_{t+1} = \underline{y}$ is a (deterministic) retirement income. We assume that households are not allowed to borrow, $a_t \geq 0$, and that consumption is restricted to be positive.

Working households. Working households solve a more general problem, and are allowed to save in both liquid assets (a_t) and illiquid pension assets (b_t). Denoting the pension fund deposits by d_t, we assume that the post-decision (or end-of-period) assets levels are given by

$$\begin{aligned} a_t &= m_t - c_t - d_t \\ b_t &= n_t + d_t + g(d_t) \end{aligned}$$

where $g(d_t)$ is a pension deposit function potentially allowing for an extra incentive to accumulate illiquid pension funds. Deposits are required to be non-negative, i.e. $d_t \geq 0$, and the assumption of no borrowing, $a_t \geq 0$, is also maintained for the working households.

The resources available for consumption and pension savings in the next period are

$$\begin{aligned} m_{t+1} &= R_a a_t + \eta_{t+1},\ \log \eta_{t+1} \sim \mathcal{N}(-.5\sigma_\eta^2, \sigma_\eta^2) \\ n_{t+1} &= R_b b_t \end{aligned}$$

where η_t is stochastic labor income, and we assume a higher return on pension assets than on liquid assets, i.e. $R_b \geq R_a$.

2.1 Recursive Formulation

Denoting the discrete choice of retirement by $z_t = 0$ and the discrete choice of working by $z_t = 1$, the Bellman equation of the illustrative model can be formulated as

$$\begin{aligned} V_t(z_{t-1}, m_t, n_t, \varepsilon_t) &= \max_{z_t \in \mathcal{Z}_t(z_{t-1})} v_t(z_t, m_t, n_t) + \sigma_\varepsilon \varepsilon(z_t) \qquad (2.1) \\ &\text{s.t.} \\ \mathcal{Z}_t(z_{t-1}) &= \begin{cases} \{0,1\} & \text{if } z_{t-1} = 1 \\ 0 & \text{if } z_{t-1} = 0 \end{cases} \end{aligned}$$

where $\varepsilon(z_t)$ is an *iid* extreme value type I taste shock across the discrete choices and σ_ε^2 is proportional to the variance of the taste shocks. Using the distributional assumption on the taste shocks, we derive a closed form expression for the expected value just before the realization of the taste shocks as

$$EV_t(z_{t-1}, m_t, n_t) \equiv \int_\varepsilon V_t(z_{t-1}, m_t, n_t, \varepsilon_t) H(d\varepsilon) = \sigma_\varepsilon \log\left(\sum_{z_t \in \mathcal{Z}_t(z_{t-1})} \exp(v_t(z_t, m_t, n_t)/\sigma_\varepsilon)\right) \qquad (2.2)$$

The discrete-choice-specific value function for the retiring (or retired) households is

$$\begin{aligned} v_t(0, m_t, n_t) &= \max_{c_t} u(c_t, 0) + \beta V_{t+1}(0, m_{t+1}, 0, 0) \\ &\text{s.t.} \\ a_t &= (m_t + n_t) - c_t \\ m_{t+1} &= R_a a_t + \underline{y} \\ c_t &\in [0, m_t + n_t] \end{aligned} \tag{2.3}$$

where we let the available resources for consumption (since that is the only choice if the consumer retires) be $m_t + n_t$, and $u(c_t, z_t)$ denotes per-period utility flow from consuming c_t in labor market state z_t.

The discrete-choice-specific value function for the working households is

$$\begin{aligned} v_t(1, m_t, n_t) &= \max_{c_t, d_t} u(c_t, 1) + \beta \underbrace{\int_\eta EV_{t+1}(1, m_{t+1}, n_{t+1}) G(d\eta)}_{\equiv w_t(a_t, b_t)} \\ &\text{s.t.} \\ a_t &= m_t - c_t - d_t \\ b_t &= n_t + d_t + g(d_t) \\ m_{t+1} &= R_a a_t + \eta_{t+1} \\ n_{t+1} &= R_b b_t \\ c_t &\geq 0 \\ d_t &\geq 0 \\ c_t + d_t &\in [0, m_t] \end{aligned} \tag{2.4}$$

where $G(\eta)$ is the probability distribution of income shocks and we denote the continuation value (defined on post-decision states) as $w_t(a_t, b_t)$. The timing is such that next-period income, η_{t+1}, is realized *after* the discrete labor market choice, z_t, has been made. This is purely for simplicity because it removes the need to include the current income realization as a state variable in the model.

We assume the following functional forms

$$u(c_t, z_t) = \frac{c_t^{1-\rho}}{1-\rho} - \alpha \mathbf{1}\{z_t = 1\} \tag{2.5}$$

$$g(d_t) = \chi \log(1 + d_t) \tag{2.6}$$

where we have chosen $g(d_t)$ such that it is increasing and concave in d_t to mimic something like a tax-deduction from pension deposits which is gradually decreasing in the level of deposits.

In our baseline parametrization we mute all stochastic elements. Particularly, we let $\sigma_\eta = \sigma_\varepsilon = 0$ such that neither income nor taste shocks are present in the baseline model.[6] We focus on the deterministic case because it is harder to solve accurately.[7] We also report results when including stochastic elements in the model.

The optimal choice functions are denoted by $c_t^\star(m_t, n_t)$ and $d_t^\star(m_t, n_t)$, and can be found using standard *value function iteration* (VFI). This entails specifying an *exogenous* grid over the states (m_t and n_t), and globally searching for the optimal choices (c_t and d_t) calculating the value-of-choice for each guess by evaluating the utility function and computing the continuation value using numerical integration over the interpolated next period value function. The repeated global searches and numerical integration are, however, time consuming, and in the next section, we therefore describe a generalized endogenous grid method, which can be used to solve the model much faster.

3 Solving the Illustrative Model Using the G^{2}EGM

In this section, we show how the illustrative model can be solved using our G^{2}EGM. For completeness, we first show how the problem for the retired households can be solved using the original EGM developed in Carroll (2006), and discuss the challenges inherent in extending the EGM to multi-dimensional models with potential non-convexities. For pedagogical reasons, we secondly explain how the problem for the retired households can alternatively be solved using our method even if we “forget” that the Euler-equation is sufficient, and our knowledge of where in the state space the borrowing constraint is binding. Finally, we show how the more complex problem of the working household can also be solved by our method, and report speed and accuracy results.

3.1 Solving the problem for retired households with the EGM

The fundamental idea in the EGM is to specify an *exogenous* grid, $\mathcal{G}_a$, over the post-decision state (end-of-period assets, a_t) instead of over the pre-decision state (resources, m_t). Let $\mathcal{G}_a$ by strictly increasing and indexed by $i \in \{1, \dots, \#_a\}$. We can then construct *nodes* containing *endogenous* resource grid points, m_t^i, with associated consumption choices, c_t^i, by inverting the Euler-equation and the budget constraint, i.e.

$$
\begin{aligned}
c_t^i &= (\beta R_a \mathbb{E}_t[c_{t+1}^\star(R a_t^i + y_{t+1})^{-\rho}]^{-\frac{1}{\rho}} && (3.1)\\
m_t^i &= a_t^i + c_t^i && (3.2)
\end{aligned}
$$

[6] In the case when $\sigma_\varepsilon = 0$, the expected value function collapses to $EV_t(z_{t-1}, m_t, n_t) = \max_{z_t \in \mathcal{Z}_t(z_{t-1})}\{v_t(z_t, m_t, n_t)\}$.

[7] See for example the discussion in Iskhakov, Jørgensen, Rust and Schjerning (2015).

where $c^{\star}_{t+1}()$ is the next-period optimal consumption function. As the Euler-equation is both *necessary* and *sufficient*, each found c^i_t must be the *unique* optimal choice at m^i_t, i.e. $c^{\star}_t(m^i_t) = c^i_t$. Additionally, it can be shown that the borrowing constraint is binding, $c^{\star}_t(m_t) = m_t$, for all m_t lower than the limit of m^i_t when $a^i_t \downarrow 0$.

The central benefit of the EGM thus is that the optimal consumption function (and therefore also the value function) can be found without any *root-finding* (as required in time iterations) or any *numerical optimization* (as required in value function iterations). This in particular implies that the expectation over next-period variables is only taken once for each grid point.

3.2 Three challenges for generalizing the EGM

In order to develop a general EGM for multi-dimensional models with non-convexities, we need to handle the following three challenges:

1) Irregular endogenous grids. Firstly, the non-linearity of the Euler-equation implies that the *endogenous* resource grid points, m^i_t, are *un*evenly spaced even if the *exogenous* end-of-period asset grid points, a^i_t, are evenly spaced. In the one-dimensional case this is not a problem because the neighboring points of a point to be interpolated can still be located efficiently using e.g. *bisection search*. In the multi-dimensional case, however, there are no simple algorithm for finding the neighboring points in a fully irregular grid. Ludwig and Schön (2014) therefore suggest to use a *Delaunay-triangulation* to divide a two dimensional irregular grid into triangles (or into simplexes in higher dimensions). A so-called *visibility walk* can then be used to find the triangle containing the point to be interpolated. Hereafter standard *barycentric interpolation* can be applied.[8] Even using highly optimized algorithms for both the triangulation and the visibility walks, these procedures are, however, time consuming, and thus a major computational burden for multi-dimensional EGM.[9]

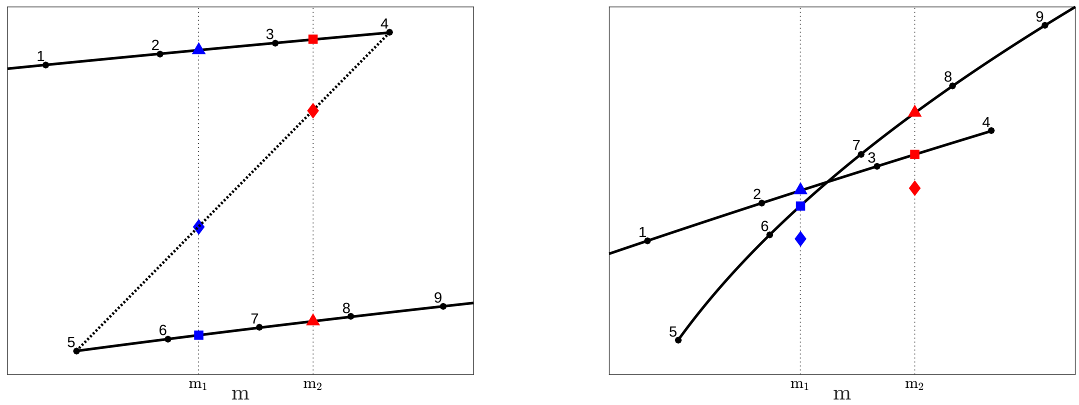
2) Non-sufficient FOCs. Secondly, in models with *non-convexities* the FOCs (and thus the Euler-equations) might not be *sufficient*, but only *necessary*. Non-sufficiency of the consumption Euler-equation can, for example, arise if an infinitesimal increase in savings today ($a_t \uparrow$) implies a change in a future discrete choice. If such a change implies a *down*ward jump in the next-period optimal consumption function then a small increase

[8] See Brumm and Grill (2014) for another application of Delaunay-triangulation and barycentric interpolation in economics.

[9] In the case without constraints and non-convexities, White (2015) shows how assuming a specific form of monotonicity can be used to construct a faster interpolation method avoiding a triangulation-type operation, but still requiring visibility walks. His procedure does not, however, extend to the general case.

Figure 3.1: Illustration: Interpolation of all solutions, 1 dimension.

(a) Consumption. (b) Value-of-choice.



Notes: Figure 3.1 plot nodes generated from nine increasing levels of post-decision resource levels a_t. Panel (a) plots pre-decision resource levels, m_t, and consumption levels, c_t. Panel (b) plots pre-decision resource levels, m_t, and values-of-choice. The vertical lines at m_1 and m_2 are resource levels, where we need to interpolate optimal consumption. The values indicated by the triangles are chosen in G^{2}EGM.

in a_t^i will lead to a *down*ward jump in c_t^i (see equation (3.1)) and an *up*ward jump in m_t^i (see equation (3.2)). An example of this is shown in panel (a) of figure 3.1, where nodes of resource levels, m_t^i, and consumption choices, c_t^i, are shown for nine increasing end-of-period asset grids points, a_t^i, and where there is a jump between the fourth and fifth node.

The challenge now becomes to construct an *upper envelope algorithm* to determine which of the solutions to the Euler-equation to respectively *keep* and *disregard*. If we incidentally had $m_t^{i_0} = m_t^{i_1}$, but $c_t^{i_0} \neq c_t^{i_0}$, then the true globally optimal consumption choice could be found by calculating the value-of-choice as

$$v_t^i = \frac{(c_t^i)^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t[v_{t+1}(Ra_t^i + y_{t+1})] \tag{3.3}$$

and determining whether $v_t^{i_0} > v_t^{i_1}$ or vice versa. In general, however, the values-of-choice are only found for *different* m_t^i as shown in panel (b) of figure 3.1.

The upper envelope algorithms presented in Fella (2014) and Iskhakov, Jørgensen, Rust and Schjerning (2015) for one-dimensional models rely on monotonicity assumptions, which have no counterpart in multi-dimensional models.[10]

[10]For a given node i_0, an intuitive upper envelope algorithm would be to loop through all the other nodes, and discard the node i_1 if there exists another node $i_1 \neq i_0$ which have choices, if feasible and made at node i_0, would imply a higher value-of-choice. This is, however, a rather costly operation because it in principle implies that we for any given node need to loop through *all the other nodes* checking this condition.

3) No prior knowledge on where the constraints are binding. Thirdly, in multi-dimensional models with *multiple constraints*, we might not have any prior knowledge on where in the state space which constraints are binding. The method presented in Hintermaier and Koeniger (2010) for handling multiple constraints utilize specific properties of their model's Kuhn-Tucker multipliers, and is a hybrid EGM with a (slow) time iteration step. The general framework for a multi-dimensional EGM presented in White (2015) does consider constraints at all.

The central benefit of our proposed solution method, solving the three challenges above, is that we are able to fully reap the benefits of the EGM in terms of limiting the use of numerical integration and avoiding global constrained searches without introducing new computationally expensive tasks such as long upper envelope loops, Delaunay-triangulations or visibility walks. This can be illustrated by solving the problem for the retired household with our G^{2}EGM "forgetting" that we know that the Euler-equation is sufficient, and where in the state space the borrowing constraint is binding.

3.3 Solving the problem for retired households with the G^{2}EGM

The first step in our G^{2}EGM is to divide the considered problem into so-called *segments*, such that it in each segment is given, which choices are constrained, and which choices are unconstrained. For the retired household problem there thus is an *unconstrained* (**ucon**) segment with $a_t > 0$, and a *constrained* (**con**) segment with $a_t = 0$.

In the *unconstrained* segment (**ucon**), we still specify an *exogenous* grid, $\mathcal{G}_a$, over the post-decision state a_t^i, and *endogenously* find resource grid points, m_t^i, with consumption choices, c_t^i, using equation (3.1) and (3.2). Let $\mathcal{G}_a$ be strictly increasing and indexed by $i \in \{1, \ldots, \#_a\}$.

In the *constrained* segment (**con**), we instead specify an *exogenous* grid, $\mathcal{G}_c$, over the constrained choice, c_t^i, and *endogenously* find resource grids points, m_t^i, using the borrowing constraint, $a_t^i = 0 \leftrightarrow m_t^i = c_t^i$. Let $\mathcal{G}_c$ be strictly increasing and indexed by $i \in \{1, \ldots, \#_c\}$.

The fundamental new idea in our solution method is to also specify an exogenous *common* resource grid, $\mathcal{G}_m$, over m_t. Let $\mathcal{G}_m$ be strictly increasing and indexed by $j \in \{1, \ldots, \#_m\}$. Interpolating the consumption choices from each of the two endogenous segment-specific irregular grids, $\mathcal{G}_a$ and $\mathcal{G}_c$, to the common regular grid, $\mathcal{G}_m$, will immediately solve the first challenge discussed above.

Let us first see how interpolation to the common grid can be performed efficiently for the *unconstrained* segment (**ucon**). Note that the end-of-period asset grid, $\mathcal{G}_a$, can be divided into a set of *line segments* given by $\{(a_t^1, a_t^2), (a_t^2, a_t^3), \ldots, (a_t^{\#_a - 1}, a_t^{\#_a})\}$. Using the associated resource grid points, m_t^i, we thus immediately have the same line segments translated from a-space into m-space, i.e. $\{(m_t^1, m_t^2), (m_t^2, m_t^3), \ldots, (m_t^{\#_a - 1}, m_t^{\#_a})\}$. For each *line segment* in m-space, we can easily find the resource levels, m_t^j's, in the common resource grid, $\mathcal{G}_m$, which are in between m_t^i and m_t^{i+1}. At these common resource grid points candidate optimal consumption choices can then be found by linear interpolation, i.e.

$$c_t^{ij} = c_t^i - \frac{c_t^{i+1} - c_t^i}{m_t^{i+1} - m_t^i}(m_t^j - m_t^i) \tag{3.4}$$

Looping through all the line segments, multiple candidate optimal consumption choices could be found at each m_t^j. These different candidates can subsequently be compared directly in terms of the values-of-choice they imply, i.e.

$$v_t^{ij} = \frac{(c_t^{ij})^{1-\rho}}{1-\rho} + \beta w_t(m_t^j - c_t^{ij}) \tag{3.5}$$

where we avoid taking the expectation repeatedly by using the post-decision value function $w_t()$. In sum, the interpolation to the common grid is thus simultaneously an *upper envelope,* which solves the second challenge discussed above. Specifically, it ensures that solutions to the Euler-equation, which are not optimal consumption choices, are not used. This can be seen most clearly by returning to the example in figure 3.1. If m_1 is a point in the common resource grid, $\mathcal{G}_m$, we see that our procedure constructs three candidate optimal consumption choices by interpolation between, respectively, node 2 and 3 (the triangle), node 6 and 7 (the square), and node 4 and 5 (the diamond). The implied values-of-choice is plotted in panel (b), and the candidate implying the highest value-of-choice (the triangle) is subsequently chosen as the optimal level of consumption at m_1.[11]

Next, we turn to the *constrained* segment (**con**), and loop through all the *line segments* in the grid over the constrained choice, $\mathcal{G}_c$, i.e. $\{(c_t^1, c_t^2), (c_t^2, c_t^3), \ldots, (c_t^{\#_c - 1}, a_t^{\#_c})\}$. As above, we interpolate to the common resource grid, $\mathcal{G}_m$. If these new candidate consumption choices overlap with some of the candidate consumption choices found from the *unconstrained* segment, we again only keep the consumption choice with the highest implied value of choice (see equation (3.5)). Hereby a *second* upper envelope is effectively applied also finding the over-arching maximum across the two segments. This solves the third, and final, challenge by determining where in the state space which constraints are binding.

[11] As illustrated, the optimal level of consumption will also be found from interpolation between node 2 and 3 for a bit higher resource levels than m_1, but at some point (close to the true kink) the optimal level of consumption will instead begin to be found from interpolations between node 6 and 7. And further to the right, such as at m_2, the optimal consumption level will be found by interpolation between node 7 and 8.

3.4 Solving the problem for working households with the G^2EGM

We now turn to the two-dimensional problem for the working households. This problem cannot be solved with standard EGM or any of the previous extensions in the literature without introducing steps with time or value function iterations. However, following the same steps as in the one-dimensional case above, it can be solved by our G^2EGM. Basically, the only difference is that we, instead of looping through the *line segments* of the exogenous grids, need to loop through easily defined *triangles* covering the exogenous grids.

Before stating the five steps of our solution method, we first divide the problem into four *segments.* In the first segment, both the consumption choice, c_t, and the pension deposit choice, d_t, are unconstrained. In the second segment, both the borrowing constraint, $a_t = 0$, and the deposit constraint, $d_t = 0$, bind. In the third and fourth segment, only one of the constraints binds, respectively.[12] For each of these segments, we in particular need to understand how we can construct node sets containing pre-decision states (m_t,n_t) and associated choices (c_t,d_t).

Firstly, we consider the *fully unconstrained* segment (**ucon**), where c_t and d_t are unconstrained choices. Using recent results from Clausen and Strub (2013) we show in the online supplemental material that despite the presence of the discrete retirement choice the following two first order conditions are *necessary*

$$c_t = (\beta w_{a,t}(a_t, b_t))^{-\frac{1}{\rho}} \tag{3.6}$$

$$d_t = \frac{\chi}{\left(\frac{w_{a,t}(a_t,b_t)}{w_{b,t}(a_t,b_t)} - 1\right)} - 1 \tag{3.7}$$

where $w_{x,t}()$ is the derivative of $w_t()$ wrt. x.

This implies that by fixing a_t and b_t we can find c_t using equation (3.6) and d_t using equation (3.7). Using the inverted budget constraints

$$m_t = a_t + c_t + d_t \tag{3.8}$$

$$n_t = b_t - d_t - g(d_t) \tag{3.9}$$

we can then find m_t and n_t.

Secondly, we consider the *fully constrained* segment (**con**), where $a_t = 0$ and $d_t = 0$. In this segment there is obviously no FOCs, and instead of fixing only the post-decision

[12] We can rule out cases with $a_t = m_t$ or $d_t = m_t$ due to $\lim_{c_t \to 0} u_c(c_t, z_t) = \infty$.

states, we also fix c_t and $d_t = 0$.[13] From the budget constraints we then have

$$\begin{aligned} m_t &= a_t + c_t + d_t = c_t \\ n_t &= b_t - d_t - g(d_t) = b_t \end{aligned}$$

Note, that fixing grids for constrained choices can, alternatively, be interpreted as fixing grids for the Kuhn-Tucker multipliers associated with the binding constraints.

Thirdly, we consider the segment with *only* $d_t = 0$ *constrained* (**dcon**). In this segment c_t is the sole choice variable with the same FOC as in equation (3.6). Fixing a_t, b_t and $d_t = 0$, we can therefore find c_t using this FOC, and from the budget constraints we then have

$$\begin{aligned} m_t &= a_t + c_t + d_t = a_t + c_t \\ n_t &= b_t - d_t - g(d_t) = b_t \end{aligned}$$

Fourthly, we consider the final segment with *only* $a_t = 0$ *constrained* (**acon**). In this segment d_t can be seen as the sole choice variable by expressing c_t as a function of d_t by $a_t = 0 \leftrightarrow c_t = m_t - d_t$. Substituting this into the original problem, we find that d_t must satisfy the segment-specific FOC

$$\begin{aligned} 0 &= u_c(c_t)\frac{\partial c_t}{\partial d_t} + \beta w_{b,t}(a_t, b_t)(1 + g_d(d_t)) \leftrightarrow \\ d_t &= \frac{\chi}{\left(\frac{c_t^{-\rho}}{\beta w_{b,t+1}(a_t,b_t)} - 1\right)} - 1 \end{aligned}$$

Fixing $a_t = 0$, b_t and c_t we can therefore find d_t using this FOC, and using the budget constraints we also have

$$\begin{aligned} m_t &= a_t + c_t + d_t = c_t + d_t \\ n_t &= b_t - d_t - g(d_t) \end{aligned}$$

In sum, we can thus construct node sets containing pre-decision states (m_t,n_t) and associated choices (c_t,d_t) for each of the four segments. Our G^{2}EGM can now be summarized in following five steps:

Step 1. Specify exogenous grids. *i)* construct a common regular (e.g. rectilinear) grid $\mathcal{G}_{m,n}$ over states m_t and n_t, *ii)* construct a common regular grid $\mathcal{G}_{a,b}$ over

[13] If we fixed a_t and b_t alone, we could derive n_t from $n_t = b_t - d_t - g(d_t) = b_t$, but otherwise we would only have $m_t = c_t$, and no more equations to determine c_t. Alternatively, we could also fix m_t, n_t, and $d_t = 0$, and derive c_t and b_t from the budget constraints.

post-decision states a_t and b_t and an interpolant of $w_t(a_t, b_t)$, and its derivatives, on this grid, *iii)* for each segment construct segment-specific regular grids over post-decision states a_t and b_t and the segment-specific constrained choices (c_t and d_t in *con*, d_t in *dcon*, and c_t in *acon*).

Step 2. Construct node sets. For each segment and for all points in the segment-specific grids over post-decision states and constrained choices, use the first order conditions, the inequality constraints, and the budget constraints, to construct node sets containing states and candidate optimal choices. (The details on this was provided separately for each segment above).

Step 3. Local triangulation. For each segment, *i)* divide the regular grids over post-decision states and constrained choices into triangles, *ii)* consider the corresponding triangles mapped into (m, n)-space, and *iii)* construct the triangle's bounding box in (m, n)-space.

Step 4. Interpolation to common state grid $\mathcal{G}_{m,n}$ and first upper envelope. For each segment and each bounding box, *i)* find the nodes in the common state grid $\mathcal{G}_{m,n}$ inside the bounding box, *ii)* for each state space node find candidate choices using barycentric interpolation, *iii)* calculate the implied value of these interpolated choices,[14] and *iv)* update the optimal choice if no previous set of choices have been found yielding a higher value-of-choice.

Step 5. Second upper envelope over segments. For each point in the common state grid $\mathcal{G}_{m,n}$, choose the optimal choice as the choice from the segment with the highest value-of-choice.

The fundamental idea in our solution method lies in the *local triangulation* in step 3. This firstly implies that we avoid computing the global Delaunay triangulation in (m,n)-space, which is a costly operation. Secondly, it implies that it is straightforward to interpolate to the common grid, and find the upper envelopes in step 4 and 5.

Details on the local triangulation. Taking the *ucon* segment as an example, the regularity of the post-decision grids in (a,b)-space imply that the local triangulation is straightforward in this space as also shown in panel (a) of figure 3.2 with rectilinear grids.[15] This further implies that we can avoid time-consuming visibility walks when

[14] Instead also interpolating/extrapolating the value-of-choide is possible and faster, but can imply large errors, especially outside the triangle when extrapolation is used.

[15] We have constructed triangles of the form lower-left (LL) and upper-right (UR) in figure 3.2. Alternatively, LR and UL triangles could be constructed or, in general, all combinations of the simplices could be used to increase robustness.

interpolating the candidate choices at the common grid nodes of m_t and n_t in step 4. The reason is that from the ABC triangle in (a, b)-space shown in panel (a) of figure 3.2, we directly have the transformed ABC triangle in (m, n)-space shown in panel (b). Panel (c) then shows that the bounding box can be constructed from the coordinates of the triangle's corners. The assumed regular structure of the common state grid, $\mathcal{G}_{m,n}$, over m_t and n_t, imply that it is easy to find the sub-grid inside the bounding box (e.g. using bisection searches). For all points in this sub-grid using *barycentric interpolation* in step 4 is then standard. Specifically, for a triangle with corners A, B, and C, the interpolated consumption choice, c, at the state point (m, n) can be found using

$$c = \omega_A c_A + \omega_B c_B + (1 - \omega_A - \omega_B) c_C$$

where the ω's are the *barycentric weights* given by[16]

$$\begin{aligned}
\omega_A &= \frac{(n_B - n_C)(m - m_C) + (m_C - m_B)(n - n_C)}{(n_B - n_C)(m_A - m_C) + (m_C - m_B)(n_A - n_C)} \\
\omega_B &= \frac{(n_C - n_A)(m - m_C) + (m_A - m_C)(n - n_C)}{(n_B - n_C)(m_A - m_C) + (m_C - m_B)(n_A - n_C)}
\end{aligned}$$

which sum to one inside and on the edge of the triangle. To limit the use of extrapolation, we do not consider points with any barycentric weights less than -0.25; in panel (c) of figure 3.2 we thus only consider the black nodes inside the bounding box.[17]

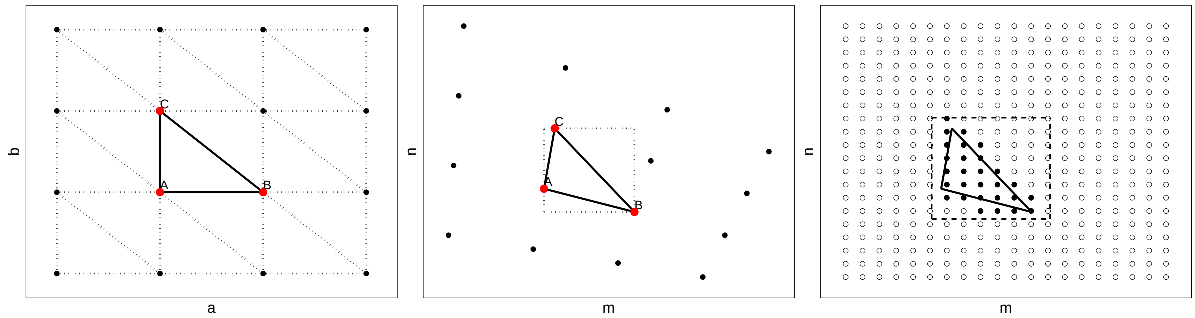
Details on upper envelope. A central feature of our method is that although the considered triangles are disjoint in (a, b)-space, there might be overlaps in (m, n)-space. Multiple (interpolated) guesses of the optimal choices will then be found at some points in the common state grid; only the set of choices implying the highest value-of-choice should therefore be saved. Overlaps in (m, n)-space will happen in regions where there is a kink in the continuation value for the optimally implied post-decision states a_t and b_t. In these regions, the derivatives of the next-period value function are not continuous, and for a small change in a_t and/or b_t it can thus change a lot, implying a large shift in c_t and/or d_t, and therefore also in m_t and/or n_t. This is illustrated in panel (a)-(c) in figure 3.3, where the triangle EDF is to the northeast of the triangle ABC in (a,b)-space (panel

[16] The coordinates are found by solving $m = \omega_A m_A + \omega_B m_B + (1 - \omega_A - \omega_B) m_C$ and $n = \omega_A n_A + \omega_B n_B + (1 - \omega_A - \omega_B) n_C$ for ω_A and ω_B. If A, B, and C should happen to lie on straight line only interpolation and extrapolation along this line is allowed.

[17] Contrary to the case plotted for illustrative purposes in panel (c) in figure 3.2, the (a, b)-grids should be more dense than the (m, n)-grid. This is necessary in order to avoid a large interpolation of interpolation error because multiple interpolations are used to arrive at the optimal choices in the (m, n)-grid. We found that around four times as many points in the (a, b)-grids yielded accurate solutions.

Figure 3.2: Illustration.

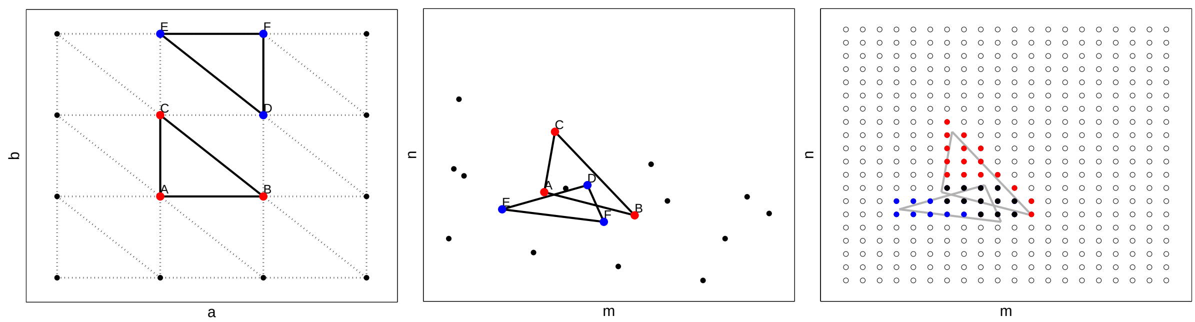
(a) Triangle in (a, b)-space. (b) Triangle in (m, n)-space. (c) Associated common grid nodes.



Notes: Figure 3.2 illustrates how a right-angled triangle in (a, b)-space (panel a) is transformed into a non-right-angled triangle in (m, n)-space (panel b), with an associated bounding box and common grid nodes (black dots) to be interpolated or extrapolated (panel c).

Figure 3.3: Illustration. With Kinks.

(a) Triangles in (a, b)-space. (b) Triangles in (m, n)-space. (c) Associated common grid nodes.



Notes: See figure 3.2.

a), but to the southwest and partly overlapping in (m,n)-space (panel b).[18] In panel (c), we consequently see that the interpolated choice candidates at the filled black nodes are calculated twice, once for each triangle. Because we only save the set of choices implying the highest value-of-choice, this constitutes an upper envelope. The underlying reason for these overlaps is that the FOCs are only necessary, such that there for given m_t and n_t exists multiple c_t and d_t (and thus multiple a_t and b_t) satisfying the FOCs.[19]

Finally, we note that our construction of triangles spanning the relevant (a,b)-space does not necessarily ensure that all the relevant nodes in the common state grid in

[18] The construction of triangles in the post-decision grid, instead of for example squares, ensures that the corresponding triangles mapped into (m, n)-space are convex despite the presence of non-convexities. We thank Matthew White for pointing this out to us.

[19] Using monotonicity requirements on the post-decision states, it would in principle be possible to *a priori* disregard certain triangles with corners on differing sides of a kink. This would be beneficial because the choices interpolated from these triangles will surely be inferior, and disregarding them would therefore speed up our solution method without a loss of accuracy. In order to focus on a simple and robust solution method, we have not explored this possibility any further.

(m, n)-space are covered. Nodes not covered will not be assigned any choices. Allowing for extrapolation (i.e. negative barycentric coordinates), this only happens rarely, but to strengthen the robustness of our method, we add a nearest neighbor interpolation step (between step 4 and 5) for all nodes without any choices assigned. The computational cost of this is negligible.[20]

Handling many segments. An apparent drawback of our solution method is that the number of segments we need to consider is exponentially increasing in the number of occasionally binding constraints.

To speed-up our solution method in the face of this curse of dimensionality, we firstly use the same grids for multiple segments. This implies that the interpolation and inversion of FOCs done in one segment, can be re-used in another segment. Specifically, if we in the *dcon* segment use the same grid over a_t and b_t as in the *ucon* segment, then we can directly copy the c_t choices found in the *ucon* segment to the *dcon* segment. In more general terms, constrained segments will always be special cases of unconstrained segments.

Secondly, we can avoid applying the upper envelope algorithm to nodes, where the constrained choices are clearly not optimal. If we in the *dcon* segment, for example, have a node with states (m_0,n_0) and choices (c_0,$d_0 = 0$), then we can disregard this node if the value-of-choice is increased by slightly deviating from the constraint, i.e. if

$$\frac{(c_0 - \epsilon)^{-\rho}}{1-\rho} + \beta w_t(m_0 - c_0, n_0 + \epsilon + g(\epsilon)) \quad > \quad \frac{c_0^{-\rho}}{1-\rho} + \beta w_t(m_0 - c_0, n_0)$$

where ϵ is a small number.

Thirdly, and finally, for any segment where all the choices are constrained, we can use the common pre-decision state grid as the exogenous grid, and directly find optimal candidate choices and implied values-of-choice.

3.5 Policy Functions

All the following results are, unless otherwise explicitly noted, based on the parameters in table 3.1. This parametrization has *no* stochastic elements smoothing out the non-concave regions of the value function, and has been chosen to illustrate the complexity of the solution, and to test the performance of our proposed solution method in a situation with many discontinuities in the policy functions. For robustness we also consider a case with smoothing in terms of $\sigma_\xi = 0.1$ and $\sigma_\eta^2 = 0.1$.

[20] Alternatively, VFI could be used to determine the optimal choices at these few nodes.

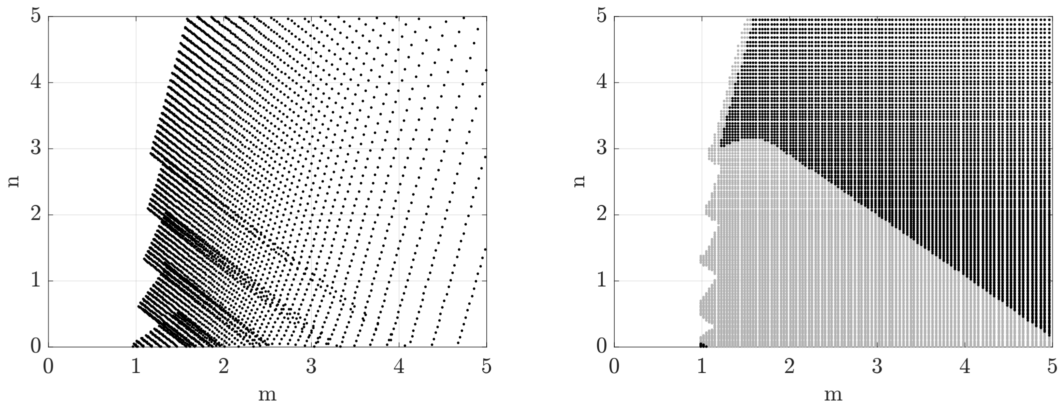
Table 3.1: Baseline Parameter Values.

R_a	R_b	β	ρ	α	χ	$\underline{y}$	σ_η^2	σ_ε
1.02	1.04	0.98	2.00	0.25	0.10	0.50	0.00	0.00

Figure 3.4: Endogenous Irregular Grid, *acon*-segment, $t = T - 5$.

(a) Irregular EGM grid.

(b) Common grid and upper envelopes.



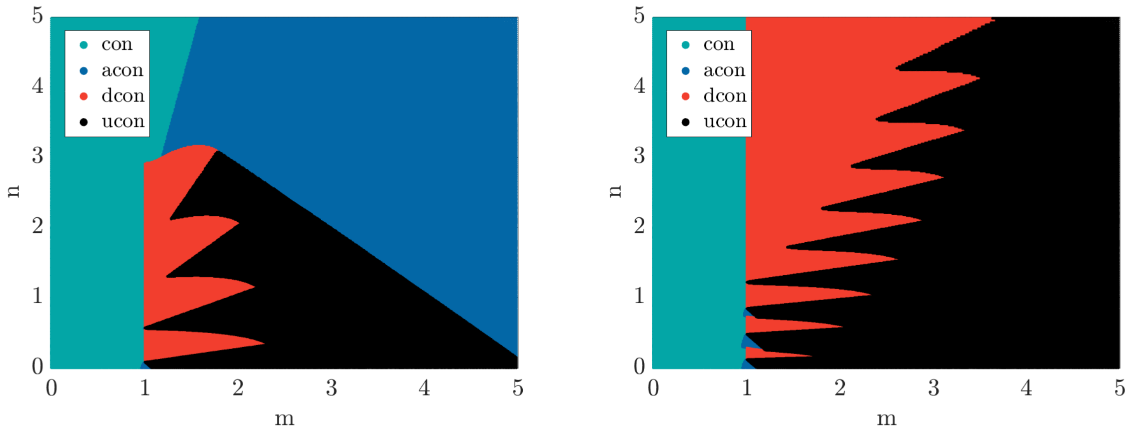
Notes: Figure 3.4 illustrates how the upper envelope algorithm removes non-optimal solutions stemming from the non-convexity of the problem due to the discrete retirement choice, and how the final segment covers the (m, n)-space after being interpolated onto a common (across sections) grid. The black region in panel (b) indicates that the acon-solution is optimal here. See also figure 3.5 panel (a). Only $\#_m = 200$ points are used here.

Figure 3.4 illustrates our upper envelope algorithm applied to the *acon*-segment, in which $a_t = 0$ while $d_t > 0$, in the illustrative model. Panel (a) shows the endogenous irregular grid. The discrete retirement choice in the illustrative model has generated non-optimal solutions to the FOCs visible in the lower left part of panel (a). These non-optimal points are removed by the upper envelope algorithm when interpolating to the common grid illustrated in panel (b). Finally, panel (b) illustrates that primarily the north-east part of the region is actually *optimal* when compared to the remaining segments in *step 5*. This part of the segment is highlighted with black dots in panel (b) while the non-optimal part of the acon segment on the common grid is gray.

Figure 3.5: Optimal segments. Working.

(a) Optimal segments, $t = T - 5$.

(b) Optimal segments, $t = T - 19$.

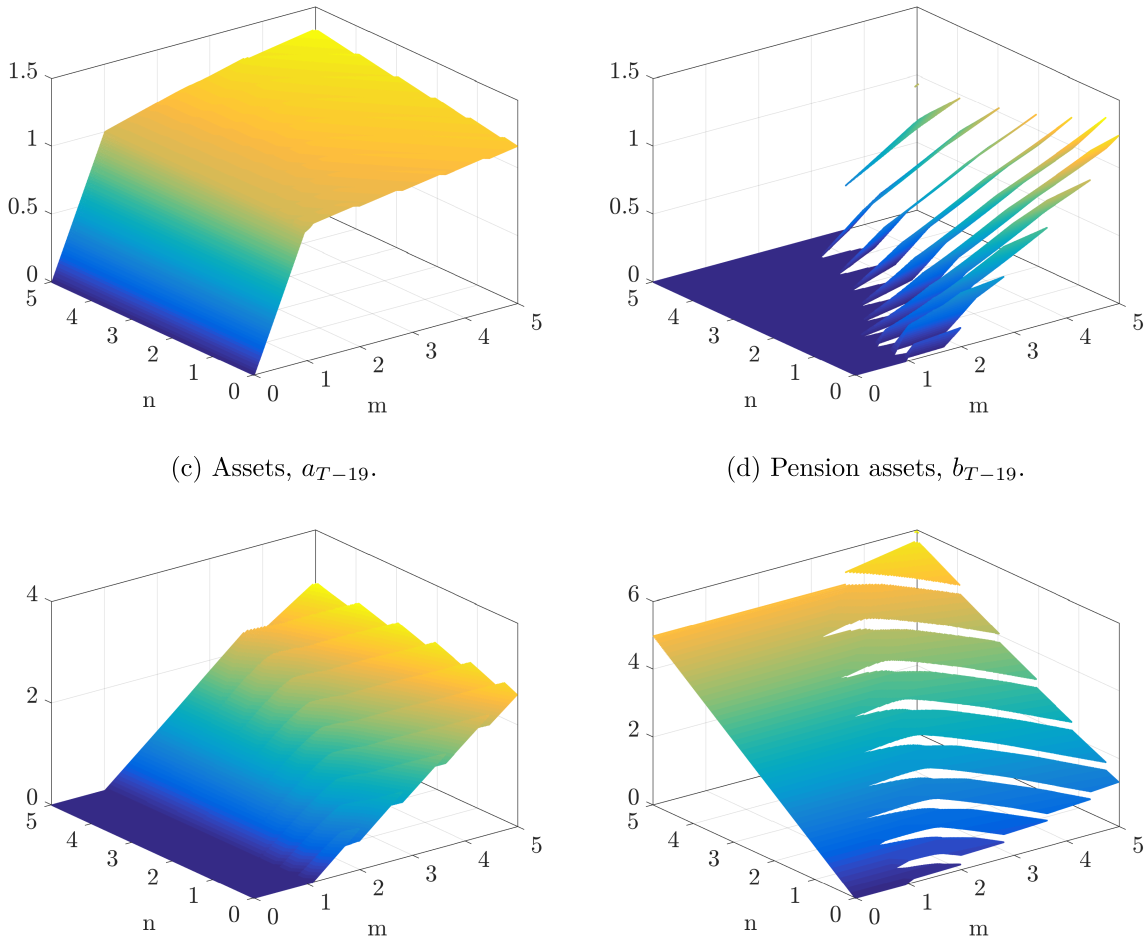


Notes: Figure 3.5 illustrates which segments (*ucon*, *con*, *dcon* and *acon*) are optimal in the (m,n)-space when working, $z_t = 1$. Solved using the G^{2}EGM with $\#_m = 600$ and parameters in table 3.1.

Figure 3.5 shows which segments (*ucon*, *con*, *dcon* and *acon*) are optimal in (m,n)-space for $t \in \{T-5, T-19\}$. This illustrates the complexity of allowing for multiple occasionally binding constraints. Figure 3.6 shows the optimal pension deposit and consumption functions in period $t = T - 19$, and the implied optimal post-decision state functions. Several discontinuities are clearly visible due to the discrete retirement choice, and are captured precisely by our solution method. The logic behind these discontinuities is that at certain points in the state space consumers are indifferent between different planned retirement ages. Infinitesimal changes in m_t and n_t can thus induce a change in the planned retirement age, implying a discontinuous jump in, for example, the optimal level of pension deposits d_t and post-decision pension deposits b_t.

The supplemental material shows the same figures when smoothing is included in the model, $\sigma_\varepsilon = 0.1$ and $\sigma_\eta^2 = 0.1$. As found in Iskhakov, Jørgensen, Rust and Schjerning (2015) for the one-dimensional case, adding smoothing via extreme value type one taste shocks and income uncertainty, reduces the complexity of the solution and removes several (and potentially all) discontinuities in the policy functions.

Figure 3.6: Policy Functions. Working.



(a) Consumption, c_{T-19}. (b) Pension deposits, d_{T-19}. (c) Assets, a_{T-19}. (d) Pension assets, b_{T-19}.

Notes: Figure 3.6 shows optimal policy functions for working households, $z_t = 1$. Solved using the G^{2}EGM with $\#_m = 600$ and parameters in table 3.1.

4 Accuracy and Speed

To illustrate how the accuracy and speed of our proposed solution method is in comparison to existing methods, we compare it to a fully optimized VFI using multi-starts of a local derivative-based optimization algorithm to *globally* search for the optimal choices.[21] The VFI is written almost fully in C++, while only the core parts of the G^{2}EGM is in *C* – both algorithms are called from MATLAB.[22] All code is available from the authors

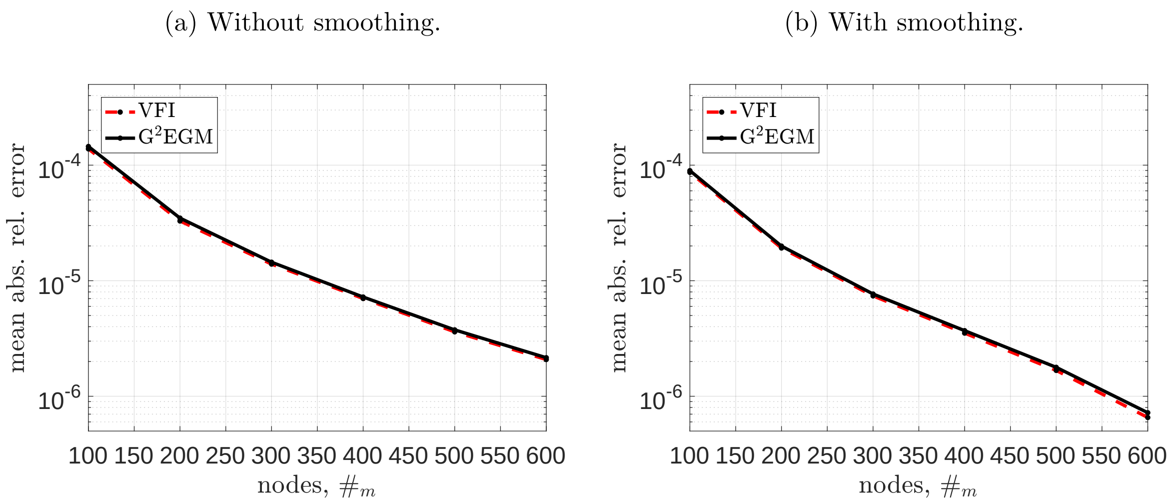
[21]Specifically, we use the *Method of Moving Asymptotes* from Svanberg (2002), implemented in `NLopt` by Johnson (2014). We set *xtol_rel* and *ftol_rel* to 10^{-6}.

[22]A standard alternative to VFI is time iterations (TI), where the optimal choices over an exogenous pre-decision state grid are found by solving the FOCs. In the presence of non-convexities and multiple constraints, TI both needs to find all the solutions to the FOCs and determine which constraints are binding. TI furthermore requires interpolation of both the next period value function and its

web pages. The problem for the retired households are always solved by standard EGM, which takes less than a second.

To speed-up both G^{2}EGM and VFI, we pre-construct an interpolant of $w_t(a_t, b_t)$ for a dense exogenous grid of a_t and b_t, such that we can avoid numerical integration when calculating the value of various candidate choices, and instead rely on interpolation of $w_t(a_t, b_t)$.[23] This construction speeds up the solution significantly, when stochastic elements are included in the model. We have found that grids of a_t and b_t approximately four times as large as the state grid over m_t and n_t is optimal in terms of speed and accuracy. For the G^{2}EGM the same grids over a_t and b_t are used when constructing nodes with candidate optimal choices for each segment.

Figure 4.1: Accuracy of G^{2}EGM and VFI.



Notes: Figure 4.1 shows the accuracy of the G^{2}EGM and VFI. The left panel refers to a model without smoothing and the right panel refers to a version of the model with smoothing ($\sigma_\varepsilon = 0.1$ and $\sigma_\eta^2 = 0.1$).

In VFI we use four different starting values to reduce the risk of reaching a non-global local optimum. Particularly, we initialize the VFI solver in the solution found for the preceding grid point (in the n_t dimension) and the three corners of the choice set.[24] The online supplemental material provides additional implementation details for both methods.

To define a measure of accuracy, we first use an alternative VFI with a slow but robust two step discretized global search algorithm. In the first step we search over a tensor grid of $\#_c = 400$ choice candidates in $c_t \in [0, m_t]$ and $d_t \in [0, m_t]$, while we in the second step

derivatives. We found that TI was not competitive with VFI, where the derivatives are not used.

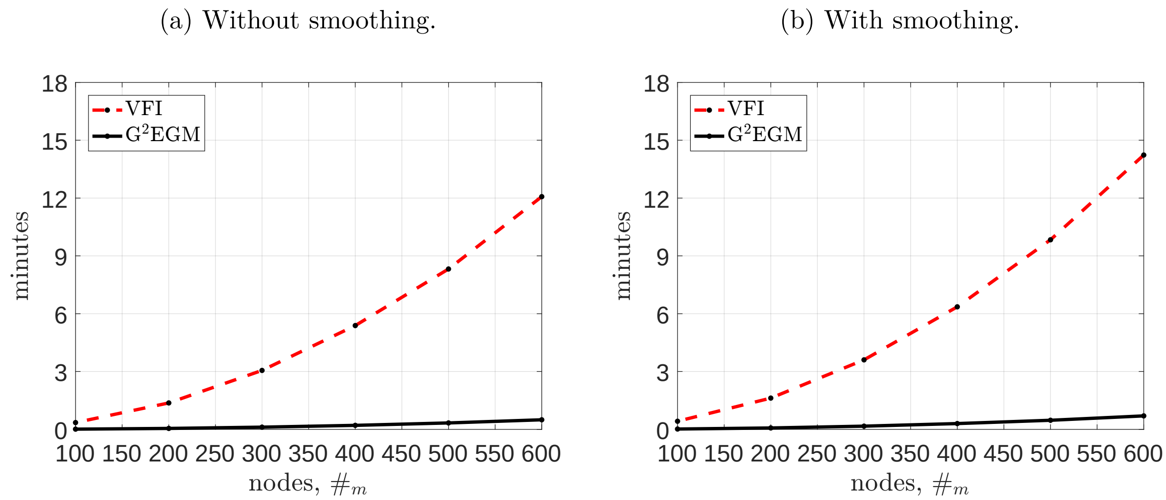
[23]This is normally not done for VFI when comparing it to EGM. Everything else equal, the speed gains of EGM, we find, should therefore be smaller than those typically found in the literature.

[24]The three corners of the choice set are *i)* low c_t, low d_t, *ii)* low c_t, high d_t, and *iii)* high c_t, low d_t. Initializing the solver at the last found solution is a good initial guess, but relying only on this starting value can produce significant errors because it tend to locate only a local maximum.

fine-tune the solution over a discretized tensor grid with 100 choices candidates in each dimension in a close neighborhood of the previously found maximum. We use $\#_m = 800$ points in each dimension of the state space of m_t and n_t, and henceforth denote the resulting solution as the *truth*.

Figure 4.1 shows the *mean absolute relative error* (MARE) in the found value function in period $t = 1$ compared to the *truth* for both G^{2}EGM and VFI when increasing the number of nodes, $\#_m$, in m_t and n_t.[25] The accuracy of G^{2}EGM and our VFI with multiple starting values are on the same order of magnitude. Figure 4.2, on the other hand, shows the associated solution time in minutes. The *G^{2}EGM is around 20 times faster than VFI.*[26] These results are also robust to adding smoothing ($\sigma_\xi = 0.1$ and $\sigma_\eta^2 = 0.1$).

Figure 4.2: Speed of G^{2}EGM and VFI.



Notes: Figure 4.2 shows the speed of the G^{2}EGM and VFI. The left panel refers to a model without smoothing (baseline) and the right panel refers to a version of the model with smoothing ($\sigma_\varepsilon = 0.1$ and $\sigma_\eta^2 = 0.1$).

The speed-up of G^{2}EGM relatively to VFI naturally depends on the number of starting values required for the numerical solver to reach the *global* maximum in VFI. To be completely sure that the global optimum has been found a discrete grid search could be applied (as we do when finding the *truth*). Such a brute force strategy is clearly very time consuming. On the other hand, if there is guaranteed to be no non-global local maxima, a single good guess of the optimum could speed up VFI significantly. However, in models with non-convexities it seems impossible to determine *a priory* whether there – for all

[25] We restrict attention to regions where either $a_t > 0$ or $d_t > 0$ because we found that in the *con*-region the VFI implementation is particularly sensitive to the placement of nodes due to the very high degree of curvature of the value function in that region.

[26] All timings have been computed on a desktop computer with an Intel i7-4770 3.50 Ghz processor. We only report results from a single threaded implementation as the VFI and G^{2}EGM are equally parallelizable.

interesting parameter values – is enough smoothing in the model to achieve this.

As an alternative measure of accuracy we follow the approach proposed by Judd (1992) and Santos (2000) and calculate the Euler residuals from simulated consumers. We simulate $N_s = 10,000$ individuals and in each of the 18 periods we initialize consumers as working ($z_{t-1} = 1$) with endowments (m_t, n_t) evenly spaced across $[0.5, 5] \times [0.01, 5]$. In each period, we calculate the consumption Euler residuals,

$$\tilde{\mathcal{E}}_{i,t} \equiv c_{i,t} - \mathbb{E}_t \left[\beta R_a c_{i,t+1}^{-\rho}\right]^{-\frac{1}{\rho}}$$

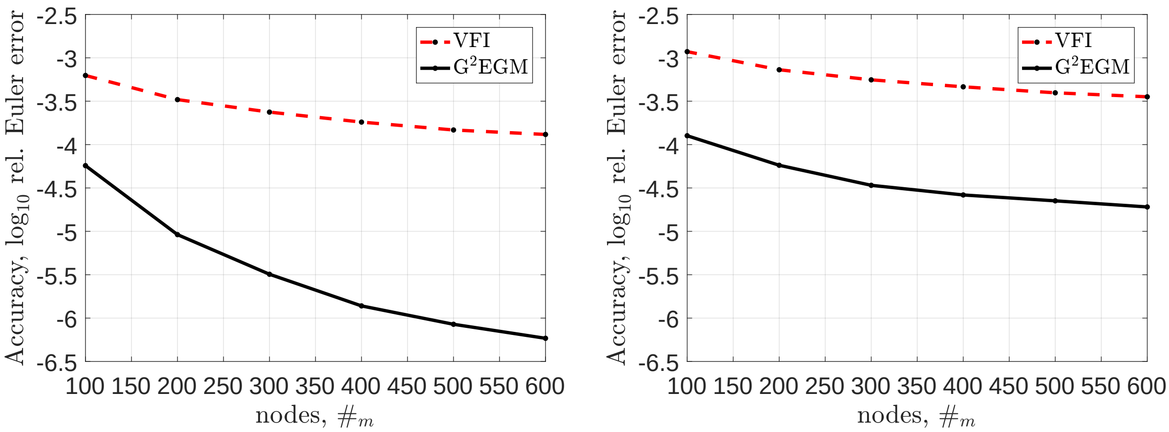
where expectations are approximated using the same Gauss-Hermite quadrature nodes as when solving the model. Using observations where it initially is optimal to keep working and leave some liquid wealth for next period, we calculate a measure of accuracy as the average $\log_{10}$ of the relative Euler error,

$$\mathcal{E} = \frac{\sum_{i=1}^{N_s} \sum_{t=1}^{T_s} \log_{10}(\tilde{\mathcal{E}}_{i,t}/c_{i,t}) \mathbf{1}_{\{z_{i,t}=1, a_t>0\}}}{\sum_{i=1}^{N_s} \sum_{t=1}^{T_s} \mathbf{1}_{\{z_{i,t}=1, a_t>0\}}} \tag{4.1}$$

As shown in figure 4.3 the accuracy increases with the number of grid points and the G^{2}EGM significantly dominates the accuracy of the VFI – again both with and without smoothing. A value of -2 and -4 of the accuracy measure indicate average approximation errors of 1 and 0.01 percent of consumption, respectively.

Figure 4.3: Accuracy: Euler Errors of G^{2}EGM and VFI.

(a) Without smoothing. (b) With smoothing.



Notes: Figure 4.3 shows the accuracy of the VFI and G^{2}EGM in terms of the average (across 10,000 simulated individuals over 18 periods) $\log_{10}$ relative Euler error, described in equation (4.1). The left panel refers to a model without smoothing (baseline) and the right panel refers to a version of the model with smoothing ($\sigma_\varepsilon = 0.1$ and $\sigma_\eta^2 = 0.1$).

4.1 Higher Dimensions

In order to asses the speed and accuracy gains of G^{2}EGM in higher dimensions we extend the illustrative model with a labor supply decision and human capital accumulation in the spirit of Imai and Keane (2004), but with an upper bound on the labor supply choice. Specifically, we augment the utility function with dis-utility of labor, l_t, on the intensive margin for working households,

$$u(c_t, 1, l_t) = \frac{c_t^{1-\rho}}{1-\rho} - \varphi \frac{l_t^{1+\gamma}}{1+\gamma} - \alpha \tag{4.2}$$

and introduce (stochastic) accumulation of human capital, k_t, according to

$$q_t = (1-\delta)k_t + l_t \tag{4.3}$$

$$k_{t+1} = \eta_{t+1} q_t \tag{4.4}$$

where δ is the depreciation rate of human capital, q_t is end-of-period human capital, and η_{t+1} is a permanent shock to human capital with a mean of one. Re-defining resources as simply $m_t = Ra_t$, and using the wage function $w(k_t) = r_k k_t$, end-of-period assets is now given by

$$a_t = m_t + r_k k_t l_t - c_t - d_t \tag{4.5}$$

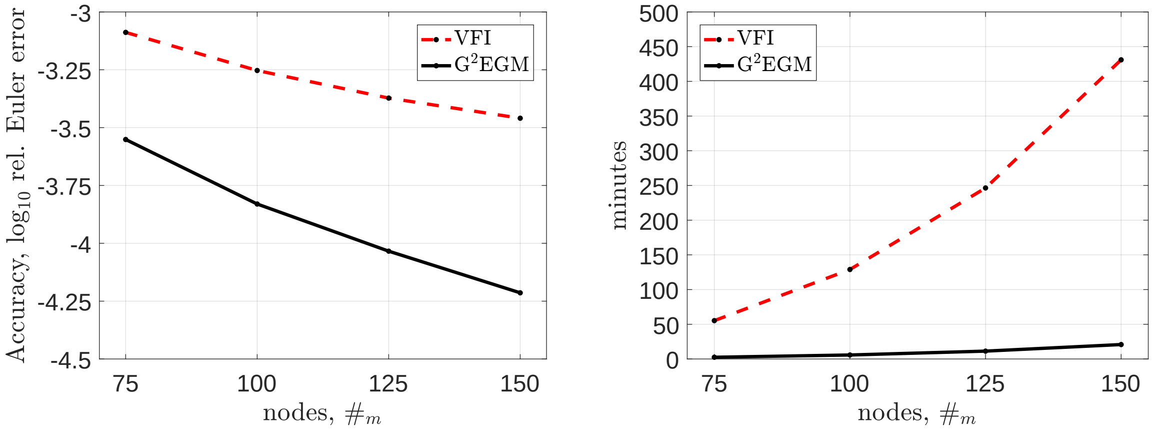
The labor supply is constrained to be in the range $[0, \bar{l}]$.

The full recursive formulation of the problem is given in the online material together with the chosen parametrization. The details on the implementation of the G^{2}EGM for solving this 3-dimensional model is also relegated to the online material. As there are also jumps in the policy function for l_t we augment the multi-start approach for VFI with a low and high labor supply choice, and then start from each corner of the resulting budget set.

G^{2}EGM is again around 20 times faster than VFI, for a given number of grid points and the Euler-errors of our method are substantially smaller. Figure 4.4 compares the accuracy and speed of G^{2}EGM and VFI in solving the extended illustrative model. Memory and time constraints imply that it is unfeasible to solve the model using very fine grids. Consequently the left panel shows the accuracy of the two methods measured in terms of the average $\log_{10}$ of the relative Euler error (see equation (4.1) above) for grids with up to 150 points in each dimension. The right panel shows the associated time (in minutes) required to solve the model.

Figure 4.4: 3-Dimensionsal Model: Accuracy and Speed of G^{2}EGM and VFI.

(a) Euler Errors. (b) Speed.



Notes: Figure 4.4 shows the accuracy of the VFI and G^{2}EGM in terms of the average (across 10,000 simulated individuals over 18 periods) $\log_{10}$ relative Euler error, described in equation (4.1) in the left panel. The right panel shows the associated time spent solving the model for both methods.

5 General Model Class

This section defines a broad class of stochastic dynamic programming models in terms of *necessary* and *sufficient* conditions for which our G^{2}EGM is applicable. In the online supplemental material, we show how the illustrative model presented in section 2 fits into this class.

5.1 Bellman Equation

We begin from a very general stochastic dynamic programming model containing both discrete and continuous states and choices. We assume that the *continuous choices* only affect a subset of the *continuous states*, and *not* the *discrete states*. This allows us to decompose the state space into two parts: The first part consists of a mixed vector of discrete and continuous states $\mathbf{s}_t \in \mathcal{S}$ whose transition is unaffected by the continuous choices. The second part consists of k continuous states $\mathbf{m}_t \in \mathcal{M}(\mathbf{s}_t) \subseteq \mathbb{R}^k$ whose transition depend on both the discrete and continuous choices. Denoting the discrete (or discretized) choices $z_t \in \mathcal{Z}(\mathbf{s}_t)$, the (stochastic) transition function for $\mathbf{s}_t$ is given by $\Gamma_s(\mathbf{s}_t, z_t)$. Denoting the l continuous choices by $\mathbf{c}_t \in \mathcal{C}(\mathbf{s}_t, z_t, \mathbf{m}_t) \subseteq \mathbb{R}^l$, where we assume that $\mathcal{C}(\mathbf{s}_t, z_t, \mathbf{m}_t)$ is bounded, the (stochastic) transition function for $\mathbf{m}_t$ is given by $\Gamma_m(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t)$, which is assumed to be differentiable except at a finite number of kinks or discontinuities.[27]

[27]Continuity and differentiability is always wrt. $(\mathbf{m}_t, \mathbf{c}_t)$ unless noted otherwise.

Denoting the per-period utility flow by $u(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t)$, which is assumed to be differentiable, and assuming exponential discounting with a factor of β, the Bellman equation for the model is[28]

$$\begin{aligned}
V_t(\mathbf{s}_t, \mathbf{m}_t, \varepsilon_t) &= \max_{z_t, \mathbf{c}_t} u(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t) + \sigma_\varepsilon \varepsilon(z_t) + \beta \mathbb{E}_t \left[V_{t+1}(\mathbf{s}_{t+1}, \mathbf{m}_{t+1}) \right] && (5.1)\\
&\text{s.t.} \\
\mathbf{s}_{t+1} &= \Gamma_s(\mathbf{s}_t, z_t) && (5.2)\\
\mathbf{m}_{t+1} &= \Gamma_m(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t) && (5.3)\\
z_t &\in \mathcal{Z}(\mathbf{s}_t) && (5.4)\\
\mathbf{c}_t &\in \mathcal{C}(\mathbf{s}_t, z_t, \mathbf{m}_t) && (5.5)
\end{aligned}$$

where $\varepsilon(z_t)$ is an *iid* extreme value type I taste shock across the discrete choices and σ_ε^2 is proportional to the variance of these shocks.

As always, we assume that there is a *unique* solution to the problem in (5.1)-(5.5), and we denote the optimal policy functions by $z_t^\star(\mathbf{s}_t, \mathbf{m}_t)$ and $c_t^\star(\mathbf{s}_t, \mathbf{m}_t)$. Given some terminal condition, our goal will be to find these optimal policy functions using backwards induction on finite compact grids $\widehat{\mathcal{S}} \subseteq \mathcal{S}$ and $\widehat{\mathcal{M}}(\mathbf{s}_t) \subseteq \mathcal{M}(\mathbf{s}_t)$, where we rely on some form of interpolation between grid points and extrapolation outside.

5.2 Segments

As our G^{2}EGM treats constrained and unconstrained choices differently, as the former satisfy FOCs while the latter does not, it is helpful to reformulate the problem by introducing an additional discrete choice over what we call *segments,* where the continuous choices have to satisfy more strict requirements. In each segment, indexed by q_t, we use $\mathbf{c}_t^+$ to denote the set of $l - r(q_t)$ unconstrained choices, and $\mathbf{c}_t^-$ to denote the set of the $r(q_t) \leq l$ constrained choices. The *un*constrained choices, $\mathbf{c}_t^+$, are without loss of generality required to belong to an open, convex and bounded set $\mathcal{C}^+(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t) \subseteq \mathbb{R}^{k-r(q_t)}$, where the derivative of $\Gamma_m(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t)$ wrt. to $\mathbf{c}_t^+$ always exists and are continuous. Non-differentiabilities in $\Gamma_m(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t)$ wrt. some elements in $\mathbf{c}_t$ can, for example, arise due to kinks or jumps in tax rates or interest rates. The constrained choices, $\mathbf{c}_t^-$, are on the other hand, almost without loss of generality, given by the differentiable function $\check{c}(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t^+, q_t) \in \mathbb{R}^{r(q_t)}$ with both states and unconstrained choices as input arguments. We denote the set of segments by $\mathcal{Q}(\mathbf{s}_t, z_t)$ such that the original choice set is

[28] For notational simplicity the Bellman equation given here does not include Epstein-Zin-Weil preferences, but our method also applies to models with this type of recursive utility.

recovered as a union over all segments,

$$\mathcal{C}(\mathbf{s}_t, z_t, \mathbf{m}_t) = \cup_{q_t \in \mathcal{Q}(\mathbf{s}_t, z_t)} \left\{ \mathbf{c}_t = c(\mathbf{c}_t^-, \mathbf{c}_t^+) \mid \begin{array}{l} \mathbf{c}_t^+ \in \mathcal{C}^+(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t), \\ \mathbf{c}_t^- = \breve{c}(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t^+, q_t) \end{array} \right\} \tag{5.6}$$

where $c(\mathbf{c}_t^-, \mathbf{c}_t^+)$ stacks the unconstrained and the constrained choices correctly. For *direct constraints*, such as $d_t = 0$ in the illustrative model from section 2, the function $\breve{c}(\bullet)$ will be a point for given $\mathbf{s}_t$, z_t and q_t, and thus independent of $\mathbf{m}_t$ and $\mathbf{c}_t^+$. For *borrowing constraints* and *collateral constraints* the function $\breve{c}(\bullet)$ will typically not be independent of $\mathbf{m}_t$ and $\mathbf{c}_t^+$. Introducing a collateral constraint $a_t \geq \theta b_t$ in the illustrative model would, for example, imply that in the segment where this collateral constraint is binding, we would have $c_t = m_t - \theta n_t - (d_t + \theta(d_t + g(d_t))$.[29]

In total, the reformulated problem can be written as

$$\begin{aligned}
V_t(\mathbf{s}_t, \mathbf{m}_t, \varepsilon_t) &= \max_{z_t, q_t, \mathbf{c}_t^+} u(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t) + \sigma_\varepsilon \varepsilon(z_t) + \beta \mathbb{E}_t \left[V_{t+1}(\mathbf{s}_{t+1}, \mathbf{m}_{t+1}) \right] && (5.7) \\
&\text{s.t.} \\
\mathbf{s}_{t+1} &= \Gamma_s(\mathbf{s}_t, z_t) && (5.8) \\
\mathbf{m}_{t+1} &= \Gamma_m(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t) && (5.9) \\
z_t &\in \mathcal{Z}(\mathbf{s}_t) && (5.10) \\
q_t &\in \mathcal{Q}(\mathbf{s}_t, z_t) && (5.11) \\
\mathbf{c}_t^+ &\in \mathcal{C}^+(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t) && (5.12) \\
\mathbf{c}_t^- &= \breve{c}(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t^+, q_t) && (5.13) \\
\mathbf{c}_t &= c(\mathbf{c}_t^-, \mathbf{c}_t^+) && (5.14)
\end{aligned}$$

Denoting the optimal segment choice by $q_t^\star(\mathbf{s}_t, \mathbf{m}_t)$, and the discrete-choice-specific optimal policy function for the unconstrained continuous choices by $\mathbf{c}_t^{+\star}(\mathbf{s}_t, \mathbf{m}_t, z_t, q_t)$, we have that the over-arching optimal policy function for the continuous choices is given by

$$\mathbf{c}_t^\star(\mathbf{s}_t, \mathbf{m}_t) = c(\breve{c}(\mathbf{s}_t, z_t^\star, \mathbf{m}_t, \mathbf{c}_t^{+\star}, q_t^\star), \mathbf{c}_t^{+\star}) \tag{5.15}$$

To simplify notation, we henceforth use the composite variable,

$$\mathbf{x}_t = (\mathbf{s}_t, z_t, q_t) \tag{5.16}$$

to denote the given discrete state ($\mathbf{s}_t$), the current discrete choice (z_t) and the current

[29] This also illustrates that there sometimes might be a freedom of choice wrt. which choices are considered to be constrained in a given segment as d_t could also alternatively be expressed as a deterministic function of m_t, n_t and c_t.

segment (q_t).

5.3 Conditions

To state the first condition for the applicability of G^{2}EGM, we first introduce the notion of *post-decision states*.

Definition 1. We say that a function of states and choices

$$\mathbf{a}_t = a(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t) \in \mathbb{R}^h \tag{5.17}$$

is a *post-decision state function* if the implied *post-decision states* $\mathbf{a}_t$ are a *sufficient* statistic in the sense that they contain all the relevant information for determining the probability distribution of future states,

$$\mathbf{m}_{t+1} = \Gamma_m(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t) \quad = \quad \Gamma_{\mathbf{m}}(\mathbf{x}_t, \mathbf{a}_t) \tag{5.18}$$

$$\mathbb{E}\left[V_{t+1}(\mathbf{s}_{t+1}, \mathbf{m}_{t+1}) \,|\, \mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t\right] \quad = \quad \mathbb{E}\left[V_{t+1}(\mathbf{s}_{t+1}, \mathbf{m}_{t+1}) \,|\, \mathbf{x}_t, \mathbf{a}_t\right] \tag{5.19}$$

such that we can define the *post-decision value function* as

$$w_t(\mathbf{x}_t, \mathbf{a}_t) \equiv \mathbb{E}\left[V_{t+1}(\Gamma_s(\mathbf{x}_t), \Gamma_{\mathbf{m}}(\mathbf{x}_t, \mathbf{a}_t)) \,|\, \mathbf{x}_t, \mathbf{a}_t\right] \tag{5.20}$$

Hereby we have:

Condition 1 (Post-decision states)**.** There exists a differentiable post-decision state function $a(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t)$.

Setting $\mathbf{a}_t = [\mathbf{m}_t, \mathbf{c}_t]$ we see that a (degenerate) post-decision state function always exists, and the requirement that it should be differentiable is weak as we are conditioning on $\mathbf{x}_t$. In the illustrate model a kink or discontinuity in $g(d_t)$ implying that b_t would not be globally differentiable as a function of d_t, could, for example, be handled by introducing segments where d_t is respectively below, at, and above this kink or discontinuity.

The efficiency of our algorithm, however, rely on the dimensionality of $\mathbf{a}_t$ being lower than the full dimensionality of the state and choices spaces because we otherwise in practice are simply doing a time iteration over a fixed grid of $\mathbf{m}_t$ with discretized guesses for the optimal choices. Limiting the dimensionality of $\mathbf{a}_t$ is, however, not always possible. In the illustrative model, we would e.g. need to include n_t and d_t separately as post-decision states (instead of combined in b_t), if we extended the model with a stochastic re-evaluation factor κ_{t+1}, not known in period t, but affecting next-period pension assets only through period t pre-decision pension assets – i.e. if $n_{t+1} = R_b(\kappa_{t+1} n_t + d_t + g(d_t))$.

Given a post-decision state function, the second condition highlighting the need for optimality conditions for the unconstrained choices can be stated as:

Condition 2 (FOCs are at least necessary)**.** $V_{t+1}(\mathbf{s}_{t+1}, \mathbf{m}_{t+1})$ is differentiable in $\mathbf{m}_{t+1}$ at optimal unconstrained choices such that we can derive $l - r(q_t)$ *necessary* FOCs

$$\begin{aligned} \mathbf{0}_{1\times(l-r)} &= u_c(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t)c_{c^+}(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+) \\ &\quad +\beta w_{a,t}(\mathbf{x}_t, \mathbf{a}_t)a_c(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t)c_{c^+}(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+) \\ &\equiv f(\mathbf{m}_t, \mathbf{c}_t^+; \mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-) \end{aligned} \tag{5.21}$$

where $u_c(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t)$ is of dimension $1 \times l$, $c_{c^+}(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+)$ is of dimension $l \times (l - r(q_t))$, $w_{a,t+1}(\mathbf{x}_t, \mathbf{a}_t)$ is of dimension $1 \times h$ and $a_c(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t)$ is of dimension $h \times l$.

The differentiability of the value function can, for example, be checked by applying the recently proposed general approach in Clausen and Strub (2013), which we apply to our illustrative model in the online supplemental material.

Based on the first and second condition, we can now define the *endogenous grid method* (EGM) in general terms for models with multiple states, choices, and constraints.

Definition 2. The *endogenous grid method* (EGM) operator $\mathcal{E}$

$$(\mathbf{m}_t, \mathbf{c}_t^+) = \mathcal{E}(\mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-) \tag{5.22}$$

takes the "parameters" $\mathbf{x}_t$, $\mathbf{a}_t$ and $\mathbf{c}_t^-$, as given, and returns a pair of states and choices $(\mathbf{m_t}, \mathbf{c}_t^+) \in \mathcal{M}(\mathbf{s}_t) \times \mathcal{C}^+(\mathbf{x}_t, \mathbf{m_t})$ by solving the equation system

$$\begin{aligned} \mathbf{0}_{1\times(l-r+h)} &= \begin{bmatrix} f\left(\mathbf{m}_t, \mathbf{c}_t^+; \mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-\right) \\ \mathbf{a}_t - a\left(\mathbf{x}_t, \mathbf{m}_t, c\left(\mathbf{c}_t^-, \mathbf{c}_t^+\right)\right) \end{bmatrix} \\ &\equiv F\left(\mathbf{m_t}, \mathbf{c_t^+}; \mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-\right) \end{aligned} \tag{5.23}$$

where F is a function returning the stacked discrepancies in the FOCs and post-decision state equations implied by a given guess of $(\mathbf{m_t}, \mathbf{c}_t^+)$.

Relying on condition 2, we know that $\mathbf{c}^+ = \mathbf{c}_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t)$ is a solution to the equation system in (5.23) for given $\mathbf{x}_t$ and $\mathbf{m}_t$ together with $\mathbf{c}_t^- = \check{c}(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+)$ and $\mathbf{a}_t = a(\mathbf{x}_t, \mathbf{m}_t, c(\mathbf{c}_t^-, \mathbf{c}_t^+))$. If the FOCs are only *necessary*, but not *sufficient*, then the reverse does not hold; though we for given $\mathbf{x}_t$, $\mathbf{a}_t$ and $\mathbf{c}_t^-$ find a pair $(\mathbf{m_t}, \mathbf{c}_t^+)$ solving the equation system (5.23), it does not follow that $\mathbf{c}_t^+$ is the optimal choice at $\mathbf{m}_t$ (given $\mathbf{x}_t$). We will therefore only refer to a found $\mathbf{c}_t^+$ as *candidate* optimal choices at $\mathbf{m}_t$.

Finding a solution to the equation system (5.23) is in general very fast because the numerical integration underlying $w_{a,t}(\mathbf{x}_t, \mathbf{a}_t)$ (in equation (5.21)) only needs to be performed once for each unique $\mathbf{a}_t$-point (even across segments), and F otherwise purely consists of known functions, which can be evaluated easily. If the inverse of F exists and is analytical, the need for root-finding can be eliminated completely.

The above definition of the EGM operator allows us to define the *EGM set* containing pairs of states and candidate optimal choices, and its *upper envelope*.

Definition 3. We say that $\mathcal{O}(\mathbf{x}_t;\mathcal{A})$ is the *EGM node set* for a given set $\mathcal{A}$, if it is given by applying the EGM operator to each element in $\mathcal{A}$

$$\mathcal{O}(\mathbf{x}_t;\mathcal{A}) = \left\{ \ (\mathbf{m}_t, \mathbf{c}_t^+) = \mathcal{E}(\mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-),\ (\mathbf{a}_t, \mathbf{c}_t^-) \in \mathcal{A} \ \right\} \tag{5.24}$$

Given the *value-of-choice* defined as

$$\tilde{v}(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+) = u(\mathbf{x}_t, \mathbf{m}_t, c(\mathbf{c}_t^-, \mathbf{c}_t^+)) + \beta w_{t+1}(\mathbf{x}_t, a(\mathbf{x}_t, \mathbf{m}_t, c(\mathbf{c}_t^-, \mathbf{c}_t^+))) \tag{5.25}$$

we further say the the corresponding *upper envelope set* is given by removing all pairs, where there exists another pair with the same states, but unconstrained choices implying a higher value-of-choice,

$$\overline{\mathcal{O}}(\mathbf{x}_t;\mathcal{A}) = \left\{ \mathbf{m}_t \in \mathcal{M}(\mathbf{s}_t), \mathbf{c}_t^+ = \arg\max_{\mathbf{c}^+} \tilde{v}(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+) \text{ s.t. } (\mathbf{m}_t, \mathbf{c}^+) \in \mathcal{O}(\mathbf{x}_t;\mathcal{A}) \right\} \tag{5.26}$$

A particularly interesting, and easily constructible, choice of $\mathcal{A}$ is the set of post-decision states and constrained choices implied by a *feasible* choice somewhere in the state space, which naturally nests the set of post-decision states and constrained choices implied by an *optimal* choice,

$$\begin{aligned}
\overline{\mathcal{A}}(\mathbf{x}_t) &= \left\{ \begin{array}{l} \mathbf{c}^- = \breve{c}(\mathbf{x}_t, \mathbf{m}, \mathbf{c}^+), \\ \mathbf{a} = a(\mathbf{x}_t, \mathbf{m}, c(\mathbf{x}_t, \mathbf{m}, \mathbf{c}^+)) \end{array}, (\mathbf{m}, \mathbf{c}^+) \in \mathcal{M}(\mathbf{s}_t) \times \mathcal{C}^+(\mathbf{x}_t, \mathbf{m}) \right\} \\
&\supseteq \mathcal{A}^\star(\mathbf{x}_t) \\
&= \left\{ \begin{array}{l} \mathbf{c}^- = \breve{c}(\mathbf{x}_t, \mathbf{m}, \mathbf{c}_t^{+\star}(\mathbf{x}_t, \mathbf{m})), \\ \mathbf{a} = a(\mathbf{x}_t, \mathbf{m}, \mathbf{c}_t^\star(\mathbf{x}_t, \mathbf{m})) \end{array}, \mathbf{m} \in \mathcal{M}(\mathbf{s}_t) \right\}
\end{aligned} \tag{5.27}$$

A central question now is whether we can ensure that there in the limit will be no pairs of states and optimal choices which $\mathcal{O}(\mathbf{x}_t;\mathcal{A})$ will *not* contain if a discrete approximation $\mathcal{A} \approx \overline{\mathcal{A}}(\mathbf{x}_t)$ becomes dense enough. To ensure this, we introduce the following uniqueness condition:

Condition 3 (Uniqueness)**.** For given "parameters" $\mathbf{x}_t$ and $(\mathbf{c}_t^-, \mathbf{a}_t) \in \mathcal{A}^\star(\mathbf{x}_t)$ the equation system (5.23) must have a *unique* solution in $(\mathbf{m}_t, \mathbf{c}_t^+) \in \mathcal{M}(\mathbf{s}_t) \times \mathcal{C}(\mathbf{x}_t, \mathbf{m}_t)$.[30]

As a parallel to the proof in Iskhakov, Jørgensen, Rust and Schjerning (2015), for the

[30] The uniqueness condition only relates to optimal choices, i.e. where $(\mathbf{c}_t^-, \mathbf{a}_t) \in \mathcal{A}^\star(\mathbf{x}_t)$; for any $(\mathbf{c}_t^-, \mathbf{a}_t) \in \overline{\mathcal{A}}(\mathbf{x}_t) \backslash \mathcal{A}^\star(\mathbf{x}_t)$ we do not need uniqueness because any solutions found will not be optimal.

one dimensional case, we now have:

Lemma 1 (All solutions)**.** *As the approximation $\widehat{\mathcal{A}}$ becomes infinitely dense on a compact subset $\subseteq \overline{\mathcal{A}}(\mathbf{x}_t)$, in the sense that the maximum distance between all adjacent points approaches zero, there are no pairs of states and optimal unconstrained choices, implying post-decision states and constrained choices in $\widehat{\mathcal{A}}$, which will not be included in $\mathcal{O}(\mathbf{x}_t; \widehat{\mathcal{A}})$ if conditions 1-3 are satisfied.*

Proof. By conditions 1-3 the equation system (5.23) constitutes a well defined parametric specification of a curve of pairs of states and all candidate unconstrained choices for given $\mathbf{x}_t$, where $(\mathbf{a}_t, \mathbf{c}_t^-)$ plays the role of the parameters. Remembering that all *optimal* unconstrained choices are *candidate* unconstrained choices, this ensures that in limit as $(\mathbf{a}_t, \mathbf{c}_t^-)$ runs through all the values in $\widehat{\mathcal{A}}$, no pairs of states and optimal unconstrained choices, implying post-decision states and constrained choices in $\widehat{\mathcal{A}}$, are not found. □

This immediately imply that the G^{2}EGM will work in the limit:

Lemma 2 (Upper envelope)**.** *As the approximations $\widehat{\mathcal{A}}$, for all $\mathbf{s}_t \in \widehat{\mathcal{S}}$, $z_t \in \mathcal{Z}(\mathbf{s}_t)$ and $q_t \in \mathcal{Q}(\mathbf{s}_t, z_t)$, becomes infinitely dense on a compact subset*

$$\widehat{\mathcal{A}} \approx \left\{ \begin{array}{l} \mathbf{c}^- = \breve{c}(\mathbf{x}_t, \mathbf{m}, \mathbf{c}^+), \\ \mathbf{a} = a(\mathbf{x}_t, \mathbf{m}, c(\mathbf{x}_t, \mathbf{m}, \mathbf{c}^+)) \end{array} , \left(\mathbf{m}, \mathbf{c}^+\right) \in \widehat{\mathcal{M}}(\mathbf{s}_t) \times \mathcal{C}^+(\mathbf{x}_t, \mathbf{m}) \right\}$$

the value function for all $\mathbf{s}_t$ and $\mathbf{m}_t$ in $\widehat{\mathcal{S}} \times \widehat{\mathcal{M}}(\mathbf{s}_t)$ is given by

$$\begin{aligned} V(\mathbf{s}_t, \mathbf{m}_t) &= \max_{z_t, q_t, \mathbf{c}_t^+} \tilde{v}(\mathbf{s}_t, z_t, q_t, \mathbf{m}_t, \mathbf{c}_t^+) \\ & \quad s.t. \\ (\mathbf{m}_t, \mathbf{c}_t^+) &\in \overline{\mathcal{O}}(\mathbf{s}_t, z_t, q_t; \widehat{\mathcal{A}}) \end{aligned}$$

and the optimal policy functions are the maximizing arguments.

5.4 Verifying Uniqueness

Verifying condition 3 directly is, however, typically not possible because it through $\mathcal{A}^\star(\mathbf{x}_t)$ relies on the endogenous function $\mathbf{c}_t^{\star+}(\mathbf{x}_t, \mathbf{m}_t)$ (see equation (5.27)). A feasible alternative is to instead consider the following *sufficient* requirement based on the *constructible* set $\overline{\mathcal{A}}(\mathbf{x}_t)$.

Lemma 3 (Sufficient requirement)**.** *Condition 3 is satisfied if F is an injection in $\mathbf{m}_t$ and $\mathbf{c}_t^+$ for all $(\mathbf{c}_t^-, \mathbf{a}_t) \in \overline{\mathcal{A}}(\mathbf{x}_t)$, i.e.*

$$\begin{aligned} (\mathbf{m}_t', \mathbf{c}_t^{+\prime}) &\neq (\mathbf{m}_t'', \mathbf{c}_t^{+\prime\prime}) \rightarrow \\ F(\mathbf{m}_\mathbf{t}', \mathbf{c}_\mathbf{t}^{+\prime}; \mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-) &\neq F(\mathbf{m}_t'', \mathbf{c}_t^{+\prime\prime}; \mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-) \end{aligned} \tag{5.28}$$

for $\mathbf{m}_t'$, $\mathbf{m}_t'' \in \mathcal{M}(\mathbf{s}_t)$, $\mathbf{c}_t^{+\prime} \in \mathcal{C}^+(\mathbf{x}_t, \mathbf{m}_t')$, *and* $\mathbf{c}_t^{+\prime\prime} \in \mathcal{C}^+(\mathbf{x}_t, \mathbf{m}_t'')$.

Proof. The required injectivity is a *sufficient* requirement because if we for a $(\mathbf{c}^-, \mathbf{a}) \in \mathcal{A}^\star(\mathbf{x}_t) \subseteq \overline{\mathcal{A}}(\mathbf{x}_t)$ find a solution $F(\mathbf{m}_\mathbf{t}', \mathbf{c}_\mathbf{t}^{+\prime}; \mathbf{x}_t, \mathbf{a}, \mathbf{c}^-) = 0$, we then know there can be no other *optimal* choices with the same $(\mathbf{c}^-, \mathbf{a})$. □

Determining whether F is an injection in $\mathbf{m}_t$ and $\mathbf{c}_t^+$ (thus fulfilling lemma 3 and therefore condition 3) can often be proven by construction because if F^{-1} exists then F is an *bijection* and therefore also an *injection*; and this is the case even if F^{-1} is not analytical.[31]

Alternatively, the injectivity of F can be proven using abstract sufficiency results on the injectivity of functions on convex sets (e.g. Gale and Nikaido (1965) where sufficiency is established by requiring that the Jacobian of F wrt. $\mathbf{m}_t$ and $\mathbf{c}_t^+$ is always positive (or negative) semi-definite), or a global inverse function theorem.[32]

The restriction of the uniqueness requirement in condition 3 can be illustrated in terms of a *necessary* requirement on the injectivity in the post-decision states and constrained choices implied by the *optimal* unconstrained choices. Specifically we have

Lemma 4 (Necessary requirement)**.** *A* necessary *requirement for condition 3 is that the optimal choice function must imply that the combined post-decision states and constrained choices is an* injection *in* $\mathbf{m}_t$*, i.e.*

$$\mathbf{m}_t' \neq \mathbf{m}_t'' \rightarrow \begin{array}{l} a(\mathbf{x}_t, \mathbf{m}_t', c_t^\star(\mathbf{x}_t, \mathbf{m}_t')) \neq a(\mathbf{x}_\mathbf{t}, \mathbf{m}_\mathbf{t}'', c_t^\star(\mathbf{x}_t, \mathbf{m}_t'')), \textit{ and/or} \\ \breve{c}(\mathbf{x}_t, \mathbf{m}_t', \mathbf{c}_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t')) \neq \breve{c}(\mathbf{x}_t, \mathbf{m}_t'', \mathbf{c}_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t'')) \end{array} \tag{5.29}$$

for $\mathbf{m}_t'$, $\mathbf{m}_t'' \in \mathcal{M}(\mathbf{s}_t)$.

Proof. The required injectivity is *necessary* because we otherwise for *some* $\mathbf{m}_t', \mathbf{m}_t'' \in \mathcal{M}(\mathbf{s}_t)$ would have a violation of the uniqueness condition by

$$\begin{aligned} (\mathbf{m}_t', c_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t')) \neq (\mathbf{m}_t'', c_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t'')) \quad &\rightarrow \quad F(\mathbf{x}_t, \mathbf{m}_t', c_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t'); \mathbf{a}, \mathbf{c}^-) \\ &= \quad F(\mathbf{x}_t, \mathbf{m}_t'', c_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t''); \mathbf{a}, \mathbf{c}^-) \\ &= \quad \mathbf{0} \end{aligned}$$

where

$$\begin{aligned} \mathbf{a} &= a(\mathbf{x}_t, \mathbf{m}_t', c_t^\star(\mathbf{x}_t, \mathbf{m}_t')) = a(\mathbf{x}_\mathbf{t}, \mathbf{m}_\mathbf{t}'', c_t^\star(\mathbf{x}_t, \mathbf{m}_t'')) \\ \mathbf{c}^- &= \breve{c}(\mathbf{x}_t, \mathbf{m}_t', \mathbf{c}_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t')) = \breve{c}(\mathbf{x}_t, \mathbf{m}_t'', \mathbf{c}_t^{+\star}(\mathbf{x}_t, \mathbf{m}_t'')) \end{aligned}$$

[31] Iskhakov (2015) presents sufficient requirements for F to be *analytically* invertible in the unconstrained case. His proof can easily be extended to cover similar cases of *non-analytical* invertibility.

[32] In principle it might be feasible to put forward necessary and sufficient requirements on the model fundamentals in order to verify condition 3, but we leave this task to future work.

□

In the one-dimensional case this result is equivalent to the requirement of a monotonic savings function discussed by Fella (2014) and Iskhakov, Jørgensen, Rust and Schjerning (2015).

The restrictiveness of the uniqueness requirement can also be elaborated on in terms of our illustrative model, where we for the unconstrained segment have

$$F(\mathbf{m_t}, \mathbf{c_t^+}; \mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-) = \begin{bmatrix} u_c(c_t) - \beta w_{a,t+1}(a_t, b_t) \\ -w_{a,t}(a_t, b_t) + w_{b,t}(a_t, b_t)(1 + g_d(d_t)) \\ a_t - (m_t - c_t - d_t) \\ b_t - (n_t + d_t + g(d_t)) \end{bmatrix}$$

We see if $g_d(d_t)$ were independent of d_t then F would not be invertible such that condition 3 would not be satisfied. Fixing a_t and b_t we could still easily find c_t, but as the second equation no longer would give us d_t, we would have three unknowns for the last two equations, and no unique solution.

6 Solution Method

The upper envelope set $\overline{\mathcal{O}}(\mathbf{x}_t; \widehat{\mathcal{A}})$, given by equation (5.26), is theoretically satisfactory, but not useful in practice. The reason is that it for finite $\widehat{\mathcal{A}}$ the set $\overline{\mathcal{O}}(\mathbf{x}_t; \widehat{\mathcal{A}})$ might contain pairs of states and choices where the choices are not optimal because the dominating pair with the same states (but the globally optimal choices) have not have been created yet. In practice, we therefore need a more robust upper envelope set construction procedure also using information from "neighboring" pairs with similar states.

In order to do so, we first define the following two objects:

Definition 4. We let $\check{\mathbf{c}}^+(\mathbf{m}_t; \mathcal{G})$ denote *interpolation* of $\mathbf{c}_t^+$ on a grid

$$\mathcal{G} = \mathcal{G}(\mathcal{T}) = \left\{ (\mathbf{m}_t, \mathbf{c}_t^+) = \mathcal{E}(\mathbf{x}_t, \mathbf{a}_t, \mathbf{c}_t^-) \,|\, (\mathbf{a}_t, \mathbf{c}_t^-) \in \mathcal{T} \right\}$$

of $(\mathbf{m}_t, \mathbf{c}_t^+)$-nodes created by the EGM, where $\mathcal{T}$ is a set of $(\mathbf{a}_t, \mathbf{c}_t^-)$-nodes.

Definition 5. We let $\mathbb{S}(\widehat{\mathcal{A}})$ denote a set where each element is collection of simplex corners such that the combined simplexes *covers* $\widehat{\mathcal{A}}$ *at least once.*

For finite grids $\widehat{\mathcal{M}}$ over $\mathbf{m}_t$ and $\widehat{\mathcal{A}}$ over $(\mathbf{a}_t, \mathbf{c}_t^-)$, this let us consider the following *alternative upper envelope set,*

$$\tilde{\mathcal{O}}(\mathbf{x}_t; \widehat{\mathcal{M}}, \widehat{\mathcal{A}}) = \left\{ \begin{array}{l} \mathbf{m}_t \in \widehat{\mathcal{M}}, \\ \mathbf{c}_t^+ = \check{\mathbf{c}}^+(\mathbf{m}_t; \mathcal{G}(\mathcal{T}^\star)), \\ \mathcal{T}^\star = \arg\max_{\mathcal{T} \in \mathbb{S}(\widehat{\mathcal{A}})} \tilde{v}(\mathbf{x}_t, \mathbf{m}_t, \check{\mathbf{c}}^+(\mathbf{m}_t; \mathcal{G}(\mathcal{T}))), \end{array} \right\} \tag{6.1}$$

where the unconstrained choices are found by local interpolation, and only the pairs with choices implying the highest value-of-choice are kept.

Given a certain degree of smoothness of the derivatives $w_{t,a}$, u_c, and a_c in the neighborhood of the optimal choices, and the implied post-decision states, the F-function will have a local inverse. For a small change in $\mathbf{a}_t$ only small changes in $(\mathbf{m}_t, \mathbf{c}_t^+)$ will thus be required to keep the equation system (5.23) satisfied. Once $\widehat{\mathcal{A}}$ is dense enough the interpolation will therefore not be across any discontinuities in the policy functions, and the implied error will be small.

In full detail, the G^{2}EGM solution method consists of the following steps:

Step 0. Construct grids and simplexes, find terminal period solution, and set time-index to $t = T - 1$.

Step 0a. For each $(\mathbf{s}_T, z_T)$ construct:

i) a common (regular) grid over (pre-decision) states $\widehat{\mathcal{M}} = \{\mathbf{m}_1, \ldots, \mathbf{m}_{\#_m}\}$.

ii) a common (regular) grid over post-decision states $\widehat{\mathcal{A}} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{\#_a}\}$.

iii) all the q_t-specific grids over tuples of post-decision states and constrained choices $\widehat{\mathcal{A}}_{q_t} = \{(\mathbf{a}_1, \mathbf{c}_1^-), \ldots, (\mathbf{a}_{\#_{qac}}, \mathbf{c}_{\#_{qac}}^-)\}$.

Step 0b. For all $(\mathbf{s}_t, z_t)$, and each q_t, divide the grid $\widehat{\mathcal{A}}_{q_t}$ into *simplexes* (2D: triangles, 3D: tetrahedra, etc.) covering its entire span at least once, $\mathbb{S}(\widehat{\mathcal{A}}_{q_t})$.

Step 0c. For all $(\mathbf{s}_T, z_T)$:

i) find the discrete-specific terminal value- and policy functions, $v_T(\mathbf{s}_T, z_T, \mathbf{m}_T)$ and $c_T^\star(\mathbf{s}_T, z_T, \mathbf{m}_T)$, by solving the fully intra-temporal problem using standard tools.

ii) compute the value function derivatives $v_{m,T}(\mathbf{s}_T, z_T, \mathbf{m}_T)$.

iii) construct interpolants for the value function $\check{v}_T\left(\mathbf{s}_T, z_T, \mathbf{m}_T; \widehat{\mathcal{M}}\right)$ and its derivatives $\check{v}_{m,T}\left(\mathbf{s}_T, z_T, \mathbf{m}_T; \widehat{\mathcal{M}}\right)$ (henceforth all interpolants are indicated with a $\check{\bullet}$).

Step 1. For all $(\mathbf{s}_t, z_t)$, construct interpolants of the expected next-period value function and associated derivatives which all take post-decision states as inputs,

$$\check{w}_t(\mathbf{s}_t, z_t, \mathbf{a}_t; \widehat{\mathcal{A}}) \;=\; \mathbb{E}_t\left[EV_{t+1}(\mathbf{s}_t, z_t, \mathbf{m}_{t+1})\right] \tag{6.2}$$

$$\check{w}_{\mathbf{a},t}(\mathbf{s}_t, z_t, \mathbf{a}_t; \widehat{\mathcal{A}}) \;=\; \mathbb{E}_t\left[\sum_{\mathcal{Z}(\mathbf{s}_{t+1})} \Pr(z|\mathbf{s}_{t+1}, \mathbf{m}_{t+1})\check{v}_{m,t+1}(\mathbf{s}_{t+1}, z, \mathbf{m}_{t+1})\right] \tag{6.3}$$

where $\mathbf{m}_{t+1} = \Gamma_m\left(\mathbf{s}_t, z_t, \mathbf{a}_t\right)$ and $\mathbb{E}_t\left[\bullet\right] = \mathbb{E}\left[\bullet \,|\, \mathbf{s}_t, z_t, \mathbf{a}_t\right]$ is computed using some form of numerical integration.[33] The closed form expectations over taste shocks in (6.2) is given by the log-sum,

$$EV_{t+1}(\mathbf{s}_{t+1}, z_{t+1}, \mathbf{m}_{t+1}) \equiv \begin{cases} \sigma_\varepsilon \log\left(\sum_{z\in\mathcal{Z}(\mathbf{s}_{t+1})} \check{v}_{t+1}(\mathbf{s}_{t+1}, z, \mathbf{m}_{t+1})/\sigma_\varepsilon\right) & \text{if } \sigma_\varepsilon > 0 \\ \max_{z\in\mathcal{Z}(\mathbf{s}_{t+1})} \check{v}_{t+1}(\mathbf{s}_{t+1}, z, \mathbf{m}_{t+1}) & \text{if } \sigma_\varepsilon = 0 \end{cases}$$

and the choice-probabilities are given by the standard multinomial logit,

$$\Pr(z|\mathbf{s}_{t+1}, \mathbf{m}_{t+1}) \equiv \begin{cases} \frac{\exp(\check{v}_{t+1}(\mathbf{s}_{t+1}, z, \mathbf{m}_{t+1})/\sigma_\varepsilon)}{\sum_{k\in\mathcal{Z}(\mathbf{s}_{t+1})}\exp(\check{v}_{t+1}(\mathbf{s}_{t+1}, k, \mathbf{m}_{t+1}))/\sigma_\varepsilon)} & \text{if } \sigma_\varepsilon > 0 \\ \mathbf{1}\left\{z = \arg\max_{z\in\mathcal{Z}(\mathbf{s}_{t+1})} \check{v}_{t+1}(\mathbf{s}_{t+1}, z, \mathbf{m}_{t+1})\right\} & \text{if } \sigma_\varepsilon = 0 \end{cases}$$

The log-sum and choice probabilities reduce to the max-operator and indicator function, respectively, if there are no taste shocks in the model, $\sigma_\varepsilon = 0$.

Step 2. For all $(\mathbf{s}_t, z_t)$, and each q_t, construct the segment-specific optimal policy functions $c^\star_t(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t)$ and the associated implied value-of-choice functions $v_t(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t)$ over the common grid $\widehat{\mathcal{M}}$.

Step 2a. For each $\mathbf{m}_t$ in $\widehat{\mathcal{M}}$ initialize $v_t(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t) = -\infty$.

Step 2b. For all nodes $(\mathbf{a}_t, \mathbf{c}^-_t)$ in $\widehat{\mathcal{A}}_{q_t}$, solve the equation system in (5.23),

$$F\left(\bullet\right) = \begin{bmatrix} f\left(\mathbf{m}_t, \mathbf{c}^+_t; \mathbf{s}_t, z_t, q_t, t_i, \mathbf{c}^-_t\right) \\ \mathbf{a}_t - a(\mathbf{s}_t, z_t, \mathbf{m}_i, c(\mathbf{c}^-_t, \mathbf{c}^+_t)) \end{bmatrix} = \mathbf{0}$$

for the $l - r(q_t)$ unconstrained choices $\mathbf{c}^+_t$ and k states $\mathbf{m}_t$ where $f\left(\bullet\right)$ contains the FOC discrepancies. (For further details on how to solve this equation system analytically or by root-finding see section 5).

Step 2c. For all $\mathcal{T} \in \mathbb{S}(\widehat{\mathcal{A}}_{q_t})$ (see step 0b):

i) construct the k-dimensional bounding box, and use bisection search in each dimension to find the sub-grid $\underline{\widehat{\mathcal{M}}} \subset \widehat{\mathcal{M}}$ inside it.

[33] Both interpolants $\check{w}_t$ and $\check{w}_{\mathbf{a},t}$ imply an interpolation of an interpolation, and in order to avoid a precision loss, it is therefore important that the grid over post-decision states $\widehat{\mathcal{A}}$, is dense relative to the grid over pre-decision states $\widehat{\mathcal{M}}$.

ii) for each $\mathbf{m}_t$ in $\widehat{\mathcal{M}}$ interpolate the *unconstrained* choices $\check{\mathbf{c}}^{\star+}$ using barycentric interpolation (or extrapolation if the coordinates are not too negative), and calculate the optimal choices and the value-of-choice[34]

$$\begin{aligned}\tilde{c} &= c(\mathbf{c}_t^-, \check{\mathbf{c}}^{\star+}(\mathbf{m}_t; \mathcal{G}(\mathcal{T}))) \\ \tilde{v} &= u(\mathbf{s}_t, z_t, \mathbf{m}_t, \tilde{c}) + \beta \check{w}_{t+1}(\mathbf{s}_t, z_t, a(\mathbf{s}_t, z_t, \mathbf{m}_t, \tilde{c}))\end{aligned}$$

and if $\tilde{v} > v_t(\mathbf{s}_t, z_t, \mathbf{m}_i, q)$ update

$$\begin{aligned}v_t(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t) &= \tilde{v} \\ c_t^\star(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t) &= \tilde{c}\end{aligned}$$

Step 2e. For all $\mathbf{m}_t \in \widehat{\mathcal{M}}$ with $v_t(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t) = -\infty$ in the "close neighborhood" of a $\mathbf{m}_t'$ with $v_t(\mathbf{s}_t, z_t, \mathbf{m}_t', q_t) \neq -\infty$ use nearest neighbor interpolation to interpolate the choices and calculate the implied value-of-choice.[35]

Step 3. For all $(\mathbf{s}_t, z_t)$, find the over-arching optimal continuous choices and construct interpolants for the value function and its derivatives.

Step 3a. Find the overarching maximum across q_t-cases using the max-operator for each $\mathbf{m}_t$ in $\widehat{\mathcal{M}}$, i.e.

$$q_t^\star(\mathbf{s}_t, z_t, \mathbf{m}_t) = \arg\max_{q_t} v_t(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t)$$

and set

$$\begin{aligned}v_t(\mathbf{s}_t, z_t, \mathbf{m}_t) &= v_t(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t^\star(\mathbf{s}_t, z_t, \mathbf{m}_t)) \\ c_t^\star(\mathbf{s}_t, z_t, \mathbf{m}_t) &= c_t^\star(\mathbf{s}_t, z_t, \mathbf{m}_t, q_t^\star(\mathbf{s}_t, z_t, \mathbf{m}_t)\end{aligned}$$

Step 3b. Terminate if $t = 1$, else:

i) compute the value function derivatives $v_{m,t}(\mathbf{s}_t, z_t, \mathbf{m}_t)$.

ii) construct the interpolants $\check{v}_t(\mathbf{s}_t, z_t, \mathbf{m}_t; \widehat{\mathcal{M}})$ and $\check{v}_{m,t}(\mathbf{s}_t, z_t, \mathbf{m}_t; \widehat{\mathcal{M}})$.

iii) decrease the time index to $t = t - 1$, and return to step 1.

[34] Interpolating the value-of-choices separately can create a large loss of precision.

[35] Alternatively, VFI could be used to determine the optimal choices at these few nodes.

7 Concluding Remarks

We have provided a generalized version of the endogenous grid method (EGM) originally proposed by Carroll (2006) to solve one-dimensional continuous choice models. While the EGM has been generalized in recent research, our parsimonious solution method is the first to simultaneously handle *i)* multiple continuous states and choices, *ii)* multiple occasionally binding constraints, and *iii)* discrete choices as well as other non-convexities. Furthermore, we explicitly provide *necessary* and *sufficient* conditions for when our solution method can be applied, and we define a general model class in terms of those conditions.

There is a vast range of models that can be solved using our proposed method. We show that our proposed generalized EGM is more than an order of magnitude faster than standard value function iteration in solving an illustrative model of liquid and illiquid assets with a discrete retirement choice, without reducing numerical accuracy. In turn, our method makes it possible to estimate and perform policy analysis from much richer models than what have typically been feasible in the existing literature. Our proposed method, we envision, will provide applied researchers with a powerful, yet relatively easy to implement, numerical tool to investigate economic questions with more realistic economic models. For example, including more preference heterogeneity into complex dynamic economic models is an interesting avenue for future research, and the speed gain from our proposed solution method would enable inclusion of such heterogeneity.

References

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

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

Berger, D. and J. Vavra (2015): "Consumption Dynamic During Recessions," *Econometrica*, 83(1), 101–154.

Bertsekas, D. P. (2012): *Dynamic Programming and Optimal Control: Approximate dynamic programming*. Athena Scientific.

Brumm, J. and M. Grill (2014): "Computing equilibria in dynamic models with occasionally binding constraints," *Journal of Economic Dynamics and Control*, 38, 142–160.

Carroll, C. D. (2006): "The method of endogenous gridpoints for solving dynamic stochastic optimization problems," *Economics Letters*, 91(3), 312–320.

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

Den Haan, W. J., K. L. Judd and M. Juillard (2011): "Computational suite of models with heterogeneous agents II: Multi-country real business cycle models," *Journal of Economic Dynamics and Control*, 35(2), 175–177.

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), 693–732.

Gale, D. and H. Nikaido (1965): "The Jacobian matrix and global univalence of mappings," *Mathematische Annalen*, 159(2), 81–93.

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.

Hong, J. H. and J.-V. Ríos-Rull (2012): "Life Insurance and Household Consumption," *American Economic Review*, 102(7), 3701–30.

Hull, I. (2015): "Approximate dynamic programming with post-decision states as a solution method for dynamic economic models," *Journal of Economic Dynamics and Control*, 55, 57–70.

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., T. H. Jørgensen, J. Rust and B. Schjerning (2015): "Estimating Discrete-Continuous Choice Models: The Endogenous Grid Method with Taste Shocks," Discussion Paper 15-19, Department of Economics, University of Copenhagen.

Johnson, S. G. (2014): "The NLopt nonlinear-optimization package," http://ab-initio.mit.edu/nlopt.

Judd, K. (1992): "Projection Methods for Solving Aggregate Growth Models," *Journal of Economic Theory*, 58, 410–452.

Kaplan, G. and G. Violante (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82(4), 1199–1239.

Ludwig, A. and M. Schön (2014): "Endogenous Grids in Higher Dimensions: Delaunay Interpolation and Hybrid Methods," Working Paper 65, University of Cologne.

Maliar, L. and S. Maliar (2014): "Numerical methods for large scale dynamic economic models," in *Handbook of Computational Economics*, ed. by K. Schmedders and K. Judd, vol. 3, chap. 7, pp. 325–477. Elsevier Science, Amsterdam.

Powell, W. B. (2011): *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey.

Santos, M. S. (2000): "Accuracy of Numerical Solutions Using the Euler Equation Residuals," *Econometrica*, 68(6), 1377–1402.

Svanberg, K. (2002): "A class of globally convergent optimization methods based on conservative convex separable approximations," *SIAM Journal on Optimization*, pp. 555–573.

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

Supplemental Material (not for publication)

A General Endogenous Grid Method for Multi-Dimensional Models with Non-Convexities and Constraints

Jeppe Druedahl

Thomas H. Jørgensen

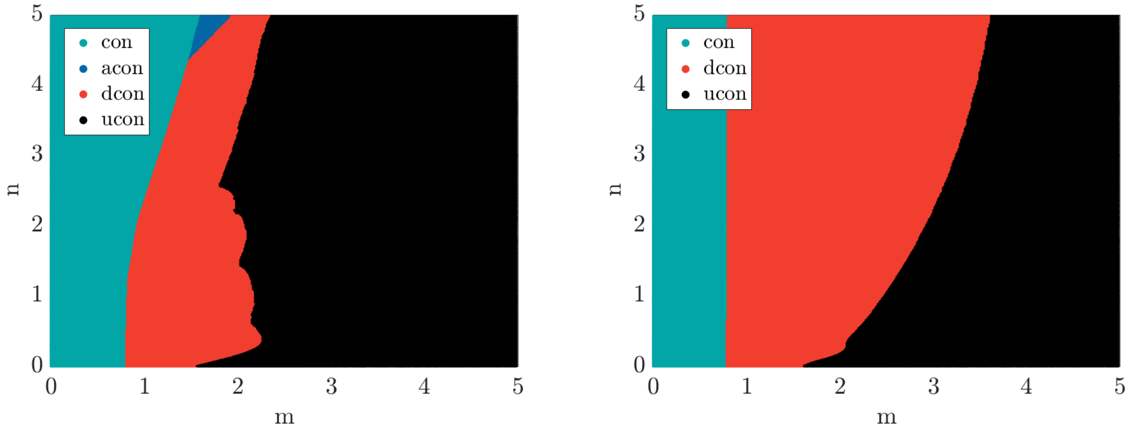
September 1, 2016

A Additional Figures and Tables

Figure A.1: Optimal segments - *with smoothing* ($\sigma_\varepsilon = 0.1$ and $\sigma_\eta^2 = 0.1$).

(a) Optimal sections, $t = T - 5$.

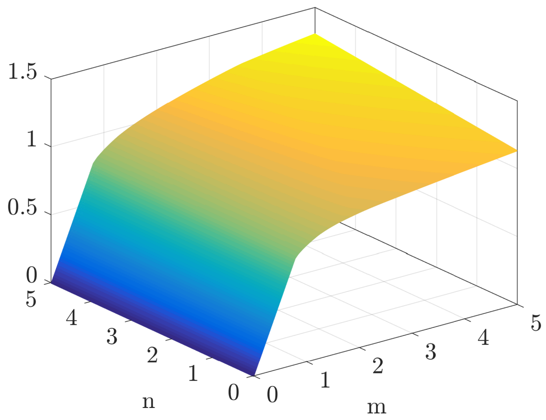
(b) Optimal sections, $t = T - 19$.



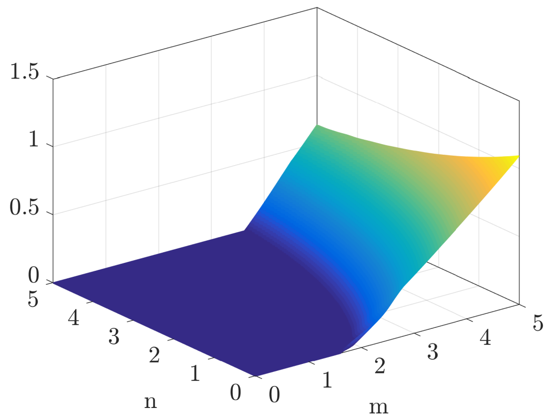
Notes: Figure A.1 illustrates which q_t-segments (*ucon*, *con*, *dcon* and *acon*) are optimal in the (m,n)-space for $t = T - 5$ and $t = T - 19$. Solved using G^{2}EGM, $\#_m = 600$.

Figure A.2: Policy functions - *with smoothing* ($\sigma_\varepsilon = 0.1$ and $\sigma_\eta^2 = 0.1$).

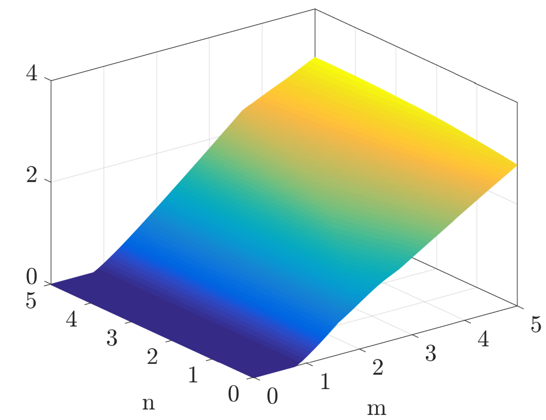
(a) Consumption, c_{T-19}.



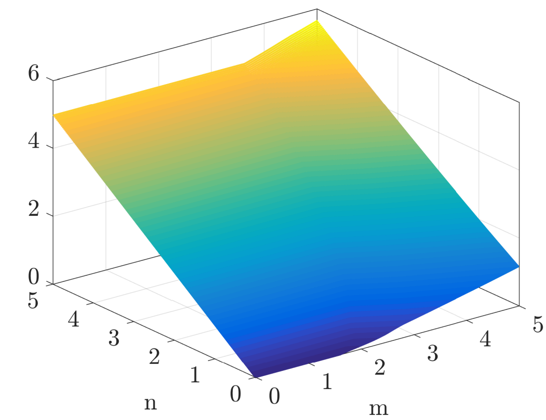
(b) Pension deposits, d_{T-19}.



(c) Assets, a_{T-19}.



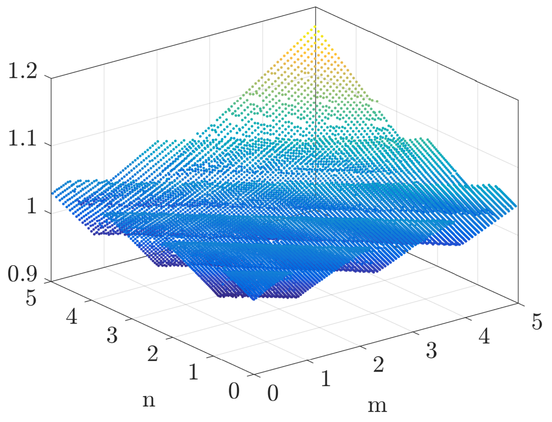
(d) Pension assets, b_{T-19}.



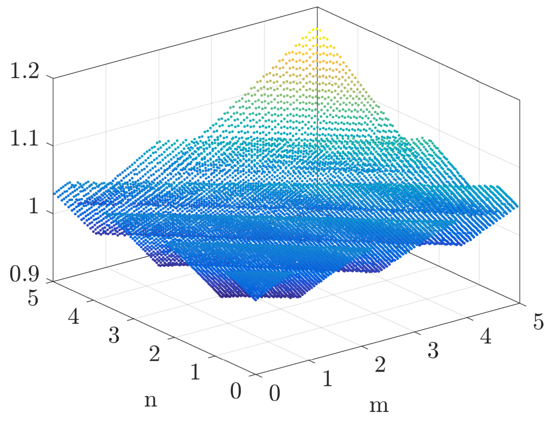
Notes: Figure A.2 shows optimal policy functions for working households. Solved using G^{2}EGM, $\#_m = 600$.

Figure A.3: Policy functions - 3-dimension model. G^{2}EGM and VFI.

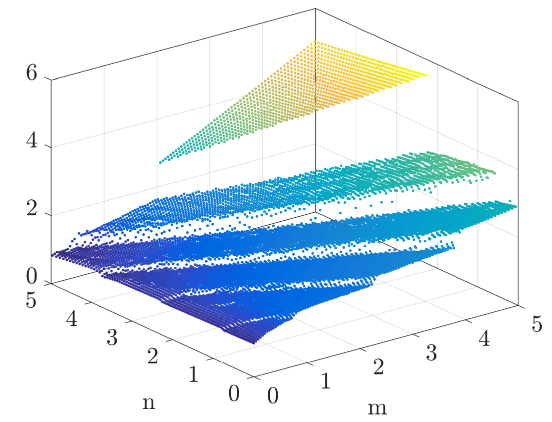
(a) Consumption, c_{T-19}, G^{2}EGM.



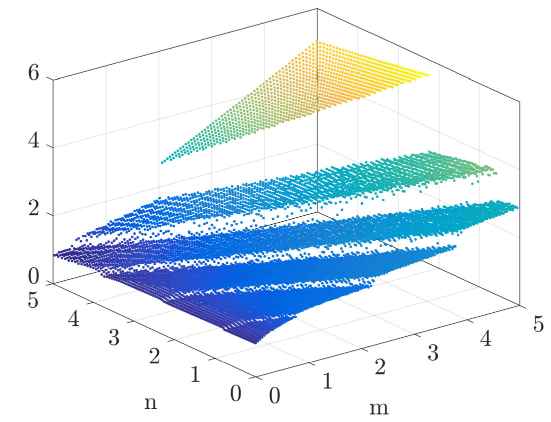
(b) Consumption, c_{T-19}, VFI.



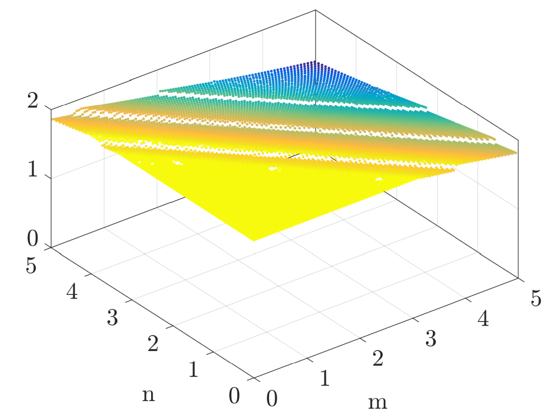
(c) Deposits, d_{T-19}, G^{2}EGM.



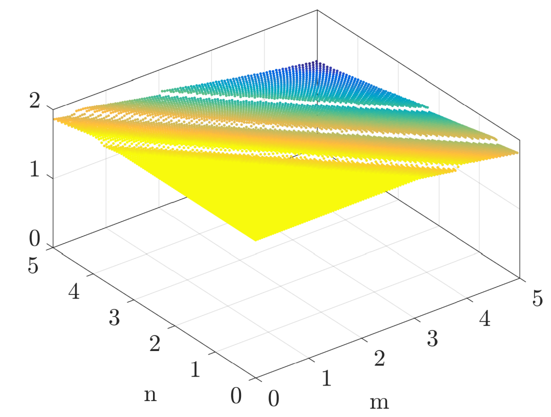
(d) Deposits, d_{T-19}, VFI.



(e) Labor, l_{T-19}, G^{2}EGM.



(f) Labor, l_{T-19}, VFI.



Notes: Figure A.3 shows optimal policy functions for working households. Solved using G^{2}EGM and VFI, $\#_m = 150$. For $k \approx 20$.

B Details on the Illustrative Model

B.1 Value Function Derivatives, $v_{m,t}(1, m_t, n_t)$ and $v_{n,t}(1, m_t, n_t)$

B.1.1 $v_{m,t}(1, m_t, n_t)$

To determine the derivative of $v_t(1, m_t, n_t)$ wrt. m_t note that

$$\begin{aligned} v_{m,t}(1, m_t, n_t) &= c^{\star}(m_t, n_t)^{-\rho} c^{\star}_m(m_t, n_t) \\ &\quad + \beta w_{a,t}(a_t, b_t)(1 - c^{\star}_m(m_t, n_t) - d^{\star}_m(m_t, n_t)) \\ &\quad + \beta w_{b,t}(a_t, b_t)(1 + g(d^{\star}(m_t, n_t)) d^{\star}_m(m_t, n_t) \end{aligned}$$

We need to consider three different cases:

1. If $c^{\star}(m_t, n_t) < m_t - d^{\star}(m_t, n_t)$ and $d^{\star}(m_t, n_t) > 0$ we can use a standard envelope argument to show that $c^{\star}_m(m_t, n_t) = d^{\star}_m(m_t, n_t) = 0$. We thus have

$$\begin{aligned} v_{m,t}(1, m_t, n_t) &= \beta w_{a,t}(a_t, b_t) \\ &= c^{\star}(m_t, n_t)^{-\rho} \end{aligned}$$

   where the second equality is due to the FOC in equation (3.6).

2. If $c^{\star}(m_t, n_t) = m_t - d^{\star}(m_t, n_t)$ and $d^{\star}(m_t, n_t) > 0$ we can still use an envelope argument to show $d^{\star}_m(m_t, n_t) = 0$. Furthermore we directly have $c^{\star}_m(m_t, n_t) = 1$ such that

$$v_{m,t}(1, m_t, n_t) = c^{\star}(m_t, n_t)^{-\rho}$$

3. If $c^{\star}(m_t, n_t) = m_t - d^{\star}(m_t, n_t)$ and $d^{\star}(m_t, n_t) = 0$ we directly have $c^{\star}_m(m_t, n_t) = 1$ and $d^{\star}_m(m_t, n_t) = 0$ such that

$$v_{m,t}(1, m_t, n_t) = c^{\star}(m_t, n_t)^{-\rho}$$

In sum, we thus have that

$$v_{m,t}(1, m_t, n_t) = c^{\star}(m_t, n_t)^{-\rho} \tag{B.1}$$

.

B.1.2 $v_{n,t}(1, m_t, n_t)$

To determine the derivative of $v_t(1, m_t, n_t)$ wrt. n_t note that

$$
\begin{aligned}
v_{n,t}(1, m_t, n_t) &= c^\star(m_t, n_t)^{-\rho} c_n^\star(m_t, n_t) \\
&\quad + \beta w_{a,t}(a_t, b_t)(c_n^\star(m_t, n_t) - d_n^\star(m_t, n_t)) \\
&\quad + \beta w_{b,t}(a_t, b_t)(1 + (1 + g(d^\star(m_t, n_t)) d_n^\star(m_t, n_t))
\end{aligned}
$$

We need to consider three different cases:

1. If $c^\star(m_t, n_t) < m_t - d^\star(m_t, n_t)$ and $d^\star(m_t, n_t) > 0$ we can use a standard envelope argument to show that $c_n^\star(m_t, n_t) = d_n^\star(m_t, n_t) = 0$. We thus have

$$
v_{n,t}(1, m_t, n_t) = \beta w_{b,t}(a_t, b_t)
$$

2. If $c^\star(m_t, n_t) = m_t - d^\star(m_t, n_t)$ and $d^\star(m_t, n_t) > 0$ we can still use an envelope argument to show $d_n^\star(m_t, n_t) = 0$ such that we also have $c_n^\star(m_t, n_t) = 0$. We thus have

$$
v_{n,t}(1, m_t, n_t) = \beta w_{b,t}(a_t, b_t)
$$

3. If $c^\star(m_t, n_t) = m_t - d^\star(m_t, n_t)$ and $d^\star(m_t, n_t) = 0$, we directly have $d_n^\star(m_t, n_t) = 0$ such that we also have $c_n^\star(m_t, n_t) = 0$. We thus have

$$
v_{n,t}(1, m_t, n_t) = \beta w_{b,t}(a_t, b_t)
$$

In sum, we thus have that

$$
v_{n,t}(1, m_t, n_t) = \beta w_{b,t}(a_t, b_t) \tag{B.2}
$$

.

B.2 Grids

All grid-vectors are constructed as recursions where the i'th element is given by

$$
\begin{aligned}
i \geq 2: \; x_i &= x_{i-1} + \frac{\overline{x} - x_{i-1}}{(n - i + 1)^\phi} \\
x_0 &= \underline{x}
\end{aligned} \tag{B.3}
$$

Given $\#_m$, and using the relative parameters in table B.1, the common and case-specific grids are then created as follows:

- **common state grid ($\widehat{\mathcal{M}}$):**

Table B.1: Grids.

$\#_\eta$	$\#_a$	$\#_{aq}$	$\#_{qacon}$	ϕ_m	ϕ_n	$\overline{m}$	$\overline{n}$	$\overline{a}$	$\overline{b}$
8	$2 \cdot \#_m$	$2 \cdot \#_a$	$\frac{1}{4}\#_a$	1.1	1.25	10	$\overline{m}+2$	$\overline{m}-2$	$\overline{n}+2$

- $\overrightarrow{m} = \{0, \ldots \overline{m} \,|\, \#_m, \phi_m\}$,
- $\overrightarrow{n} = \{0, \ldots \overline{n} \,|\, \#_m, \phi_n\}$,
- $\widehat{\mathcal{M}} = \overrightarrow{m} \times \overrightarrow{n}$.

- **common post-decision state grid ($\widehat{A}$):**
  - $\overrightarrow{a} = \{0, \ldots \overline{a} \,|\, \#_a, \phi_n\}$,
  - $\overrightarrow{b} = \left\{0, \ldots \overline{b} \,|\, \#_a, \phi_n\right\}$,
  - $\widehat{\mathcal{A}} = \overrightarrow{a} \times \overrightarrow{b}$.

- **post-decision state grids for sections ($\widehat{\mathcal{A}}_{q_t}$):**
  - **common:** $\overrightarrow{a_q} = \{0, \ldots \overline{a} \,|\, \#_{qa}, \phi_m\}$, $\overrightarrow{b_q} = \left\{0, \ldots \overline{b} \,|\, \#_{qb}, \phi_n\right\}$.
  - **ucon:** $\widehat{\mathcal{A}}_{ucon} = \overrightarrow{a_q} \times \overrightarrow{b_q}$.
  - **con:** $\widehat{\mathcal{A}}_{con} = \widehat{\mathcal{M}}$.
  - **dcon:** $\widehat{\mathcal{A}}_{dcon} = \widehat{\mathcal{A}}_{ucon} \times \{0\}$.
  - **acon:** $\widehat{\mathcal{A}}_{acon} = \{0\} \times \left\{0, \ldots \overline{b} \,|\, \#_{qacon}, \phi_n\right\} \times \{\underline{c}(b) \ldots, \overline{c}(b) \,|\, \#_{qacon}, \phi_m\}$, $\underline{c}(b) = ((\chi+1)\beta w_{b,t}(0,b))^{-\frac{1}{\rho}}$ $\overline{c}(b) = (\beta w_{b,t}(0,b))^{-\frac{1}{\rho}}$.

B.3 Value Function Iteration

Step 0. Construct grids, find terminal period solution, and set time-index to $t = T-1$.

Step 0a. For each $(\mathbf{s}_T, z_T)$, construct:
i) a (rectilinear) grid over (pre-decision) states $\widehat{\mathcal{M}} = \{\mathbf{m}_1, \ldots, \mathbf{m}_{\#_m}\}$.
ii) a (rectilinear) grid over post-decision states $\widehat{\mathcal{A}} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{\#_a}\}$.

Step 0b. For all $(\mathbf{s}_T, z_T)$:
i) find the choice-specific terminal value- and policy functions, $v_T(\mathbf{s}_T, z_T, \mathbf{m}_T)$ and $c_T^\star(\mathbf{s}_T, z_T, \mathbf{m}_T)$, by solving the fully intra-temporal problem using standard tools.
ii) construct an interpolant for the value function $\check{v}_T\left(\mathbf{s}_T, z_T, \mathbf{m}_T; \widehat{\mathcal{M}}\right)$ (henceforth all interpolants are indicated with a $\check{\bullet}$).

Step 1. For all $(\mathbf{s}_t, z_t)$ construct an interpolant of the expected next-period value function which post-decision states as input,

$$\check{w}_t(\mathbf{s}_t, z_t, \mathbf{a}_t \widehat{\mathcal{A}}) \quad = \quad \mathbb{E}_t \left[EV_{t+1}(\mathbf{s}_t, z_t, \mathbf{m}_{t+1}) \right] \tag{B.4}$$

where $\mathbf{m}_{t+1} = \Gamma_m(\mathbf{s}_t, z_t, \mathbf{a}_t)$ and $\mathbb{E}_t[\bullet] = \mathbb{E}[\bullet \,|\, \mathbf{s}_t, z_t, \mathbf{a}_t]$ is computed using some form of numerical integration. The closed form expectations over taste shocks in (B.4) is given by the log-sum,

$$EV_{t+1}(\mathbf{s}_{t+1}, z_{t+1}, \mathbf{m}_{t+1}) \equiv \begin{cases} \sigma_\varepsilon \log \left(\sum_{z \in \mathcal{Z}(\mathbf{s}_{t+1})} \check{v}_{t+1}(\mathbf{s}_{t+1}, z, \mathbf{m}_{t+1}) / \sigma_\varepsilon \right) & \text{if } \sigma_\varepsilon > 0 \\ \max_{z \in \mathcal{Z}(\mathbf{s}_{t+1})} \check{v}_{t+1}(\mathbf{s}_{t+1}, z, \mathbf{m}_{t+1}) & \text{if } \sigma_\varepsilon = 0 \end{cases}$$

Step 2. For all $(\mathbf{s}_t, z_t)$ find the over-arching optimal continuous choices and construct the interpolant for the value function.

Step 2a. For each $\mathbf{m}_t$ in $\widehat{\mathcal{M}}$ search for the global constrained optimal choices using a numerical optimization routine,

$$\begin{aligned} v_t(\mathbf{s}_t, z_t, \mathbf{m}_t) &= \max_{\mathbf{c}} u(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}) + \beta \check{w}_t(\mathbf{s}_t, z_t, a(\mathbf{s}_t, \mathbf{m}_t, \mathbf{c})) \\ \mathbf{c}_t^\star(\mathbf{s}_t, z_t, \mathbf{m}_t) &= \arg\max_{\mathbf{c}} u(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}) + \beta \check{w}_t(\mathbf{s}_t, z_t, a(\mathbf{s}_t, \mathbf{m}_t, \mathbf{c})) \end{aligned}$$

where unfeasible choices are not considered.

Step 2c. Terminate if $t = 1$, else:

i) construct the interpolant $\check{v}_t \left(\mathbf{s}_t, z_t, \mathbf{m}_t; \widehat{\mathcal{M}} \right)$.

ii) decrease the time index to $t = t - 1$, and return to step 1.

B.4 Relation to the General Model Class

In terms of the defined broad model class, the illustrative model is given by

$$
\begin{aligned}
\mathbf{s}_t &= z_{t-1} \in \{0,1\} = \mathcal{S} \\
z_t &= z_t \in \begin{cases} \{0,1\} & \text{if } z_{t-1} = 1 \\ 0 & \text{if } z_{t-1} = 0 \end{cases} = \mathcal{Z}(\mathbf{s}_t) \\
\mathbf{m}_t &= \begin{bmatrix} m_t \\ n_t \end{bmatrix} \in \mathbb{R}^2_+ = \mathcal{M}(\mathbf{s}_t) \\
\mathbf{a}_t &= \begin{bmatrix} a_t \\ b_t \end{bmatrix} \in \mathbb{R}^2_+ \\
\mathbf{c}_t &= \begin{bmatrix} c_t \\ d_t \end{bmatrix} \in \{c_t, d_t \,|\, c_t, d_t \geq 0,\ c_t + d_t \leq m_t\} = \overline{\mathcal{C}}(\mathbf{s}_t, z_t, \mathbf{m}_t) \\
\Gamma_s(\mathbf{s}_t, z_t) &= z_t \\
\Gamma_m(\mathbf{s}_t, z_t, \mathbf{m}_t, \mathbf{c}_t) &= \begin{bmatrix} R_a a_t + \eta_{t+1} \\ R_b b_t \end{bmatrix}
\end{aligned}
$$

and we have the four q_t-cases *uncon*, *con*, *dcon* and *acon*. The derivative of the continuous choices with respect to the unconstrained choices are respectively

$$
\begin{aligned}
c_{c^+}\left(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+, q_t = 0\right) &= \begin{bmatrix} \frac{\partial c_t}{\partial c_t} & \frac{\partial c_t}{\partial d_t} \\ \frac{\partial d_t}{\partial c_t} & \frac{\partial d_t}{\partial d_t} \end{bmatrix}_{d_t > 0, c_t + d_t < m_t} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
c_{c^+}\left(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+, q_t = 1\right) &= \text{"}\emptyset\text{"} \\
c_{c^+}\left(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+, q_t = 2\right) &= \begin{bmatrix} \frac{\partial c_t}{\partial c_t} \\ \frac{\partial d_t}{\partial c_t} \end{bmatrix}_{d_t = 0, c_t + d_t < m_t} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
c_{c^+}\left(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+, q_t = 3\right) &= \begin{bmatrix} \frac{\partial c_t}{\partial d_t} \\ \frac{\partial d_t}{\partial d_t} \end{bmatrix}_{d_t > 0, c_t + d_t = m_t} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}
\end{aligned}
$$

where $c_{c^+}\left(\mathbf{m}_t, \mathbf{c}_t^+, q = 2\right) = \text{"}\emptyset\text{"}$ indicates that in this case there are no unconstrained choices.

The derivative of the post-decision state function is

$$
a_c\left(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t\right) = \begin{bmatrix} \frac{\partial a_t}{\partial c_t} & \frac{\partial a_t}{\partial d_t} \\ \frac{\partial b_t}{\partial c_t} & \frac{\partial b_t}{\partial d_t} \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 0 & 1 + g_d(d_t) \end{bmatrix}
$$

and we get the FOCs:

$$
\begin{aligned}
f(\bullet, q=0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}' &= \begin{bmatrix} \frac{\partial u}{\partial c_t} \\ \frac{\partial u}{\partial d_t} \end{bmatrix}' \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \beta \begin{bmatrix} w_{a,t}(a_t, b_t) \\ w_{b,t}(a_t, b_t) \end{bmatrix}' \begin{bmatrix} -1 & -1 \\ 0 & 1 + g_d(d_t) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&\Updownarrow \\
\begin{bmatrix} 0 \\ 0 \end{bmatrix}' &= \begin{bmatrix} u_c(c_t) \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -w_{a,t}(a_t, b_t) \\ -w_{a,t}(a_t, b_t) + w_{b,t}(a_t, b_t)(1 + g_d(d_t)) \end{bmatrix}
\end{aligned}
$$

$$
f(\bullet, q=1) = \text{"}\emptyset\text{"}
$$

$$
\begin{aligned}
f(\bullet, q=2) = 0 &= \begin{bmatrix} u_c(c_t) \\ 0 \end{bmatrix}' \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} w_{a,t}(a_t, b_t) \\ w_{b,t}(a_t, b_t) \end{bmatrix}' \begin{bmatrix} -1 & -1 \\ 0 & 1 + g_d(d_t) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&\Updownarrow \\
0 &= u_c(c_t) + \beta w_{a,t+1}(a_t, b_t)
\end{aligned}
$$

$$
\begin{aligned}
f(\bullet, q=3) = 0 &= \begin{bmatrix} u_c(c_t) \\ 0 \end{bmatrix}' \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} w_{a,t}(a_t, b_t) \\ w_{b,t}(a_t, b_t) \end{bmatrix}' \begin{bmatrix} -1 & -1 \\ 0 & 1 + g_d(d_t) \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\
&\Updownarrow \\
0 &= -u_c(c_t) + \beta w_{b,t}(a_t, b_t)(1 + g_d(d_t))
\end{aligned}
$$

B.5 Minor Extension: A Collateral Constraint

Our solution method can also handle *collateral constraints*. Assume that we have the collateral constraint

$$
a_t \geq \theta b_t \tag{B.5}
$$

Choosing c_t as the constrained choice, this can be re-written as

$$
\begin{aligned}
a_t &= \theta b_t \leftrightarrow \\
m_t - c_t - d_t &= \theta n_t + \theta(d_t + g(d_t)) \leftrightarrow \\
c_t &= m_t - \theta n_t - (d_t + \theta(d_t + g(d_t))) \\
&= \check{c}\left(\mathbf{x}_t, \mathbf{m}_t, \mathbf{c}_t^+\right)
\end{aligned}
$$

The first order condition for d_t then is

$$
u_c(c_t) \frac{\partial c_t}{\partial d_t} + \beta w_{b,t}(a_t, b_t)(1 + g_d(d_t)) = 0 \tag{B.6}
$$

with

$$
\frac{\partial c_t}{\partial d_t} = -(1 + \theta(1 + g_d(d_t)))
$$

Fixing b_t and c_t and setting $a_t = \theta b_t$, we can find d_t from (B.6), and hereafter use the budget constraints to find $m_t = a_t + c_t + d_t = c_t + d_t$ and $n_t = b_t - d_t - g(d_t)$.

B.6 Extension with Labor Supply and Human Capital

The new choice-specific value function for the working household is now given by

$$\begin{aligned}
v_t(1, m_t, n_t, k_t) &= \max_{c_t, d_t, l_t} u(c_t, 1, l_t) + \beta \underbrace{\int_\eta EV_{t+1}(1, m_{t+1}, n_{t+1}, k_{t+1}) G(d\eta)}_{\equiv w_t(a_t, b_t, q_t)} \\
& \text{s.t.} \\
a_t &= m_t + r_k k_t l_t - c_t - d_t \\
b_t &= n_t + d_t + g(d_t) \\
q_t &= (1-\delta)k_t + l_t \\
m_{t+1} &= R_a a_t \\
n_{t+1} &= R_b b_t \\
k_{t+1} &= \eta_{t+1} q_t \\
l_t &\in [0, \bar{l}] \\
c_t &\geq 0 \\
d_t &\geq 0 \\
c_t + d_t &\in [0, m_t + r_k k_t l_t]
\end{aligned}$$

where q_t is the post-decision state associated with end-of-period human capital after depreciation of existing human capital, but including current labor supply. For simplicity, we require that the households retire in the terminal period choosing $l_t = 0$.

The problem for the retired households are unchanged, but to strictly bound the marginal value of human capital from below, we assume that retiring households receive a one time payment equal to 5 percent of their human capital.

In order to ensure that the optimal choices are not always constrained, we had to change some of the baseline parameters. We did this to ensure that the speed and accuracy measures was not an artifact of the parametrization leading to an (uninteresting) special version of the model. The full set of parameters are given in table B.2.

Table B.2: Baseline Parameter Values.

R_a	R_b	β	ρ	α	χ	$\underline{y}$	γ	φ	δ	r_k	$\bar{l}$	σ_η^2	σ_ε
1.02	1.025	0.98	2.00	0.25	0.10	0.50	1.0	0.6	0.1	0.05	2.00	0.00	0.00

B.6.1 Segments and FOC for l_t

The choice of l_t is always above zero (as $\lim_{l_t \to 0} u_l = 0$ and $u_c > 0$), and we thus have the same segments as in the baseline model plus copies of these with $l_t = \bar{l}$.

In the *unconstrained* case (**ucon**), the FOCs related to consumption and pension deposits are unchanged, and the FOC wrt. labor supply, l_t, is given by

$$l_t = \frac{\frac{\beta r_k}{1-\delta} q_t w_{a,t}(a_t, b_t, q_t) + \beta w_{q,t}(a_t, b_t, q_t)}{\frac{\beta r_k}{1-\delta} w_{a,t}(a_t, b_t, q_t) + \varphi}$$

For the constrained case **(con)**, we have that $a_t = 0 \leftrightarrow c_t = m_t + r_k k_t l_t$ and $d_t = 0$ such that the FOC for the unconstrained choice of labor supply, l_t, is

$$l_t = \frac{\frac{r_k}{1-\delta} q_t c_t^{-\rho} + \beta w_{q,t}(a_t, b_t, q_t)}{\frac{r_k}{1-\delta} c_t^{-\rho} + \varphi}$$

For the segment where only pension deposits are constrained, $d_t = 0$, (**dcon**) the optimal consumption is found as in the baseline model, and l_t is found as in the *unconstrained* segment above. Finally, for the segment where only end-of-period assets are constrained, $a_t = 0$ (**acon**), the optimal consumption choice is found as in the baseline model, and l_t is found as in the *constrained* segment above.

B.6.2 Value Function Derivatives

The derivatives of the value function wrt. m_t and n_t are unchanged as q_t is independent of m_t, n_t, c_t and d_t, and the choice of l_t is always either unconstrained or equal to a constant such that

$$l_m^\star(m_t, n_t, k_t) = l_n^\star(m_t, n_t, k_t) = l_k^\star(m_t, n_t, k_t) = 0$$

To determine the derivate of $v_t(1, m_t, n_t, k_t)$ wrt. k_t note that

$$\begin{aligned}
v_{k,t}(1, m_t, n_t, k_t) = \; & c^\star(m_t, n_t)^{-\rho} c_k^\star(m_t, n_t, k_t) \\
& + \beta w_{a,t}(a_t, b_t, q_t) \\
& (r_k l^\star(m_t, n_t, k_t) + r_k k_t l_k^\star(m_t, n_t, k_t) - c_k^\star(m_t, n_t, k_t) - d_k^\star(m_t, n_t, k_t)) \\
& + \beta w_{b,t}(a_t, b_t, q_t)(1 + g(d^\star(m_t, n_t)) d_k^\star(m_t, n_t) \\
& + \beta w_{q,t}(a_t, b_t, q_t)((1-\delta) + l_k^\star(m_t, n_t, k_t))
\end{aligned}$$

This always reduces to

$$v_{k,t}(1, m_t, n_t) = c^\star(m_t, n_t)^{-\rho} r_k l_k^\star(m_t, n_t, k_t) + \beta(1-\delta) w_{q,t}(a_t, b_t, q_t) \tag{B.7}$$

C Deriving necessary FOCs

This appendix presents methods for proving that FOCs (and Euler equations) are *necessary* at optimal interior continuous choices even in the presence of discrete choices which imply that the value function is neither necessarily globally *concave* nor globally *differentiable*. The methodology builds upon the theoretical results in Clausen and Strub (2013) which we begin by recapping in brief. Hereafter we focus on a specific model including the illustrative model from the main text as a special case.

C.1 Existing lemmas from Clausen and Strub (2013)

Following Clausen and Strub (2013), we define differentiable *lower* and *upper* support functions as follows:

Definition C.1. We say that $L : X \to \mathbb{R}$ is a *differentiable* *lower* *support function* for $F : X \to \mathbb{R}$ at $\widehat{x} \in \text{int}(X)$ if

$$
\begin{aligned}
\forall x \in X : \quad & L \text{ is differentiable} && \text{(C.1)}\\
& L(x) \leq F(x) && \text{(C.2)}\\
x = \widehat{x} : \quad & L(x) = F(x) && \text{(C.3)}
\end{aligned}
$$

Definition C.2. We say that $U : X \to \mathbb{R}$ is a *differentiable* *upper* *support function* for $F : X \to \mathbb{R}$ at $\widehat{x} \in \text{int}(X)$ if

$$
\begin{aligned}
\forall x \in X : \quad & U \text{ is differentiable} && \text{(C.4)}\\
& U(x) \geq F(x) && \text{(C.5)}\\
x = \widehat{x} : \quad & U(x) = F(x) && \text{(C.6)}
\end{aligned}
$$

This leads us to the following definition:

Definition C.3. We say that $F : X \to \mathbb{R}$ is *differentiable sandwiched* between L and U if they are respectively differentiable *lower* and *upper* support functions.

Clausen and Strub (2013) then prove the following lemma:

Lemma C.1. (Differentiable Sandwich Lemma). *If $F : X \to \mathbb{R}$ is differentiable sandwiched between L and U at $\widehat{x}$ for an $\mathcal{X} \subseteq X$ with $\widehat{x} \in int(\mathcal{X})$ then F is differentiable at $\widehat{x}$ with*

$$F'(\widehat{x}) = L'(\widehat{x}) = U'(\widehat{x}) \tag{C.7}$$

They also further prove a a *maximum lemma*:

Lemma C.2. (Maximum Lemma). *Let $\phi : X \to \mathbb{R}$ be a function. If $\widehat{x} \in int(X)$ maximizes ϕ then the constant $U(x) = \phi(\hat{x})$ is a differential upper support function for ϕ at $\widehat{x}$.*

and a so-called *reverse calculus lemma*:

Lemma C.3. (Reverse Calculus). *Suppose $F : X \to \mathbb{R}$ and $G : X \to \mathbb{R}$ have differentiable lower support functions at $\widehat{x}$ then*

1. *If $H(x) = F(x) + G(x)$ is differentiable at $\widehat{x}$, then F is differentiable at $\widehat{x}$.*
2. *If $H(x) = F(x)G(x)$ is differentiable at $\widehat{x}$ and $F(\widehat{x}) > 0$ and $G(\widehat{x}) > 0$, then F is differentiable at $\widehat{x}$.*
3. *If $H(x) = \max\{F(x), G(x)\}$ is differentiable at $\widehat{x}$ and $F(\widehat{x}) = H(\widehat{x})$ then F is differentiable at $\widehat{x}$.*

C.2 New lemma

An almost immediate implication of the reverse calculus lemma C.3 is

Lemma C.4. *Suppose $H(x)$ is differentiable at $\widehat{x}$ and*

$$H(x) = \max_{z\in\mathcal{Z}} \{J(x,z)\} \tag{C.8}$$

$$J(x,z) = G(x,z) + \beta \sum_{k=1}^{K} \pi_k F(x,k,z), \quad \beta, \pi_k \in \mathbb{R} \tag{C.9}$$

where $\mathcal{Z}$ is a finite set. If $F : X \times \{1,2,\ldots,K\} \times \mathcal{Z} \to \mathbb{R}$ and $G : X \times \mathcal{Z} \to \mathbb{R}$ have lower support functions at $\widehat{x}$ for all $k \in \{1,2,\ldots,K\}$ and all z, then $F(x,k,z)$ is differentiable at $\widehat{x}$ for all $k \in \{1,2,\ldots,K\}$ when $z = z^\star(x) = arg\max_{z\in\mathcal{Z}} J(x,z)$.

Proof. Defining

$$\begin{aligned}
J^\star(x) &\equiv J(x, z^\star(x)) && \text{(C.10)}\\
&= G(x, z^\star(x)) + \overline{F}(x) && \text{(C.11)}\\
\overline{F}(x) &\equiv \beta \sum_{k=1}^{K} \pi_k F(x,k,z^\star(x)) && \text{(C.12)}\\
&= \beta \tilde{F}(x) && \text{(C.13)}\\
\tilde{F}(x) &= \sum_{k=10}^{K} \pi_k F(x,k,z^\star(x)) && \text{(C.14)}\\
&= \sum_{k=1}^{K} \tilde{F}_k(x) && \text{(C.15)}\\
\tilde{F}_k(x) &\equiv \pi_k F(x,k,z^\star(x)) && \text{(C.16)}
\end{aligned}$$

and noting that we from the lower support functions for F and G can construct lower support functions for $\tilde{F}_k$, $\tilde{F}$, $\overline{F}$, $J^\star$ and J, we have

1. $J^\star(x)$ in (C.10) is differentiable at $\hat{x}$ due to (3) in lemma C.3.

2. $\overline{F}(x)$ in (C.11) is then differentiable at $\widehat{x}$ due to (1) in lemma C.3.

3. $\tilde{F}(x)$ in (C.13) is then differentiable at $\hat{x}$ due to (2) in lemma C.3.

4. $\tilde{F}_k(x)$ in (C.15) is then differentiable at $\widehat{x}$ (for all k) due to (1) in lemma C.3.

5. $F(x, k, z^\star(x))$ in (C.16) is then differentiable at $\widehat{x}$ (for all k) due to (2) in lemma C.3.

$\square$

C.3 Model

We consider household decision problems on the following form

$$
\begin{aligned}
v_t(s_t, \mathbb{S}_t) &= \max_{z_t, a_t, b_t} u(c_t) + \mathbb{E}_t\left[\beta v_{t+1}(s_{t+1}, \mathbb{S}_{t+1})\right] \\
&\text{s.t.} \\
c_t &= c(m_t, n_t, a_t, b_t) \\
d_t &= d(n_t, b_t) \\
s_{t+1} &= \Gamma_s(s_t, z_t) \\
m_{t+1} &= \Gamma_m(a_t, s_{t+1}) \\
n_{t+1} &= \Gamma_n(b_t, s_{t+1}) \\
z_t &\in \mathcal{Z}(s_t) \\
a_t &\in [0, m_t] \subseteq \mathbb{R}_+ \\
b_t &\in \left[n_t, \bar{b}(n_t, m_t)\right] \subseteq \mathbb{R}_+ \\
c_t &\in [0, m_t] \\
d_t &\in [0, m_t]
\end{aligned}
\tag{C.17}
$$

where $\mathbb{S}_t = (m_t, n_t)$ and

- **Discrete elements:** s_t is a vector of discrete states in the finite set $\mathcal{S}$, z_t is a vector of discrete choices in the finite set $\mathcal{Z}(s_t)$, and $\Gamma_s : \mathcal{S} \times \mathcal{Z} \to \mathcal{S}$ is the (stochastic) law of motion.

- **Liquid resources:** m_t is pre-decision liquid resources, a_t is post-decision liquid ressources, and $\Gamma_m : \mathbb{R}_+ \times \mathcal{S} \to \mathbb{R}_+$ is the continuous and differentiable (stochastic) law of motion. We use the notation

$$\Gamma'_{t,m,a} \equiv \frac{\partial \Gamma_m (a_t, b_t, s_{t+1})}{\partial a_t}$$

- **Illiquid resources:** n_t is pre-decision illiquid resources, b_t is post-decision illiquid resources, and $\Gamma_n : \mathbb{R}_+ \times \mathcal{S} \to \mathbb{R}_+$ is the continuous and differentiable (stochastic) law of motion. We use the notation

$$\Gamma'_{t,n,b} \equiv \frac{\partial \Gamma_n (b_t, s_{t+1})}{\partial b_t}$$

- **Utility:** $u : \mathbb{R}^2 \to \mathbb{R}$ is the continuous and twice differentiable utility function satisfying

$$u'_{t,c} \equiv \frac{\partial u (c_t)}{\partial c_t} > 0, \lim_{c_t \downarrow 0} u'_c = \infty$$

$d : \mathbb{R}^2 \to \mathbb{R}$ is the implied continuous and differentiable deposit function where

$$\begin{aligned} \lim_{b_t \to n_t} d (n_t, b_t) &= 0 \\ \lim_{b_t \to \bar{b}(n_t, m_t)} d (n_t, b_t) &= m_t \end{aligned}$$

$c : \mathbb{R}^4 \to \mathbb{R}$ is the implied continuous and differentiable consumption function where

$$\begin{aligned} \lim_{a_t \to m_t} c (m_t, n_t, a_t, b_t) &= 0 \\ \lim_{a_t \to 0, d_t \to 0} c (m_t, n_t, a_t, b_t) &= m_t \\ \lim_{d_t \to m_t} c (m_t, n_t, a_t, b_t) &= 0 \\ c'_{t,m} &\equiv \frac{\partial c (m_t, n_t, a_t, b_t)}{\partial m_t} > 0 \\ c'_{t,n} &\equiv \frac{\partial c (m_t, n_t, a_t, b_t)}{\partial n_t} > 0 \\ c'_{t,a} &\equiv \frac{\partial c (m_t, n_t, a_t, b_t)}{\partial a_t} < 0 \\ c'_{t,b} &\equiv \frac{\partial c (m_t, n_t, a_t, b_t)}{\partial b_t} < 0 \end{aligned}$$

- **Choice set:** $\bar{b} : \mathbb{R}^4 \to \mathbb{R}$ is a continuous and differentiable function.

The *terminal* value function is given by

$$v_T(s_T, \mathbb{S}_T) = u(m_T) \tag{C.18}$$

We denote the optimal *discrete choice* by $z_t^\star = z_t^\star(s_t, \mathbb{S}_t)$, and the *conditional* optimal *continuous choices* as $a_t^\star = a_t^\star(s_t,\mathbb{S}_t, z_t^\star)$ and $b_t^\star = b_t^\star(s_t,\mathbb{S}_t, z_t^\star)$. We also construct the optimal *consumption function* as

$$c_t^\star \quad = \quad c(m_t, n_t, a_t^\star, b_t^\star) \tag{C.19}$$

We assume that all the choice functions exist and are unique for $\forall s_t \in \mathcal{S}$, $\forall n_t \geq 0$ and $m_t > 0$.

Lemma C.5. *The choices of a_t and b_t are never "upward constrained", i.e.*

$$\begin{array}{c} a_t < m_t \\ b_t < \overline{b}(n_t, m_t) \end{array} \quad \rightarrow \quad c_t > 0$$

Proof. This is a consequence of $\lim_{c_t \downarrow 0} u'_c = \infty$ and $c_t \downarrow 0$ as $a_t \uparrow m_t$ or $b_t \uparrow \overline{b}(n_t, m_t)$. □

C.4 *Lazy-schizophrenic* households

Definition C.4. We say that a household is *lazy-schizophrenic* at $\breve{s}_0$ around the initial continuous state $\breve{\mathbb{S}}_0 = (\breve{m}_0, \breve{n}_0)$ with certain implied (stochastic) paths of optimal choices $(\breve{z}_k,\breve{a}_k,\breve{b}_k, \breve{c}_k, \breve{d}_k)$ and states $(\breve{s}_k,\breve{\mathbb{S}}_k)$ if it in the current and all future periods always, even if $(m_{t+k}, n_{t+k}) \neq \breve{\mathbb{S}}_k$, chooses $z_{t+k} = \breve{z}_k$ and (a_{t+k}, b_{t+k}) "as close as possible" to $\left(\breve{a}_k, \breve{b}_k\right)$.

Lemma C.6. *If $\breve{b}_0 > \breve{n}_0$ a differentiable lazy-schizophrenic value function is*

$$\begin{aligned} L_t\left(\mathbb{S}_t; \breve{s}_0, \breve{\mathbb{S}}_0\right) \quad = \quad & u\left(c\left(m_t, n_t, \breve{a}_0, \breve{b}_0\right)\right) \\ & +\beta \cdot \mathbb{E}_t\left[u\left(c\left(\breve{m}_1, \breve{n}_1, \breve{a}_1, \breve{b}_1\right)\right)\right] \\ & +\beta^2 \cdot \mathbb{E}_t\left[u\left(c\left(\breve{m}_2, \breve{n}_2, \breve{a}_2, \breve{b}_2\right)\right)\right] \\ & \ldots \end{aligned}$$

where

$$\begin{aligned} n_t \quad &= \quad \breve{n}_0 + \Delta_n \\ m_t \quad &= \quad \breve{m}_0 + \Delta_m \end{aligned}$$

and with

$$\begin{aligned} L'_{t,m} &= c'_{t,m}u'_{t,c} \\ L'_{t,n} &= c'_{t,n}u'_{t,c} \end{aligned}$$

Proof. For small enough $\Delta \equiv (\Delta_m, \Delta_n)$, consider the following behavior:

1. Choose $a_t = \breve{a}_0$ and $b_t = \breve{q}_0$. This implies:

   (a) $c_t = c\left(\breve{m}_0 + \Delta_m, \breve{n}_0 + \Delta_n, \breve{a}_0, \breve{b}_0\right)$, which is feasible as $\breve{c}_0 > 0$.

   (b) $d_t = d\left(\breve{m}_0 + \Delta_n, \breve{b}_0\right)$, which is feasible as $\breve{d}_0 > 0$ because $\breve{b}_0 > \breve{n}_0$.

   (c) $m_{t+1} = \Gamma_m\left(\breve{a}_0, \breve{b}_1, \breve{s}_1\right) = \breve{m}_1$.

   (d) $n_{t+1} = \Gamma_n\left(\breve{b}_1, \breve{s}_1\right) = \breve{n}_1$.

2. Back on known path.

The differentiability and derivatives can be proven using

$$\begin{aligned} L'_{t,m} &= \lim_{\Delta_m \to 0} \frac{L\left(\left(\breve{m}_0 + \Delta_m, \breve{n}_0\right); \breve{s}_0, \breve{\mathbb{S}}\right)}{\Delta_m} \\ L'_{t,n} &= \lim_{\Delta_n \to 0} \frac{L\left(\left(\breve{m}_0, \breve{n}_0 + \Delta_n\right); \breve{s}_0, \breve{\mathbb{S}}_0\right)}{\Delta_n} \end{aligned}$$

□

Lemma C.7. *If $\breve{b}_0 = \breve{n}_0$ (i.e. $\breve{d}_0$) a differentiable lazy-schizophrenic value function is*

$$\begin{aligned} L_t\left(\mathbb{S}_t; s_0, \breve{\mathbb{S}}_0\right) &= u\left(c\left(m_t, n_t, \breve{a}_1, b_t\right), 0\right) \\ &+\beta\mathbb{E}_t\left[\begin{cases} u\left(c\left(m_t, n_t, \breve{a}_1, \breve{b}_1\right)\right) & \text{if } \breve{b}_1 > \breve{n}_1 \\ u\left(c\left(m_t, n_t, \breve{a}_1, b_t\right), 0\right) & \text{else} \end{cases}\right] \\ &+\beta^2\mathbb{E}_t\left[\begin{cases} u\left(\breve{m}_2, \breve{n}_2, \breve{a}_2, \breve{b}_2\right) & \text{if } I_{t+1} = 0 \\ u\left(c\left(m_2, n_2, \breve{a}_2, \breve{b}_2\right)\right) & \text{if } I_{t+1} = 1 \\ u\left(c\left(m_2, n_2, \breve{a}_2, b_2\right), 0\right) & \text{if } I_{t+1} = 2 \end{cases}\right] \\ &+\beta^3\mathbb{E}_t\left[\begin{cases} u\left(\breve{m}_3, \breve{n}_3, \breve{a}_3, \breve{b}_3\right) & \text{if } I_{t+1} = 0 \\ u\left(c\left(m_3, n_3, \breve{a}_3, \breve{b}_3\right)\right) & \text{if } I_{t+1} = 1 \\ u\left(c\left(m_3, n_3, \breve{a}_3, b_3\right), 0\right) & \text{if } I_{t+1} = 2 \end{cases}\right] \\ &\ldots \end{aligned}$$

where

$$\begin{aligned}
k>1:\ I_{t+k} &= \begin{cases}0 & \textit{if } \exists j<k:\ \breve{b}_j > \breve{n}_j\\ 1 & \textit{if } \breve{b}_k > \breve{n}_k\\ 2 & \textit{else}\end{cases}\\
m_t &= \breve{m}_0 + \Delta_m\\
n_t &= \breve{n}_0 + \Delta_n\\
b_t &= n_t\\
m_{t+1} &= \Gamma_m\left(\breve{a}_0, b_t, \breve{s}_1\right)\\
n_{t+1} &= \Gamma_n\left(b_t\right)\\
b_{t+1} &= n_{t+1}\\
m_{t+2} &= \Gamma_m\left(\breve{a}_1, b_{t+1}, \breve{s}_2\right)\\
n_{t+2} &= \Gamma_n\left(b_{t+1}\right)\\
&\ldots
\end{aligned}$$

and with

$$\begin{aligned}
L'_{t,m} &= c'_{t,m}u'_{t,c}\\
L'_{t,n} &= \textit{complex}
\end{aligned}$$

Proof. For simplicity we assume $\breve{b}_1 = \breve{n}_1$ and $\breve{b}_2 > \breve{n}_2$, but generalizing the proof is straightforward. For small enough $\Delta \equiv (\Delta_m, \Delta_n)$, consider the following behavior:

1. Choose $a_t = \breve{a}_0$ and $b_t = \breve{n}_0 + \Delta_n$. This implies:
   (a) $c_t = c\left(\breve{m}_0 + \Delta_m, \breve{n}_0 + \Delta_n, \breve{a}_0, \breve{n}_0 + \Delta_n\right)$, which is feasible as $\breve{c}_0 > 0$.
   (b) $d_t = 0$, which is always feasible.
   (c) $m_{t+1} = \Gamma_m\left(\breve{a}_0, b_t, \breve{s}_1\right)$.
   (d) $n_{t+1} = \Gamma_n\left(b_t\right)$
2. Choose $a_{t+1} = \breve{a}_1$ and $b_{t+1} = n_{t+1}$. This implies:
   (a) $c_{t+1} = c\left(m_{t+1}, n_{t+1}\ ,\breve{a}_1, b_t\right)$, which is feasible as $\breve{c}_1 > 0$.
   (b) $d_{t+1} = 0$, which is always feasible.
   (c) $m_{t+2} = \Gamma_m\left(\breve{a}_1, b_{t+1}, \breve{s}_1\right)$.
   (d) $n_{t+2} = \Gamma_h\left(b_{t+1}\right)$.
3. Choose $a_{t+2} = \breve{a}_2$ and $b_{t+1} = \breve{b}_2$.

(a) $c_{t+2} = c\left(m_{t+2}, n_{t+2}, \breve{a}_2, \breve{b}_2\right)$, which is feasible as $\breve{c}_2 > 0$.

(b) $d_{t+2} = d\left(n_{t+2}, \breve{b}_2\right)$, which is feasible as $\breve{d}_2 > 0$.

(c) $m_{t+3} = \Gamma_m\left(\breve{a}_2, \breve{b}_2, \breve{s}_2\right) = \breve{m}_3$.

(d) $n_{t+3} = \Gamma_h\left(\breve{b}_2\right) = \breve{n}_3$.

4. Back on known path.

The differentiability and derivatives can be proven using

$$\begin{aligned}L'_{t,m} &= \lim_{\Delta_m \to 0} \frac{L\left((\breve{m}_0 + \Delta_m, \breve{n}_0)\,; \breve{s}_0, \breve{\mathbb{S}}\right)}{\Delta_m} \\ L'_{t,n} &= \lim_{\Delta_n \to 0} \frac{L\left((\breve{m}_0, \breve{n}_0 + \Delta_n)\,; \breve{s}_0, \breve{\mathbb{S}}_0\right)}{\Delta_n}\end{aligned}$$

□

Lemma C.8. *The lazy-schizophrenic value function $L_t\left(\mathbb{S}_t; s_t, \breve{\mathbb{S}}_0\right)$ is a differentiable lower support function for $v_t\left(s_t, \mathbb{S}_t\right)$ at $\breve{\mathbb{S}}_0$.*

Proof. The lazy value function $L\left(\mathbb{S}_t; s_t, \breve{\mathbb{S}}_0\right)$ is obviously differentiable, and it satisfies the remaining conditions for being a differentiable lower support function due to

$$\mathbb{S}_t \neq \breve{\mathbb{S}}_0 : \; L_t\left(\mathbb{S}_t; s_t, \breve{\mathbb{S}}_0\right) \leq v_t\left(s_t, \mathbb{S}_t\right) \tag{C.20}$$

$$\mathbb{S}_t = \breve{\mathbb{S}}_0 : \; L_t\left(\mathbb{S}_t; s_t, \breve{\mathbb{S}}_0\right) = v_t\left(s_t, \mathbb{S}_t\right) \tag{C.21}$$

□

C.5 Necessary FOCs

Proposition C.1. *If $a_t^\star > 0$ and $b_t^\star > n_t$ an optimal choice must satisfy*

$$c'_{t,a} u'_{t,c} = -\beta \mathbb{E}_t\left[\Gamma'_{t,m,a} v'_{t+1,m}\right] \tag{C.22}$$

$$c'_{t,b} u'_{t,c} = -\beta \mathbb{E}_t\left[\Gamma'_{t,n,b} v'_{t+1,n}\right] \tag{C.23}$$

Proof. Define the *value-of-choice function* $\phi\left(a_t, b_t; \mathbb{S}_t\right)$ conditional on the optimal discrete choice as

$$\begin{aligned} \phi\left(a_t, b_t; \mathbb{S}_t\right) & \equiv u\left(c_t\right)+\beta \mathbb{E}_t\left[v_{t+1}\left(\mathbb{S}_{t+1}\right)\right] \\ & \text{s.t.} \\ c_t & = c\left(m_t, n_t, a_t, b_t\right) \\ d_t & = d\left(n_t, b_t\right) \\ s_{t+1} & = \Gamma_s\left(s_t, z_t^{\star}\right) \\ m_{t+1} & = \Gamma_m\left(a_t, b_t, s_{t+1}\right) \\ n_{t+1} & = \Gamma_n\left(b_t, s_{t+1}\right) \end{aligned} \tag{C.24}$$

Due to lemma (C.2) we immediately have a *differentiable* *upper* *support function* at $(a_t, b_t) = (a_t^{\star}, b_t^{\star})$. We can construct a *differentiable* *lower* *support function* at $(a_t^{\star}, b_t^{\star})$ as

$$\begin{aligned} \underline{\phi}\left(a_t, b_t; \mathbb{S}_t\right) & \equiv u\left(c_t\right)+\beta \mathbb{E}_t\left[L_{t+1}\left(\mathbb{S}_{t+1}; s_{t+1}, \mathbb{S}_{t+1}^{\star}\right)\right] \\ & \text{s.t.} \\ c_t & = c\left(m_t, n_t, a_t, b_t\right) \\ d_t & = d\left(n_t, b_t\right) \\ s_{t+1} & = \Gamma\left(s_t, z_t^{\star}\right) \\ m_{t+1} & = \Gamma_m\left(a_t, b_t, s_{t+1}\right) \\ n_{t+1} & = \Gamma_n\left(b_t, s_{t+1}\right) \end{aligned} \tag{C.25}$$

where $\mathbb{S}_{t+1}^{\star} = \left(m_{t+1}^{\star}, b_{t+1}^{\star}\right) = \left(\Gamma_m\left(a_t^{\star}, b_t^{\star}, s_{t+1}\right), \Gamma_n\left(b_t^{\star}\right)\right)$.

This is a differentiable lower support function for $\phi\left(a_t, b_t; \mathbb{S}_t\right)$ at $(a_t^{\star}, b_t^{\star})$ because the first terms are the same (and differentiable in a_t and b_t), and because we showed in lemma (C.8) that the lazy value function $L_{t+1}\left(\mathbb{S}_{t+1}; s_{t+1}, \mathbb{S}_{t+1}^{\star}\right)$ is a *differentiable lower support function* for $v_{t+1}\left(\mathbb{S}_{t+1}\right)$ at $\mathbb{S}_{t+1}^{\star}$.

Using the *differentiable sandwich lemma* (C.1) we can now conclude that $\phi\left(a_t, b_t; \mathbb{S}_t\right)$ is differentiable in a_t and b_t at $(a_t^{\star}, b_t^{\star})$, and by using the *reverse calculus lemma* (C.3) as in lemma (C.4) we can conclude that $v_{t+1}\left(\mathbb{S}_{t+1}\right)$, is differentiable in a_t and b_t at $(a_t^{\star}, b_t^{\star})$. Equation (C.22) and (C.23) are now the FOCs. □

Corollary C.1. *If $a_t^{\star} > 0$ and $b_t^{\star} = n_t$ then an optimal choice must satisfy*

$$c_{t,a}' u_{t,c}' = -\beta \mathbb{E}_t\left[\Gamma_{t,m,a}' v_{t+1,m}'\right]$$

Proof. Follow from the proof of proposition C.1. □

Corollary C.2. *If $a_t^\star = 0$ and $b_t^\star > n_t$ then an optimal choice must satisfy*

$$c'_{t,b}u'_{t,c} = -\beta\mathbb{E}_t\left[\Gamma'_{t,n,b}v'_{t+1,n}\right]$$

Proof. Follow from the proof of proposition C.1. □

C.6 Euler equations

Corollary C.3. *If $a_t^\star > 0$ and $b_t^\star = n_t$ then an optimal choice must satisfy*

$$c'_{t,a}u'_{t,c} = -\beta\mathbb{E}_t\left[\Gamma'_{t,m,a}c'_{t+1,m}u'_{t+1,c}\right] \tag{C.26}$$

Proof. Continuing from the proof of proposition C.1 we have using lemma C.6 and C.7

$$\begin{aligned} 0 = \overline{\phi}'_a\left(a_t, b_t; \mathbb{S}_t\right) &= \underline{\phi}'_a\left(a_t, b_t; \mathbb{S}_t\right) \\ &= c'_{t,a}u'_{t,c} + \beta\mathbb{E}_t\Gamma'_{t,m,a}\left[L'_{t+1,m}\right] \\ &= c'_{t,a}u'_{t,c} + \beta\mathbb{E}_t\left[\Gamma'_{t,m,a}c'_{t+1,m}u'_{t+1,c}\right] \end{aligned}$$

□

Note that a similar simple Euler equation is not obviously available for the b_t-choice due to the complexity of the derivative $L'_{t+1,n}$.

References

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