# A generalized endogenous grid method for non-smooth and non-concave problems*

Giulio Fella
Queen Mary University of London†

14 June 2013

**Abstract**

This paper extends Carroll's (2006) endogenous grid method and its combination with value function iteration by Barillas and Fernández-Villaverde (2007) to a class of dynamic programming problems, such as problems with both discrete and continuous choices, in which the value function is non-smooth and non-concave.

The method is illustrated using a consumer problem in which the consumer chooses both durable and non-durable consumption subject to a borrowing constraint. The durable choice is discrete and subject to non-convex adjustment costs.

The algorithm yields substantial gains in accuracy and computational time relative to value function iteration, the standard solution choice for problems in which the value function is non-smooth or non-concave.

**JEL classification:** C63

**Keywords:** Endogenous grid method, non-concavity

*I am grateful to Giovanni Gallipoli for his extensive feedback and to Jesús Fernández-Villaverde, Josep Pijoan-Mas, Victor Rios-Rull and two anonymous referees for useful comments. I am indebted to the Department of Economics at the University of Pennsylvania for its generous hospitality while working on this paper.

†School of Economics and Finance, Queen Mary University of London, Mile End Road, London E1 4NS, UK. E-mail: `g.fella@qmul.ac.uk`

# 1 Introduction

Most stochastic, dynamic optimization problems cannot be solved analytically and their numerical solution is often computationally intensive. Computational costs are compounded for problems with non-convex choice sets (*non-concave problems* in what follows), such as problems that involve both discrete and continuous choices, or fixed adjustment costs.[1] Since, in general, the value function of these problems is non-concave or has kinks, short of introducing sufficient uncertainty to guarantee smoothness and concavity of the value function,[2] one cannot apply more efficient solution methods relying on, necessary and sufficient, first-order conditions. Instead, one has to resort to value function iteration (VFI hereafter), which is notoriously slow.[3]

This paper develops an algorithm which is much more efficient and accurate than standard VFI to solve a class of dynamic programming problems—including problems with mixed discrete-continuous choices and fixed adjustment costs[4]—with non-smooth or non-concave value functions. The algorithm applies to problems in which the per-period objective function is strictly concave and twice-differentiable with respect to the continuous choices. Clausen and Strub (2012) prove that a milder requirement implies that first-order conditions are still necessary for a local maximum for a continuous choice, away from its bounds.[5] This paper exploits their result to generalize the endogenous grid method (EGM hereafter) first proposed by Carroll (2006), and its extension to value function iteration by Barillas and Fernández-Villaverde (2007), to the class of problems considered. The algorithm is illustrated for a consumer problem with one continuous (wealth) and one discrete (durables) endogenous state variable and fixed durable adjustment costs. Compared to the VFI benchmark, the algorithm yields substantial savings in computational time and has higher accuracy.

Consider the problem of solving for the optimal continuous choice *for a given value* of the discrete choice. Standard solution methods fix a grid for the continuous endogenous state variable—wealth in our application—at the beginning of the period and solve *forward* for the optimal choice of end-of-period wealth. Carroll's (2006) EGM instead fixes a grid for end-of-period wealth and solves the first-order condition *backward* for the associated values of wealth at the beginning of the period. The EGM approach is much faster as the first-order condition is often linear in current assets (or an appropriate auxiliary variable), but non-linear in next period's ones.

The standard EGM, though, cannot be used when the value function is non-concave as the first-order condition is necessary but not sufficient for an interior optimum. The generalization proposed here addresses this problem, by supplementing the EGM step

[1]Discrete choices arise naturally in quantitative analyses of retirement behavior (Rust 1989), labor supply (Gomes, Greenwood and Rebelo 2001), education (Abbott, Gallipoli, Meghir and Violante 2013). Fixed adjustment cost are found in quantitative studies of investment (Khan and Thomas 2008, Bloom 2009) and consumer-durables choice (Luengo-Prado 2006, Bajari, Chan, Krueger and Miller 2009).

[2]This is the approach followed, e.g., in Gomes et al. (2001) and Khan and Thomas (2008).

[3]To be precise, most papers—e.g. Rust (1989), Luengo-Prado (2006), Bloom (2009), Bajari et al. (2009)—use VFI on a discretized state space.

[4]Since inaction may be optimal in the presence of fixed adjustment costs, they induce a discrete, extensive (action/inaction) margin.

[5]In other words, the value function is differentiable in the endogenous continuous state variables at the *maximum*, if interior, but not necessarily everywhere.

with a discretized-VFI step to verify whether the grid point is not only a local but also a global maximum. The solution is very accurate because the EGM step determines the initial value of the continuous state variable for which the first-order condition holds *exactly.* The discretized-VFI step is used only to verify whether the candidate global maximum is indeed so or to discard it.

The generalized EGM algorithm described above is used to obtain the value functions for each *given* value of the discrete choice. A second VFI maximization step over the discrete choice variable then recovers the discrete policy function by taking the upper envelope over the discrete-choice-specific value functions.

Barillas and Fernández-Villaverde (2007) were the first to propose this kind of nesting of EGM within VFI for problems with more than one continuous control. Their algorithm uses a (standard) EGM step to optimize with respect to one continuous choice variable while keeping the other policy functions fixed. These are then updated by the VFI step. The important difference between their paper and the present one is the generalization of the internal EGM step to deal with problems with a non-concave and non-smooth value function. Hintermaier and Koeniger (2010) extend EGM to the case in which there are two continuous endogenous state variables and occasionally binding constraints on both.

Both Barillas and Fernández-Villaverde (2007) and Hintermaier and Koeniger (2010) apply only to problems with a concave and differentiable value function and are, therefore, not suitable to solve the class of problems considered here. Conversely, the algorithm in this paper is less accurate—for the same execution time—than those in Barillas and Fernández-Villaverde (2007) and Hintermaier and Koeniger (2010), within the class of problems to which they apply, as it relies on the discretization of one of the two continuous choices.

The paper is structured as follows. Section 2 introduces the problem and the properties of the solution that underpin the solution algorithm. Section 3 describes the basic algorithm while Section 4 discusses how to modify it to exploit monotonicity and deal with problems with occasionally-binding borrowing constraints. Section 5 reports the numerical results, while Section 6 concludes.

# 2 The problem

The algorithm in this paper applies to a class of dynamic programming problems that includes problems with one continuous and an arbitrary number of discrete choices. In what follows, we illustrate it for an optimal consumption problem involving one continuous and one discrete choice.[6]

[6]Since values for multiple discrete choices can always be stacked into a single vector, assuming a single discrete choice is without loss of generality.

## 2.1 The model

A household makes a continuous non-durable consumption choice $c_t \in C$ and a discrete durable consumption choice[7] $d_{t+1} \in D$ in each period $t$ to maximize

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^s u(c_t, d_{t+1}), \tag{1}$$

where $\beta \in (0,1)$ is a discount factor and the function $u$ is strictly increasing in both its arguments and continuous in $c_t$. More formally $C$ is an interval in $\mathbb{R}_+$ while $D$ is a countable, compact subset of $\mathbb{R}_+$, with smallest element normalized to zero and cardinality larger than one.

The relative price of durables in terms of non-durables is normalized to one. The durable stock is subject to an adjustment cost $h(d_t, d_{t+1}) = \mathbb{I}_{d_{t+1} \neq d_t} \phi d_{t+1}$, with $\phi > 0$ and $\mathbb{I}_{d_{t+1} \neq d_t}$ an indicator function equal to one if $d_{t+1} \neq d_t$ and one otherwise. The discreteness of $D$, together with the fact that the adjustment cost has a fixed component,[8] imply that the choice set is non-convex.

In each period, the household earns a stochastic labor income $y_t$ which follows an $m$-state Markov chain with transition matrix $P$ and state space $Y = \{y^1, \ldots, y^m\}$, with $y^1 \geq 0$ and $y^i > y^{i-1}, i = 2, \ldots, m$. The household also earns capital income $rw_t$, where $r > 0$ is the risk-free rate of return and $w_t$ financial wealth at the beginning of period $t$.

It follows that the household dynamic budget identity can be written as

$$c_t + w_{t+1} + d_{t+1} + h(d_t, d_{t+1}) = y_t + (1+r)w_t + d_t. \tag{2}$$

The non-durable consumption choice is bounded below by a non-negativity constraint

$$c_t \geq 0 \tag{3}$$

and above by a borrowing constraint

$$w_{t+1} \geq -\gamma y_1 - \xi d_{t+1}, \tag{4}$$

where $\gamma \in [0, r^{-1}]$ is the fraction of minimum labor income and $\xi \in [0, (1+r)^{-1}]$ the fraction of the durable stock that can be collateralized.

The restrictions on the two parameters $\gamma$ and $\xi$ require that the lowest feasible wealth level is never lower than the natural (in Aiyagari's (1994) sense) borrowing limit that obtains when both parameters are at their upper bounds.[9] The restrictions imply that the household choice set is bounded.

[7] I adopt the notational convention of indexing durable consumption at *time t* by $t+1$ to simplify the notation in the recursive problem.

[8] To see this, note that when $d_{t+1} \neq d_t$, the adjustment cost can be rewritten as $h(d_t, d_{t+1}) = \phi d_t + \phi(d_{t+1} - d_t)$, where the first addendum is the (fixed, hence non-convex) cost component unrelated to the investment flow $(d_{t+1} - d_t)$.

[9] To see this, note that when $\gamma = r^{-1}$ and $\xi = (1+r)^{-1}$ the borrowing limit for a household with current durable stock $d_{t+1}$ equals the present value of the durable stock next period plus human wealth along the worst possible income history. This is the maximum amount that can be repaid with probability one next period, without violating the non-negativity constraint on consumption (3), by downsizing the durable stock to zero and keeping it at zero forever after.

The household maximizes (1) subject to the constraints (2)-(4) and $d_{t+1} \in D$.

Let

$$a_{t+1} = w_{t+1} + \gamma y_1 + \xi d_{t+1} \tag{5}$$

and let

$$z(d_{t+1}; a_t, d_t, y_t) = y_t + (1+r)a_t - (1-\xi)(d_{t+1} - d_t) - h(d_t, d_{t+1}) - r(\gamma y_1 + \xi d_t) \tag{6}$$

denote total resources available for consumption after (conditional on) the durable choice $d_{t+1}$. The dynamic budget identity (2) and the borrowing constraint (4) become[10]

$$c_t + a_{t+1} = z(d_{t+1}; a_t, d_t, y_t) \tag{7}$$

and

$$a_{t+1} \geq 0. \tag{8}$$

Let $\Omega(\cdot,\cdot;\cdot,\cdot,\cdot)$ denote the feasibility set, with $\Omega(a_{t+1}, d_{t+1}; a_t, d_t, y_t)$ one specific point in it and

$$\Omega(\cdot,\cdot; a_t, d_t, y_t) = \{a_{t+1}, d_{t+1} : a_{t+1} \in [0, z(d_{t+1}; a_t, d_t, y_t)], d_{t+1} \in D\}$$

the set of feasible end-of-period values for the endogenous state variables if the current state is $(a_t, d_t, y_t)$.

The household sequence problem can then be written in the canonical form

$$\max_{\{a_{t+1}, d_{t+1}\}_{t=0}^{\infty}} \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^s u(z(d_{t+1}; a_t, d_t, y_t) - a_{t+1}, d_{t+1}), \tag{9}$$
$$\text{s.t. } (a_{t+1}, d_{t+1}) \in \Omega(\cdot,\cdot; a_t, d_t, y_t),\ t = 0, 1, \ldots,$$
$$(a_0, d_0, y_0) \in A \times D \times Y \text{ given.}$$

## 2.2 The recursive problem

Standard results imply that Bellman's principle of optimality applies to problem (9). Dropping time indices, its recursive representation can be written as

$$\mathbb{V}(a, d, y) = \max_{(a',d') \in \Omega(\cdot,\cdot;a,d,y)} u(z(d'; a, d, y) - a', d') + \tilde{V}(a', d'; y), \tag{10}$$

where $\mathbb{V}(a, d, y)$ is the value function in state $(a, d, y)$ and

$$\tilde{V}(a', d'; y) = \beta \mathrm{E} \mathbb{V}(a', d', y') \tag{11}$$

denotes the expectation of the continuation value.

The following, rather standard, assumptions characterize the class of problems considered in this paper. For comparability with Clausen and Strub (2012), it is useful to state them with reference to a general utility functional $U(a', d'; a, d, y)$,[11] where

$$U(a', d'; a, d, y) = u(z(d'; a, d, y) - a', d') \tag{12}$$

in the present set up.

[10] Similarly to Aiyagari (1994), the change of variable defined by equation (5) usefully implies that the grid for the continuous state variable in the solution algorithm is independent of the borrowing limit and here, in particular, of the current durable choice.

[11] Stating the assumptions in terms of a general utility functional also highlights that the algorithm applies to any problem in which the objective function satisfies the assumptions; e.g., the investment problem of a risk-neutral firm with a concave production function and fixed adjustment costs.

**Assumption 1.** *The per-period utility function $U(a', d'; a, d, y)$ is differentiable with respect to $a'$ on the interior of $\Omega(\cdot, d'; a, d; y)$ and with respect to $a$ on the interior of $\Omega(a', d'; \cdot, d; y)$.*

**Assumption 2.** *The second derivative $U_{a'a'}(a', d'; a, d, y)$ is strictly negative.*

It follows from equation (12) that assumptions 1 and 2 are satisfied if $u(c, d')$ is twice differentiable and strictly concave with respect to $c$ and if $z(d'; a, d, y)$ is a smooth function of $a$. They imply concavity, and differentiability of the value function away from bounds, *if* the choice set $\Omega(\cdot, \cdot; a, d, y)$ is convex (see Theorem 4.11 in Stokey, Lucas and Prescott 1989). This would be the case, for example, if there were no discrete choice. As a result, the maximand in equation (10) would be the sum of two concave and differentiable functions and the first-order condition

$$- u_c(z(d'; a, d, y) - a', d') + \tilde{V}_a(a', d'; y) = 0 \tag{13}$$

would be necessary and sufficient for a maximum for $a'$ on the interior of $\Omega(\cdot, d'; a, d; y)$—an *interior maximum* for $a'$ in what follows. The necessity and sufficiency of the first-order condition (13) lies at the heart of Carroll's (2006) original EGM, whose contribution consists in providing an efficient algorithm for solving it.

Here, though, $\Omega(\cdot, \cdot; a, d, y)$ is non-convex as $d'$ is a non-degenerate discrete variable. It follows that, in general, the value function

$$\mathbb{V}(a, d, y) = \max_{d' \in D} \mathbb{V}^{d'}(a, d, y), \tag{14}$$

the upper envelope of the $d'$-contingent value functions

$$\mathbb{V}^{d'}(a, d, y) = \max_{a' \in \Omega(\cdot, d'; a, d, y)} u(z(d'; a, d, y) - a', d') + \tilde{V}(a', d'; y), \tag{15}$$

has kinks at those values of $a$ for which the discrete choice $d'$ switches, and is globally non-concave. As a result, the continuation value $\tilde{V}(a', d', y)$ is also non-smooth and non-concave, in general, and the first-order condition (13) is no longer sufficient for an interior solution to equation (15).

Yet, Theorem 3 in Clausen and Strub (2012) establishes that, if Assumption 1 and a technical regularity condition hold, at an optimum $(\bar{a}', \bar{d}')$ the maximand in equation (10) is differentiable with respect to $a'$, away from bounds. The intuition is the following. On the interior of the choice set for $a'$, kinks in the maximand can only be due to *downward*-pointing kinks in the continuation value $\tilde{V}(a', d', y)$ at those value of $a'$ for which the *future* discrete choice changes. Since the derivative of the maximand with respect to $a'$ jumps up at a downward kink, an interior maximum cannot be located at a kink; namely the agent is never indifferent between two discrete choices at an optimum. It follows that the first-order condition (13) is still necessary for a interior local maximum, and therefore for a candidate interior global maximum, of equation (15).

The generalized EGM algorithm in this paper exploits the fact that the first-order condition is still necessary, and the standard EGM insight is still valid, for an interior optimal *$d'$-contingent* saving choice in equation (15). The non-concavity of the continuation value $\tilde{V}(a', d', y)$ with respect to $a'$, though, implies the EGM algorithm cannot be applied in its original form as equation (13) is not sufficient; a zero of equation (13) is not necessarily an interior *global* maximum for (15).

Two final remarks are in order before proceeding. First, the regularity condition required for Clausen and Strub's (2012) result is somewhat technical. It is implied (see p. 15 in Clausen and Strub 2012), though, by the following standard assumption which is maintained throughout.

**Assumption 3.** *The felicity function* $u(c, d')$ *satisfies* $\lim_{c\to 0} u_c = +\infty$.

Secondly, Assumption 2 is not necessary for Clausen and Strub's (2012) differentiability result, but is exploited by the algorithm in this paper. It implies that any non-concavity with respect to $a'$ of the maximand in equation (10) is due to the continuation value $\tilde{V}(a', d'; y)$.

The Bellman equation (15) for the $d'$-contingent value functions can also be written in the alternative state space $(d'; z, y)$ as

$$V^{d'}(z, y) = \max_{a'\in[0,z]} u(z - a', d') + \tilde{V}(a', d'; y). \tag{16}$$

In fact, given that $\Omega(\cdot, d'; a, d, y) = [0, z(d'; a, d, y)]$, it is easily verified that the value functions for problems (15) and (16) satisfy

$$\mathbb{V}^{d'}(a, d, y) = V^{d'}(z(d'; a, d, y), y). \tag{17}$$

The alternative representation (16) in terms of cash at hand is crucial to the EGM algorithm and will be exploited in what follows.

Finally, the following Proposition establishes that the $d'$-contingent optimal saving correspondence $a'^{d'}(z, y)$ is increasing in $z$ in the following sense.

**Proposition 1.** *Let* $z_H > z_L$*. For any* $a'_L \in a'^{d'}(z_L, y)$ *and* $a'_H \in a'^{d'}(z_H, y)$*,* $a'_H \geq a'_L$ *holds. Furthermore,* $a'_H > a'_L$ *if* $a'_H \in (0, z)$*.*

*Proof.* See Appendix. □

The result is standard for problems in which the choice set $\Omega(\cdot, \cdot; a, d, y)$ is convex and the saving correspondence $a'^{d'}(z, y)$ is a function; i.e. the set $a'^{d'}(z, y)$ is a singleton for all $(d'; z, y)$. Proposition 1, though, generalizes the result to the present setup in which the non-convexity of $\Omega(\cdot, \cdot; a, d, y)$ may imply that the set $a'^{d'}(z, y)$ is not a singleton for all $(d'; z, y)$, in which case the policy correspondence is not a function. Proposition 1 exploits a monotone-comparative-statics result to establish that the correspondence is still strictly increasing in $z$ off corners which implies that its inverse with respect to $z$ is a function. The saving correspondence in figure 2 provides a graphical illustration of the property implied by Proposition 1.

# 3 The solution algorithm

To simplify the exposition of the algorithm, we assume the following.

**Assumption 4.** *The parametric restriction* $\gamma = r^{-1}$ *and* $\xi = (1 + r)^{-1}$ *holds.*

The assumption implies that the borrowing constraint $a' \geq 0$ is the natural borrowing constraint and therefore never binds. We relax this in Section 4.1.

It also follows from Proposition 1 that one can exploit the monotonicity of the $d'$-contingent saving correspondence $a'^{d'}(z, y)$ to accelerate the computation of the solution. It turns out that monotonicity is even more powerful when the value function has downward kinks. A refined version of the algorithm exploiting monotonicity is described in Section 4.2.

## 3.1 Relationship to the standard endogenous grid method

The result in Proposition 1 is at the heart of Carroll's (2006) EGM. It implies that in solving the first-order condition (13) for $a'^{d'}(z, y)$ as a function of $z$ for given $(d'; y) \in D \times Y$ one can, interchangeably, proceed in one of two ways.

The usual way is to construct an ordered grid $G_z = \{z_i\}_{i=1}^m$ for total *initial* resources $z$ and solve the Euler equation (13) *forward* for the associated $a'$, to obtain the policy correspondence $a'^{d'}(z_i, y)$, on the chosen grid $G_z$.

Carroll's (2006) EGM instead defines a, fixed, ordered grid $G_{a'} = \{a'_i\}_{i=1}^m$ for *end-of-period* assets and solves the Euler equation (13) *backward* for the value of total resources $\hat{z}_i(d'; y)$ for each $a'_i \in G_{a'}$. Therefore, the EGM algorithm returns the policy correspondence $a'^{d'}(\hat{z}_i(d'; y), y)$ on the, endogenously-determined, grid $G_{\hat{z}}(d'; y) = \{z_i(d'; y)\}_{i=1}^m$ implied by $a'^{d'}(\hat{z}_i(d'; y), y) = a'_i$ for all $a'_i \in G'_a$.

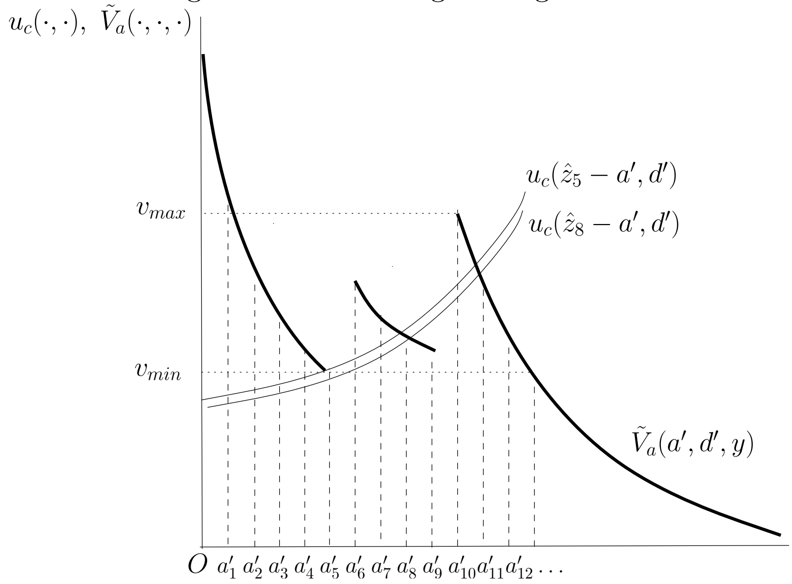
The disadvantage of standard, forward-solving methods is that the Euler equation is non-linear in $a'$. Solving for $a'$ involves evaluating the Euler equation multiple times for each grid point $z_i$. Conversely, the computational cost of solving the Euler equation backward for $z$ given $a'$ is usually very low. This can be easily seen in the case in which the felicity function is separable in $c$ and $d'$; e.g., it satisfies $u(c, d') = \theta \log(c) + (1 - \theta) \log(g(d'))$. In such a case, the Euler equation can be written as $z - a' = \theta / \tilde{V}_a(a', d'; y)$, which is linear in $z$.[12]

Given, the policy correspondence $a'^{d'}(\hat{z}_i(d'; y), y)$, on the endogenous grid $G_{\hat{z}}(d'; y)$, one can construct an interpolating function and evaluate it at any arbitrary point for total resources $z$ to obtain the associated value $a'^{d'}(z, y)$. Conversely, standard, forward-solving methods solve *directly* for the policy correspondence at any chosen point for $z$. Therefore, EGM trades off the cost of evaluating an interpolating function against the cost of solving a non-linear equation, a very advantageous trade-off.

If the maximand in equation (15) is concave and differentiable with respect to $a'$, maximization of (15) is equivalent to solving for the unique zero of equation (13) and one can use the standard EGM to solve for the saving function. This is not the case if the maximand in equation (15) is either non-concave or non-differentiable in $a'$. As discussed in Section 2.2, though, the first-order condition (13) is still necessary for an interior maximum, for the class of problems satisfying Assumptions 1 to 3. Therefore, EGM is still useful to locate an interior local maximum. Since, given the non-concavity of the maximand in equation (15), a local maximum is not necessarily a global one, the

[12]Even if $u_c(c, d')$ does not have a closed-form inverse with respect to $c$, given $a'$ some variant of Newton method converges to the unique solution for $z$ at a quadratic (or close to quadratic) rate. A similar method cannot be applied to solve for $a'$ if, as in the present model, $\tilde{V}_a(a', d'; y)$ is discontinuous, let alone non-differentiable, with respect to $a'$.

Figure 1: Illustrating the algorithm.




generalized algorithm in this paper modifies the standard EGM step in the following way. First, it partitions the set of grid points for future assets $G_{a'}$ into the region in which the continuation value $\tilde{V}(a', d', y)$ is non-concave (*non-concave region*) in $a'$ and its set complement. Secondly, for all $a'_i$ in the non-concave region, the algorithm supplements the EGM step with a standard discretized-VFI maximization step. Since the non-concave region is a subset of the choice set $G_{a'}$, identifying the non-concave region allows to confine the, costly, application of the discretized-VFI step to such region rather than the whole of $G_{a'}$.

The next two subsections describe respectively how to identify the non-concave region and the details of the modified EGM algorithm.

## 3.2 Identifying the non-concave region

Understanding how the algorithm delimits the region over which $\tilde{V}(a', d', y)$ is non-concave in $a'$ (*non-concave region*) is easier with the help of Figure 1 which draws the marginal utility of present consumption and of future assets as functions of $a'$. The thick non-monotonic and discontinuous curve plots the marginal utility of future assets $\tilde{V}_a(a', d', y)$, for given *current* discrete durable choice $d'$ and income realization $y$. The curve is discontinuous at those values of $a'$ for which some *future* discrete choice changes along some continuation history. The thinner upward sloping curve is the marginal utility of present consumption for a given value of total resources $z$ and current durable choice $d'$. A point where the two curves intersect is a zero of the first-order condition (13).

In terms of Figure 1, for each abscissa $a'_i \in G_{a'}$ EGM finds the value of total resources $z_i$ for which an upward sloping curve intersects the thick $\tilde{V}_a(a', d', y)$ curve at $a' = a'_i$; namely for which $a'_i$ is a zero of the first-order condition (13). The first-order condition is sufficient for $a'_i$ to be a global maximum if $a'_i$ is the unique intersection between the upward sloping curve $u_c(z_i - a', d')$ through it and the curve $\tilde{V}_a(a', d', y)$. A sufficient condition for the intersection at $a'_i$ to be unique on some subset $S^G \subseteq G_{a'}$ is that $\tilde{V}_a(a'_j, d', y) >$

$\tilde{V}_a(a'_i, d', y)$ for all $a'_j < a'_i \in S^G$ and $\tilde{V}_a(a'_j, d', y) < \tilde{V}_a(a'_i, d', y)$ for all $a'_j > a'_i \in S^G$. In Figure 1, this is the case in the regions where $\tilde{V}_a(a', d', y)$ is above $v_{max}$ and below $v_{min}$, or equivalently for any value of assets outside the set $G^{nc}_{a'}(d'; y) = \{a'_2, \ldots, a'_{11}\}$.[13] Note, that as $\tilde{V}_a(a', d', y)$ is a function of $(d'; y)$ so is the non-concave region $G^{nc}_{a'}(d'; y)$.

Assuming the function $\tilde{V}_a(a', d', y)$ is known, the bounds $v_{min}$ and $v_{max}$ can be computed, for each given $(d'; y)$, as respectively the lowest value of $\tilde{V}_a(a'_i, d', y)$ and the highest value of $\tilde{V}_a(a'_{i+1}, d', y)$ for all $i$ such that $\tilde{V}_a(a'_{i+1}, d', y) > \tilde{V}_a(a'_i, d', y)$. Given $v_{min}$ and $v_{max}$, one can compute $\underline{i}$—the largest $i$ such that $\tilde{V}_a(a'_i, d', y) > v_{max}$—and $\bar{i}$—the smallest $i$ such that $\tilde{V}_a(a'_i, d', y) < v_{min}$.

By construction, the first-order condition (13) is necessary and sufficient for a maximum for $a' \leq a'_{\underline{i}}$ and $a' \geq a'_{\bar{i}}$. The first-order condition is only necessary though for $a'_i \in G^{nc}_{a'}(d'; y) = \{a'_{\underline{i}+1}, \ldots, a'_{\bar{i}-1}\}$.

## 3.3 The algorithm

Given $(d'; y)$ and the associated non-concave region $G^{nc}_{a'}(d'; y)$ identified in the previous section, the generalized EGM algorithm proceeds in the following way. For each $a'_i \in G_{a'}$ it applies the standard EGM algorithm and uses equation (13) to solve for $\hat{z}_i(d'; y)$. If $a'_i$ lies outside the non-concave region—e.g., $a'_i = a'_1$ in Figure 1—the algorithm stores the pair $\{\hat{z}_i(d'; y), a'_i\}$ and moves to the next point in $G_{a'}$. If instead, as is the case for $a'_4$ in Figure 1, $a'_i$ belongs to the non-concave region, the algorithm verifies whether $a'_i$ is also a *global* maximum for $z = \hat{z}_i(d'; y)$. To do so, for given $\hat{z}_i(d'; y)$, the algorithm constructs the discretized Bellman maximand for all $a'_j$ in the non-concave region $G^{nc}_{a'}(d'; y)$ and finds the maximum of the discretized problem

$$a'_g = \arg \max_{a' \in G^{nc}_{a'}(d'; y)} u(\hat{z}_i(d'; y) - a', d') + \tilde{V}(a', d', y). \tag{18}$$

If $a_g = a_i$, $a_i$ is both a local and global maximum given $\hat{z}_i(d'; y)$ and, again, the pair $\{\hat{z}_i(d'; y), a'_i\}$ is stored. If instead $a'_g$ is different from $a'_i$ the algorithm discards point $a'_i$ and moves to the next grid point $a'_{i+1}$. Evaluating all grid points $a'_i \in G_{a'}$ yields the saving correspondence $a'^{d'}(\hat{z}_{i_l}(d'; y), y) = a'_{i_l}$ on the endogenous grid points, where $i_l$, $l = 1, 2, \ldots$ indexes the—ordered—subset $M(d'; y) \subseteq G_{a'}$ of grid points $a'_i$ that have not been discarded. Replacing the set of pairs $\{\hat{z}_{i_l}(d'; y), a'_{i_l}\}$ in equation (16) one obtains the $d'$-contingent value functions $V^{d'}(\hat{z}_{i_l}(d'; y), y)$ on the endogenous grid points $\hat{z}_{i_l}(d'; y)$.

Before presenting the pseudo-code for the complete algorithm, it is useful to tie a few loose ends. First, one has to select an ordered grid $G_{a'}$ for next-period's assets $a'$. Second, it is useful to store in memory the grid for total resource $G_z(d'; d, y) = \{z_i(d'; d, y)\}_{i=1}^m$ implied by $z(d'; a, d, y)$ defined in equation (6), for all $a \in G_a$, as a function of $(d'; d, y) \in D \times D \times Y$. Note that $G_a$ is the grid for *current* assets that is used in the VFI step. It is convenient to choose the same grid as the grid for *next period's* assets $G_{a'}$, keeping in mind, though, that the two are conceptually different. Third, since the function $\tilde{V}(a', d', y)$ is unknown it has to be found by repeated backward iteration of the system formed by equations (11), (14), (17) and (16) starting from equation (16) and some initial guess

[13] The fact that $\tilde{V}_a(a', d', y) > v_{max}$ for $a'$ low enough follows from Assumption 4, maintained in this section, that implies that the borrowing constraint is always slack. This not true in general. Section 4.1 discusses how the algorithm needs to be modified when it is not.

$\tilde{V}(a',d',y)^0$. The initial choice of guess $\tilde{V}(a',d',y)^0$ has to be continuous and increasing in $a'$. Its wealth derivative $\tilde{V}_a(a',d',y)^0$ can be computed by finite differences.

At all subsequent iterations $n > 0$, the wealth derivative $\tilde{V}_a(a',d',y)^n$ at the points of the grid $G_{a'}$, can be approximated either by taking finite differences of $\tilde{V}(a',d',y)^n$ or using the envelope condition[14]

$$\tilde{V}_a(a',d',y)^{n+1} = (1+r)\mathrm{E}u_c(c(a',d',y')^n, d'(a',d',y')^n) \tag{19}$$

where $c(a,d,y)^n = z - a'^{d'}(z,y)$ with $z = z(d';a,d,y)$ and $d' = d'(a,d,y)^n$.

The corresponding pseudo code is the following.

1. Initialize the iteration index $n = 0$. Guess a function $\tilde{V}(a',d',y)^0$ and compute its wealth derivative $\tilde{V}_a(a',d',y)^0$.

2. For all $(d';y)$ apply the modified EGM step as follows:

   2.1. Solve for the bounds $\underline{i}$, $\bar{i}$ of the non-concave region $G^{nc}_{a'}(d';y)$ as derived in Section 3.2.

   2.2. Compute the endogenously-determined level of total resources $\hat{z}_i(d';y)$ that solves equation (13) for all $a'_i \in G_{a'}$.

   2.3. For all $\underline{i} < i < \bar{i}$, find the solution $a'_g$ to the discretized saving problem using (18) evaluated at $z = \hat{z}_i(d';y)$. If $a'_g \neq a'_i$, discard the pair $\{\hat{z}_i(d';y), a'_i\}$.

   2.4. The set of pairs $\{\hat{z}_{i_l}(d';y), a'_{i_l}\}$ that have not been discarded is the policy function $a'^{d'}(\hat{z}_{i_l}(d';y),y)$ on the endogenous grid points. Use each pair to replace for $z$ and $a'$ in equation (14) to obtain the associated $d'$-contingent value functions $V^{d'}(\hat{z}_{i_l}(d';y),y)$.

   2.5. The $d'$-contingent value functions $V^{d'}(\hat{z}_{i_l}(d';y),y)$ need to be evaluated on a common grid in order to solve for the optimal discrete choice. Therefore, interpolate $a'^{d'}(\hat{z}_{i_l}(d';y),y)$ and $V^{d'}(\hat{z}_{i_l}(d';y),y)$ on the grid $G_z(d';d,y)$ to obtain $\mathbb{V}^{d'}(a,d,y) = V^{d'}(z(d';a,d,y),y)$ and $a'^{d'}(z(d';a,d,y),y)$ on the grid $G_{a'} \times D \times Y$.

3. For all $(a,d,y) \in G_a \times D \times Y$ compute the optimal discrete choice $d'(a,d,y)^n$, and the value function $\mathbb{V}(a,d,y)^{n+1}$ by solving (14).

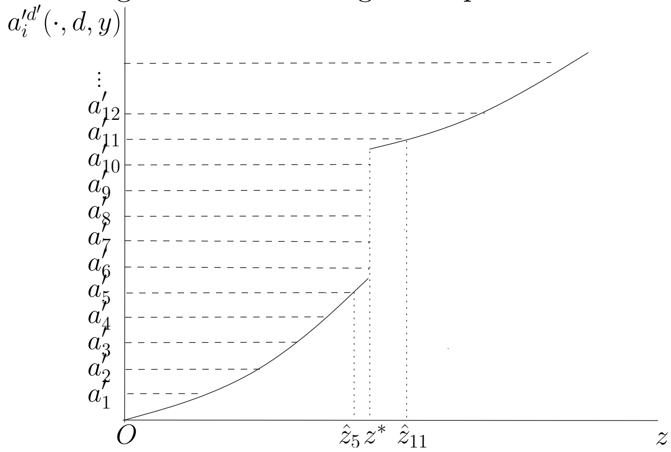
4. Compute $\tilde{V}(a,d,y)^{n+1}$ using equation (11).

5. If $||\tilde{V}(a,d,y)^{n+1} - \tilde{V}(a,d,y)^n||_\infty > 10^{-5}$, with $||\cdot||_\infty$ the sup norm over $G_a \times D \times Y$, use the envelope condition (19) to obtain $\tilde{V}_a(a,d,y)^{n+1}$ and start a new iteration.

It should be clear from the above description that the application of the global maximization step to the discretized problem is used only to verify whether a local extremum is a global maximum. It is not used to actually solve for a point on the *$d'$-contingent* saving correspondence. If the solution $a'_g$ differs from the original point $a'_i$, the algorithm

[14]In our numerical experiments using the envelope condition results in a slight improvement in accuracy relative to finite differences.

Figure 2: The saving correspondence.

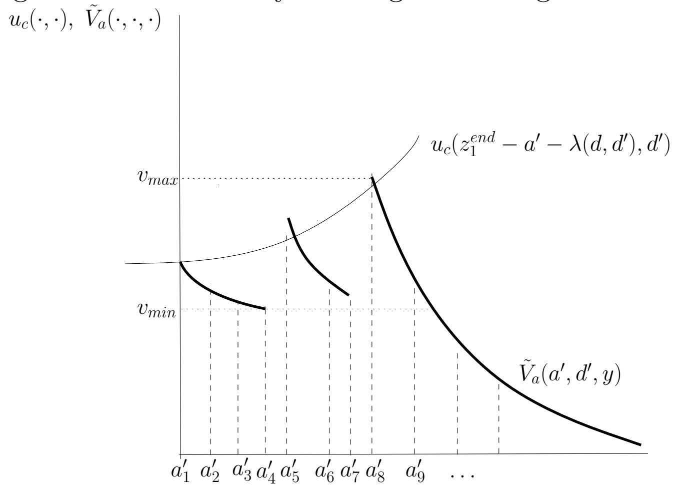



does not replace $a'_i$ with the global maximum $a'_g$ for the discretized problem; it just discards $a'_i$. If $a'_g > a'_i$ the same procedure will be repeated when $a'_g$ is reached and $a'_g$ will be stored only if it is a fixed-point of the procedure. For any point $a'_i$ belonging to the range of the saving correspondence the solution is very accurate because the algorithm determines the value of total resources for which a given grid point $a'_i$ for future assets solves the first-order condition (13) *exactly.* The imprecise, discretized, global maximization step is used only to confirm that the candidate local extremum is indeed a global maximum or to discard it.

On the other hand, as other methods that solve a discretized version of the true problem, the maximization step in equation (18) may incorrectly identify a pair $(\hat{z}_i, a'_i)$ as an optimal one when $a'_i$ does *not* belong to the range of the saving correspondence for the continuous problem. For example, suppose that the algorithm is evaluating point $a'_8$ in Figure 1 and that $a'_8$ does not belong to the saving correspondence for the true—continuous—problem; namely the global maximum associated with $\hat{z}_8$ is a point in the interval $(a'_{10}, a'_{11})$. Yet, if the grid is not fine enough—i.e., if the two grid points $a'_{10}$ and $a'_{11}$ in Figure 1 are not close enough—the algorithm may wrongly accept the pair $(\hat{z}_8, a'_8)$ as optimal. It is therefore important that the grid $G_{a'}$ is sufficiently fine to minimize this type of errors. Once again, though, note that this is a problem that applies to *all* methods that discretize continuous variables; i.e. all methods commonly used to solve non-concave problems (see footnote 3).

There is also a second reason, having to do with the discontinuity of the policy correspondence, for which the asset grid has to be sufficiently fine. To see this consider again the policy correspondence plotted in Figure 2, where the discontinuous curve represents the true solution which the numerical algorithm has recovered on the discrete grid on the vertical axis. The algorithm discards points $\{a'_6, \ldots, a'_{10}\}$ and the associated values of cash at hand that solve the first-order condition. Compared to an algorithm that solves forward for $a'$ given some exogenous grid for $a$, and hence for $z$, the modified EGM loses five interpolating nodes. The loss of some of those nodes, namely those for which $\hat{z}_i$ lies outside the interval $(\hat{z}_5, \hat{z}_{11})$ in Figure 2—e.g., $\hat{z}_9 > \hat{z}_{11}$ from Figure 1—is not very costly as the value of the policy at such points can be recovered by interpolation, provided the the grid is sufficiently fine outside $(\hat{z}_5, \hat{z}_{11})$. On the other hand, one would want a high accuracy in bracketing the set of values of $a'$ that do not belong to the policy cor-

Figure 3: Occasionally-binding borrowing constraint.




respondence. Since the algorithm identifies the points in the grid $G_{a'}$ that bound such a set—namely $a'_6$ and $a'_{10}$— in Figure 2, it provides information on where to efficiently place the extra grid points—namely in the intervals $(a'_5, a'_6)$ and $(a'_{10}, a'_{11})$.[15]

# 4 Generalizations and refinements

## 4.1 Borrowing constraints

Up to now I have maintained, only for expositional reasons, the assumption that the borrowing constraint is never binding. Carroll (2006) though shows that EGM can accommodate occasionally binding borrowing constraints extremely effectively. I now relax the assumption that the borrowing constraint is never binding, by assuming the following.

**Assumption 5.** *The parametric restrictions $\gamma < r^{-1}$ or $\xi < (1+r)^{-1}$ hold.*

The assumption implies that the expected marginal utility of future consumption is finite at the borrowing constraint $a'_1 = 0$ and the constraint binds with positive probability. Figure 3, effectively the counterpart of Figure 1, illustrates how EGM deals with the borrowing constraint.

For given $(d'; y)$ EGM calculates the value of total resources $\hat{z}_1(d'; y)$ for which the first-order condition (13) is satisfied as an equality. The are two possible cases to distinguish.

In the first case, $a'_1 = 0$ is both a local and global maximum given $z = \hat{z}_1(d'; y)$, namely $a'^{d'}(\hat{z}_1(d'; y), y) = a'_1 = 0$ in point 2.4 in the pseudo-code in Section 3.3. Therefore $\hat{z}_1(d'; y)$ is the first interpolation node for the $d'$-contingent saving and value functions. Since $a'^{d'}(\hat{z}_1(d'; y), y)$ satisfies the first-order condition, the borrowing constraint is on the verge of being binding at $\hat{z}_1(d'; y)$ and, from monotonicity, the constraint is strictly binding for any $z < \hat{z}_1(d'; y)$.

[15] I am indebted to Victor Rios-Rull for this insight. The refinement is not implement in the code to avoid additional complexity and given the relatively low cost of having a sufficiently fine grid over the entire range.

The value of the $d'$-*contingent* saving correspondence for all points for which the borrowing constraint is binding—namely those for which $z \leq \hat{z}_1(d';y)$ is just $a'^{d'}(z,y) = 0$.

Replacing in (10), the associated value of the $d'$-*contingent* value function can be recovered as

$$V^{d'}(z,y) = u(z,d') + \tilde{V}(0,d',y).^{16} \tag{20}$$

As first pointed out by Carroll (2006), the fact that EGM returns the *exact* value of total resources for which the borrowing constraint is weakly binding implies that the solution is analytic *everywhere* to the left of the first endogenous grid point. For the same reason, in the presence of occasionally-binding borrowing constraints, the accuracy of the model simulation is improved by using the endogenous grid.

The previous case is the only one that applies to concave problems. A second possibility exists for non-concave problems, though.

Consider the case in which $a'_1 = 0$ is not a global maximum for $z = \hat{z}_1(d';y)$; i.e., the solution to equation (18) for $z = \hat{z}_1(d';y)$ is some $a'_g > 0$. Therefore, the borrowing constraint is not binding for $z = \hat{z}_1(d';y)$. In fact, the EGM steps 2.1-2.4 in Section 3.3 would return an interpolating function whose first point is $a'^{d'}(\hat{z}_{i_1}(d';y),y) = a'_{i_1} > 0$; e.g. $a'_{i_1} = a'_5$ in Figure 3. Therefore, one cannot conclude that the household chooses to be borrowing constrained for $z$ in a left neighborhood of $\hat{z}_{i_1}(d';y)$. The EGM algorithm no longer necessarily determines the lower bound on total resources below which the household is borrowing constrained.

Yet, because the $d'$-contingent saving correspondence is monotonic such a lower bound exists. An approximation to it can be recovered by finding the value of $\hat{z}_{i_0}(d';y)$ that solves the following equation

$$u(\hat{z}_{i_0}(d';y),d') + \tilde{V}(0,d',y) = u(\hat{z}_{i_0}(d';y) - a'_{i_1},d') + \tilde{V}(a'_{i_1},d',y), \tag{21}$$

where $a'_{i_1} = a'^{d'}(\hat{z}_{i_1}(d';y),y)$.

The solution $\hat{z}_{i_0}(d';y)$ is the value of total resources for which the global optimum switches from $a'_1 = 0$ to $a'_{i_1}$. Adding the point $(\hat{z}_{i_0},0)$ as the first point of the vector of interpolating nodes for the $d'$-contingent saving correspondence, and the associated value $V^{d'}(\hat{z}_{i_0}(d';y),y) = u(z_{i_0}(d';y),d') + \tilde{V}(0,d',y)$, allows to use the same interpolation procedure as in the first case considered.

## 4.2 Monotonicity

Proposition 1 implies that the $d'$-*contingent* saving correspondence $a'^{d'}(z,y)$ is increasing in $z$. As for concave problems, monotonicity can be usefully exploited to economize on the number of comparisons in the global, discretized-VFI maximization step 2.3 in Section 3.3. Since, as discussed in Section 3.3 the global maximization step applies only in the non-concave region $G^{nc}_{a'}(d';y)$ it is only in such region that one needs to exploit monotonicity.

While the optimal saving correspondence is increasing in $z$ (see, e.g., Figure 2) the mapping of *candidate* points $\{\hat{z}_i(d',y),a'_i\}_{i=1}^m$ that satisfy the Euler equation (13) is not necessarily so in the non-concave region, where the marginal utility of future wealth is not monotonic. To see this, consider again Figure 3, where the thick broken line plots

[16] Alternatively , $V^{d'}(z,y) = u(z,d') - u(\hat{z}_1(d';y),d') + V^{d'}(\hat{z}_1(d';y),y)$, as the continuation value $\tilde{V}(0,d',y)$ in equation (16) is the same for all $z \leq \hat{z}_1(d';y)$.

the expected marginal utility $\tilde{V}_a(a', d', y)$ and the upward sloping line plots the marginal utility of current consumption, both as a function of $a'$. Since higher values of $z$ shift the latter curve down, it follows that $\hat{z}_5(d', y) < \hat{z}_1(d', y)$ in Figure 3, although $a'_5 > a'_1$.

Note that the current marginal utility locus through a point $a'_i$ partitions the non-concave region $G^{nc}_{a'}(d'; y)$ into the set of points associated with $\hat{z}_j(d', y) > \hat{z}_i(d', y)$ and lying to the right of the curve and the complementary set of points to the left of the curve for which $\hat{z}_j(d', y) < \hat{z}_i(d', y)$. To exploit monotonicity, it is convenient to sort the set of points $\{\hat{z}_i(d', y), a'_i\}_{i=1}^m$ in increasing order of $z$ and apply the discretized-VFI maximization step 2.3 in Section 3.3 to the permuted sequence in the sequence's order. Let $\{\hat{z}_{oi}(d', y), a'_{oi}\}_{i=1}^m$, denote the reordered set of pairs where $oi$ indexes the permutation of the sequence $i = 1, \cdots, m$ associated with the reordering. In terms of Figure 3, we have $o1 = 5, o2 = 8, o3 = 1, \ldots$

Suppose that one is considering whether some point $a'_{oi}$—e.g. point $a'_{o3} = a'_1$ in Figure 3—is a global maximum for $z = \hat{z}_{oi}(d', y)$. Finding the global maximum associated with $z = \hat{z}_{oi}(d', y)$ for the "true" problem would require comparing the value of the maximand in equation (18) at the three intersections of the upward-sloping curve through $a'_1$ and the thick downward-sloping curve. In fact, though, the value of the maximand can be computed only at the points of the discrete grid. Therefore, the value of the maximand at $a'_1$ has to be compared to its value at the right and left bracketing points of the two intersections to the north-east.

Consider first the right bracketing points; i.e. points such $\hat{z}_{oj}(d', y) > \hat{z}_{oi}(d', y)$. By construction they are points $a'_{oj}$ with $j > i$. Bracketing the intersections implies locating points $a'_{oj}$ lying on upward-sloping curves close to the upward-sloping curve through $\hat{z}_{oi}(d', y)$; i.e., points associated with values of cash at hand $\hat{z}_{oj}(d', y)$ close to $\hat{z}_{oi}(d', y)$. It follows from the ordering of points that their index $oj$ is relatively close to $oi$. Therefore, one needs to compare $a'_{oi}$ only to a small number of points to its right.

One can of course conceive of pathological cases in which considering a (too) small number of points may not bracket all the intersections; e.g. a grid similar to that in Figure 3 that places lots of points close to the first two intersections, but such that the right bracketing point for the third intersection— point $a'_9$—is far from it. Even that may not be problematic though. The very fact that $a'_9$ is far from the intersection is likely to imply that it cannot improve upon $a'_{oi}$ even if the global maximum associated with $z = \hat{z}_{oi}(d', y)$ in the "true," continuous problem is the point associated with the highest intersection.[17,18]

Consider now the left bracketing points $a'_{oj}$ with $j < i$. There are two cases to distinguish. In the first case, all points $a'_{oj}$, $j < i$, have been discarded. Since, they have been considered sequentially starting from the first one, they must have been improved upon by points bracketing the curve through them from the right. It follows, by induction, that all of them are improved upon by $a_{oi}$ and, by monotonicity, cannot improve upon $a_{oi}$ for $z = \hat{z}_{oi}(d', y)$. Therefore, one does not need to consider any $a'_{oj}$ with $j < i$ if all of them have been discarded. Suppose, instead, some $a'_{oj}$, $j < i$, has already been established as a global maximum. Among these, let $a'_{ok}$ be the point such that $\hat{z}_{ok}(d', y)$ is closest to

[17]This is another aspect of the need for a fine enough grid discussed in Section 3.3.

[18]In the code used for the numerical part, we compare each candidate point to the closest ten to its right. Even reducing this number down to four produced the same results as applying the discretized VFI step to the whole non-concave region.

$\hat{z}_{oi}(d', y)$; e.g $a'_{ok} = a'_5$ in Figure 3. A first, and standard[19], implication of monotonicity is that in solving equation (18) for $z = \hat{z}_{oi}(d', y)$ it is not necessary to compare $a'_{oi}$ to any point other than $a'_{ok}$. All points $a'_{oj}$ with $j < k$ have either been discarded or are optima for $\hat{z}_{oj}(d', y) < \hat{z}_{ok}(d', y)$. Either way they cannot improve upon $a'_{ok}$ for $z > z_{ok}(d', y)$.

A final implication of monotonicity is that if a point $a'_{ok}$—e.g. $a'_{ok} = a'_5$— has been established as a global maximum, one can discard, without even solving (18), all points $a'_{oi} < a'_{ok}$—e.g. points $a'_1$ to $a'_4$ in Figure 3—since, by construction, $z_{oi}(d', y) > \hat{z}_{ok}(d', y)$.

# 5 Results

## 5.1 Parameterization

The parameterization follows Bajari et al. (2009) along a number of dimensions. The chosen felicity function is

$$u(c, d') = \frac{1}{\tau} \log(\theta c^{\tau} + (1 - \theta)\kappa(d' + \iota)^{\tau}), \tag{22}$$

where $\iota$ is a number small enough to be irrelevant for our quantitative exercises, but makes the utility function finite for $d' = 0$.[20] As in Bajari et al. (2009), the durable flow equivalent is $\kappa = 0.075$, the non-durable share $\theta = 0.77$ and the fractions of human and durables wealth that can be collateralized are respectively $\gamma = 0$ and $\xi = 0.2$. Note, that the latter two parameter values satisfy Assumption 5 and imply that the borrowing limit binds with positive probability. The intermediation fee is set to $\phi = 0.06$.

The income process is a discrete approximation to a log-normal process with a persistent and transitory components as in Storesletten, Telmer and Yaron's (2000)[21]. Namely,

$$\begin{aligned} \log y_t &= z_t + \epsilon_t \\ z_t &= \rho z_{t-1} + \eta_t, \end{aligned}$$

with $\epsilon_t, \eta_t$ distributed independently according to $N(0, \sigma_\epsilon),\ N(0, \sigma_\eta)$.

The Markov chain approximation to the process follows Tauchen (1986). The number of grid points for both the transitory and persistent components is 7 which implies that $y$ can take 49 discrete states.

I choose seven uniformly-spaced points for the durable choice stock and a double exponential grid for assets $a$. The upper bounds on $a$ and $d$ equal approximately 25 and 10 times unconditional average income.[22] These values are large enough to ensure: (1) that the upper bound of the stationary distribution for $a$ is below the highest grid point, and (2) that the upper bound on $d$ does not constrain the durable choice.

[19] This is the usual way monotonicity is used to speed up the solution of concave problems.

[20] The, fully equivalent, alternative of choosing a strictly positive lower bound for the discrete choice set $D$ would have implied slightly less neat parametric restrictions in Assumption 4.

[21] The estimates are from row D. in their Table 1. The permanent, individual-specific random effect is not included as it would play no role in the present set up.

[22] It follows from equation (5) and the zero lower bound on the durable choice that the upper bound on non-normalized wealth $w$ also equals 25 times unconditional average income.

Table 1: Chosen parameters.

| Parameter | $\beta$ | $r$ | $\kappa$ | $\phi$ | $\xi$ | $\sigma_y$ | $\rho$ | $\sigma_\eta$ |
|---|---|---|---|---|---|---|---|---|
| Value | 0.93 | 0.06 | 0.075 | 0.06 | 0.20 | 0.063 | 0.977 | 0.024 |

Finally, the interest rate is set to $r = 0.06$, roughly in line with average real mortgage rates, and the discount rate is set to $\beta = 0.93$ to ensure boundedness of the wealth distribution. The chosen values for parameters are collected in Table 1.

The parameter $\tau$ governing the elasticity of substitution takes different values in the simulations. In most of the simulations it equals zero, which implies the Cobb-Douglas specification

$$u(c, d') = \theta \log(c) + (1 - \theta) \log(\kappa(d' + \iota)),$$

used in Fernández-Villaverde and Krueger (2010). Under this specification the first-order condition (13) is linear in total resources as discussed in Section 3. The parameter $\iota$ is set to 0.01.

I also experiment with $\tau = 0.2435$, as estimated in Bajari et al. (2009), to assess how the speed of the algorithm is affected by the non-linearity in total resources of the first-order condition.

## 5.2 Numerical results

VFI is the standard method of choice for non-concave and/or non-differentiable problems. It is, therefore, natural to compare the accuracy and speed of the modified EGM algorithm to those of VFI. Though, most papers solving problems in this class discretize all continuous state variables, there is just one such variable in the above model. For this reason, it seems appropriate to compare the modified EGM algorithm to a version on VFI that does not discretize the continuous variable. A natural choice is linearly-interpolated VFI which deals with the non-concavity by bracketing the maximum over the discrete grid and then switching to linear interpolation on the bracketing interval.[23]

There are two possible ways to generate the grid for total resources $z$ for the VFI algorithm. The first, "brute force" approach is to solve directly problem (15) at all points $(d'; a, d, y) \in D \times G_a \times D \times Y$. The alternative is to choose an exogenous grid for total resources $G_z$ and solve the intermediate problem (16) for all points $(d'; z, y) \in G_z \times D \times Y$, using interpolation to recover $\mathbb{V}^{d'}(a, d, y)$ on the grid $D \times G_a \times D \times Y$. In fact, the second approach solves the same problem as EGM, with the only difference that the grid $G_z$ is exogenous. If $N_I$ denotes the cardinality of a set $I$, the first method solves the maximization problem at $N_{G_a} \times N_D^2 \times N_Y$ against $N_{G_z} \times N_D \times N_Y$ points. That is the two choices, denoted respectively VFI1 AND VFI2 in what follows, imply a trade-off between accuracy and speed. For this reason, I report results for both of them. The sizes of the two grids $G_a$ and $G_z$ are chosen to be the same for ease of comparison,

[23]This is also the VFI algorithm used in Barillas and Fernández-Villaverde (2007).

Finally, since the algorithm in this paper exploits the monotonicity of the policy correspondence, monotonicity is also exploited when solving the model using VFI, so as not to bias the comparison between EGM and VFI.

To compare the accuracy of the two algorithms I compute Euler equation errors following Judd (1992). If $s = (a, d, y)$ denotes the state vector, the Euler equation

$$u_c[c(s), d'(s)] = \beta(1+r)\mathrm{E}u_c[c(a'(s), d'(s), y'), d''(a'(s), d'(s), y')] \tag{23}$$

should hold exactly for the true policy correspondences off corners. Given that the computed policy correspondences are only approximations, equation (23) does not hold exactly when evaluated with the computed policies.

Let $c^*(s)$ denote the solution to

$$u_c[c^*(s), \bar{d}'(s)] = \beta(1+r)\mathrm{E}u_c[\bar{c}(\bar{a}'(s), \bar{d}'(s), y'), \bar{d}'(\bar{a}'(s), \bar{d}'(s), y')], \tag{24}$$

where bars over variables denote the approximate policy correspondences. The (absolute) Euler equation error measured in units of current consumption can then be written as

$$EE(s) = \left| 1 - \frac{c^*(s)}{\bar{c}(s)} \right| \tag{25}$$

for any point of the state space $s$.

An Euler error $EE(s)$ equal to one per cent means that the agent is making a mistake of one cent for each dollar spent. Following Judd, I report the base 10 logarithm of the Euler error. Therefore, a one per cent error in (25) corresponds to an Euler error of -2.

As standard in the literature, I report both the maximum and the average of the Euler errors along a simulated path. To construct the two measures, I draw a 50,000-period income history. This together with the policy correspondences generates a history for the whole state vector. Since the Euler equation does not have to be satisfied at the borrowing constraint, the maximum and average Euler errors are computed over the set of points in the state space where the borrowing constraint is slack.

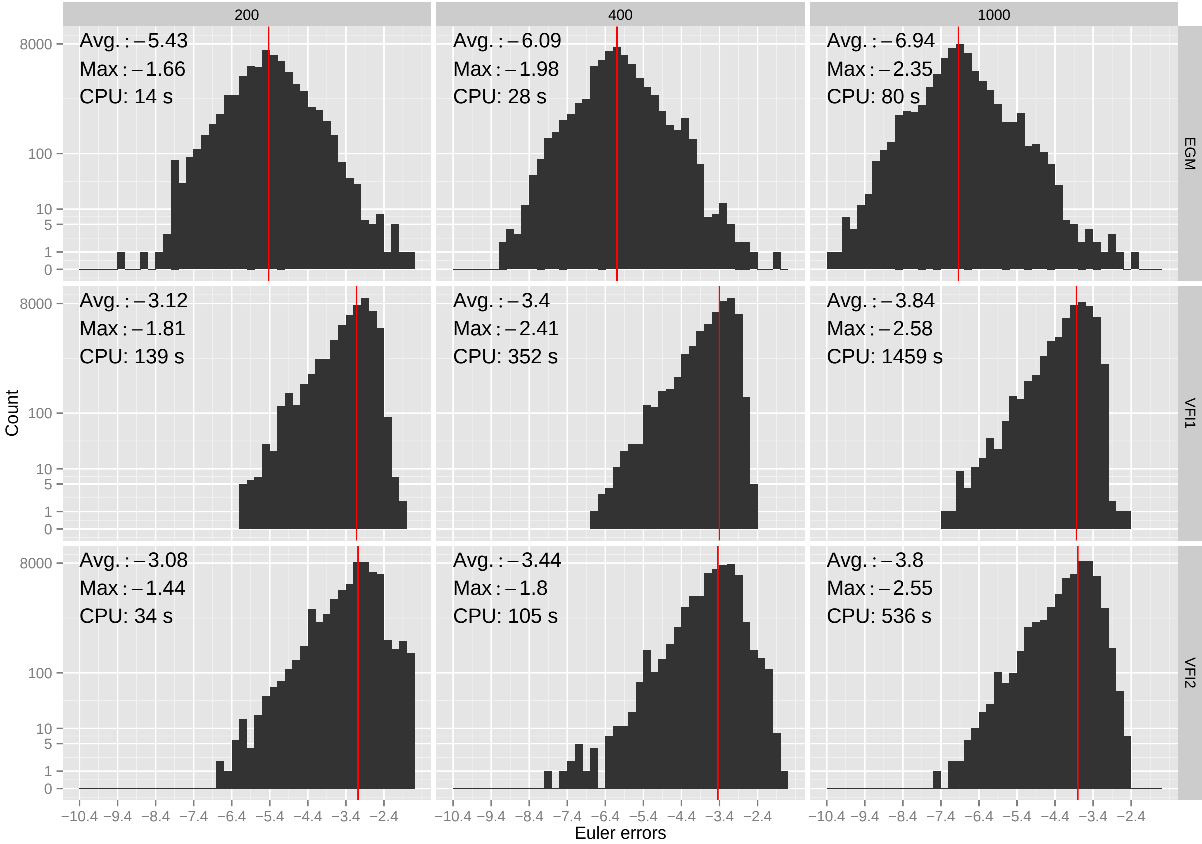
The chosen initial conditions are $a_0 = d_0 = 0$ and the unconditional average of the income process. All the computations were carried out on a single core of a Xeon X5570 processor. The programs were written in `Fortran 95`. The code is available at http://webspace.qmul.ac.uk/gfella/research/research.html for download.

The model is first simulated with $\tau = 0$ and a grid of 200, 400 and 1000 for the continuous state variable $a$. Figure 4 reports execution time, the two error metrics and plots the size distribution and the average of the Euler errors over the simulated history, for each of the nine combination of the asset grid dimensions and solution methods.

In terms of computational time, EGM is from 2.5 to 6.5 faster, depending on the grid size, than VFI2 and from 10 to 18 times faster than VFI1.[24] This is very encouraging and not completely expected. The EGM algorithm has to use a global comparison method over the subset of the state space where the problem is non-concave. Over such a subset the method has no computational advantage compared to VFI2. It turns out that, when the policy correspondence is discontinuous, applying monotonicity to the Euler equation,

[24]All three methods converge in the same number of iterations, 117 when the asset grid has 1000 points.

Figure 4: Accuracy and speed of the algorithms for various asset-grid sizes.[a]




[a]Note: Avg. and Max are respectively the average and maximum Euler error, off corners, on a stochastic simulation of 50,000 observations. CPU is CPU time in seconds.

as discussed in Section 4.2, allows to discard a larger number of candidate points than in the standard application of monotonicity in VFI.

In terms of accuracy, EGM produces an average approximation error of an order between about $-5$ and $-7$, two to three orders of magnitude smaller than either version of VFI.

Yet, comparing the maximum Euler errors along the simulated path does not show a clear superiority of EGM. In particular, on all grid sizes VFI1 produces a smaller maximum error.

The reason why EGM is not necessarily more accurate according to this last metric is apparent once one realizes that the true consumption and saving correspondences are discontinuous and that they are approximated by interpolation. As long as the true policy correspondences jump between two interpolating nodes, the Euler equation evaluated at their approximations may be significantly violated at any point in between. This is true independently from the algorithm used. Therefore, the last summary statistics may not be very informative about the upper tail of the error size distribution, in the presence of discontinuities in the policy correspondences. This can be better understood by looking at the histograms of the size distribution of the Euler errors in Figure 4. The histograms are constructed by dividing a, common, error-size range into bins of size 0.2. Note that the number of error realizations falling into each bin is plotted on a logarithmic vertical

Table 2: CPU time and accuracy for $\tau = .2435$ (1000 asset grid points).

| Model | CPU (s) | Average . Euler error | Maximum . Euler error | # of Euler errors $> -3$ (‱) |
|---|---|---|---|---|
| $\sigma = 1,\ \tau = .2435$ | | | | |
| EGM | 131 | -6.98 | -4.29 | 0 |
| VFI1 | 3589 | -3.83 | -2.97 | 0 |
| VFI2 | 861 | -3.88 | -2.94 | 6 |

scale, to better visualize what happens at the upper bound of the error distribution. The fact that the total histogram area may appear to differ across sub-figures, despite the number of errors being about the same—around 40,000 errors off-corners—in all the nine cases, is an artifact of the logarithmic scale.

Consider the largest difference in the maximum Euler error between EGM and VFI1 which obtains in the case of a 400-point asset grid case. Looking at the whole error distribution reveals that it is only one error out of 40,000 where EGM underperforms VFI1. The same is true for the 200-point grid. Apart from this error, the EGM error distribution is first-order stochastically dominated by the error distributions of VFI1 and VFI2. Not only this is true for the same asset-grid size. The error distribution for EGM with 200 grid points compares extremely favorably with those for VFI1 and VFI2 with 1000 points, with the exception of less than 10 errors out of 40,000. This against a difference in computational time of 38 and 100 times.

Finally, Table 2 conducts the same analysis for $\tau = 0.2435$, the value estimated in Bajari et al. (2009). When $\tau$ differs from zero the consumption aggregator is no longer a power function. Therefore, the Euler equation (13) is non-linear in total resources $z$, as the marginal utility of consumption has no analytic inverse. While non-linear, the Euler equation is twice differentiable with respect to total resources $z$—but not with respect to $a'$—and can be solved for $z$ using Newton method.

The chosen size of the asset grid is 1000 points to minimize the impact of overheads. Table 2 reports computation time and the maximum and average Euler error as in Figure 4. It also reports the number of Euler errors (per 10,000 observations) exceeding -3, as a way of summarizing, the right tail of error size distribution.

The results in Table 2 make clear that, while, as expected, the change in $\tau$ increases significantly computational time for all methods, it leaves their performance, as measured by the average error and the number of errors exceeding -3, virtually unaffected. If anything, the relative advantage of EGM in terms of computational time increases, in particular relative to VFI1.

Finally, though the generalized EGM algorithm yields significant benefits, relative to VFI, in terms of accuracy and speed, one may wonder about its implementation costs. A standard, although, imperfect measure of implementation complexity is the number of lines of source code. The natural comparison is against VFI2 which entails the same state space. I report results just for the core code solving the decision problems, all the rest of the code being virtually identical between the two algorithms. The code for solving the decision problem in VFI2 contains 247 lines. The counterpart for EGM is 269 lines

Table 3: Comparison of solution methods with two continuous choices.

| Model | CPU (s) | Average Euler error | Maximum Euler error | # of Euler errors $> -3$ (‱) |
|---|---|---|---|---|
| EGM | 341 | -5.41 | -1.80 | 8 |
| VFI1 | 7190 | -3.14 | -1.81 | 4640 |
| VFI2 | 860 | -3.20 | -1.56 | 3930 |
| Discretized VFI1 | 3046 | -2.54 | -1.78 | 8629 |
| Discretized VFI2 | 743 | -2.70 | -1.68 | 7939 |

of original code and 197 lines of, off-the-shelf, Newton solver and sorting subroutine.

## 5.3 Continuous durable choice

This section evaluates the performance of the generalized EGM algorithm when applied to non-concave problems with more than one continuous choice. Instances of these problems are studied in Luengo-Prado (2006), Bloom (2009) and Bajari et al. (2009). Typically, these problems are solved by discretizing *all* state variables and applying VFI to the discretized problem.

Discretization requires a fine enough grid to ensure a reasonable accuracy and runs therefore quickly into the curse of dimensionality. The generalized EGM algorithm in this paper does away with discretization of *one* of the continuous variables, requiring a coarser grid for that variable, for the same accuracy.

In what follows, the durable choice is assumed to be continuous, but the fixed-adjustment-cost component still implies that the problem is non-concave.

The EGM algorithm is applied by discretizing the durable choice alone. Compared to Section 5.2 the difference is that now the size of the grid for the durable choice is relative large—50 rather than 7 points. A grid of 200 points for the continuous asset variable is chosen to facilitate the comparison with Section 5.2. For the same reason, the parameterization is also unchanged.

As in Section 5.2, the performance of the EGM algorithm is compared to that of the two versions of VFI that treat the asset variable as continuous. In addition, EGM is also compared to their discretized-state-space counterparts that constitute the typical solution method for this class of problems.

Table 3 reports computation time and the usual measures of accuracy. The first three lines of the table can be understood by comparing them to the 200-grid column in Figure 4. With only 200 asset grid points, EGM does significantly better than both version of VFI also in terms of number of errors exceeding -3. The discrete-state-space versions of VFI perform even worse both in terms of average Euler error and of the number of errors in excess of -3. Somewhat surprisingly, both versions of VFI2 are more accurate than their VFI1 counterparts in terms of the number of Euler errors exceeding -3.

# 6 Conclusion

This paper has presented an extension of Carroll's (2006) EGM, and its combination with VFI by Barillas and Fernández-Villaverde (2007), to non-concave, and possibly non-differentiable problems. The proposed algorithm yields substantial gains in accuracy and computational time compared to VFI.

I have illustrated the algorithm in the context of a problem with one continuous non-durable and one discrete durable choice and fixed adjustment costs. Yet, the generalized EGM is also a much faster and accurate alternative to discretized VFI to solve non-concave problems with multiple continuous variables. It improves accuracy by avoiding discretization in one dimension, while at the same time increasing computation speed thanks to the efficiency of the EGM step.

# References


Abbott, Brant, Gallipoli, Giovanni, Meghir, Costas and Violante, Giovanni L. (2013), Education policy and intergenerational transfers in equilibrium. NBER Working Paper 18782.

Aiyagari, S. R. (1994), 'Uninsured idiosyncratic risk, and aggregate saving', *Quarterly Journal of Economics* **109**(3), 659–684.

Bajari, Patrick, Chan, Phoebe, Krueger, Dirk and Miller, Daniel P. (2009), A dynamic structural model of housing demand: Estimation and policy implications. Mimeo.

Barillas, Francisco and Fernández-Villaverde, Jésus (2007), 'A generalization of the endogenous grid method', *Journal of Economic Dynamics and Control* **31**, 2698–2712.

Bloom, Nicholas (2009), 'The impact of uncertainty shocks', *Econometrica* **77**, 623–685.

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

Clausen, Andrew and Strub, Carlo (2012), Envelope theorems for non-smooth and non-concave optimization. Mimeo, University of Pennsylvania.

Edlin, Aaron S. and Shannon, Chris (1998), 'Strict monotonicity in comparative statics', *Journal of Economic Theory* **81**, 201–219.

Fernández-Villaverde, Jesús and Krueger, Dirk (2010), 'Consumption and saving over the life cycle: How important are consumer durables?', *Macroeconomic Dynamics* . forthcoming.

Gomes, Joao, Greenwood, Jeremy and Rebelo, Sergio (2001), 'Equilibrium unemployment', *Journal of Monetary Economics* **48**(1), 109–152.

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

Judd, Kenneth L. (1992), 'Projection methods for solving the aggregate growth model', *Journal of Economic Theory* **58**, 410–452.

Khan, Aubhik and Thomas, Julia K. (2008), 'Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics', *Econometrica* **76**(2), 395–436.

Luengo-Prado, María José (2006), 'Durables, nondurables, down payments and consumption excesses', *Journal of Monetary Economics* **53**, 1509–1539.

Rust, John P. (1989), A dynamic programming model of retirement behavior, *in* David A.Wise, ed., 'The Economics of Aging', University of Chicago Press, pp. 359–398.

Stokey, Nancy L., Lucas, Robert E. and Prescott, Edward C. (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press.

Storesletten, Kjetil, Telmer, Chris I. and Yaron, Amir (2000), Consumption and risk sharing over the life cycle. NBER Working Paper 7995.

Tauchen, George (1986), 'Finite state Markov-chain approximations to univariate and vector autoregressions', *Economics Letters* .

# A Proofs

**Proof of Proposition 1**. Theorem 1 in Edlin and Shannon (1998) implies that an interior maximizer $x^*(t) \in \arg\max_x g(x,t)$ of a function $g(x,t)$ is strictly increasing in $t$ if $\partial g/\partial x$ is increasing in $t$ at $x^*(t)$. Under the maintained assumptions, Theorem 3 in Clausen and Strub (2012) implies that, for given $(d, y, d')$, the objective function on the right hand side of (16) is differentiable in $a'$ at an interior optimum, with partial derivative with respect to $a'$ equal to the right hand side of (13). Since the right hand side of (13) is strictly increasing in $a$, Theorem 1 in Edlin and Shannon (1998) applies. $\square$

# B Sensitivity analysis

This section reports a sensitivity analysis for a number of parameters. When not explicitly reported, parameter values are the same as in Section 5.2.

All three methods converge in 87 iterations when the convergence tolerance is 1e-4, as opposed to 117 iterations when the tolerance is 1e-5. Reducing the tolerance by an order of magnitude has hardly any effect on accuracy.

Table 4: Sensitivity analysis (1000 asset grid points).

| Model | CPU (s) | Average Euler error | Maximum Euler error | # of Euler Errors $> -3$ (‱) |
|---|---|---|---|---|
| $\beta = .92, \sigma = 1, \tau = 0$ | | | | |
| EGM | 69 | -6.93 | -2.09 | 10 |
| VFI1 | 1414 | -3.74 | -2.22 | 3 |
| VFI2 | 499 | -3.85 | -2.00 | 30 |
| $\beta = .94, \sigma = 1, \tau = 0$ | | | | |
| EGM | 82 | -6.85 | -2.83 | 0 |
| VFI1 | 1270 | -3.98 | -2.51 | 0 |
| VFI2 | 527 | -3.99 | -1.94 | 5 |
| $\beta = .93, \sigma = 2, \tau = .2435$ | | | | |
| EGM | 153 | -6.79 | -2.50 | 1 |
| VFI1 | 4182 | -3.88 | -2.77 | 0 |
| VFI2 | 934 | 3.82 | -2.16 | 4 |
| $\beta = .93, \sigma = 3, \tau = .2435$ | | | | |
| EGM | 148 | -6.75 | -2.23 | 2 |
| VFI1 | 3897 | -3.94 | -2.31 | 1 |
| VFI2 | 881 | -3.92 | -2.21 | 96 |
| Convergence tolerance=1e-4 | | | | |
| EGM | 58 | -6.93 | -2.34 | 1 |
| VFI1 | 1101 | -3.84 | -2.67 | 1 |
| VFI2 | 406 | -3.81 | -2.58 | 12 |