Optimal consumption under uncertainty, liquidity constraints, and bounded rationality*

Ömer Özak†
Department of Economics
Southern Methodist University

October 9, 2013

Abstract

I study how boundedly rational agents can learn a "good" solution to an infinite horizon optimal consumption problem under uncertainty and liquidity constraints. Using an empirically plausible theory of learning I propose a class of adaptive learning algorithms that agents might use to choose a consumption rule. I show that the algorithm always has a globally asymptotically stable consumption rule, which is optimal. Additionally, I present extensions of the model to finite horizon settings, where agents have finite lives and life-cycle income patterns. This provides a simple and parsimonious model of consumption for large agent based models.

Key Words: Adaptive learning models, bounded rationality, dynamic programming, consumption function, behavioral economics, saving behavior
JEL classification: C6, D8, D9, E21

---

*I would especially like to thank Peter Howitt for his comments, encouragement, and support during the writing of this paper. Additionally, I would like to thank Herbert Dawid, Kfir Eliaz, Glenn Loury, Santanu Roy, Tim Salmon, David Weil, and two anonymous referees for comments and helpful discussions.

†Email: omer@omerozak.com. Tel: +1 (214) 768-2755. Fax: +1 (214) 768-1821. Send correspondence to: PO Box 0496, Southern Methodist University, Dallas, TX 75275-0496, USA.

1 Introduction

Rationality is one of the main tenets of modern economics and though it has proven fruitful in all areas of economics, it has recently been subject to attacks both on theoretical and empirical grounds. In particular, the modern theory of consumption under liquidity constraints and uncertainty, which is one of the main building blocks of modern macroeconomics, has been criticized for its rationality requirements. For example, Carroll (2001) presents this theory and argues that "when there is uncertainty about the future level of labor income, it appears to be impossible under plausible assumptions about the utility function to derive an explicit solution for consumption as a direct (analytical) function of the model's parameters". Similarly, Allen and Carroll (2001) admit that "finding the exact nonlinear consumption policy rule (as economists have done) is an extraordinarily difficult mathematical problem".

In order to answer this line of critiques, economists have tried to provide bounded rationality foundations to optimal behavior, especially within game theory (Fudenberg and Levine, 1998) and macroeconomics (Evans and Honkapohja, 2001; Sargent, 1993). Still, the study of how agents learn the optimal policy to an infinite horizon dynamic programming problem under uncertainty, and the consumption function in particular, has been ignored except for a few exceptions (Allen and Carroll, 2001; Evans and McGough, 2009; Howitt and Özak, 2009; Lettau and Uhlig, 1999).[1] While theoretical results

[1]There is a large literature which studies dynamic programming problems in which agents do not hold Rational Expectations, but are otherwise fully rational. The objective of this literature is to understand the conditions under which the expectational mechanism held by agents converges to Rational Expectations (Sargent, 1993). This is not the problem I am alluding to here. In this setting agents are not able or willing to solve the optimal

have been mixed, empirical and experimental evidence suggests that agents *do learn to behave as if* they had solved the optimal consumption problem (see Brown, Chua and Camerer, 2009, and references therein).

In this paper I study an infinite horizon optimal consumption problem under uncertainty, liquidity constraints, and bounded rationality. I follow the previous literature in assuming that boundedly rational agents use a consumption rule that is linear in wealth.[2] I endow agents with a learning algorithm, which I call the HO-algorithm, that is a generalization of the one studied in the numerical exercise of Howitt and Özak (2009). I show that the behavior of agents using this learning scheme converges to a unique consumption rule that has good welfare and stability properties. In particular, in a steady state agents would not have an incentive to choose a different linear consumption function, since their welfare under the current rule would be maximal. In contrast to the previous literature, these results are based on analytical and not numerical tools. Additionally, I extend the analysis to situations where the time horizon is finite. In particular, I study the properties of applying this algorithm during a finite period of time, when agents have a life-cycle profile of income. The results suggest that the algorithm keeps its properties in this setting.

---

consumption problem, even if they had the correct expectational mechanism.

A related literature studies the problem of convergence of the computational methods applied to solve numerically dynamic programming problems, for example Puterman and Brumelle (1979) and Santos and Rust (2004). Although I am not studying this problem, one could apply the methods of this paper to find the optimal partition of the state space or to approximate the optimal solution.

[2]Gabaix (2011) has suggested that boundedly rational agents only use "sparse" rules of behavior. In this paper, the assumption is that agents focus only on wealth and disregard all other variables. As can be seen from the results and proofs below, they can be extended to include linearity in other variables, without affecting the results.

My analytical results generalize the numerical ones found by Howitt and Özak (2009), while overcoming the two main drawbacks of their setting. In particular, it is not clear if one should expect their results to hold in other settings. Additionally, they assume consumers can perform some complex mathematical operations, which might not be a desirable assumption in a bounded rationality setting. I solve these problems by showing that in a general class of consumption problems under uncertainty and liquidity constraints, agents using variations of the HO-algorithm learn an optimal consumption rule. I show that the HO-algorithm converges to the unique asymptotically stable point of a particular ordinary differential equation (ODE) and that this stationary point is "optimal". In particular, this stationary consumption rule maximizes her steady state expected life-time utility under the stationary distribution generated by her consumption rule, so that she has no incentives to change it in a steady state.

This implies that applying the HO-algorithm to *any* initial linear consumption rule in an uncountable and compact set, for different assumptions about an agents' rationality, her level of risk-aversion or impatience, or her income process, will with probability one converge to the globally asymptotically stable stationary point of this ODE, which is optimal. This not only solves the critiques encountered by Howitt and Özak (2009) and some of the problems raised in the literature, but might provide new approaches to the study of models with bounded rationality.

The approach to learning that I follow is based on Euler-equations, where agents change their behavior in response to differences between their experienced marginal utility in one period and next period's discounted marginal utility. Thus, in this paper agents react to mistakes in their Euler equation, and adjust their consumption rule if it failed to equalize the marginal utilities of consumption between yesterday and today. This follows from the idea that agents regret their consumption decision if their Euler equation is not satisfied.

This approach is close in spirit to "learning direction theory" (Selten and Buchta, 1999; Selten and Stoecker, 1986) and "regret theory" (Hart, Mas-Colell and Babichenko, 2013; Loomes and Sugden, 1982), which have been proposed as an explanations for behavior observed in various experimental settings. According to these theories, an agent's success or failure changes her behavior in the direction that increases her expected payoff in the following opportunity she has for action. While both learning direction theory, regret theory, and the HO-algorithm explain the direction of change, the HO-algorithm also tells agents by how much they ought to change their behavior. This might suggest the results obtained by Howitt and Özak (2009) depend on the specific details of their implementation. I show bellow that this is not the case and that the algorithm can be varied in many dimensions without affecting its main optimality and convergence properties.

There are various reasons why the HO-algorithm seems like a good candidate for behavior under bounded rationality. First, it does not require complex optimizing behavior by agents, which is a fundamental requirement of any theory of bounded rationality (Selten, 2001). Agents in this theory are only required to compare two marginal utilities in order to make changes. In this aspect it is similar to aspiration adaptation theory (Selten, 1998), reinforcement learning (Börgers and Sarin, 1997), and learning direction theory (Selten and Stoecker, 1986). This simplicity lowers the cognitive capabilities required of agents. Second, it has low informational and computational requirements, which are *independent* of the number of states or possible rules. This is extremely important, since both requirements put a heavy burden on agents' cognitive abilities. In particular, agents only need to remember a small amount of information and know basic algebraic operations. Third, in order to apply the algorithm, agents do not need to fully understand the economic environment in which they are embedded, nor the effects of changes in it. Thus, it gives them guidance even in unfamiliar situations. Fourth, unlike some models of bounded rationality, which are qualitative in nature (e.g. learning direction theory), the quantitative nature of the HO-algorithm allows its use in applied macroeconomic models. Fifth, its similarity to learning direction and regret theory gives it empirical relevance. Finally, as I will show below, various versions of the algorithm can easily accommodate different levels of rationality.

My approach differs from the one used by Lettau and Uhlig (1999) and Allen and Carroll (2001), who use the accumulated performance of a rule as measured by the discounted sum of utilities as a base for their learning mechanisms. In these papers, agents estimate the value function of their respective problem in order to select the best rule. Regrettably, the algorithms put forward in these papers do not converge to the optimal rule or only converge very slowly. Thus they are "not an adequate description of the process by which consumers learn about consumer behavior" (Allen and Carroll, 2001, p.268)

The approach in these two papers has three main drawbacks. First, they require the set of rules and states to be finite. Second, the memory, processing, and rationality requirements increase in the number of rules and states. Third, they cannot determine the welfare properties of the rules that are learnt, especially when the rational rule is not available or if the rational rule is not equivalent to a mix of the available rules.

This paper is most closely related to Evans and McGough (2009), who use Euler equation and shadow price learning schemes. In their approach, agents are forward looking and forecast either shadow prices or the control variable and then choose the control variable optimally according to the first order conditions in their problem. In particular, they show that agents using Euler equation or shadow price learning can learn the optimal solution to a linear quadratic dynamic programming under the same conditions required for a fully rational solution to be found. Their results are encouraging for the whole research agenda. Still, they are obtained under strong rationality assumptions, since agents need to understand their dynamic programming problem well, keep track of many parameters, and know econometric techniques. Furthermore, the optimality results hold only in this case. It is not clear what kind of properties these learning schemes have in the general case when the fully rational solution is not linear or if the agents have a misspecified model.

The paper proceeds as follows: Section 2 presents the model, section 3 introduces the learning algorithm, section 4 studies the convergence, stability, and optimality properties of the algorithm, section 5 presents various extensions and variations of the model, as well as simulations for the finite horizon case, and section 6 concludes. All technical proofs are left for the online-appendix.

2 The model

Time evolves discretely and is indexed by t. Agents are infinitely lived and born with an initial wealth $w_0 > 0$. Every period t they consume an amount c_t out of their current wealth w_t, receive interest on their savings and before taking the next consumption decision, get an income y_{t+1}. So, their wealth evolves according to

$$w_{t+1} = R\big(w_t - c_t\big) + y_{t+1}, \tag{2.1}$$

where $R > 1$ is the interest rate factor on wealth held at the end of each period.

Assumption A. The income process, $\{y_t\}$ is such that $y_t \in \mathcal{Y} \equiv [\underline{y}, \bar{y})$, where $0 < \underline{y} < \bar{y} < \infty$, is identically and independently distributed across periods with distribution Γ that is absolutely continuous with a lower semicontinuous density γ_y with full support on $\mathcal{Y}$.

Each period's consumption level c_t gives her a per period level of utility $u(c_t)$. The utility function $u(\cdot)$ is continuous, strictly increasing, concave, and three times continuously differentiable. I also assume that the utility function and its derivatives are ϕ-bounded, i.e. there exists constants K, $n^* \in \mathbb{N}$ and some continuous function $\phi : \mathbb{R}_+ \to \mathbb{R}_{++}$,[3] such that

$$\begin{aligned} \|u\|_\phi &= \sup \frac{|u(c)|}{\phi(c)} < K, \quad \text{and} \\ \left\|u^{(n)}\right\|_\phi &= \sup \frac{\left|\frac{d^n u(c)}{dc^n}\right|}{\phi(c)} < K, \text{for all } n \leq n^*,\ n^* \geq 3, \end{aligned} \tag{2.2}$$

where $\phi(\cdot)$ is a continuous function for which the level sets $C_\phi = \{w \in \mathbb{R}_+ \mid \phi(w) \leq d,\ d \in \mathbb{R}_+\}$ are compact. Clearly, any utility function that is bounded and has bounded derivatives satisfies these conditions. The agent discounts her per period utility at a rate $\beta \in (0,1)$, so that her lifetime utility level is given by $U = \sum_{t=0}^{\infty} \beta^t u(c_t)$.

I assume agents use a linear consumption rule and may be liquidity constrained, which implies that their current consumption is given by

$$c_t^b(w) \equiv c^b(\alpha_t, w) \equiv c^b(\alpha_t^0, \alpha_t^1, w) = \min\left\{\alpha_t^0 + \alpha_t^1 w, w\right\} \tag{2.3}$$

where $(\alpha_t^0, \alpha_t^1) \in \Lambda \equiv [0, \bar{y}/R] \times [(R-1)/R, 1]$.[4] Below I will endow agents with a learning algorithm that will tell them how to change their consumption rule across periods based on their consumption experience.

For each α let $\tilde{w}_\alpha$ denote the level of wealth below which the consumer is liquidity constrained, so that $\tilde{w}_\alpha = \alpha_0/(1-\alpha_1)$. Also, let $\underline{w}_\alpha$ and $\bar{w}_\alpha$ denote the unique asymptotically stable stationary levels of wealth the consumer would

[3] I assume the same K and ϕ satisfy the conditions below, but one can allow for different functions ϕ to satisfy each of them.

[4] Although one could allow any finite upper bound on α_t^0, $\bar{y}/R$ ensures that the agent will not always be liquidity constrained. To see this, notice that if $\alpha_t^0 > \bar{y}/R$, then $\alpha_t^0/(1-\alpha_t^1) > \bar{y}$, so that after some finite time period the agent will become liquidity constrained forever. This clearly cannot be optimal.

attain if her income was always equal to $\underline{y}$ or $\bar{y}$ respectively. Thus,

$$\underline{w}_\alpha = \underline{y}\mathbb{I}_{\tilde{w}_\alpha \geq \underline{y}} + \frac{\underline{y} - R\alpha_0}{1 - R(1-\alpha_1)}\mathbb{I}_{\tilde{w}_\alpha < \underline{y}}, \qquad \bar{w}_\alpha = \frac{\bar{y} - R\alpha_0}{1 - R(1-\alpha_1)} < \infty, \tag{2.4}$$

where $\mathbb{I}_x$ equals one if condition x is met and zero otherwise.

Figure 1: Dynamics of wealth under a linear consumption function

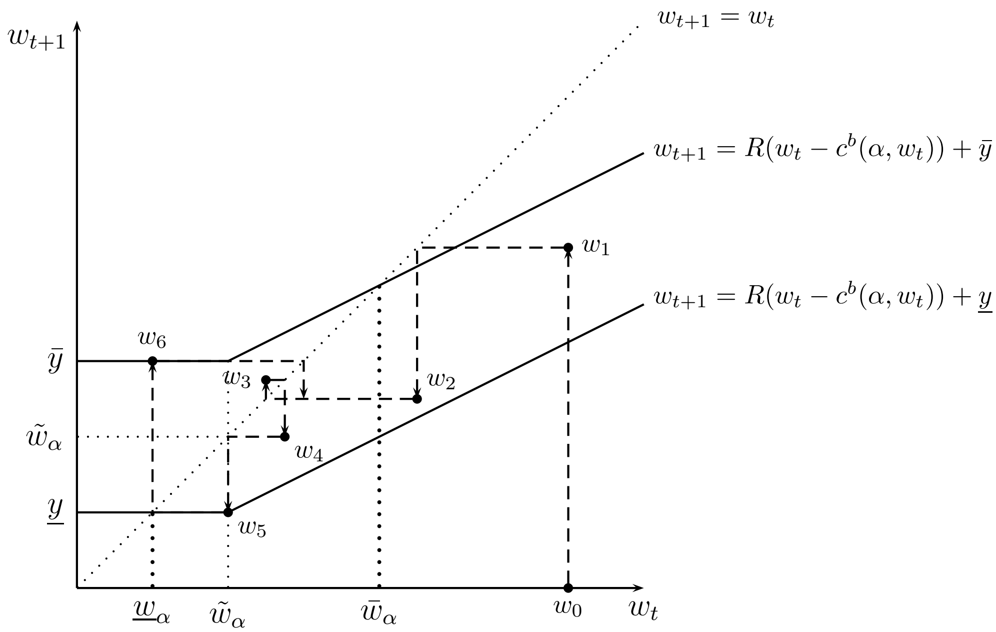


Figure 1 shows the dynamics of wealth when the agent uses a fixed consumption rule such that $\tilde{w}_\alpha \in (\underline{y}, \bar{y})$. In particular, the lines representing the dynamics of wealth when income is fixed at $\underline{y}$ and $\bar{y}$ delimit the region where her wealth can evolve. The points where these lines intersect the 45 degree line determine the location of $\underline{w}_\alpha$ and $\bar{w}_\alpha$. Clearly, for each fixed level of income, the wealth dynamics generated by the budget constraint imply the existence of a unique asymptotically stable stationary level of wealth that belongs to the set $[\underline{w}_\alpha, \bar{w}_\alpha]$. Thus, as can be seen in the figure, starting at any wealth level $w_0 \notin [\underline{w}_\alpha, \bar{w}_\alpha]$, the stochastic dynamics quickly move wealth towards the set $[\underline{w}_\alpha, \bar{w}_\alpha]$, from which it can never leave. In particular, notice that wealth can never become smaller than $\underline{w}_\alpha$, since at this level of income, the agent will consume all her current wealth, and receive a level of income at least as big as $\underline{w}_\alpha$. On the other hand, if she has a level of wealth $\bar{w}_\alpha$, she will choose an amount of consumption such that for any level of income, her wealth level next period will be at most $\bar{w}_\alpha$ again.

This result implies the following theorem:[5]

Theorem 2.1. *If assumption A holds and the consumer uses a fixed consumption rule with parameter α, then there exists a unique stationary distribution of wealth $\pi_\alpha(w)$, which for any $A \subseteq \mathbb{R}_+$ satisfies*

$$\pi_\alpha(A) = \begin{cases} > 0 & \text{if } A \cap [\underline{w}_\alpha, \bar{w}_\alpha] \neq \emptyset \\ = 0 & \text{if } A \cap [\underline{w}_\alpha, \bar{w}_\alpha] = \emptyset \end{cases}, \qquad \pi_\alpha\Big([\underline{w}_\alpha, \bar{w}_\alpha]\Big) = 1. \tag{2.5}$$

Additionally, as $t \to \infty$, the consumer's wealth distribution converges to π_α.

If agents are learning to consume using some algorithm, then the parameters of their consumption function will vary across time, i.e. $\alpha_{t+1} = g(\alpha_t)$, and so will the invariant distribution and all the parameters that depend on α_t. Clearly, if the learning algorithm converges to some $\alpha^e \in \Lambda$, i.e. $\alpha_t \to \alpha^e$, then $\pi_{\alpha_t} \to \pi_{\alpha^e}$, $\tilde{w}_{\alpha_t} \to \tilde{w}_{\alpha^e}$, $\underline{w}_{\alpha_t} \to \underline{w}_{\alpha^e}$, and $\bar{w}_{\alpha_t} \to \bar{w}_{\alpha^e}$. In the following

[5] In the appendix I prove a more general version of this theorem. The strength of that theorem is that it *does not require* wealth to be bounded for the existence of a unique ergodic distribution and it provides an easily verifiable condition for the existence of a unique invariant distribution that can be applied to other types of consumption functions or to more general Markov processes.

two sections I introduce and study the convergence properties of a learning algorithm and the welfare properties of its stationary consumption rule.

3 Learning

I assume that agents adapt their rules in response to past mistakes in their behavior. The main difficulty boundedly rational agents face is that they do not know the value function of following a certain rule α_t. This implies that they need some other measure of adjustment.

As a first approximation, I assume that agents are sophisticated enough to comprehend that if they were not liquidity constrained, they could have raised their utility in period $t+1$, by lowering their consumption in period t by following a different rule. In particular, once agents have consumed in period $t+1$, they compare the discounted marginal utility generated by their *actual* consumption in period $t+1$ using their consumption rule α_t, $\beta R u'(c_t^b(w_{t+1}))$, with the marginal utility they would have gotten with that rule in the previous period, i.e. $u'(c_t^b(w_t))$. An agent's "regret" of using the consumption rule α_t is measured as the difference between those two measures, $\beta R u'(c_t^b(w_{t+1})) - u'(c_t^b(w_t))$.

So, if an agent has positive regret, i.e. $\beta R u'(c_t^b(w_{t+1})) > u'(c_t^b(w_t))$, she understands she should have decreased her consumption last period and increased it in this one. Similarly, if she has a negative level of regret, she knows she should have increased her consumption in the previous period and consumed less in this one. So, she understands she should change her rule's parameters if her regret is not zero. So, as in direction learning theory (Selten, 2004), reinforcement learning (Börgers and Sarin, 1997), and regret theory (Loomes and Sugden, 1982) agents react to mistakes committed in the past.[6,7]

Since her level of regret only tells her the direction in which she should change her consumption, the agent still has to decide how much to change each parameter of her consumption function. Letting M be a positive definite matrix and $\{\kappa_t\}$ a decreasing sequence of positive real numbers, I assume she updates her consumption rule using the following equation:

$$\begin{pmatrix} \alpha^0_{t+1} \\ \alpha^1_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha^0_{t} \\ \alpha^1_{t} \end{pmatrix} + \kappa_t M \left[\left(\beta R u'(c^b_t(w_{t+1})) - u'(c^b_t(w_t)) \right) u''(c^b_t(w_t)) \right] \begin{pmatrix} 1 \\ w_t \end{pmatrix} \mathbb{I}_U \tag{DG}$$

for all $t \geq 0$, where $\mathbb{I}_U$ is equal to one if the agent was unconstrained and zero otherwise. Multiplying her level of regret by the second derivative of the utility function, ensures she takes into account how her marginal utility changes when she updates her rule. This ensures she does not overreact to regret levels, so that she only changes her rule by little if her marginal utility might change a lot with small changes in consumption. Additionally, if M is not the identity matrix, she allows both rule's parameters to change in reaction to her level of regret weighted by her level of wealth.

Clearly, equation (DG) might cause the new parameters to be outside the set $\Lambda = [0, \bar{y}/R] \times [(R-1)/R, 1]$. If that is the case, I assume agents use

[6] Notice that agents look to the past, but do not fully use their past experience. Below I generalize this behavior to allow agents to react to their *actual* regret levels.

[7] As in Börgers and Sarin (1997) and Loomes and Sugden (1982), in this paper experienced regret drives behavior, since it has information regarding the effects of the choice made by the agent. Kahneman (2011) presents other ways in which regret might affect behavior. For example, when comparing two situations, an agent might choose an action that minimizes her expected regret from the action. Kahneman suggests that norms affect the perceived levels of regret. Furthermore, for him perceived regret levels might not convey any information about the problem at hand.

a "projection facility" to select the new parameters, i.e. some ad hoc rule to choose an element of Λ. For example, the agent could choose not to update her parameters whenever the new parameters are outside Λ, or choose a random rule in Λ, or just choose a fixed element in the interior of Λ. The effect of this projection facility is to dismiss the last observation or restart the learning process from a new point. Clearly, the assumption that the agent knows the parameter should remain in the set Λ implies that she understands, that outside Λ her rule would force her to always be liquidity constrained or to accumulate an infinite amount of wealth, none of which can be optimal.

A nice quality of the algorithm is that agents require very little information and keep track of only a few values, *independently* of the number of states or of the possible number of linear consumption rules. Additionally, it requires very low memory, and computational and cognitive abilities on the part of agents, especially when compared with methods that require an estimate of the value function as in Allen and Carroll (2001), which compare many rules simultaneously as in Lettau and Uhlig (1999), or which use econometric techniques to forecast (Evans and McGough, 2009). Of course, this also means the agent does not use all the information she has available, e.g. the full distribution of income, nor has any forward looking behavior that might suggest some kind of precautionary behavior.

4 Convergence and Optimality

The literature on stochastic approximations to recursive algorithms (Benveniste, Métivier and Priouret, 1990; Kushner and Yin, 2003) studies the dynamics of recursive algorithms like (DG) by using a differential equation obtained by averaging the dynamics of the algorithm as time evolves. In this case, the following ordinary differential equation related to the algorithm (DG) is of interest:

$$\begin{pmatrix} \frac{d\alpha_0}{\partial \tau} \\ \\ \frac{d\alpha_1}{\partial \tau} \end{pmatrix} = M \int_{\mathcal{W}} \left[\left(\beta R E_t u'(c^b(w')) - u'(c^b(w)) \right) u''(c^b(w)) \right] \cdot \begin{pmatrix} 1 \\ \\ w \end{pmatrix} \mathbb{I}_U \pi_\alpha(dw) \quad \text{(ODE-DG)}$$

where E_t denotes the expectation under Γ, $c^b(w) = \min\{w, \alpha_0 + \alpha_1 w\}$ and $w' = R(w - c(w)) + y$. Let α^e denote an equilibrium of (ODE-DG). In the following section I study conditions for the convergence of the algorithm to α^e, as well as the optimality properties of this stationary point.

The approach I follow in order to study the convergence and optimality properties of the equilibrium α^e is based on the concept of $\mathcal{D}_\phi$-continuity and $\mathcal{D}_\phi$-differentiability of measures (Heidergott and Vázquez-Abad, 2008). In particular, letting $\mathcal{D}_\phi$ be the set of all continuous and ϕ-bounded functions, the transition probability $P_\alpha(w, A) = \Gamma(w - c_t^b(w) \in A \cap \mathcal{Y})$ is $\mathcal{D}_\phi$-*continuous* if for all w and $u \in \mathcal{D}_\phi$, the expected value of u for given w is a continuous function of α. Similarly, $P_\alpha(w, A)$ is $\mathcal{D}_\phi$-*Lipschitz continuous* if the expected value of u given w is Lipschitz continuous. Additionally, $P_\alpha(w, A)$ is $\mathcal{D}_\phi$-*differentiable* with respect to α_i if for any w there exists a finite signed measure $P'_{\alpha_i}(w, A)$ such that for any $u \in \mathcal{D}_\phi$

$$\frac{d}{d\alpha_i} \int P_\alpha(w, ds) u(s) = \int P'_{\alpha_i}(w, ds) u(s) \in \mathcal{D}_\phi.$$

Heidergott and Vázquez-Abad (2008) show that if $P_\alpha(w, A)$ is $\mathcal{D}_\phi$-differentiable with respect to α_i, then this derivative can be written as $k_i(P_i^+ - P_i^-)$ where P_i^+, P_i^- are probability measures. Clearly, the following assumption is required in order to ensure the $\mathcal{D}_\phi$-continuity and differentiability of P_α.

Assumption B. γ_y is continuously differentiable and Lipschitz continuous.

In the appendix I prove that under assumptions A and B for every α the invariant probability distribution π_α is $\mathcal{D}_\phi$-continuous and differentiable with respect to each α_i. Furthermore, the $\mathcal{D}_\phi$ derivative with respect to α_0, $\partial\pi_\alpha/\partial\alpha_0 = k_0(\pi_0^+ - \pi_0^-)$, and with respect to α_1, $\partial\pi_\alpha/\partial\alpha_1 = k_1(\pi_1^+ - \pi_1^-)$, satisfy $\bar{w}k_0(\pi_0^+ - \pi_0^-) = k_1(\pi_1^+ - \pi_1^-)$ almost everywhere. Notice that since π_α is a probability distribution, then

$$\int_{\underline{w}_\alpha}^{\bar{w}_\alpha} \frac{\partial\pi_\alpha}{\partial\alpha_i}(dw) = 0 \quad i = 0, 1.$$

Before presenting the paper's main results I introduce some additional notation and assumptions. First, for a given w_0 and a fixed consumption rule $c^b(\alpha, w)$ the expected life-time utility of an agent is

$$U(\alpha, w_0) = \sum_{t=0}^{\infty} E_0\Big[\beta^t u\Big(c^b(\alpha, w_t)\Big)\Big], \tag{4.1}$$

where w_t evolves according to (2.1). Clearly, $U(\alpha, w) \in \mathcal{D}_\phi$ and is strictly concave. The agent's ex-ante expected life-time utility of using rule α when initial income is distributed according to π_σ is

$$\int U(\alpha, w)\pi_\sigma(dw). \tag{4.2}$$

A rule α is *σ-optimal* if it maximizes (4.2), i.e. it maximizes the agent's ex-ante expected life-time utility from using α when initial income is distributed according to the invariant distribution generated by rule σ. Additionally, a rule α is *optimal* if it is α-optimal. Optimality here requires stability of the optimal rule in the following sense: if an agent uses an optimal rule α, then in a steady state when wealth is distributed according to the invariant distribution generated by that rule, π_α, an optimizing agent would have no incentive to change her behavior by selecting another rule. Clearly, this property is also satisfied by the fully rational solution. Let

$$EV_\alpha \equiv \int U(\alpha, w)\pi_\alpha(dw) \tag{4.3}$$

denote the agent's ex-ante expected life-time utility from using rule α when initial income is distributed according to the invariant distribution of α.

Second, the agent's σ-expected squared regret is given by

$$\int \Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big)^2 \pi_\sigma(dw). \tag{4.4}$$

Third, let $c^*(w)$ and $v(w)$ denote the fully rational consumption function and the expected life-time utility generated by it. Let $\tilde{w}^*$ be the level of wealth below which a fully rational agent would be liquidity constrained, i.e. $c^*(w) = w$ for all $w \le \tilde{w}^*$. Finally,

Assumption C. The sequence $\{\kappa_t\}_t$ satisfies $\sum_t \kappa_t = \infty$ and $\sum_t \kappa_t^2 < \infty$.

Under these assumptions it follows that:

Theorem 4.1. *There exists a unique rule $\alpha^* \in \Lambda$ such that:*

(i) α^* is α^*-optimal, i.e. optimal.

(ii) α^* minimizes the agent's α^*-expected squared regret.

(iii) α^* is an asymptotically stable equilibrium of (ODE-DG).

(iv) The learning algorithm converges to α^* almost surely.

(v) $\tilde{w}_{\alpha^*} = \tilde{w}^*$ and $\underline{w}_{\alpha^*} = \underline{y}$.

Proof. The full proof requires various intermediate steps and the verification of some conditions, all of which I do in the appendix. Here I explain the proof taking those results as given. First, notice that from the continuity of $U(\alpha, w)$, Weierstrass' theorem ensures the existence of a rule $\alpha_\sigma \in \Lambda$ that is σ-optimal for each $\sigma \in \Lambda$. Strict concavity of the utility function ensures that α_σ is unique. By the $\mathcal{D}_\phi$-continuity of π_σ, α_σ is a continuous function of σ. Thus, Brouwer's fixed point theorem ensures the existence of α^* that is α^*-optimal. Second, one can show that any α that is σ-optimal belongs to the interior of Λ and satisfies $\tilde{w}_\alpha = \tilde{w}^*$.[8] Thus, α^* belongs to the interior of Λ. Given α_σ it is easy to construct a contracting map that shares the same fixed points as α_σ. Then Banach's fixed point theorem ensures uniqueness. Third, using the properties of the fully rational consumption function and its value function it is not difficult to show that α^* is α^*-optimal if and only if it minimizes α^*-expected squared regret (see appendix). Fourth, since α^* is an

[8] This follows from the fact that if α is not in the interior, then $\tilde{w}_\alpha \notin [\underline{y}, \bar{y}]$, and one can find a rule that generates a higher expected life-time utility. Furthermore, this result implies $\underline{w}_{\alpha^*} = \underline{y}$.

interior solution it must satisfy the condition that the first derivatives of the σ-expected squared regret minimization problem with respect to α_0 and α_1 are simultaneously equal to zero. Fifth, using the properties of the invariant distribution of π_{α^*}, one can show that α^* minimizes the α^*-expected squared regret if and only if it is an equilibrium of (ODE-DG). Sixth, in this case the α^*-expected squared regret is a Lyapunov function for equation (ODE-DG), and thus α^* is globally asymptotically stable (Hirsch, Smale and Devaney, 2004, p. 205). Thus, the algorithm converges to α^* (Benveniste *et al.*, 1990). Finally, since α^* is α^*-optimal, it is optimal. ■

This theorem ensures that agent's consumption behavior will converge to a steady state in which wealth is distributed according to the invariant distribution π_{α^*}. In this steady state, agent's expected life-time utility under that wealth distribution is maximal and agents would have no incentives to change their behavior.

Notice that this result holds for a wide class of consumption problems. The most stringent assumption is that $\underline{y} > 0$, which is required in order to ensure Lipschitz continuity of equation (ODE-DG). Thus, $\underline{y} = 0$ can be accommodated, as long as regret levels are Lipschitz continuous. For example, this requirement is satisfied if $u'(c^b(w))$ and $u''(c^b(w))$ are bounded for all $\alpha \in \Lambda$ and $w \in [\underline{y}, \bar{w}_\alpha]$.

Additionally, the matrix M does not play a fundamental role in this result. Thus, it could have been replaced by some other matrix that maintained the "directional" properties of the algorithm. That is, as long as it altered the consumption rule in the opposite direction of the agent's regret level. Thus, the algorithm studied by Howitt and Özak (2009) is a special case of this more general algorithm.

Although convergence to the optimal linear rule is guaranteed, it is useful to have an idea of the speed of convergence to the optimum. In particular:

Corollary 4.2. *If the sequence $\{\kappa_t\}$ satisfies*

$$\liminf_{n\to\infty} 2\delta\frac{\kappa_t^\eta}{\kappa_{t+1}} + \frac{\kappa_{t+1}^\eta - \kappa_t^b}{\kappa_{t+1}^2} > 0 \tag{4.5}$$

where $\eta \leq 1$ and $\delta > 0$ is a lower bound of the norm of the Hessian matrix of the agent's α^*-expected squared regret, then

$$E_{\alpha_0}\Big(\|\alpha_t - \alpha^*\|\Big) \leq \lambda(\alpha_0)\kappa_t^\eta \tag{4.6}$$

for some suitable constant $\lambda(\alpha_0)$. In particular, this result holds with $\eta < 1$ if

$$\kappa_t = \frac{A}{t^\zeta + B}, \quad 0 \leq \zeta \leq 1. \tag{4.7}$$

Thus, under these conditions the algorithm converges to the optimal rule at the same rate as κ_t converges to zero.

5 Extensions and Variations

The previous section showed that the algorithm converges to the unique asymptotically stable equilibrium of (ODE-DG) and that this unique equilibrium is also optimal. In this section I use this result to analyze variations and extensions of this algorithm. In particular, notice that the previous results provide a simple way to analyze variants of the algorithm. Any algorithm that converges to the same equilibrium will have the same properties. Furthermore, any algorithm, which has the same related differential equation (ODE-DG) will have the same asymptotically stable equilibrium.

5.1 Some Variations

In the original algorithm (DG) agents are backward looking, but they use introspection instead of using their past experience directly. A first variant of the algorithm imposes a zero-intelligence behavior, in which agents use their *actual* past experience.

Corollary 5.1 (Zero-Intelligence)**.** *Assume agents' regret is given by past experience, so that they adjust their rule given their* experienced *marginal utilities. Then the algorithm is given by*

$$\begin{pmatrix} \alpha^0_{t+1} \\ \alpha^1_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha^0_{t} \\ \alpha^1_{t} \end{pmatrix} + \kappa_t M \left[\left(\beta R u'(c^b_t(w_{t+1})) - u'(c^b_{t-1}(w_t)) \right) u''(c^b_t(w_t)) \right] \begin{pmatrix} 1 \\ w_t \end{pmatrix} \mathbb{I}_U, \quad (5.1)$$

which has the same properties as the original algorithm. In particular, it converges to the optimal linear rule α^* of theorem 4.1 at the same rate.

One problem with both this version and the original one, is that agents may not be able to start learning, or learning can be slow if they are liquidity constrained. A first step to overcome this problem is to assume that agents start with very low levels of α_0, so that in period 0 they are not liquidity constrained. Second, assume that when agents are determining what their marginal utility last period would have been, they assume they would not have been liquidity constrained. Then:

Corollary 5.2 (Unconstrained). *The algorithm is given by*

$$\begin{pmatrix} \alpha^0_{t+1} \\ \alpha^1_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha^0_t \\ \alpha^1_t \end{pmatrix} + \kappa_t M \left[\left(\beta R u'(c^b_t(w_{t+1})) - u'(\hat{c}^b_{t-1}(w_t)) \right) u''(\hat{c}^b_t(w_t)) \right] \begin{pmatrix} 1 \\ w_t \end{pmatrix}, \qquad (5.2)$$

which has the same properties as the original algorithm. But, it converges to the optimal linear rule α^* of theorem 4.1 at a faster rate.

One problem of these backward looking algorithms is that they do not use all available information. In particular, they do not use the information about the distribution of their income. Clearly, forward looking behavior would take this into account, although it would require more rationality and sophistication by agents.

Corollary 5.3 (Forward Looking). *Assume agents' regret is based on their expected discounted levels of marginal utility, so that*

$$\begin{pmatrix} \alpha^0_{t+1} \\ \alpha^1_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha^0_t \\ \alpha^1_t \end{pmatrix} + \kappa_t M \left[\left(E_t[\beta R u'(c^b_t(w_{t+1}))] - u'(c^b_t(w_t)) \right) u''(c^b_t(w_t)) \right] \begin{pmatrix} 1 \\ w_t \end{pmatrix} \mathbb{I}_U. \qquad (5.3)$$

This algorithm has the same properties as the original algorithm. In particular, it converges to the optimal linear rule α^* of theorem 4.1 at the same rate.

Clearly, in this setting, agents would have precautionary savings. But, this implies that all agents that use these algorithms behave as though they had precautionary savings motives.

Another possible variation of the algorithm is to allow for social learning. In particular, assume that there are N agents in the economy and denote the set of agents by N. Each agent i has a set $N_i \subseteq N$ of friends whose consumption rules she can observe.[9] Let A_{ij} denote the importance agent i gives to agent j, where $A_{ij} \in [0, 1]$, $A_{ij} > 0$ if and only if $j \in N_i$, and $\sum_{j \in N} A_{ij} = 1$. Also, let q_{it} denote the elements after the matrix M in one of the previous algorithms, and α_{it} be the parameters of the consumption rule of agent i in period t. I assume agents now use the following algorithm

$$\alpha_{it+1} = \sum_{j \in N} A_{ij} \alpha_{jt} + \kappa_t M q_t^i, \quad i \in N. \tag{5.4}$$

Thus, in a sense every agent is sampling their friends' parameters, averaging them, and changing this average according to their own regret.

Theorem 5.4 (Social Learning)**.** *If the conditions of corollary 4.2 hold, $A = (A_{ij})_{i,j \in N}$ is regular and diagonizable, and for some $x \in (0, 1)$, $(1 + x) < N^{\eta}$, then the algorithm with social learning converges to the optimal rule α^* at higher rates than the other versions of the algorithm.*

Proof. Reorder the system of equations and denote $\alpha_t^0 = (\alpha_{it}^0)_{i \in N}$ and $\alpha_t^1 = (\alpha_{it}^1)_{i \in N}$ the column vectors of agents' consumption rules. Let $q_t = (q_{it})_{i \in N}$ denote the column vector of agents' regrets. Additionally, let M^0 and M^1 denote the rows of matrix M. Then the algorithm for the population can be

[9] In order to simplify notation I assume every agent is a friend of herself.

written as

$$\begin{pmatrix} \alpha_{t+1}^0 \\ \alpha_{t+1}^1 \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \cdot \begin{pmatrix} \alpha_t^0 \\ \alpha_t^1 \end{pmatrix} + \kappa_t \begin{pmatrix} I_{N\times N} \otimes M^0 \\ I_{N\times N} \otimes M^1 \end{pmatrix} \cdot q_t$$

Since A is regular, its largest eigenvalue is $r^* = 1$ and it is the only eigenvalue with modulus 1. Additionally, any other eigenvalue has modulus less than 1. Also, the vector $e = (1, \ldots, 1)'$ is an eigenvector associated with r^*. As A is diagonizable, it can be decomposed so that $A = \tilde{A}\Lambda\tilde{A}^{-1}$, where Λ is the diagonal matrix of eigenvalues and $\tilde{A}$ the matrix of eigenvectors of A. Let $\tilde{A}$ be such that the first column is e. Then, the first element on the diagonal of Λ is $r^* = 1$ and A^{-1} can be rewritten as $A^{-1} = (\frac{1}{N}e, X'^{-1})'$. Premultiplying on both sides by $I_{2\times 2} \otimes \tilde{A}^{-1}$ the previous equation can be written as

$$\begin{aligned}
\begin{pmatrix} \tilde{\alpha}_{t+1}^0 \\ \tilde{\alpha}_{t+1}^1 \end{pmatrix} &= \begin{pmatrix} \Lambda\tilde{A}^{-1} & 0 \\ 0 & \Lambda\tilde{A}^{-1} \end{pmatrix} \cdot \begin{pmatrix} \alpha_{t+1}^0 \\ \alpha_{t+1}^1 \end{pmatrix} \\
&= \begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda \end{pmatrix} \begin{pmatrix} \tilde{\alpha}_t^0 \\ \tilde{\alpha}_t^1 \end{pmatrix} + \kappa_t \begin{pmatrix} \tilde{A}^{-1} & 0 \\ 0 & \tilde{A}^{-1} \end{pmatrix} \cdot \begin{pmatrix} I_{N\times N} \otimes M^0 \\ I_{N\times N} \otimes M^1 \end{pmatrix} \cdot \tilde{q}_t,
\end{aligned}$$

where $\tilde{q}_t(\tilde{\alpha}_t) = q_t(\alpha_t)$. But, this implies that as $t \to \infty$, $\tilde{\alpha}_{it} \to 0$ if $i \neq r^*$. So, by construction as $t \to \infty$, $\alpha_{it} \to \tilde{\alpha}_{r^*t}$ for all $i \in N$. Thus, the system of equations for $\tilde{\alpha}_{r^*}$,

$$\begin{pmatrix} \tilde{\alpha}_{r^*t+1}^0 \\ \tilde{\alpha}_{r^*t+1}^1 \end{pmatrix} = \begin{pmatrix} \tilde{\alpha}_{r^*t}^0 \\ \tilde{\alpha}_{r^*t}^1 \end{pmatrix} + \frac{\kappa_t}{N} M \tilde{q}_{r^*t},$$

is asymptotically equivalent to (DG). Thus, it is associated with the following differential equation

$$\begin{pmatrix} \frac{d\tilde{\alpha}_{r^*0}}{\partial\tau} \\ \\ \frac{d\tilde{\alpha}_{r^*1}}{\partial\tau} \end{pmatrix} = M \int_{\mathcal{W}} \tilde{q}_{\tilde{\alpha}_{r^*}} \pi_{\tilde{\alpha}_{r^*}}(dw),$$

which has the same asymptotically stable equilibrium as (ODE-DG). But then $E_{\alpha_0}\Big(\|\tilde{\alpha}_{r^*t} - \alpha^*\|\Big) \leq \lambda(\alpha_0)(\kappa_t/N)^\eta < \lambda(\alpha_0)(\kappa_t)^\eta$, which implies

$$\begin{aligned} E_{\alpha_0}\Big(\|\alpha_{it} - \alpha^*\|\Big) \leq & E_{\alpha_0}\Big(\|\tilde{\alpha}_{r^*t} - \alpha^*_{it}\|\Big) + E_{\alpha_0}\Big(\|\tilde{\alpha}_{r^*t} - \alpha^*\|\Big) \\ \leq & (1+x)\lambda(\alpha_0)(\kappa_t/N)^\eta < \lambda(\alpha_0)(\kappa_t)^\eta. \end{aligned}$$

■

This result might seem surprising, since in this social learning scheme agents are not copying the "best" rule used by their friends. Nor is the result based on some agent learning the optimal rule and "infecting" their friends with it. As there is no way for agents to know if their friends' rules are better than theirs, the effect of social learning seems to be more related to the aggregation of learning experience across agents. One way to see this is as follows: in an economy with many independent agents each learning individually, the behavior of the average agent will be close to the optimal rule. With social learning, each agent is using some weighted average of the behavior of the friends. This effect is similar to the effect of the average agent in the first case.

5.2 Some Extensions

This subsection extends the model in order to analyze the effect of changing some of the underlying hypothesis. As stated above, one can extend the analysis for example to allow for $\underline{y} = 0$ if one is willing to assume bounded first and second derivatives of the utility function. It is also not difficult to see that including a random interest rate or discount factor is easily accommodated as long as they are identically and independently distributed with a distribution that is Lipschitz continuous and bounded.

Corollary 5.5. *Assume β_t and/or R_t are each identically independently distributed with $\beta_t \in [\underline{\beta}, \bar{\beta}] \subset \mathbb{R}_{++}$ and $R_t \in [\underline{R}, \bar{R}] \subset \mathbb{R}_{++}$. Additionally assume that the distributions of β_t and R_t satisfy the technical conditions in assumptions A and B, and redefine $\Lambda = [0, w_0 + \bar{y}] \times [(\bar{R} - 1)/\bar{R}, 1]$. Then the results of theorem 4.1 hold.*

This suggests a more interesting extension of the model. Consider the case in which income is not identically independently distributed, but it is composed of the product of a permanent and a transitory components. In particular, let $y_{t+1} = y^P_{t+1} y^T_{t+1}$, where the transitory component y^T_{t+1} is identically and independently distributed with mean 1 and variance σ^2_T. On the other hand, assume that the logarithm of the permanent component follows a random walk with drift, so that $y^P_{t+1} = y^P_t \epsilon_{t+1}$, where $\epsilon_{t+1} = \epsilon \epsilon^P_{t+1}$, $\epsilon \in \mathbb{R}_+$, and ϵ^P_{t+1} is identically independently distributed with mean 1 and variance σ^2_ϵ. Additionally, assume that all the identically independently distributed random variables satisfy the conditions of assumptions A and B. Also, assume that the utility function satisfies $u(x \cdot C_t) = u(x)u(C_t)$ for any $x \in \mathbb{R}_{++}$. Finally, assume all the parameters satisfy the conditions required for the existence of a fully rational solution.[10]

Clearly, in this setting it is not optimal for the agent to have wealth growing at a different rate than income, since she could adjust her consumption and obtain a higher level of utility without ever getting to a situation where she is liquidity constrained infinitely often or never liquidity constrained. Thus, under these conditions, any good rule requires agents' wealth to be growing at the same rate as her income. But this can only happen if her consumption grows at the same rate also. So, if consumers are sufficiently sophisticated to deduce this, they can apply the same algorithm to a similar problem and learn the optimal linear rule. In particular:

Theorem 5.6. *Assume agents comprehend that normalizing all variables, which affect their decision problem, by the permanent component of income eliminates underlying growth in their problem. Then they can learn the optimal rule α^* by applying algorithm* (DG) *to their consumption problem based on normalized values.*

Proof. In this case a consumer's wealth evolves according to

$$W_{t+1} = R(W_t - C_t) + y^P_{t+1} y^T_{t+1}, \tag{5.5}$$

[10]While the conditions that ensure the algorithm can be analyzed using the ODE are satisfied by my assumptions, in order to obtain optimality it is required that such a fully rational solution exists. This is an implicit assumption throughout the paper. See e.g. Carroll (2004) for the conditions required when the utility function has the constant relative risk aversion property.

where W_t and C_t are their wealth and consumption levels. But, dividing by y^P_{t+1} on both sides, this is equivalent to

$$w_{t+1} = R_{t+1}(w_t - c_t) + y^T_{t+1}, \tag{5.6}$$

where $w_t = W_t/y^P_t$, $R_{t+1} = R/\epsilon_{t+1}$, and $c_t = C_t/y^P_t$, which has the same structure as equation (2.1) in the original problem studied above, but with a random rate of interest.

Additionally, notice that in this case $\beta u(C_t) = \beta u(\epsilon)u(\epsilon^P_{t+1})u(c_t)$. Defining $\beta_t = \beta u(\epsilon)u(\epsilon^P_{t+1})$, this problem is identical to the original one with a random discount rate. Thus, my previous results imply agents can learn the optimal rule α^* of the normalized problem using any version of the algorithm. ■

Notice that in this scenario, the optimal rule α^* that is found by the algorithm is also normalized. In particular, since normalized consumption is $c^b_t = \min\{w_t, \alpha^0_t + \alpha^1_t w_t\}$, then the consumption function is given by $C^b_t = \min\left\{W_t, \alpha^0_t y^P_t + \alpha^1_t W_t\right\}$. Thus, agents would internalize the fact the income is growing by increasing the intercept of their consumption rule by the same growth rate as income. Clearly, this requires a lot of sophistication on the part of the agents, who must understand that consumption needs to be growing at the same rate as income. But additionally, it requires them to understand what the effect of normalization is on the effective interest and discount rates.

But what if agents do not realize that consumption should be growing at the same rate as income? That is if they measure their regret based on the *actual non-normalized* values of wealth and consumption. In this case, if the utility function satisfies $u(\epsilon_{t+1}/\epsilon_t) = u(\epsilon_{t+1})/u(\epsilon_t)$, then the ODE related to the algorithm can be written as a function of consumption, wealth, and income normalized by the permanent component of income. Namely, as

$$\begin{pmatrix} \frac{d\tilde{\alpha}}{\partial\tau} \\ \\ \frac{d\tilde{\alpha}}{\partial\tau} \end{pmatrix} = \tilde{M} \int_{\mathcal{W}} \left[\left(E_t \beta_{t+1} R_{t+1} u'(c^b(w')) - u'(c^b(w)) \right) u''(c^b(w)) \right] \cdot \begin{pmatrix} 1 \\ \\ w \end{pmatrix} \mathbb{I}_U \pi_\alpha(dw),$$

where $\tilde{M} = M \cdot (u(\epsilon)^2/\epsilon^3) \cdot E(u(\epsilon^P)^2/(\epsilon^P)^3)$, ϵ^P has the same distribution as ϵ_t^P, and β_t and R_t are as defined in the previous proof. Thus,

Corollary 5.7. *Agents learn the optimal linear consumption rule α^*.*

These results show that agents using the H-O algorithm can learn the optimal linear consumption rule in more complex environments without requiring more sophistication, rationality, memory or computing abilities. Clearly, these results apply mutatis mutandis to frameworks where agents have a fixed probability of dying every period and a finite expected life.[11]

5.3 Asymptotics, shocks, and finite lives

Although encouraging, these results have been based on the asymptotic behavior of the algorithm when $\kappa_t \to 0$ as $t \to \infty$. This generates two kinds of problems: first, agents will stop learning as $t \to \infty$, which implies they will fail to react even if their environment undergoes large abrupt changes. Second,

[11] While my results are based on the normalization of variables, one could use the methods of Benveniste *et al.* (1990, part I, ch.4) for tracking non-stationary parameters in order to allow the income process to follow other more complex stochastic processes.

asymptotic results may be too far into the future given agents' lifespan, thus undermining their importance.

In order to address the first problem one can assume that agents restart the learning process whenever their environment changes if $\kappa_t \approx 0$. Another would be to let agents change their consumption behavior in ways that would be suggested by the algorithm. As my previous result suggest agents might react to a change in the average income by changing the intercept α^0 in the same quantity. Finally, one could assume that agents do not stop learning, i.e. that $\lim_{t\to\infty} \kappa_t = \kappa > 0$.

On the other hand, in order to address the second problem, two central questions need to be answered. First, how well does the learning algorithm perform in $T \ll \infty$ periods? And, second, how well does it perform when agents have finite lives? Clearly, if infinitely lived agents who use the algorithm perform poorly in finite time, this learning scheme will not be useful in situations where agents have finite lives. Thus, to answer these questions one needs to know if within T periods the algorithm is close to its stationary state.

Reassuringly, results can be established for a finite horizon T, when $\lim_{t\to\infty} \kappa_t = \kappa \geq 0$ is small enough. In particular,

Corollary 5.8 (Benveniste *et al.* (1990, Corollary 2, p.43)). *Let* $\bar{\kappa} = \max_{t\geq 0} \kappa_t$, $\epsilon > 0$*, and* T *be such that the trajectory of the ODE is close to the optimal rule, i.e.* $\|\alpha(T) - \alpha^*\| \leq \epsilon/2$*, where* $\alpha(T)$ *is the value at* T *of the solution of* (ODE-DG)*. Then* $P\left(\|\alpha_T - \alpha^*\| > \epsilon\right) \to 0$ *as* $\bar{\kappa} \to 0$.

This implies that if $\bar{\kappa}$ is small enough and the ODE converges fast enough to the optimal rule, so will the algorithm. This result implies that the results with decreasing gain can be generalized to the constant gain case, as long as the constant gain is small enough. Additionally, it indicates that if infinitely lived agents had to stop learning at T and were forced to live forever using their current rule, their ex-ante expected lifetime utility would be close to the one generated by the optimal rule.

Notice that this corollary also provides an answer to the first problem raised. If $\lim_{t\to\infty} \kappa_t = \kappa > 0$, then the agent never stops learning. So, any change in her environment, which changes α^*, will quickly affect her learning, which will track the new optimal rule. Thus, this "constant gain" version is extremely useful in settings where the agent may face a changing environment. Furthermore, notice that the agent will be learning and adapting to the new environment *even* if she does not know the environment changed! This is an interesting property of this family of algorithms, which solves the problems raised by the Lucas critique, even though agents are not fully rational.

All the optimality results presented above are based on the comparison of expected discounted life-time utilities in stationary states. But, if agents have a finite live, then these optimality criteria are not useful nor can the analysis be easily adapted. The problem here is that stationarity played an *essential* role in all my previous analyses. But unlike the infinite horizon case, the problem cannot be transformed into a variant that is stationary. On the contrary, the agent's problem varies with her age.[12]

While analytical results seem to be difficult to obtain, the ability of the

[12] Perhaps the use of social intertemporal learning, where agents of different ages "teach" each other or share past experience might allow the problem to be analyzed as a stationary one.

algorithm to track changes in the system, suggests it could still be useful in this setting. To analyze this possibility, I endow finitely lived agents with the HO-algorithm and compare their *actual* lifetime utility from using this algorithm with the utility they would obtain if they had used the fully rational consumption rule.

As in Allen and Carroll (2001) and Howitt and Özak (2009) I consider agents who have constant relative risk aversion (CRRA) utilities with a CRRA coefficient $\theta = 3$, a discount factor $\beta = 0.9$, and face a zero constant rate of interest. In order to understand the effect of a finite horizon, I consider 3 cases in all of which I follow agents for 60 periods. Case 1 assumes agents have finite lives and have the same income process as the two aforementioned papers, where each period agents get identically independently distributed income shocks. In particular, I assume $y_t \in \{0.7, 1, 1.3\}$ with the probabilities $\{0.2, 0.6, 0.2\}$ respectively. Case 2 assumes agents have finite lives and an inverse-U-shaped income profile, where $y_t = y_t^P \cdot y^T$, where y^T has the same distribution as the previous process, and $y_t^P = 1 - (t/60 - .5)^2$. In case 3 I assume agents live infinitely and their income follows the same process as case 1, but I only follow their utility for the first 60 periods of their lives. I assume all agents in all cases start to learn from the same initial linear rule $\alpha_0 = (0, 0.5)$ and initial wealth $w_0 = 2.5$. For each case I simulate the behavior of 100,000 agents who use the unconstrained and the zero-intelligence versions of the algorithm. Additionally, I assume that if the new rule $\alpha_{t+1} \notin \Lambda$, then the agent does not update the rule and simply keeps her previous one for one more period.[13] Finally, I assume the agent uses $\kappa_t = 0.35$ and the matrix

$$M = 10 \cdot \begin{pmatrix} 1 & 0.975 \\ 0.975 & 1 \end{pmatrix}. \tag{5.7}$$

Figures 2-7 summarize the results of these simulations. For each case I present the evolution of the mean, median, 25th and 75th quantile, minimum and maximum of the parameters of the consumption rules α_t in panels (a) and (b). As can be seen there, and as should be expected from the theory, in all cases most of the movement in the mean value of the parameters occurs in the first 20 periods. This suggests that the distribution of parameters converges *very* fast. In cases 1 and 3, presented in figures 2, 3, 6, and 7, the average consumption rule remains practically constant after 30 periods. On the other hand, in case 2 (figures 4-5), the parameters follow the non-monotonicity of the income process. In particular, notice that the average intercept α^0 has a similar inverse-U-shaped evolution, which could be expected, since that is precisely the shape of average income. On the other hand, the marginal propensity to consume α^1 starts with a movement towards the optimal rule α^*, but then follows a U-shaped trajectory.

Panels (c) and (d) of these figures present the distribution of realized lifetime discounted utilities under the linear and fully rational consumption rules, and the distribution of differences in these realized utilities. The figures show clearly that agents' discounted lifetime utilities would have been higher under

[13]Similar results are obtained if instead the agent chooses a random rule in Λ, a fixed rule in Λ or the closest rule that would have resulted in her not having regret.

the fully rational rule. In all cases the distribution of utilities under bounded rationality has a very long left tail, showing that there are many individuals that get very low lifetime utilities. This is caused mostly by underconsumption during their lifetime. Since all agents start from the same initial conditions, this suggests that a series of bad draws prevents some agents from learning. In particular, notice that if agents live finite lives (case 1 and 2) the median deviation of lifetime utility is less than one unit. On the other hand, if agents are infinitely lived, the median difference lies between 2 and 1.5.

Since differences in utilities cannot be easily compared, I also compute the certainty equivalent level of consumption, which would have generated that level of utility in the same $T = 60$ periods. In particular, for every agent I compute

$$CE = \frac{1-\beta}{1-\beta^{T+1}} u^{-1}\left(\sum_{t=0}^{T} \beta^t u(c_t)\right). \tag{5.8}$$

Panels (e) and (f) show the distribution of these certainty equivalents and the difference between the certainty equivalents generated by the boundedly and fully rational rules. In all cases the median boundedly rational CE is above the minimum fully rational CE. Thus, in this setting boundedly rational agents, who use the HO-algorithm, have at least a 50% probability of having similar levels of CE than fully rational ones within 60 periods. Similarly, the probability that boundedly rational agents' discounted lifetime utility levels are in the range obtained by fully rational ones is close to 50% within 60 periods. Notice that these results do not depend on the agent having a finite life. This follows basically from the fact that unless their lifespan T is extremely short, the rational consumption rule for every age is pretty close to the rational consumption rule of the infinitely lived agent.

Although these results are only suggestive, they do support the use of the algorithm in life cycle models with boundedly rational agents.

6 Conclusions

The assumption of complete and perfect rationality has increasingly been criticized due, in part, to the high complexity of many solutions in economic models under this assumption. In response, models of bounded rationality and learning have recently flourished in economics, though the study and application of these ideas to approximate solutions of stochastic dynamic programming problems is still an emerging area. In particular, the study of consumption-saving decisions under uncertainty and liquidity constraints has been pursued by only a couple of papers with limited or negative results.

In this paper I have shown that boundedly rational agents, who use a linear consumption function, face liquidity constraints, and have uncertain income, can learn to behave "optimally" by following a generalization of the learning scheme proposed by Howitt and Özak (2009). In particular, using a novel theoretical framework I have shown that in a general consumption problem under liquidity constraints and uncertainty, agents using this algorithm will learn to behave optimally. In particular, there is a unique stable consumption rule that maximizes agents' ex-ante expected life-time utility. The stability and optimality properties of this rule imply that in the steady state agents do not have an incentive to deviate from their behavior. This follows from the fact they are getting the maximum expected life-time utility under the wealth distribution generated by their consumption function.

Using the general theory I provided various extensions of the model to incorporate different income processes, random interest and discount rates, and levels of rationality and sophistication by agents. Additionally, I studied the effects of social social learning. Finally, using numerical simulations I find that the algorithm has similar properties to the infinite horizon case when agents live only a finite number of periods and have a life-cycle pattern of income.

The analytical and numerical techniques used suggest that similar results could be obtained in other dynamic programming settings. This would allow the study of bounded rationality in other macroeconomic models. Clearly, learning the fully rational solution might only be attainable in very specific settings, like the linear-quadratic one (Evans and McGough, 2009) or when the boundedly rational rule belongs to the same family of functions as the optimal one. Still, unlike other learning schemes, the one proposed in this paper is *known* to have stability and optimality properties which make its use more desirable. In particular, this framework can be applied in large agent based macroeconomic models, where millions of agents need to be simulated simultaneously. By using this learning scheme, macroeconomic modelers can incorporate boundedly rational agents in their models, without being subject to extremely suboptimal behavior by the modeled agents nor to the Lucas critique.

References

Allen, T. W. and Carroll, C. D. (2001). Individual learning about consumption. *Macroeconomic Dynamics*, **5** (02), 255–271. 1, 5, 13, 31

Benveniste, A., Métivier, M. and Priouret, P. (1990). *Adaptive algorithms and stochastic approximations*, vol. 22. Berlin: Springer-Verlag. 13, 18, 28, 29, 60, 66

Börgers, T. and Sarin, R. (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, **77** (1), 1–14. 5, 12

Brown, A. L., Chua, Z. E. and Camerer, C. F. (2009). Learning and visceral temptation in dynamic saving experiments. *Quarterly Journal of Economics*, **124** (1), 197–231. 2

Carroll, C. D. (2001). A theory of the consumption function, with and without liquidity constraints. *The Journal of Economic Perspectives*, **15** (3), 23–45. 1

— (2004). Theoretical foundations of buffer stock saving, national Bureau of Economic Research Working Paper No. 10867. 26

Evans, G. W. and Honkapohja, S. (2001). *Learning and expectations in macroeconomics*. Princeton: Princeton University Press. 1

— and McGough, B. (2009). Learning to optimize, unpublished. 1, 6, 13, 35

Fudenberg, D. and Levine, D. K. (1998). *The theory of learning in games*, vol. 2. Cambridge, Mass.: MIT Press. 1

Gabaix, X. (2011). A sparsity-based model of bounded rationality. *NBER Working Papers*. 2

Hart, S., Mas-Colell, A. and Babichenko, Y. (2013). *Simple adaptive strategies: from regret-matching to uncoupled dynamics*. World Scientific Publishing Company. 4

Heidergott, B., Hordijk, A. and Weisshaupt, H. (2006). Measure-valued differentiation for stationary markov chains. *Mathematics of Operations Research*, **31** (1), 154–172. 49

— and Vázquez-Abad, F. J. (2006). Measure-valued differentiation for random horizon problems. *Markov Processes and Related Fields*, **12**, 509–536. 49

— and Vázquez-Abad, F. J. (2008). Measure-valued differentiation for markov chains. *Journal of Optimization Theory and Applications*, **136** (2), 187–209. 14, 49, 53

Hirsch, M. W., Smale, S. and Devaney, R. L. (2004). *Differential equations, dynamical systems, and an introduction to chaos*, *Pure and applied mathematics; a series of monographs and textbooks*, vol. v. 60. San Diego, CA: Academic Press, 2nd edn. 18

Howitt, P. and Özak, Ö. (2009). Adaptive consumption behavior. *NBER Working Papers*, (15427). 1, 2, 3, 4, 19, 31, 34

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux, 1st edn. 12

Krasovskiĭ, N. N. (1963). *Stability of motion: applications of Lyapunov's second method to differential systems and equations with delay.* Stanford, Calif.: Stanford University Press. 59

Kushner, H. J. and Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications*, vol. 35. New York: Springer. 13

Lettau, M. and Uhlig, H. (1999). Rules of thumb versus dynamic programming. *American Economic Review*, **89** (1), 148–174. 1, 5, 13

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, **92** (368), 805–824. 4, 12

Meyn, S. P. and Tweedie, R. L. (1993). *Markov chains and stochastic stability.* Springer. 46, 47, 48, 49, 53

Puterman, M. and Brumelle, S. (1979). On the convergence of policy iteration in stationary dynamic programming. *Mathematics of Operations Research*, pp. 60–69. 2

Rudin, W. (1966). *Real and complex analysis.* McGraw-Hill series in higher mathematics, New York: McGraw-Hill. 49

Santos, M. and Rust, J. (2004). Convergence properties of policy iteration. *SIAM Journal on Control and Optimization*, **42** (6), 2094–2115. 2

Sargent, T. J. (1993). *Bounded rationality in macroeconomics: Thomas J. Sargent*. Oxford: Clarendon Press. 1

Selten, R. (1998). Aspiration adaptation theory. *Journal of Mathematical Psychology*, **42** (2), 191–214. 4

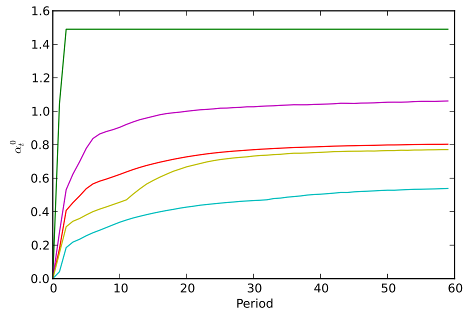
— (2001). What is bounded rationality. In G. Giderenzer and R. Selten (eds.), *Bounded rationality: The adaptive toolbox*, MIT Press, pp. 13–36. 4

— (2004). Learning direction theory and impulse balance equilibrium. In D. Friedman and A. Cassar (eds.), *economics Lab: An Intensive Course in Experimental Economics*, N.Y., USA: Routledge, pp. 133–140. 12

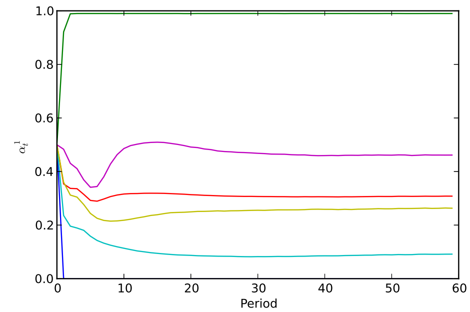
— and Buchta, J. (1999). Experimental sealed bid first price auctions with directly observed bid functions. In R. Z. David V. Budescu, Ido Erev (ed.), *Games and human behavior: Essays in honor of Amnon Rapoport*, *5*, Mahwah NJ: Psychology Press, pp. 79–102. 4

— and Stoecker, R. (1986). End behavior in sequences of finite prisoner's dilemma supergames a learning theory approach. *Journal of Economic Behavior & Organization*, **7** (1), 47–70. 4, 5

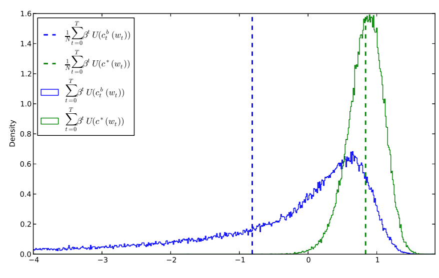
Figure 2: Dynamics of (backward-looking) algorithm with finite horizon (case 1).



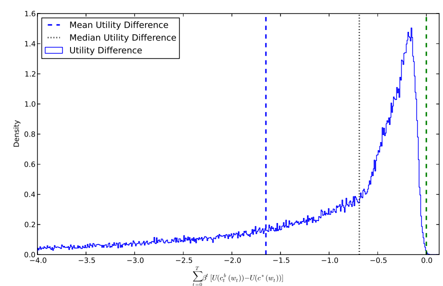
(a) Intercept α_t^0



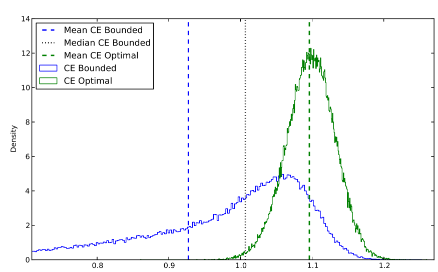
(b) Marginal propensity to consume α_t^1



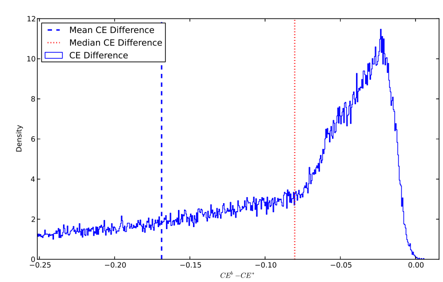
(c) Distribution actual lifetime utilities under bounded and fully rational consumption rules.



(d) Distribution differences between actual lifetime utilities (bounded vs fully rational).

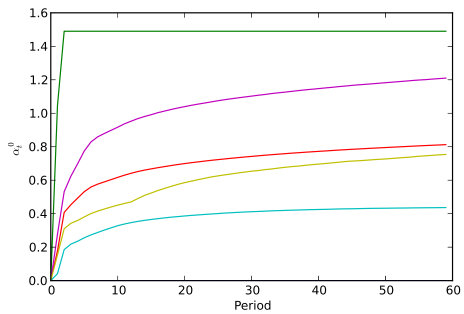


(e) Distribution CE consumption (bounded vs fully rational).

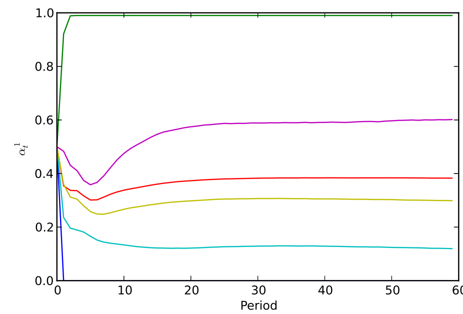


(f) Distribution differences CE consumption (bounded vs fully rational).

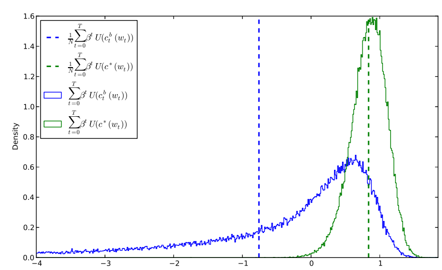
Figure 3: Dynamics of (zero-intelligence) algorithm with finite horizon (case 1).



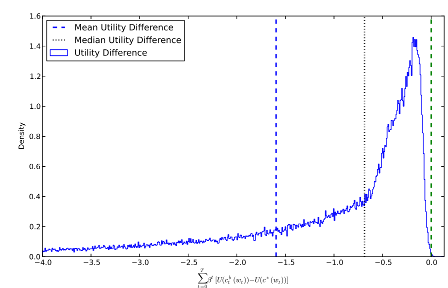
(a) Intercept α_t^0



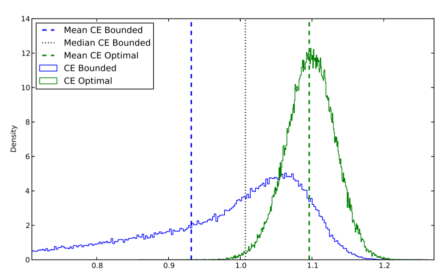
(b) Marginal propensity to consume α_t^1



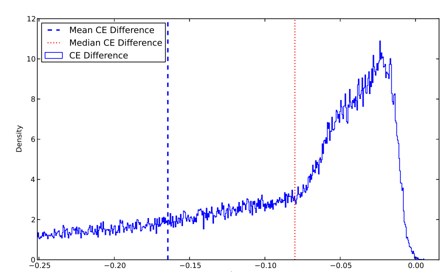
(c) Distribution actual lifetime utilities under bounded and fully rational consumption rules.



(d) Distribution differences between actual lifetime utilities (bounded vs fully rational).

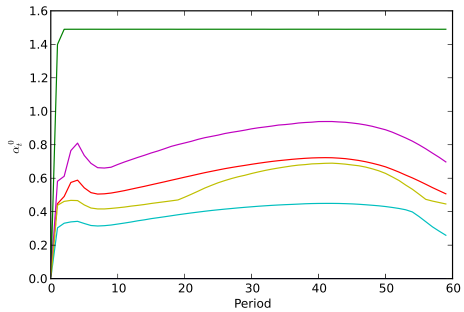


(e) Distribution CE consumption (bounded vs fully rational).

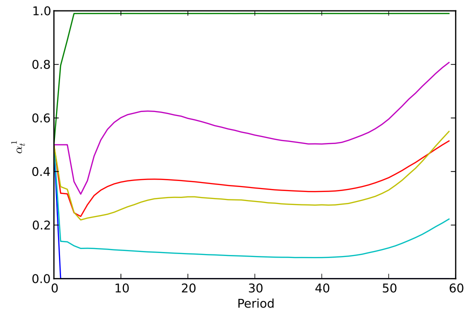


(f) Distribution differences CE consumption (bounded vs fully rational).

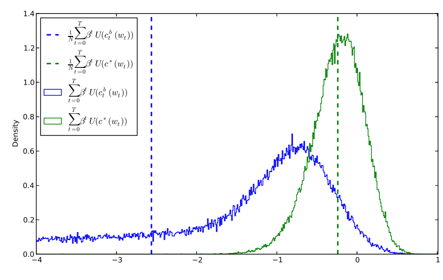
Figure 4: Dynamics of (backward-looking) algorithm with finite horizon (case 2).



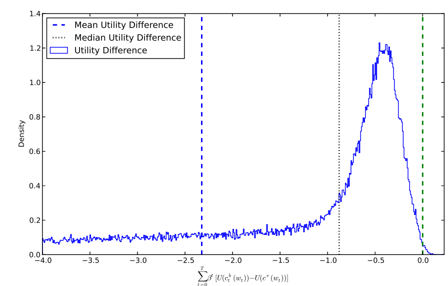
(a) Intercept α_t^0



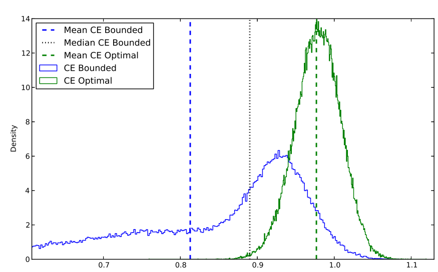
(b) Marginal propensity to consume α_t^1



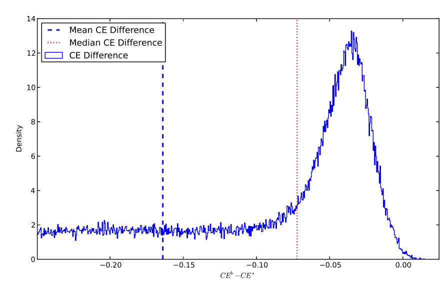
(c) Distribution actual lifetime utilities under bounded and fully rational consumption rules.



(d) Distribution differences between actual lifetime utilities (bounded vs fully rational).

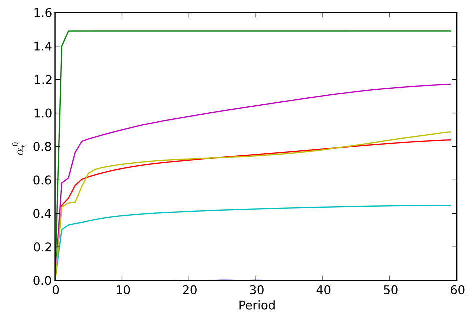


(e) Distribution CE consumption (bounded vs fully rational).

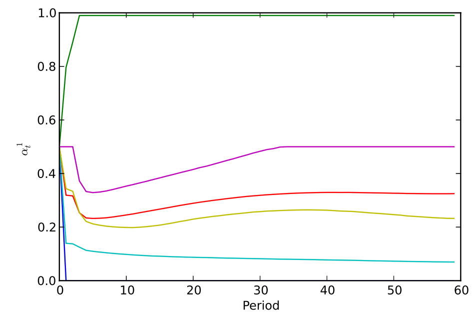


(f) Distribution differences CE consumption (bounded vs fully rational).

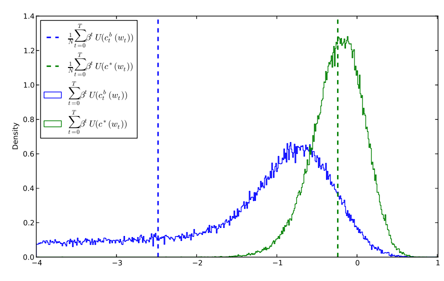
Figure 5: Dynamics of (zero-intelligence) algorithm with finite horizon (case 2).



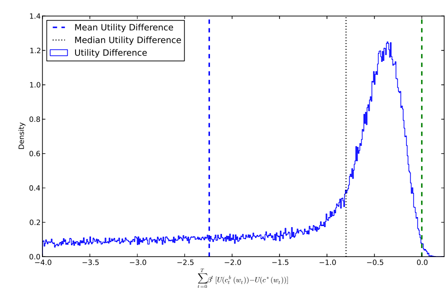
(a) Intercept α_t^0



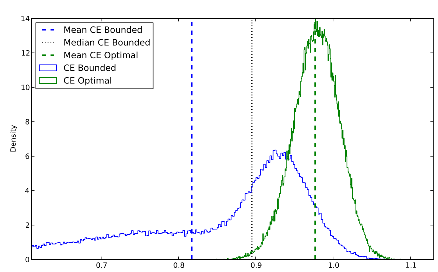
(b) Marginal propensity to consume α_t^1



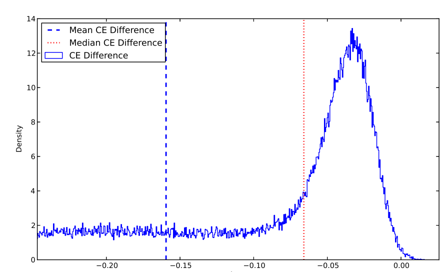
(c) Distribution actual lifetime utilities under bounded and fully rational consumption rules.



(d) Distribution differences between actual lifetime utilities (bounded vs fully rational).

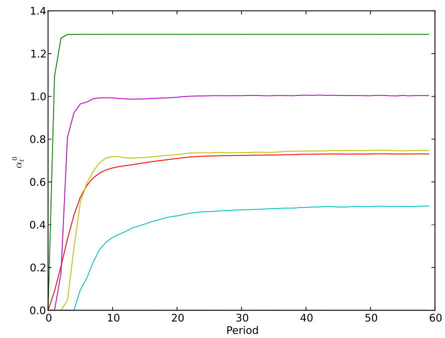


(e) Distribution CE consumption (bounded vs fully rational).

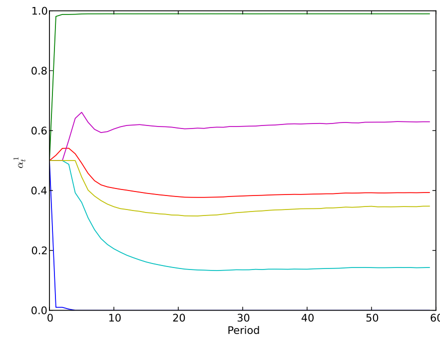


(f) Distribution differences CE consumption (bounded vs fully rational).

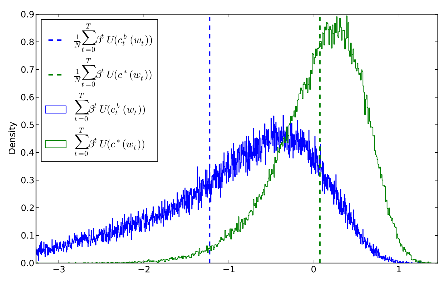
Figure 6: Dynamics of (backward-looking) algorithm with infinite horizon (case 3).



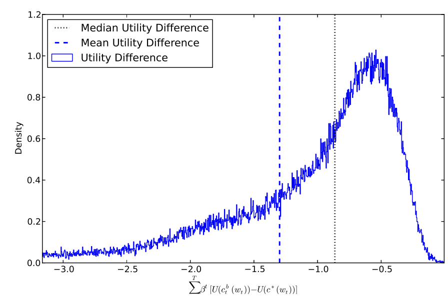
(a) Intercept α_t^0



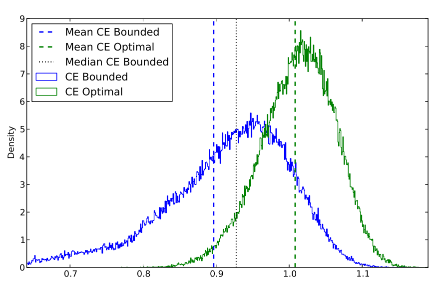
(b) Marginal propensity to consume α_t^1



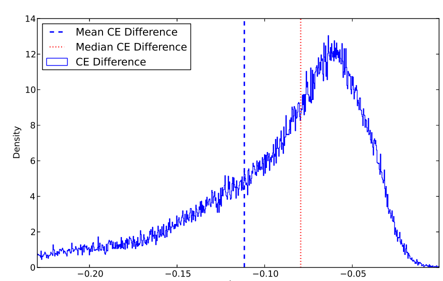
(c) Distribution actual lifetime utilities under bounded and fully rational consumption rules.



(d) Distribution differences between actual lifetime utilities (bounded vs fully rational).

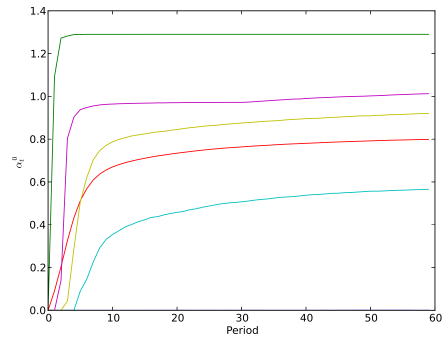


(e) Distribution CE consumption (bounded vs fully rational).

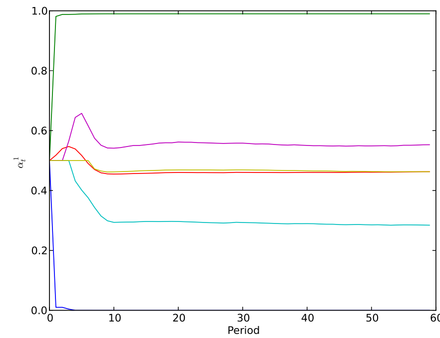


(f) Distribution differences CE consumption (bounded vs fully rational).

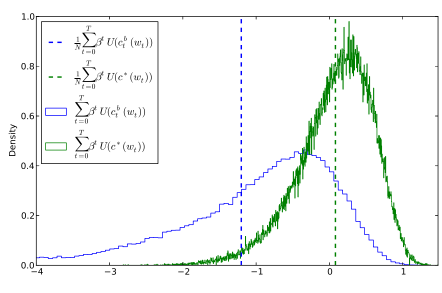
Figure 7: Dynamics of (zero-intelligence) algorithm with infinite horizon (case 3).



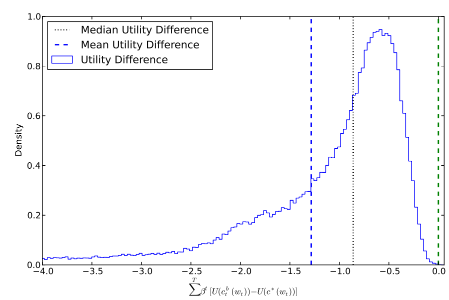
(a) Intercept α_t^0



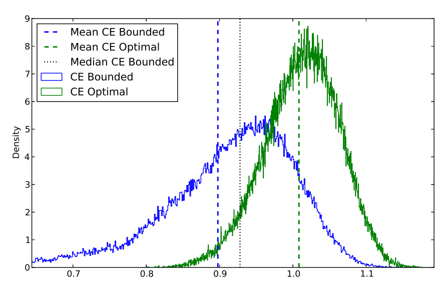
(b) Marginal propensity to consume α_t^1



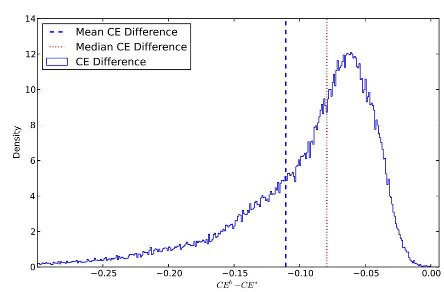
(c) Distribution actual lifetime utilities under bounded and fully rational consumption rules.



(d) Distribution differences between actual lifetime utilities (bounded vs fully rational).



(e) Distribution CE consumption (bounded vs fully rational).



(f) Distribution differences CE consumption (bounded vs fully rational).

Online Appendix (not for publication)

A Proofs

Proof of Theorem 2.1. Clearly, under assumption A, the wealth process $\{w_t\}$ is Markov with state space $(\mathcal{W}, \mathfrak{B}(\mathcal{W})) = (\mathbb{R}_+, \mathfrak{B}(\mathbb{R}_+))$ and transition probability kernel P defined by

$$P(w, \mathcal{A}) = \Gamma\Big(\big\{\mathcal{A} - R\big(w - c(w)\big)\big\} \cap \mathcal{Y}\Big), \qquad \forall w \in \mathcal{W},\ \mathcal{A} \in \mathfrak{B}(\mathcal{W}). \tag{A.1}$$

$P(w, \mathcal{A})$ gives the probability of going from state w to a set $\mathcal{A}$ in one period. Notice that if $w \in [0, \tilde{w}]$, then $P(w, \mathcal{A}) = \Gamma(\mathcal{A} \cap \mathcal{Y})$ is independent of w, i.e. $[0, \tilde{w}]$ is an atom (Meyn and Tweedie, 1993, p.103). If $\underline{y} \leq \tilde{w}$, then this set is an accessible atom.

Here I present and prove a more general version of the theorem. For that, let $\mathcal{Y} = [\underline{y}, \bar{y})$ and let $\{w_t\}$ be a Markov process with state space $(\mathcal{W}, \mathfrak{B}(\mathcal{W})) = (\mathbb{R}_+, \mathfrak{B}(\mathbb{R}_+))$, defined as

$$w_{t+1} = h(w_t) + y_{t+1}$$

for some increasing and convex (or concave) function $h(w)$ and transition probability kernel P defined by

$$P(w, A) = \Gamma\Big(\big\{A - h(w)\big)\big\} \cap \mathcal{Y}\Big), \qquad \forall w \in \mathcal{W},\ A \in \mathfrak{B}(\mathcal{W}). \tag{A.2}$$

$P(w, A)$ gives the probability of going from state w to a set A in one period.

Let $\underline{\mathcal{A}}$ denote the set of stationary wealth levels generated by the stochastic difference equation, when income is equal to $\underline{y}$ in every period, and let $\bar{\mathcal{A}}$ denote the set of stationary wealth levels when income is $\bar{y}$ in every period, i.e. $\underline{\mathcal{A}} = \big\{w \in \mathbb{R}_+ \mid w = h(w) + \underline{y}\big\}$ and $\bar{\mathcal{A}} = \{w \in \mathbb{R}_+ \mid w = h(w) + \bar{y}\}$. Let $\underline{w}$ be the infimum of $\underline{\mathcal{A}}$, $\bar{w}$ be the infimum of $\bar{\mathcal{A}}$, and $\bar{\bar{w}}$ be the supremum of $\bar{\mathcal{A}}$.

Theorem A.1. *If assumption A holds, then:*

(i) If $\bar{\bar{w}} < \infty$, then there exists a unique invariant probability measure π on $\mathcal{W}$ and a π-null

set N such that for any initial distribution of initial wealth λ,[14] *such that $\lambda(N) = 0$,*

$$\left\| \int \lambda(dw) P^m(w, \cdot) - \pi(\cdot) \right\| \to 0, \quad m \to \infty.$$

(ii) If $\bar{\bar{w}} = \infty$, then $P(w_t \to \infty) = 1$.

Proof. We present the proof for the case when $h(w)$ is convex. For the other case, just revert the roles of w^1 and $\bar{\bar{w}}$.

(i) Assume that $\bar{\bar{w}} < \infty$. We need to analyze two cases:

(a) If $\underline{w} < w^1$, let $\mathcal{A}^* = [\underline{w}, \bar{w}]$, $\mathcal{A}^0 = [0, \underline{w})$, $\mathcal{A}^1 = (\bar{w}, w^1)$ and $\mathcal{A}^\infty = [w^1, \infty)$. Then

$$P(w, \mathcal{A}^*) \begin{cases} = 1 & \text{if } w \in \mathcal{A}^* \cup \mathcal{A}^0 \\ = 0 & \text{if } w \in \mathcal{A}^\infty \end{cases}, \qquad P(w, \mathcal{A}^\infty) \begin{cases} = 1 & \text{if } w \in \mathcal{A}^\infty \\ = 0 & \text{if } w \in \mathcal{A}^0 \cup \mathcal{A}^* \end{cases} \tag{A.3a}$$

$$P^m(w, \mathcal{A}^*) > 0 \ \text{ if } w \in \mathcal{A}^1 \qquad \text{or} \qquad P^{m'}(w, \mathcal{A}^\infty) > 0 \ \text{ if } w \in \mathcal{A}^1 \tag{A.3b}$$

for some $1 \leq m, m' < \infty$. So, for any $w \in \mathcal{W}$ and $\mathcal{A}$ such that $\mathcal{A}^* \subseteq \mathcal{A}$ and $\mathcal{A} \cap \mathcal{A}^\infty = [\hat{w}, \infty)$ for some $\hat{w} > w^1$, there exists $m < \infty$ such that $P^m(w, \mathcal{A}) > 0$. Letting

$$\varphi(\mathcal{A}) = \begin{cases} \Gamma(\mathcal{A} \cap \mathcal{Y}) & \text{if } \mathcal{A}^* \subseteq \mathcal{A} \ \text{ and } \mathcal{A} \cap \mathcal{A}^\infty = [\hat{w}, \infty) \text{ for some } \hat{w} \geq w^1 \\ 0 & \text{Otherwise} \end{cases}$$

I have that the process $\{w_t\}$ is φ-irreducible and thus there exists a maximal irreducibility measure ψ on $\mathfrak{B}(\mathcal{W})$ (Meyn and Tweedie, 1993, theorem 4.0.1). Furthermore, $\psi(\mathcal{A}^* \cup \mathcal{A}^\infty) > 0$, so that ψ has support with non-empty interior, and since the process is Feller, it is a T-chain by proposition 7.1.2 and theorem 6.0.1(iii) in Meyn and Tweedie (1993). Furthermore, theorem 6.0.1.(ii) ensures that every compact set is petite, and since $\mathcal{A}^*$ is compact and absorbent, theorem 8.3.6(i) ensures the process $\{w_t\}$ is recurrent. Thus, by theorem 10.4.4 I have the existence of

[14] Here $\|\cdot\|$ denotes the total variation norm, i.e. if λ is a signed measure on $\mathfrak{B}(\mathcal{W})$ then

$$\|\lambda\| := \sup_{f : |f| \leq 1} |\lambda(f)|.$$

a unique invariant measure $\bar{\pi}$, which, since $\mathcal{A}^*$ is petite and absorbing, by theorem 10.4.10 is finite, and equivalent to a probability measure π, so that the process is positive recurrent. Now, it is not hard to prove that the process is aperiodic, so that theorem 13.3.4(ii) in Meyn and Tweedie (1993) gives the desired result.

(b) If $\underline{w} = w^1$, then also $\bar{w} = \bar{\bar{w}}$. Let $\mathcal{A}^* = [w^1, \bar{\bar{w}}]$, $\mathcal{A}^0 = [0, w^1)$ and $\mathcal{A}^\infty = [\bar{\bar{w}}, \infty)$. Then

$$P(w, \mathcal{A}^*) = 1 \qquad \text{if } w \in \mathcal{A}^* \cup \mathcal{A}^0 \qquad \text{and} \qquad P^m(w, \mathcal{A}^*) \quad > 0 \qquad \text{if } w \in \mathcal{A}^\infty \tag{A.4}$$

for some $1 \leq m < \infty$. So, for any $w \in \mathcal{W}$ and $\mathcal{A}$ such that $\mathcal{A}^* \subseteq \mathcal{A}$, there exists $m < \infty$ such that $P^m(w, \mathcal{A}) > 0$. Letting

$$\varphi(\mathcal{A}) = \begin{cases} \Gamma\left(\mathcal{A} \cap \mathcal{Y}\right) & \text{if } \mathcal{A}^* \subseteq \mathcal{A} \\ 0 & \text{Otherwise} \end{cases}$$

I have that the process $\{w_t\}$ is φ-irreducible and thus there exists a maximal irreducibility measure ψ on $\mathfrak{B}(\mathcal{W})$ (Meyn and Tweedie, 1993, theorem 4.0.1). Furthermore, since $\psi(\mathcal{A}^*) > 0$, the result follows from the same arguments as in (a).

(ii) If $\bar{\bar{w}} = \infty$, for any $w \in \mathcal{W}$ there exists $m < \infty$ such that $\epsilon_w = P^m(w, \mathcal{A}^\infty) > 0$, where $\mathcal{A}^\infty = [w^1, \infty)$. Define $\epsilon = \sup_{w \in [0, w^1)} \epsilon_w$. Clearly, $P(w, \mathcal{A}^\infty) = 1$ if $w \in \mathcal{A}^\infty$ and $P(w_t \to \infty \mid w_0 \in \mathcal{A}^\infty) = 1$, then for any $w \in [0, w^1)$,

$$P(w_t < \infty, \forall\, t) = 1 - P(w_t \to \infty) = P(w_t \in (\mathcal{A}^\infty)^C, \forall\, t) = \lim_{t \to \infty} (1 - \epsilon_{w_t})^t \leq \lim_{t \to \infty} (1 - \epsilon)^t = 0$$

so that $P(w_t \to \infty) = 1$. ■

It is easy to see that if $h(w)$ is concave, then the set N in the previous theorem is given by $N = \emptyset$. This theorem can easily be extended for other types of increasing functions, but this would require us to change the notation and analyze more subcases, so I do not pursue it here. ■

Proof of Theorem 4.1. Let's start by proving that π_α is $\mathcal{D}_\phi$-continuous and differentiable

(Heidergott, Hordijk and Weisshaupt, 2006; Heidergott and Vázquez-Abad, 2006, 2008). In order to prove this, I need to show that under our assumptions, the proof of Theorem 4.2 of Heidergott *et al.* (2006) holds, so that π_{α_0,α_1} is $\mathscr{D}_\phi$-Lipschitz continuous and differentiable. For simplicity I will follow the notation and definitions used by Heidergott *et al.* (2006). We present the proof in steps.

Step 1: Let $\mathcal{A}^*$ be as in (i)(a) in the proof of Theorem 2.1. This set is petite and the first return time to $\mathcal{A}^*$ for every $w \in \mathcal{A}^*$, $\tau_{\mathcal{A}^*}$ is 1. So, theorem 14.0.1 of Meyn and Tweedie (1993) implies that for any function $f : \mathcal{W} \to [1, \infty)$ I have that

$$\int_{\mathcal{W}} f(w)\pi_{\alpha_0,\alpha_1}(dw) < \infty.$$

Clearly, for any continuous function $g : \mathcal{W} \to \mathbb{R}$, $|g| + 1$ satisfies the conditions above, so that $|g| + 1$ is π_{α_0,α_1}-integrable. By theorem 11.27 of Rudin (1966) g is also π_{α_0,α_1}-integrable.

Step 2: For fixed α, let $P(w, \mathcal{A}; \alpha)$ denote the one-step transition kernel. By (2.1) $w' \in \mathcal{W}(\alpha, w) \equiv [R(w - c^b(\alpha, w)) + \underline{y}, R(w - c^b(\alpha, w)) + \bar{y})$ with probability one, so that the density function of w' is

$$\gamma_{w'}(w'; \alpha, w) = \begin{cases} \gamma_y(w' - R(w - c^b(\alpha, w))) & \text{if } w' \in \mathcal{W}(\alpha, w) \\ 0 & \text{Otherwise} \end{cases}$$

and

$$P(w, \mathcal{A}; \alpha) = \int_{\mathcal{A}} \gamma_{w'}(w'; \alpha, w)dw'.$$

Under (a), this implies that for any continuous function $g : \mathcal{W} \to \mathbb{R}$,

$$\begin{aligned}\int_{\mathcal{W}} |g(w')|\,\gamma_{w'}(w'; \alpha, w)dw' =& \int_{\mathcal{W}(\alpha,w)} |g(w')|\,\gamma_{w'}(w'; \alpha, w)dw' \\ \leq& (\bar{y} - \underline{y}) \cdot \sup_{w' \in \mathcal{W}(\alpha,w)} |g(w')|\,\gamma_{w'}(w'; \alpha, w) < \infty\end{aligned}$$

for all α and $w \in \mathcal{W}$. Let $\mathscr{D}$ denote the set of continuous functions $g : \mathcal{W} \to \mathbb{R}$ and $\mathscr{D}_\phi$ be the set ϕ-bounded continuous functions, where ϕ is defined in equation (2.2).

Clearly, $\mathscr{D}_\phi \subseteq \mathscr{D}$.

Under (b), I have that for any ϕ-bounded continuous function $g : \mathcal{W} \to \mathbb{R}$,

$$\int_{\mathcal{W}} |g(w')|\, \gamma_{w'}(w'; \alpha, w) dw' \leq K \int_{\mathcal{W}(\alpha,w)} \phi(w') \gamma_{w'}(w'; \alpha, w) dw' < \infty.$$

For this case, let $\mathscr{D} = \mathscr{D}_\phi$ be the set of ϕ-bounded continuous functions.

Then by assumption D, $P(w, \mathcal{A}; \alpha)$ is $\mathscr{D}$-Lipschitz continuous at α. To see this, notice that for $\alpha' \neq \alpha$ with α' chosen in such a way that either both $c^b(\alpha, w), c^b(\alpha', w) < w$ or $c^b(\alpha, w), c^b(\alpha', w) \geq w$. We can do this, since both $\mathscr{D}$-Lipschitz continuity and $\mathscr{D}$-differentiability need to hold on an open set around α. Then

$$\begin{aligned}
&\left| \int_{\mathcal{W}} g(w') \gamma_{w'}(w'; \alpha, w) dw' - \int_{\mathcal{W}} g(w') \gamma_{w'}(w'; \alpha', w) dw' \right| \\
&\qquad = \left| \int_{\mathcal{W}} g(w') \Big(\gamma_{w'}(w'; \alpha, w) - \gamma_{w'}(w'; \alpha', w) \Big) dw' \right| \\
&\qquad \leq \int_{\mathcal{W}} |g(w')|\, |\gamma_{w'}(w'; \alpha, w) - \gamma_{w'}(w'; \alpha', w)|\, dw'.
\end{aligned}$$

On $\mathcal{W}(\alpha, w) \cap \mathcal{W}(\alpha', w)$ I have that

$$\begin{aligned}
|\gamma_{w'}(w'; \alpha, w) - \gamma_{w'}(w'; \alpha', w)| \leq& M_1(w) \left| R(w - c^b(\alpha, w)) - R(w - c^b(\alpha', w)) \right| \\
=& R\, M_1(w) \left| c^b(\alpha, w) - c^b(\alpha', w) \right| \\
\leq& R\, M_1(w) \Big(|\alpha_0 - \alpha_0'| + w\, |\alpha_1 - \alpha_1'| \Big).
\end{aligned}$$

On the other hand, let

$$\begin{aligned}
\hat{w} &= \max \left\{ R(w - c^b(\alpha, w)) + \bar{y}, R(w - c^b(\alpha', w)) + \bar{y} \right\}, \\
\underline{\hat{w}} &= \min \left\{ R(w - c^b(\alpha, w)) + \bar{y}, R(w - c^b(\alpha', w)) + \bar{y} \right\}, \\
\tilde{w} &= \max \left\{ R(w - c^b(\alpha, w)) + \underline{y}, R(w - c^b(\alpha', w)) + \underline{y} \right\} \\
\underline{\tilde{w}} &= \min \left\{ R(w - c^b(\alpha, w)) + \underline{y}, R(w - c^b(\alpha', w)) + \underline{y} \right\}
\end{aligned}$$

so that

$$
\begin{aligned}
&\int_{\mathcal{W}\setminus\Big(\mathcal{W}(\alpha,w)\cap\mathcal{W}(\alpha',w)\Big)} |g(w')|\,|\gamma_{w'}(w';\alpha,w)-\gamma_{w'}(w';\alpha',w)|\,dw' \\
&= \int_{\underline{\tilde{w}}}^{\tilde{w}} |g(w')| \max\{\gamma_{w'}(w';\alpha,w),\gamma_{w'}(w';\alpha',w)\}\,dw' \\
&\quad + \int_{\underline{\hat{w}}}^{\hat{w}} |g(w')| \max\{\gamma_{w'}(w';\alpha,w),\gamma_{w'}(w';\alpha',w)\}\,dw' \\
&\leq \Big(M_2(w)+M_3(w)\Big)\Big(|\alpha_0-\alpha_0'|+w\,|\alpha_1-\alpha_1'|\Big).
\end{aligned}
$$

So,

$$
\begin{aligned}
&\int_{\mathcal{W}} |g(w')|\,|\gamma_{w'}(w';\alpha,w)dw'-\gamma_{w'}(w';\alpha',w)|\,dw' = \\
&= \int_{\mathcal{W}(\alpha,w)\cap\mathcal{W}(\alpha',w)} |g(w')|\,|\gamma_{w'}(w';\alpha,w)dw'-\gamma_{w'}(w';\alpha',w)|\,dw' \\
&\quad + \int_{\mathcal{W}\setminus\Big(\mathcal{W}(\alpha,w)\cap\mathcal{W}(\alpha',w)\Big)} |g(w')|\,|\gamma_{w'}(w';\alpha,w)dw'-\gamma_{w'}(w';\alpha',w)|\,dw' \\
&\leq \Big(K_g\,R\,M_1(w)+M_2(w)+M_3(w)\Big)\Big(|\alpha_0-\alpha_0'|+w\,|\alpha_1-\alpha_1'|\Big)
\end{aligned}
$$

where $K_g = \int_{\mathcal{W}(\alpha,w)\cap\mathcal{W}(\alpha',w)} |g(w')|\,dw'$.

Similarly, if α, α' and w are such that either $\alpha_0+\alpha_1 w \leq w \leq \alpha_0'+\alpha_1' w$ or $\alpha_0'+\alpha_1' w \leq w \leq \alpha_0+\alpha_1 w$, then

$$
\begin{aligned}
&\int_{\mathcal{W}} |g(w')|\,|\gamma_{w'}(w';\alpha,w)dw'-\gamma_{w'}(w';\alpha',w)|\,dw' = \\
&= \int_{\mathcal{W}(\alpha,w)\cap\mathcal{W}(\alpha',w)} |g(w')|\,|\gamma_{w'}(w';\alpha,w)dw'-\gamma_{w'}(w';\alpha',w)|\,dw' \\
&\quad + \int_{\mathcal{W}\setminus\Big(\mathcal{W}(\alpha,w)\cap\mathcal{W}(\alpha',w)\Big)} |g(w')|\,|\gamma_{w'}(w';\alpha,w)dw'-\gamma_{w'}(w';\alpha',w)|\,dw' \\
&\leq \Big(K_g\,R\,M_1(w)+M_2(w)+M_3(w)\Big)\Big(|\alpha_0-\alpha_0'|+w\,|\alpha_1-\alpha_1'|\Big).
\end{aligned}
$$

This follows from a proof similar as before, one just need to notice that if $\alpha_0+\alpha_1 w \leq w \leq \alpha_0'+\alpha_1' w$, then

$$0 \leq w(1-\alpha_1)-\alpha_0 \leq (\alpha_0-\alpha_0')+w(\alpha_1-\alpha_1'),$$

so that

$$\begin{aligned}\left|c^b(\alpha,w)-c^b(\alpha',w)\right| &= |w(1-\alpha_1)-\alpha_0| \\ &\leq |(\alpha_0-\alpha_0')+w(\alpha_1-\alpha_1')| \\ &\leq |\alpha_0-\alpha_0'|+w\,|\alpha_1-\alpha_1'|\,,\end{aligned}$$

and similarly for the case when $\alpha_0'+\alpha_1'w \leq w \leq \alpha_0+\alpha_1 w$.

Step 3: Since $\bar{w}_\alpha < \infty$, so

$$\sup_\alpha \|P_\alpha\|_\phi = \sup_\alpha \sup_w \int_{\mathcal{W}} \gamma_{w'}(w';\alpha,w)\frac{\phi(w')}{\phi(w)}dw' < \infty. \tag{A.5}$$

Step 4: Since $\phi(\cdot)$ is continuous and the level sets $C_\phi = \{w \mid \phi(w) \leq d\}$ for some $d \in \mathbb{R}_+$ are compact, they are also petite. This implies that $\phi(\cdot)$ is unbounded off petite sets.

Notice that $-R\alpha_0-(1-R(1-\alpha_1))w \in \mathscr{D}_\phi$ and I can find a $\lambda \in (0,1)$ such that

$$\sup_w \frac{-R\alpha_0-(1-R(1-\alpha_1))w}{\phi(w)} < (\lambda-1) < 0.$$

If $w \leq \tilde{w}$, then

$$\int_{\mathcal{Y}} \phi(y)\gamma_y(y)dy \leq L_1 \leq L_1+\lambda\phi(w).$$

On the other hand, if $w > \tilde{w}$, then

$$\begin{gathered}\int_{\mathcal{Y}} \phi\Big(R(1-\alpha_1)w-R\alpha_0+y\Big)\gamma_y(y)dy-\phi(w) = \\ \int_{\mathcal{Y}} \phi'(\xi)\Big(R(1-\alpha_1)w-R\alpha_0+y-w\Big)\gamma_y(y)dy < \\ \int_{\mathcal{Y}} y\gamma_y(y)dy-R\alpha_0-(1-R(1-\alpha_1))w\end{gathered}$$

so that

$$\begin{aligned}&\int_{\mathcal{Y}} \phi\Big(R(1-\alpha_1)w-R\alpha_0+y\Big)\gamma_y(y)dy \leq \\ &L_2+\lambda\phi(w).\end{aligned}$$

Let $L = \max\{L_1, L_2\}$, then $L, \lambda, \phi(w)$ satisfy lemma 15.2.8 in Meyn and Tweedie (1993), and by their theorem 16.0.1

$$\|P_\alpha^n - \pi_\alpha\|_\phi \le K_\alpha \rho_\alpha^n \tag{A.6}$$

for some $K_\alpha < \infty$ and $0 < \rho_\alpha < 1$.

But equations (A.5) and (A.6) are the same as in lemma 4.1 and equation (20) in theorem 4.1 in Heidergott and Vázquez-Abad (2008), so the proof of their theorem 4.2 holds, and π_α is $\mathscr{D}_\phi$-Lipschitz continuous.

Additionally, notice that if $g \in \mathscr{D}_\phi$, then

$$\frac{d}{d\alpha_0}\int_{\mathcal{W}} g(w')\gamma_{w'}(w';\alpha,w)dw' = \begin{cases} R\Big(g(R((1-\alpha_1)w-\alpha_0)+\underline{y})\gamma_y(R((1-\alpha_1)w-\alpha_0)+\underline{y})- \\ \quad g(R((1-\alpha_1)w-\alpha_0)+\bar{y})\gamma_y(R((1-\alpha_1)w-\alpha_0)+\bar{y})\Big) \\ \quad + \int_{\mathcal{W}} g(w')\frac{d}{d\alpha_0}\gamma_{w'}(w';\alpha,w)dw' & \text{if } w \ge \frac{\alpha_0}{1-\alpha_1} \\ 0 & \text{Otherwise} \end{cases}$$

$$\frac{d}{d\alpha_1}\int_{\mathcal{W}} g(w')\gamma_{w'}(w';\alpha,w)dw' = \begin{cases} Rw\Big(g(R((1-\alpha_1)w-\alpha_0)+\underline{y})\gamma_y(R((1-\alpha_1)w-\alpha_0)+\underline{y})- \\ \quad g(R((1-\alpha_1)w-\alpha_0)+\bar{y})\gamma_y(R((1-\alpha_1)w-\alpha_0)+\bar{y})\Big) \\ \quad + \int_{\mathcal{W}} g(w')\frac{d}{d\alpha_1}\gamma_{w'}(w';\alpha,w)dw' & \text{if } w \ge \frac{\alpha_0}{1-\alpha_1} \\ 0 & \text{Otherwise} \end{cases}$$

are well defined and are the $\mathscr{D}_\phi$-derivatives of $P(w,\mathcal{A};\alpha)$. Thus, π_α is also $\mathscr{D}_\phi$-differentiable.

Since Weierstrass' Theorem ensures the existence of a σ-optimal rule, and strict concavity of the utility function ensures that for each $\sigma \in \Lambda$, α_σ is unique, i.e. a function and not a correspondence of σ. Now, the maximization problem is also continuous with respect to σ, ensuring that α_σ is continuous in σ. This allows the application of Brouwer's fixed point theorem to prove existence of the solution α^* that is α^*-optimal.

For the next step, we need a couple of additional results and characterizations of the α^*-optimal rule and of the fully rational solution. It is well known that in my setting the fully rational solution $c^*(w)$, can be analyzed as the solution to

$$v(w) = \max_{s.t.\ 0\le c\le w} u(c) + \beta E\Big[v(h(c,w,y)) \mid c, w\Big] \tag{A.7a}$$

$$\text{i.e. } c^*(w) = \underset{s.t.\ 0\le c\le w}{\operatorname{argmax}}\ u(c) + \beta E\Big[v(h(c,w,y)) \mid c, w\Big] \tag{A.7b}$$

Notice that for α^* and any α,

$$EV_{\alpha^*} - EV^* \equiv \int U(\alpha^*, w) - v(w)\ \pi_{\alpha^*}(dw) \ge \int U(\alpha, w) - v(w)\ \pi_{\alpha^*}(dw)$$

Proposition A.2. *If* $\lim_{t\to\infty} \beta^t E_0\Big[U(\alpha, w_t) - v(w_t)\Big] = 0$ *for* $\alpha \in \Lambda$ *then:*

(i)

$$\begin{aligned} &U(\alpha, w_t) - v(w_t) = \\ &E_0\left[\sum_{j=0}^{\infty} \mu(w_{t+j})(c^b(w_{t+j}) - c^*(w_{t+j})) - k_{t+j}(c^b(w_{t+j}) - c^*(w_{t+j}))^2\right]. \end{aligned} \tag{A.8}$$

(ii)

$$\begin{aligned} EV_\alpha - &EV^* = \\ &\sum_{j=0}^{\infty}\beta^j \int_{\mathcal{W}} \left[\mu(w_{t+j})(c^b(w_{t+j}) - c^*(w_{t+j}))\right. \\ &\qquad \left. - k_{t+j}(c^b(w_{t+j}) - c^*(w_{t+j}))^2 \pi_\alpha(dw_{t+j})\right]. \end{aligned} \tag{A.9}$$

Proof of proposition A.2. (i) By Taylor's theorem

$$\begin{aligned} u(c^b(w_t)) + &\beta E_0 v(R(w_t - c^b(w_t)) + y_t) = u(c^*(w_t)) + u'(c^*(w_t))(c^b(w_t) - c^*(w_t)) \\ &+ \frac{1}{2}u''(\xi_t)(c^b(w_t) - c^*(w_t))^2 + \beta E_0[v(R(w_t - c^*(w_t)) + y_t)] \\ &- \beta R E_0[v'(R(w_t - c^*(w_t)) + y_t)](c^b(w_t) - c^*(w_t)) \\ &+ \frac{1}{2}\beta R E_0[v''(\zeta_t)](c^b(w_t) - c^*(w_t))^2 \\ = v(w_t) + &\Big(u'(c^*(w_t)) - \beta R E_0[v'(R(w_t - c^*(w_t)) + y_t)]\Big)(c^b(w_t) - c^*(w_t)) \\ &+ \frac{1}{2}\Big(u''(\xi_t) + \beta R E_0[v''(\zeta_t)]\Big)(c^b(w_t) - c^*(w_t))^2 \\ = v(w_t) + &\mu(w_t)(c^b(w_t) - c^*(w_t)) + \frac{1}{2}\Big(u''(\xi_t) + \beta R E_0[v''(\zeta_t)]\Big)(c^b(w_t) - c^*(w_t))^2 \end{aligned}$$

where $\mu(w_t) \ge 0$ is the Lagrange multiplier associated to the first order condition of

(A.7a), $\xi_t \in (c^*(w_t), c^b(w_t))$ and $\zeta_t \in (R(w_t - c^*(w_t)) + y_t, R(w_t - c^b(w_t)) + y_t)$. Thus,

$$\begin{aligned}
U(\alpha, w_t) - v(w_t) =& u(c^b(w_t)) + \beta E_0[U(\alpha, R(w_t - c^b(w_t)) + y_t)] - v(w_t) \\
=& \mu(w_t)(c^b(w_t) - c^*(w_t)) + \frac{1}{2}\Big(u''(\xi_t) + \beta R E_0[v''(\zeta_t)]\Big)(c^b(w_t) - c^*(w_t))^2 \\
&+ \beta E_0[U(\alpha, R(w_t - c^b(w_t)) + y_t) - v(R(w_t - c^b(w_t)) + y_t)].
\end{aligned}$$

Iterating I get

$$\begin{aligned}
U(\alpha, w_t) - v(w_t) = & E_0\left[\sum_{j=0}^{T} \big(\mu(w_{t+j})(c^b(w_{t+j}) - c^*(w_{t+j}))\big)\right] \\
&+ E_0\left[\sum_{j=0}^{T} \frac{1}{2}\Big(\Big(u''(\xi_{t+j}) + \beta R E_0[v''(\zeta_{t+j})]\Big)(c^b(w_{t+j}) - c^*(w_{t+j}))^2\Big)\right] \\
&+ \beta^T E_0[U(\alpha, R(w_{t+T} - c^b(w_{t+T})) + y_{t+T}) - v(R(w_{t+T} - c^b(w_{t+T})) + y_{t+T})]
\end{aligned}$$

which under our hypothesis gets the desired result.

(ii) Notice that $U(\alpha, w_t) - v(w_t)$ is continuous in w_t and under the stationary distribution $w_t \in [\underline{y}, \bar{w}]$. So, $U(\alpha, w_t) - v(w_t) \leq \bar{V}$, for some $0 < \bar{V} < \infty$, and thus, $\lim_{t\to\infty} \beta^t E_0\Big[U(\alpha, w_t) - v(w_t)\Big] = 0$ holds for $w_t \in \mathcal{W}$ π_{α^*}-a.e. Replacing the previous equation in the definition of $EV_\alpha - EV^*$ gives the result. ■

Corollary A.3. *α^*, belongs to the set*

$$\tilde{\Lambda} = \{\alpha \in \Lambda \mid \tilde{w}_{\alpha^*} = \tilde{w}^*\},$$

Thus, it is a solution to

$$\min_{\alpha\in\Lambda} \sum_{j=0}^{\infty} \beta^j \int_{\mathcal{W}} \left[k_{t+j}(c^b(w_{t+j}) - c^*(w_{t+j}))^2 \pi_{\alpha^*}(dw_{t+j})\right], \tag{A.10}$$

which is equivalent to

$$\min_{\alpha\in\Lambda} \int_{\mathcal{W}} \left[k_s(c^b(w_s) - c^*(w_s))^2 \pi_{\alpha^*}(dw_s)\right] \tag{A.11}$$

Proof of Corollary A.3. Assume on the contrary that $\tilde{w}_{\alpha^*} < \tilde{w}^*$, then, since $\mu(w_{t+j}) = 0$ for all $w_{t+j} \geq \tilde{w}^*$, I can find $\alpha \in \Lambda$ such that $\bar{w}_\alpha = \bar{w}_{\alpha^*}$ and $\tilde{w}_\alpha = \tilde{w}^*$ for which

$$(c^b(\alpha, w_{t+j}) - c^*(w_{t+j}))^2 < (c^b(\alpha^*, w_{t+j}) - c^*(w_{t+j}))^2 \qquad \forall w_{t+j}$$
$$\mu(w_{t+j})(c^b(\alpha, w_{t+j}) - c^*(w_{t+j})) = 0 \qquad \forall w_{t+j}$$

so that

$$\int U(\alpha, w)\pi_{\alpha^*} - EV^* > EV_{\alpha^*} - EV^*$$

which is a contradiction. If on the other hand, $\tilde{w}_{\alpha^*} > \tilde{w}$, then I can find $\alpha \in \Lambda$ such that $\tilde{w}_\alpha = \tilde{w}^*$ and both consumption rules cross $c^*(w)$ at the same point. In this case the same contradiction follows.

That $\tilde{w}_\alpha = \tilde{w}^*$ implies that the first part of the integrand in (A.9) is always zero, so that α^* solves the problem

$$\min_{\alpha \in \Lambda} \sum_{j=0}^{\infty} \beta^j \int_{\mathcal{W}} \left[k_{t+j}(c^b(w_{t+j}) - c^*(w_{t+j}))^2 \pi_{\alpha^*}(dw_{t+j})\right]. \tag{A.12}$$

But a change of variable and the fact that w_t follows the stationary distribution implies that it also solves (A.11). ∎

This implies that $\underline{w}_{\alpha^*} = \underline{y}$.

Proposition A.4. *There exists a unique α^* that is α^*-optimal.*

Proof of proposition A.4. Let $0 < \epsilon < \min\{\bar{y}/R, (1+R)/2R\}$,

$$T_1(\sigma) = \mathrm{e}^{\int U(\sigma, w)\pi_\sigma - \int U(\alpha_\sigma, w)\pi_\sigma},$$

and define $T : \Lambda \to \Lambda$ as

$$T(\sigma) = \left(\max\left\{T_1(\sigma)\sigma_0, \frac{1}{2}\sigma_0 + \epsilon\right\}, \max\left\{T_1(\sigma)\sigma_1, \frac{1}{2}\left(\sigma_1 - \frac{R-1}{R}\right) + \epsilon\right\}\right).$$

Notice that T_1 is continuous, $T_1(\sigma) \leq 1$ for all $\sigma \in \Lambda$ and $T_1(\sigma) = 1$ if and only if $\sigma = \alpha_\sigma$. This implies that $T(\sigma) = \sigma$ if and only if $\sigma = \alpha_\sigma$. So, $T(\sigma)$ and α_σ have the same set of fixed points. Since $T(\Lambda) \subset \Lambda$ and is contracting, Banach's fixed point theorem ensures there exists a unique fixed point. Thus, α^* is unique. ■

So, under our assumptions, the unique α^*-optimal rule minimizes the expected squared difference from the optimal consumption function under the stationary distribution of wealth π_{α^*}. This allows some additional and useful characterizations. In particular,

Proposition A.5. α^*, *solves the following problem*

$$\min_{(\alpha_0,\alpha_1)\in\Lambda} \int_{\mathcal{W}} \Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big)^2 \pi_{\alpha^*}(dw_t). \tag{A.13}$$

Proof of proposition A.5. From the previous corollary, α^* is such that $c^b(w) = c^*(w)$ for $w \leq \tilde{w}^*$ and the expected squared difference between $c^b(w)$ and $c^*(w)$ is minimal under π_{α^*}. Since,

$$\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t)) \underset{\text{Taylor}}{=} \tilde{k}_t(c^b(w_t) - c^*(w_t)), \tag{A.14}$$

then (α_0^*, α_1^*) must also solve (A.13). Since $c^b(\alpha, w)$ is an increasing function of α, the objective function is strictly convex, and the solution is interior, then the first derivatives of (A.11) with respect to α_0 and α_1 must equal zero at α^*. Now, consider the problem of minimizing A.14, but choosing two different set of parameters, one for $c^b(w_{t+1})$ and one for $c^b(w_t)$. The first order condition requires

$$\int_{\mathcal{W}} u''(c^b(w_t))\Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big)\begin{pmatrix}1\\ w_t\end{pmatrix}\mathbb{I}_U \pi_{\alpha^*}(dw_t) = 0 \tag{A.15}$$

$$\int_{\mathcal{W}} \Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big) E_t\left[u''(c^b(w_{t+1}))\begin{pmatrix}1\\ w_{t+1}\end{pmatrix}\right]\mathbb{I}_U \pi_{\alpha^*}(dw_t) = 0 \tag{A.16}$$

Notice that

$$\begin{aligned}
&\int_{\mathcal{W}} \Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big) E_t\left[u''(c^b(w_{t+1}))\begin{pmatrix}1\\ w_{t+1}\end{pmatrix}\right]\mathbb{I}_U \pi_{\alpha^*}(dw_t) = \\
&\int_{\mathcal{W}} E_t u''(c^b(w_t))\Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big)\begin{pmatrix}1\\ w_t\end{pmatrix}\mathbb{I}_U \pi_{\alpha^*}(dw_t) \\
&+ \int_{\mathcal{W}} \hat{k}_t E_t(w_{t+1} - w_t)\Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big)\begin{pmatrix}1\\ w_t\end{pmatrix}\mathbb{I}_U \pi_{\alpha^*}(dw_t) =
\end{aligned}$$

$$\int_{\mathcal{W}} E_t u''(c^b(w_t))\Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big)\begin{pmatrix}1\\ w_t\end{pmatrix}\mathbb{I}_U \pi_{\alpha^*}(dw_t)$$
$$+\int_{\mathcal{W}} \hat{k}_t\tilde{k}_t E_t(w_{t+1}-w_t)\Big(c^b(w_t)) - c^*(w_t)\Big)\begin{pmatrix}1\\ w_t\end{pmatrix}\mathbb{I}_U \pi_{\alpha^*}(dw_t) =$$
$$\int_{\mathcal{W}} E_t u''(c^b(w_t))\Big(\beta R E_t u'(c^b(w_{t+1})) - u'(c^b(w_t))\Big)\begin{pmatrix}1\\ w_t\end{pmatrix}\mathbb{I}_U \pi_{\alpha^*}(dw_t).$$

Thus, if (A.15) holds, so does (A.16). But (A.15) is (ODE-DG) pre-multiplied by M. That is, α^* is α^*-optimal if, and only if, it is an equilibrium of the differential equation related to the algorithm. ■

Corollary A.6. *The unique α^*-optimal rule is the unique globally asymptotically stable equilibrium of* (ODE-DG).

Proof of corollary A.6. This follows from the fact that α^* is the unique minimizer of the expected α^*-expected squared regret. Rewrite (ODE-DG) as

$$\begin{pmatrix}\frac{d\alpha_0}{\partial\tau}\\ \frac{d\alpha_1}{\partial\tau}\end{pmatrix} = M\cdot h(\alpha). \tag{ODE-DG'}$$

Pre -multiplying (ODE-DG) by M^{-1} generates a gradient system. Since α^* is the unique minimizer it is also the unique asymptotically stable equilibrium of this dynamic system. But, (ODE-DG) equals zero if and only if $M\cdot h(\alpha)=0$. Since M positive definite, this can only happen if and only if $h(\alpha)=0$. Thus, (ODE-DG) and the gradient system have the same asymptotically stable equilibrium α^*. ■

Now we need to prove the relation between equation (ODE-DG) and the algorithm. For this it is only necessary to verify that the problem satisfies the conditions required by the theorems of the theory of stochastic recursive algorithms. Before doing so, notice that the algorithm can be rewritten as

$$\begin{pmatrix}\alpha^0_{t+1}\\ \alpha^1_{t+1}\end{pmatrix} = \begin{pmatrix}\alpha^0_{t}\\ \alpha^1_{t}\end{pmatrix} + \kappa_t M\left[\int \big(\beta R E_t u'(c^b_t(w_{t+1})) - u'(c^b_t(w_t))\big)u''(c^b_t(w_t))\begin{pmatrix}1\\ w_t\end{pmatrix}\mathbb{I}_U\pi_{\alpha_t}\right]$$
$$+\kappa_t M\left[\big(\beta R E_t u'(c^b_t(w_{t+1})) - u'(c^b_t(w_t))\big)u''(c^b_t(w_t))\begin{pmatrix}1\\ w_t\end{pmatrix}\mathbb{I}_U\right.$$

$$- \int \left(\beta R E_t u'(c_t^b(w_{t+1})) - u'(c_t^b(w_t))\right) u''(c_t^b(w_t)) \begin{pmatrix} 1 \\ w_t \end{pmatrix} \mathbb{I}_U \pi_{\alpha_t} \Bigg]$$

Notice that last two terms are a Martingale difference, so that their expected value under the invariant distribution generated by α^* is zero.

Let $\mathcal{Q}^e \subseteq \Lambda$ be the domain of attraction of α^*. Then the global asymptotic stability of α^* implies:

Corollary A.7 (Krasovskiĭ (1963)). *There exists a function L on $\mathcal{Q}^e$ of class $\mathscr{C}^2$ such that*

(i) $L(\alpha^*) = 0$, $L(\alpha) > 0$ *for all* $\alpha \in \mathcal{Q}^e$, $\alpha \neq \alpha^*$.

(ii) $\nabla L(\alpha) \cdot M \cdot h(\alpha) < 0$ *for all* $\alpha \in \mathcal{Q}^e$, $\alpha \neq \alpha^*$.

(iii) $L(\alpha) \to \infty$ *if* $\alpha \to \partial \mathcal{Q}^e$ *or* $\|\alpha\| \to \infty$.

For any $b \in \mathbb{R}_+$, let $K(b) = \{(\alpha, M) \in \mathcal{Q}^e \mid L(\alpha, M) \leq b\}$, and $\tau(b) = \inf\{n \in \mathbb{N} \mid (\alpha_n, M_n) \notin K(b)\}$. Also let $\mathcal{Q}_1 \subset \mathcal{Q}$ and $\mathcal{Q}_2 \subseteq \mathcal{Q}^e$ be compact sets and

$$\Omega(\mathcal{Q}_1, \mathcal{Q}_2) = \{(\alpha_n, M_n) \in \mathcal{Q}_1 \text{ for all } n, (\alpha_n, M_n) \in \mathcal{Q}_2 \text{ for infinitely many } n\}.$$

Theorem A.8. *Let $b < b_1 < b_2 < \infty$. Then:*

(i) There exist constants B_3 and s such that for all $\alpha_0 \in K(b_1)$ and all $w \in \mathcal{W}$,

$$P_{w,\alpha_0}(\{\tau(b_2) < \infty\}) \leq B_3(1 + |w|^s) \sum_{k=1}^{\infty} \kappa_k^2. \tag{A.17}$$

(ii) For all $\alpha_0 \in K(b)$ and all $w \in \mathcal{W}$, $\alpha_n \to \alpha^*$ P_{w,α_0}-a.s. on $\{\tau(b_2) = \infty\}$.

(iii) There exist constants B_4 and s such that for all $n \geq 0$, all $\alpha_0 \in \mathcal{Q}_2$ and all $w \in \mathcal{W}$

$$P_{n,w,\alpha_0}(\{\alpha_n \to \alpha^*\}) \geq 1 - B_4(1 + |w|^s) \sum_{k=n+1}^{\infty} \kappa_k^2. \tag{A.18}$$

(iv) For all $w \in \mathcal{W}$, $\alpha_0 \in \mathcal{Q}_2$, $\alpha_n \to \alpha^*$ P_{w,α_0}-a.s. on $\Omega(\mathcal{Q}_1, \mathcal{Q}_2)$.

P_{w,α_0} denotes the distribution of $\{(w_k, \alpha_k)\}_{k \geq 0}$ starting from (w, α_0) and P_{n,w,α_0} denotes the distribution of $\{(w_{n+k}, \alpha_{n+k})\}_{k \geq 0}$ starting from (w, α_0).

Proof of Theorem A.8. This theorem is simply an application of (i) proposition 10, (ii) proposition 11, (iii) theorem 13 and (iv) theorem 15 in chapter 1 of part II of Benveniste *et al.* (1990). I only need to show that their assumptions A.1-A.7 hold for this problem.

A.1 This is given by assumption C.

A.2 Follows directly from the definition of $\mathcal{Q}$ and (2.1).

A.3 In their notation, the function $H(\alpha, w)$ is given by the elements after κ_t in our equation (DG), while their function ρ is equal to the zero function in our case. There are four cases to consider:

(a) If $w < \alpha_0 + \alpha_1 w$ and $w' < \alpha_0 + \alpha_1 w'$, then $c(w') = y$.

(b) If $w < \alpha_0 + \alpha_1 w$ and $w' \geq \alpha_0 + \alpha_1 w'$, then $c(w') = \alpha_0 + \alpha_1 y$.

(c) If $w \geq \alpha_0 + \alpha_1 w$ and $w' < \alpha_0 + \alpha_1 w'$, then $c(w') = R(1 - \alpha_1)w - R\alpha_0 + y$.

(d) If $w \geq \alpha_0 + \alpha_1 w$ and $w' \geq \alpha_0 + \alpha_1 w'$, then $c(w') = \alpha_0 + R\alpha_1(1 - \alpha_1)w - R\alpha_0\alpha_1 + \alpha_1 y$.

Let $U_1(w, \alpha) = \big(\beta R E_t u'(c(w')) - u'(\alpha_0 + \alpha_1 w)\big)u''(\alpha_0 + \alpha_1 w)$, then

(a)

$$\begin{aligned}|U_1(w, \alpha)| =& \left|\big(\beta R E_t u'(y) - u'(\alpha_0 + \alpha_1 w)\big)u''(\alpha_0 + \alpha_1 w)\right| \\ \leq& \big(|\beta R E_t u'(y)| + |u'(\alpha_0 + \alpha_1 w)|\big)\, |u''(\alpha_0 + \alpha_1 w)| \\ <& \left(\left|\beta R u'(\underline{y})\right| + \left|u'(\alpha_0 + \alpha_1 \underline{y})\right|\right) \left|u''(\alpha_0 + \alpha_1 \underline{y})\right| \\ =& \bar{U}_1(\alpha)\end{aligned}$$

(b)

$$\begin{aligned}|U_1(w, \alpha)| =& \left|\big(\beta R E_t u'(\alpha_0 + \alpha_1 y) - u'(\alpha_0 + \alpha_1 w)\big)u''(\alpha_0 + \alpha_1 w)\right| \\ \leq& \big(|\beta R E_t u'(\alpha_0 + \alpha_1 y)| + |u'(\alpha_0 + \alpha_1 w)|\big)\, |u''(\alpha_0 + \alpha_1 w)| \\ <& \left(\left|\beta R u'(\alpha_0 + \alpha_1 \underline{y})\right| + \left|u'(\alpha_0 + \alpha_1 \underline{y})\right|\right) \left|u''(\alpha_0 + \alpha_1 \underline{y})\right|\end{aligned}$$

$$=\bar{U}_1(\alpha)$$

(c)

$$\begin{aligned}
|U_1(w,\alpha)| =& \left|\big(\beta R E_t u'(R(1-\alpha_1)w - R\alpha_0 + y) - u'(\alpha_0+\alpha_1 w)\big)u''(\alpha_0+\alpha_1 w)\right| \\
\leq& \big(|\beta R E_t u'(R(1-\alpha_1)w - R\alpha_0 + y)| + |u'(\alpha_0+\alpha_1 w)|\big)\,|u''(\alpha_0+\alpha_1 w)| \\
<& \big(\left|\beta R u'(R(1-\alpha_1)\underline{y} - R\alpha_0 + \underline{y})\right| + \left|u'(\alpha_0+\alpha_1 \underline{y})\right|\big)\,\left|u''(\alpha_0+\alpha_1 \underline{y})\right| \\
=& \bar{U}_1(\alpha)
\end{aligned}$$

(d)

$$\begin{aligned}
|U_1(w,\alpha)| =& \left|\big(\beta R E_t u'(\alpha_0 + R\alpha_1(1-\alpha_1)w - R\alpha_0\alpha_1 + \alpha_1 y) - u'(\alpha_0+\alpha_1 w)\big)u''(\alpha_0+\alpha_1 w)\right| \\
\leq& \big(|\beta R E_t u'(\alpha_0 + R\alpha_1(1-\alpha_1)w - R\alpha_0\alpha_1 + \alpha_1 y)| + |u'(\alpha_0+\alpha_1 w)|\big)\,|u''(\alpha_0+\alpha_1 w)| \\
<& \big(\left|\beta R u'(\alpha_0 + R\alpha_1(1-\alpha_1)\underline{y} - R\alpha_0\alpha_1 + \alpha_1\underline{y})\right| + \left|u'(\alpha_0+\alpha_1 \underline{y})\right|\big)\,\left|u''(\alpha_0+\alpha_1 \underline{y})\right| \\
=& \bar{U}_1(\alpha)
\end{aligned}$$

Thus, $|U_1(w,\alpha| \leq \bar{U}_1$, where $\bar{U}_1 = \max_{\text{cases }(a)-(d)} \sup_{\alpha\in\mathcal{Q}_2} \bar{U}_1(\alpha)$. So,

$$\left\| M \cdot U_1(w,\alpha) \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \right\| \leq \|M\| \cdot |U_1(w,\alpha)| \cdot \left\| \begin{pmatrix} 1 \\ w \end{pmatrix} \right\| \leq K_1 \bar{U}_1 (1+w^2),$$

where $K_1 = \|M\|$.

A.4 Their function $h(\alpha)$ is given by (ODE-DG). We have that

$$\begin{aligned}
&\left\| \int M \cdot U_1(w,\alpha) \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \pi_\alpha(dw) - \int M' \cdot U_1(w,\alpha') \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \pi_{\alpha'}(dw) \right\| \\
\leq& \left\| \int M \cdot U_1(w,\alpha) \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \pi_\alpha(dw) - \int M' \cdot U_1(w,\alpha) \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \pi_\alpha(dw) \right\| \\
&+ \left\| \int M' \cdot U_1(w,\alpha) \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \pi_\alpha(dw) - \int M' \cdot U_1(w,\alpha') \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \pi_\alpha(dw) \right\| \\
&+ \left\| \int M' \cdot U_1(w,\alpha') \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \pi_\alpha(dw) - \int M' \cdot U_1(w,\alpha') \cdot \begin{pmatrix} 1 \\ w \end{pmatrix} \pi_{\alpha'}(dw) \right\| \\
\leq& \bar{U}_1 \int (1+w^2)\pi_\alpha(dw) \cdot \|M - M'\| + K_1 \int |U_1(w,\alpha) - U_1(w,\alpha')|\,(1+w^2)\pi_\alpha(dw)
\end{aligned}$$

$$+ K_1\Big(K_3 \left|\alpha_0 - \alpha_0'\right| + K_4 \left|\alpha_1 - \alpha_1'\right|\Big)$$

where K_3 and K_4 are given by the Lipschitz continuity of π_α. We only need to show now that $U_1(w, \alpha)$ is Lipschitz continuous. Since

$$\begin{aligned}
|u''(\alpha_0 + \alpha_1 w) - u''(\alpha_0' + \alpha_1' w)| &= \Big|u'''(\xi)\Big((\alpha_0 - \alpha_0') + (\alpha_1 - \alpha_1')w\Big)\Big| \\
&\leq |u'''(\xi)| \Big(|\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\, w\Big) \\
&\leq \sup_{\alpha_0, \alpha_1} \left|u'''(\min\left\{\alpha_0 + \alpha_1 \underline{y}, \underline{y}\right\})\right| \Big(|\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\, w\Big), \\
|u'(\alpha_0 + \alpha_1 w) - u'(\alpha_0' + \alpha_1' w)| &= \Big|u''(\xi')\Big((\alpha_0 - \alpha_0') + (\alpha_1 - \alpha_1')w\Big)\Big| \\
&\leq |u''(\xi')| \Big(|\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\, w\Big) \\
&\leq \sup_{\alpha_0, \alpha_1, w} |u''(\alpha_0 + \alpha_1 w)| \Big(|\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\, w\Big), \\
\left|u'(c_{t+1}) - u'(c_{t+1}')\right| &= \left|u''(\xi'')(c_{t+1} - c_{t+1}')\right| \\
&\leq |u''(\xi'')| \left|c_{t+1} - c_{t+1}'\right| \\
&\leq \sup_{\alpha_0, \alpha_1, w} |u''(\min\left\{\alpha_0 + \alpha_1 w, w\right\})| \left|c_{t+1} - c_{t+1}'\right|.
\end{aligned}$$

There are 10 different cases of $\left|c_{t+1} - c_{t+1}'\right|$ to analyze:

(a) If $c_{t+1} = y = c_{t+1}'$, then $\left|c_{t+1} - c_{t+1}'\right| = 0$.

(b) If $c_{t+1} = \alpha_0 + \alpha_1 y$ and $c_{t+1}' = y$, then by assumption, $\alpha_0' + \alpha_1' y > y \geq \alpha_0 + \alpha_1 y$, so that

$$\begin{aligned}
\left|c_{t+1} - c_{t+1}'\right| &= |\alpha_0 + (\alpha_1 - 1)y| = |(1 - \alpha_1)y - \alpha_0| < |(\alpha_0' - \alpha_0) + (\alpha_1' - \alpha_1)y| \\
&\leq |\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\, \bar{y}.
\end{aligned}$$

(c) If $c_{t+1} = R(1 - \alpha_1)w - \alpha_0 R + y$ and $c_{t+1}' = y$, then $\alpha_0' + \alpha_1' y > y \geq \alpha_0 + \alpha_1 y$, so that

$$\begin{aligned}
\left|c_{t+1} - c_{t+1}'\right| &= |R(1 - \alpha_1)w - \alpha_0 R| = R\,|(1 - \alpha_1)w - \alpha_0| < R\,|(\alpha_0' - \alpha_0) + (\alpha_1' - \alpha_1)w| \\
&\leq R\Big(|\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\, w\Big).
\end{aligned}$$

(d) If $c_{t+1} = \alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y$ and $c'_{t+1} = y$, then

$$y < \alpha_0 + \alpha_1 y, \qquad \alpha'_0 + \alpha'_1 w > w \geq \alpha_0 + \alpha_1 w,$$

and $R(1-\alpha_1)w - \alpha_0 R + y \geq \alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y$, so that

$$\begin{aligned}\left|c_{t+1} - c'_{t+1}\right| =& \left|\alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y - y\right| \\ \leq& R\left|(1-\alpha_1)w - \alpha_0\right| \leq R\Big(\left|\alpha_0 - \alpha'_0\right| + \left|\alpha_1 - \alpha'_1\right| w\Big).\end{aligned}$$

(e) If $c_{t+1} = \alpha_0 + \alpha_1 y$ and $c'_{t+1} = \alpha'_0 + \alpha'_1 y$, then

$$\left|c_{t+1} - c'_{t+1}\right| = \left|\alpha_0 + \alpha_1 y - \alpha'_0 - \alpha'_1 y\right| \leq \left|\alpha_0 - \alpha'_0\right| + \left|\alpha_1 - \alpha'_1\right| \bar{y}.$$

(f) If $c_{t+1} = R(1-\alpha_1)w - \alpha_0 R + y$ and $c'_{t+1} = \alpha'_0 + \alpha'_1 y$, then

$$y \geq \alpha'_0 + \alpha'_1 y, \qquad \alpha'_0 + \alpha'_1 w > w \geq \alpha_0 + \alpha_1 w,$$

and $\alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y \geq R(1-\alpha_1)w - \alpha_0 R + y$, so that

$$\begin{aligned}\left|c_{t+1} - c'_{t+1}\right| =& \left|R(1-\alpha_1)w - \alpha_0 R + y - \alpha'_0 - \alpha'_1 y\right| \\ =& R(1-\alpha_1)w - \alpha_0 R + y - \alpha'_0 - \alpha'_1 y \\ \leq& \alpha_0 + \alpha_1\Big(R(1-\alpha_1)w - \alpha_0 R + y\Big) - \alpha'_0 - \alpha'_1 y \\ =& (\alpha_0 - \alpha'_0) + (\alpha_1 - \alpha'_1)y + R\alpha_1\Big((1-\alpha_1)w - \alpha_0\Big) \\ \leq& (\alpha_0 - \alpha'_0) + (\alpha_1 - \alpha'_1)\bar{y} + R\bar{\alpha}_1\Big((\alpha'_1 - \alpha_1)w - (\alpha'_0 - \alpha_0)\Big) \\ \leq& (1 + R\bar{\alpha}_1)\left|\alpha_0 - \alpha'_0\right| + (\bar{y} + R\bar{\alpha}_1 w)\left|\alpha_1 - \alpha'_1\right|.\end{aligned}$$

(g) If $c_{t+1} = \alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y$ and $c'_{t+1} = \alpha'_0 + \alpha'_1 y$, then

$$y \geq \alpha'_0 + \alpha'_1 y, \qquad \alpha'_0 + \alpha'_1 w > w \geq \alpha_0 + \alpha_1 w,$$

and $\alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y \leq R(1-\alpha_1)w - \alpha_0 R + y$, so that

$$\left|c_{t+1} - c'_{t+1}\right| = \left|\alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y - \alpha'_0 - \alpha'_1 y\right|$$

$$
\begin{aligned}
&= R(1-\alpha_1)w - \alpha_0 R + y - \alpha_0' - \alpha_1' y \\
&\le |\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\,\bar{y} + R\alpha_1\,|(1-\alpha_1)w - \alpha_0| \\
&\le |\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\,\bar{y} + R\bar{\alpha}_1\Big(|\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\,w\Big) \\
&= (1 + R\bar{\alpha}_1)\,|\alpha_0 - \alpha_0'| + (\bar{y} + R\bar{\alpha}_1 w)\,|\alpha_1 - \alpha_1'|\,.
\end{aligned}
$$

(h) If $c_{t+1} = R(1-\alpha_1)w - \alpha_0 R + y$ and $c'_{t+1} = R(1-\alpha_1')w - \alpha_0' R + y$, then

$$
\left|c_{t+1} - c'_{t+1}\right| \le R\Big(|\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\,w\Big).
$$

(i) If $c_{t+1} = R(1-\alpha_1)w - \alpha_0 R + y$ and $c'_{t+1} = \alpha_0' + R\alpha_1'(1-\alpha_1') - R\alpha_0'\alpha_1' + \alpha_1' y$, then

$$
\begin{aligned}
&w \ge \alpha_0 + \alpha_1 w \qquad R(1-\alpha_1)w - \alpha_0 R + y < \alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y \\
&w \ge \alpha_0' + \alpha_1' w \qquad R(1-\alpha_1')w - \alpha_0' R + y \ge \alpha_0' + R\alpha_1'(1-\alpha_1') - R\alpha_0'\alpha_1' + \alpha_1' y
\end{aligned}
$$

so that, if $c_{t+1} - c'_{t+1} \ge 0$, then

$$
\begin{aligned}
\left|c_{t+1} - c'_{t+1}\right| &= R(1-\alpha_1)w - \alpha_0 R + y - \alpha_0' - R\alpha_1'(1-\alpha_1') + R\alpha_0'\alpha_1' - \alpha_1' y \\
&\le \alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y - \alpha_0' - R\alpha_1'(1-\alpha_1') + R\alpha_0'\alpha_1' - \alpha_1' y \\
&= (\alpha_0 - \alpha_0') + (\alpha_1 - \alpha_1')y + (\alpha_1 - \alpha_1')R\Big((1-\alpha_1)w - \alpha_0\Big) \\
&\quad + \alpha_1' R\Big((\alpha_0' - \alpha_0) + (\alpha_1' - \alpha_1)w\Big) \\
&\le |\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\,\bar{y} + |\alpha_1 - \alpha_1'|\,Rw + R\bar{\alpha}_1\Big(|\alpha_0' - \alpha_0| + |\alpha_1' - \alpha_1|\,w\Big) \\
&= (1 + R\bar{\alpha}_1)\,|\alpha_0 - \alpha_0'| + (\bar{y} + R(1+\bar{\alpha}_1)w)\,|\alpha_1 - \alpha_1'|\,.
\end{aligned}
$$

On the other hand, if $c_{t+1} - c'_{t+1} \le 0$, then

$$
\begin{aligned}
\left|c'_{t+1} - c_{t+1}\right| &= \alpha_0' + R\alpha_1'(1-\alpha_1') - R\alpha_0'\alpha_1' + \alpha_1' y - R(1-\alpha_1)w + \alpha_0 R - y \\
&\le R(1-\alpha_1')w - \alpha_0' R + y - R(1-\alpha_1)w + \alpha_0 R - y \\
&\le R\Big(|\alpha_0 - \alpha_0'| + |\alpha_1 - \alpha_1'|\,w\Big).
\end{aligned}
$$

(j) If $c_{t+1} = \alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y$ and $c'_{t+1} = \alpha_0' + R\alpha_1'(1-\alpha_1') - R\alpha_0'\alpha_1' + \alpha_1' y$, then

$$\left|c_{t+1} - c'_{t+1}\right| = \left|\alpha_0 + R\alpha_1(1-\alpha_1) - R\alpha_0\alpha_1 + \alpha_1 y - \alpha'_0 - R\alpha'_1(1-\alpha'_1) + R\alpha'_0\alpha'_1 - \alpha'_1 y\right|$$
$$\leq (1 + R\bar{\alpha}_1)\left|\alpha_0 - \alpha'_0\right| + (\bar{y} + R(1+\bar{\alpha}_1)w)\left|\alpha_1 - \alpha'_1\right|.$$

So, since $R \geq 1$, $\bar{y} > 0$ and $\bar{\alpha}_1 > 0$, I have that in general

$$\left|c_{t+1} - c'_{t+1}\right| \leq (1 + R\bar{\alpha}_1)\left|\alpha_0 - \alpha'_0\right| + (\bar{y} + R(1+\bar{\alpha}_1)w)\left|\alpha_1 - \alpha'_1\right|.$$

$$\begin{aligned}
|U_1(w\alpha) - U_1(w,\alpha')| =& \Big|\Big(\beta R E_t u'(c_{t+1}) - u'(\alpha_0 + \alpha_1 w)\Big)u''(\alpha_0 + \alpha_1 w) \\
& - \Big(\beta R E_t u'(c'_{t+1}) - u'(\alpha'_0 + \alpha'_1 w)\Big)u''(\alpha'_0 + \alpha'_1 w)\Big| \\
=& \Big|\Big(\beta R E_t u'(c_{t+1}) - u'(\alpha_0 + \alpha_1 w)\Big)\Big(u''(\alpha_0 + \alpha_1 w) - u''(\alpha'_0 + \alpha'_1 w)\Big) \\
+u''(\alpha'_0 + \alpha'_1 w) & \Big(\beta R E_t u'(c_{t+1}) - u'(\alpha_0 + \alpha_1 w) - \beta R E_t u'(c'_{t+1}) + u'(\alpha'_0 + \alpha'_1 w)\Big)\Big| \\
\leq & \left|\beta R E_t u'(c_{t+1}) - u'(\alpha_0 + \alpha_1 w)\right| \left|u''(\alpha_0 + \alpha_1 w) - u''(\alpha'_0 + \alpha'_1 w)\right| \\
+ \left|u''(\alpha'_0 + \alpha'_1 w)\right| & \Big(\beta R E_t \left|u'(c_{t+1}) - u'(c'_{t+1})\right| + \left|u'(\alpha_0 + \alpha_1 w) - u'(\alpha'_0 + \alpha'_1 w)\right|\Big) \\
\leq & (1+\beta R)\sup_{\alpha_0,\alpha_1}\left|u'(\min\left\{\alpha_o + \alpha_1 \underline{y}, \underline{y}\right\})\right| \sup_{\alpha_0,\alpha_1}\left|u'''(\min\left\{\alpha_0 + \alpha_1 \underline{y}, \underline{y}\right\})\right| \\
& \cdot \Big(\left|\alpha_0 - \alpha'_0\right| + \left|\alpha_1 - \alpha'_1\right| w\Big) + \sup_{\alpha_0,\alpha_1}\left|u''(\min\left\{\alpha_o + \alpha_1 \underline{y}, \underline{y}\right\})\right|^2 \\
& \cdot \Big(\beta R\Big[(1 + R\bar{\alpha}_1)\left|\alpha_0 - \alpha'_0\right| + (\bar{y} + R(1+\bar{\alpha}_1)w)\left|\alpha_1 - \alpha'_1\right|\Big] \\
& + \Big[\left|\alpha_0 - \alpha'_0\right| + \left|\alpha_1 - \alpha'_1\right| w\Big]\Big) \\
=& K_5\left|\alpha_0 - \alpha'_0\right| + K_6\left|\alpha_1 - \alpha'_1\right| + K_7\left|\alpha_1 - \alpha'_1\right| w
\end{aligned}$$

where

$$\begin{aligned}
K_5 =& (1+\beta R)\sup_{\alpha_0,\alpha_1}\left|u'(\min\left\{\alpha_o + \alpha_1 \underline{y}, \underline{y}\right\})\right| \sup_{\alpha_0,\alpha_1}\left|u'''(\min\left\{\alpha_0 + \alpha_1 \underline{y}, \underline{y}\right\})\right| \\
& + \sup_{\alpha_0,\alpha_1}\left|u''(\min\left\{\alpha_o + \alpha_1 \underline{y}, \underline{y}\right\})\right|^2 \Big(\beta R(1 + R\bar{\alpha}_1) + 1\Big), \\
K_6 =& \sup_{\alpha_0,\alpha_1}\left|u''(\min\left\{\alpha_o + \alpha_1 \underline{y}, \underline{y}\right\})\right|^2 \bar{y}, \\
K_7 =& (1+\beta R)\sup_{\alpha_0,\alpha_1}\left|u'(\min\left\{\alpha_o + \alpha_1 \underline{y}, \underline{y}\right\})\right| \sup_{\alpha_0,\alpha_1}\left|u'''(\min\left\{\alpha_0 + \alpha_1 \underline{y}, \underline{y}\right\})\right|
\end{aligned}$$

$$+\sup_{\alpha_0,\alpha_1}\left|u''(\min\left\{\alpha_o+\alpha_1\underline{y},\underline{y}\right\})\right|^2\Big(R(1+\bar{\alpha}_1)+1\Big).$$

Thus, I have the Lipschitz continuity of (ODE-DG), i.e. $h(\alpha, M)$.

Now, define $v(\alpha,w)=\sum_n(P_\alpha^n-\pi_\alpha)H(\alpha,w)$. Let's see that it is well defined. For that, since $\|P_\alpha^n-\pi_\alpha\|_\phi\leq K_\alpha\rho_\alpha^n$, as was established before (see the proof of $\mathcal{D}_\phi$-continuity and differentiability), I have that $|(P_\alpha^n-\pi_\alpha)H(\alpha,w)|\leq K_\alpha\left\|H\right\|_\phi\rho_\alpha^n\phi(w)$. Thus,

$$\sum_n|(P_\alpha^n-\pi_\alpha)H(\alpha,w)|\leq\frac{c_\alpha\left\|H\right\|_\phi}{1-\rho_\alpha}\phi(w)<\infty.$$

Furthermore, I have that $(I-\pi_\alpha)v(\alpha,w)=H(\alpha,w)-h(\alpha)$,

$$|v(\alpha,w)|\leq\frac{K_\alpha\left\|H\right\|_\phi}{1-\rho_\alpha}\phi(w)\leq K_8(1+w)$$

and $P_\alpha v(\alpha,w)$ is Lipschitz continuous.

A.5 If $w\leq\bar{\bar{w}}$, then $w_t\leq\bar{\bar{w}}$ for all $t\geq0$, while if $w>\bar{\bar{w}}$, then $w_t\leq w$, so that

$$E_{w,\alpha}(I((\alpha)\in\mathcal{Q}_2,\ k\leq t)\left|w_{t+1}\right|^q)\leq K_8(1+w^q),$$

where $K_8\geq\max\left\{1,\bar{\bar{w}}\right\}$.

A.6 This holds by assumption C.

A.7 Corollary A.7 ensures this. ■

Finally, corollary 16 of Benveniste *et al.* (1990) ensures that:

Corollary A.9. *α_t converges to α^* a.s.* ■

Proof of corollary 4.2. This results is a direct application of theorem 22 of Benveniste *et al.* (1990, p.244). The previous proof showed that their conditions (1.10.2)-(1.10.4) are satisfied. Additionally, (1.10.6) is satisfied by assumption. It is only necessary to show that their condition (1.10.5) also holds. To see this, notice that the agent's α^*-expected squared regret can be considered a function of three different set of consumption rule parameters. In particular rewrite it as

$$V(\alpha, \alpha', \sigma) = \int \Big(\beta R E_t u'(c^b(\alpha', w_{t+1})) - u'(c^b(\alpha, w_t))\Big)^2 \pi_\sigma(dw),$$

which is a strictly convex function of α. This and our previous results imply that

$$\begin{aligned}
V(\alpha^*, \alpha^*, \alpha^*) =& V(\alpha, \alpha, \alpha) - (1 - R\beta)(\alpha^* - \alpha)h(\alpha) + (\alpha^* - \alpha)V_\sigma + \frac{1}{2}(\alpha^* - \alpha)V_{\alpha\alpha}(\alpha^* - \alpha)^T \\
\geq& V(\alpha, \alpha, \alpha) - (1 - R\beta)(\alpha^* - \alpha)h(\alpha) + \frac{1}{2}(\alpha^* - \alpha)V_{\alpha\alpha}(\alpha^* - \alpha)^T \iff \\
0 \geq V(\alpha^*, \alpha^*, \alpha) - V(\alpha, \alpha, \alpha) \geq& -(1 - R\beta)(\alpha^* - \alpha)h(\alpha) + \frac{1}{2}(\alpha^* - \alpha)V_{\alpha\alpha}(\alpha^* - \alpha)^T \iff \\
(1 - R\beta)(\alpha^* - \alpha)h(\alpha) \geq& \frac{1}{2}(\alpha^* - \alpha)V_{\alpha\alpha}(\alpha^* - \alpha)^T \iff \\
(\alpha - \alpha^*)h(\alpha) \leq& -\frac{1}{2(1 - R\beta)}(\alpha^* - \alpha)V_{\alpha\alpha}(\alpha^* - \alpha)^T \\
<& -\frac{1}{2(1 - R\beta)} \inf \|V_{\alpha\alpha}\| \, \|\alpha - \alpha^*\|^2 \\
<& -\delta \, \|\alpha - \alpha^*\|^2
\end{aligned}$$

for some $\delta > 0$. ■