Adaptive Consumption Behavior[☆]

Peter Howitt[a,*], Ömer Özak[b]

[a] *Department of Economics, Brown University*
[b] *Department of Economics, Southern Methodist University*

Abstract

In this paper we propose and study a theory of adaptive consumption behavior under income uncertainty and liquidity constraints. We assume that consumption is governed by a linear function of wealth, whose coefficients are revised each period by a procedure that places few informational or computational demands on the consumer. We show that under a variety of settings the procedure converges quickly to a set of coefficients with low welfare cost relative to a fully optimal nonlinear consumption function.

Keywords: Learning, consumption function, liquidity constraint, dynamic programming
JEL classification: C63, E21

1. Introduction

The standard theory of lifetime utility maximization under uncertainty and liquidity constraints places enormous informational and computational demands on the consumer. Carroll (2001) 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". This problem raises two closely related questions. First, are there simpler rules that have low welfare costs? And, second, can consumers learn the optimal rule or a simple low-cost rule? While the answer to the first question has been generally positive (see Akerlof and Yellen, 1985a,b; Allen and Carroll, 2001; Cochrane, 1989), the second one has only been addressed in a few papers with negative results (see e.g. Allen and Carroll, 2001; Lettau and Uhlig, 1999).

In this paper, we provide a positive answer to the second question by proposing an novel adaptive theory of consumption behavior. Our approach, in the spirit of Simon (1990), Arthur (1994) and Clark (1997), places limited computational and rationality demands on the consumer. Using the basic setup of Allen and Carroll

[☆]We would like to thank an anonymous referee, participants at Brown University's theory and macroeconomics seminars, participants at the Econometric Society's LAMES 2008 and NASM 2009 meetings, and especially Kfir Eliaz, George Evans, Axel Leijonhufvud, Glenn Loury, and David Weil for helpful comments and discussion.

[*]Address correspondence to: Peter Howitt, Department of Economics, Brown University, Box B, Providence, RI 02912, USA
Email addresses: Peter_Howitt@brown.edu (Peter Howitt), ozak@smu.edu (Ömer Özak)
URL: http://www.econ.brown.edu/fac/peter_howitt/ (Peter Howitt), http://omerozak.com (Ömer Özak)

(2001) we show that consumers who use this adaptive mechanism are able to learn a rule with a low welfare cost after a few periods. In particular, we show that constant-relative-risk-aversion consumers who follow a linear consumption rule in wealth and use our proposed algorithm, which adaptively adjusts the parameters of their rule, lose less than 0.5% of the equivalent consumption of the fully rational consumption rule within 500 periods with a probability higher than 0.9. Furthermore, we show that under social learning, the time required to attain a loss of 0.5% falls to less than 100 periods for some parametrizations. Additionally, the mean and median welfare losses in a population, under both individual and social learning, fall to around 1% in less than 25 periods.[1]

Our approach and results stand in sharp contrast with the previous literature. Lettau and Uhlig (1999) use a variant of Holland's (1992) classifier system, in which they call the measure of a rule's success its "strength". In each period after a rule has been used, its strength is adjusted partially towards the sum of the immediate utility attained under the rule last period plus the discounted strength of the rule that has succeeded it. They show however that the classifier system does a poor job of approximating optimal consumption behavior, even when the optimal consumption rule is available to the consumer. Allen and Carroll (2001), on the other hand, assume that the consumer is able to perform what amounts to a Monte Carlo simulation to evaluate each rule. They show that this procedure, instead of being quick and simple, actually needs 4 million periods in order to find the optimal linear rule in their parameterization.[2]

We follow Allen and Carroll in restricting consumers to rules that are linear in current wealth (for their parameterization they show that the optimal linear rule is almost as good as the fully optimal rule).[3] But instead of responding to a measure of cumulative discounted utility, we assume that the consumer adjusts her rule gradually in response to the difference between the immediate marginal utility implied by the rule and the discounted marginal utility of the consumption experienced next period. In effect our criterion of success is the ex post Euler equation error, and our algorithm operates like a stochastic approximation (see Robbins and Monro (1951), Ljung (1977) or Kushner and Yin (2003)) for solving the consumption Euler equation.[4]

Although our approach presumes an awareness of sophisticated notions of Euler equations, nevertheless the

[1]In an independent paper, Evans and McGough (2009) also address the question of adaptive approximation to optimal intertemporal choice. They present a procedure for updating expectations that is asymptotically fully optimal in a linear-quadratic environment, given that the decision maker knows enough about her environment to specify the correct functional form of her policy function. By contrast, we have found an adaptive procedure for updating the parameters of the policy function that works "reasonably well" even outside a linear-quadratic environment when the decision-maker does not know the correct functional form. Özak (2013) studies our algorithm under more general assumptions and shows that convergence occurs quite generally.

[2]Allen and Carroll use a 5% threshold in order to assume the optimal linear rule has been found successfully. Under this assumption they require 1 million periods in order to get a success rate of 0.75 and of 4 million to get at least a success rate of 0.85.

[3]Padula (2010) presents a family of two parameter non-linear consumption functions that are "almost" fully optimal. His purpose is to use this functional form in order to simplify the computation of the rational solution by economists. His framework might provide a new avenue to learning the optimal consumption rule.

[4]The idea that the opportunity cost of current spending could be learned adaptively through experience rather than calculated ex ante was suggested by Leijonhufvud (1993) in the context of Marshallian demand theory. The "micro-based" model in Howitt (1992), as well as the learning models in the New Keynesian DSGE literature (Woodford (2003), pp.261-9) also suppose that people revise their planned consumption in direct response to Euler equation errors.

informational and computational requirements of our algorithm are very low. Moreover, these requirements are independent of the size of the set of rules or states, which make it a good candidate for an adaptive procedure under bounded rationality for the problem at hand. In contrast, both requirements are increasing in the number of rules and states for Lettau and Uhlig and for Allen and Carroll. The reason for this difference is that in our algorithm a consumer needs to revise the two parameters of her linear rule each period, based only on its performance last period, whereas in the other papers she must keep track of the past performance of a large number of rules under all states.

In addition to our baseline simulations, we generalize our mechanism in order to study the effects of social learning, relaxation of credit constraints, and changes to the consumer's income process. Our brief analysis of social learning suggests that when imitation is allowed for, both the mean and median welfare losses become small even faster. On the other hand, we find that relaxing the credit constraint diminishes the incentives for consumers to learn a good rule, and thus slows down their learning process, an effect observed by Brown, Chua, and Camerer (2009) in their experiments on saving behavior. Finally, we deal with the problem of changes in consumer's environment, in particular, her income process. We show how our procedure can deal with the Lucas critique by making the consumer aware of regime changes. Furthermore, since for some parameterizations she uses a constant gain adjustment procedure, her ex post reaction to a regime change will often be quick and in the appropriate direction even when she is not aware of the change.

Our results show that the use of non-rational/behavioral consumption theories does not necessarily imply costly behavior by agents.[5] This is especially important for agent-based computational economics (ACE), where the macroeconomy is studied by endowing each agent with a set of adaptive behavioral algorithms.[6] Rejection of the standard paradigm of the rational agent has left ACE practitioners without a "good" model for consumption decisions, thus creating a proliferation of consumption behavior assumptions in the ACE literature.[7] Regrettably, besides being "simple", very little is known about the welfare properties of these behavioral rules. Furthermore, the literature has not used adaptive algorithms, like ours, which will allow agents to adapt to their environment through learning. On the contrary, modeling consumption decisions based

[5] In a similar vein, Feigenbaum, Caliendo, and Gahramanov (2011) show that irrational behavior can generate higher welfare in certain market settings.

[6] Tesfatsion (2006a) provides a much fuller account of ACE methodolgy. Advocates of ACE macro include Leijonhufvud (1993), Chen (2003), Axtell (2006), Leijonhufvud (2006), Tesfatsion (2006b), Colander, Howitt, Kirman, Leijonhufvud, and Mehrling (2008), LeBaron and Tesfatsion (2008), Howitt (2008), Farmer and Foley (2009) and Buchanan (2009). ACE macro models have been developed by Albin and Foley (1992), Basu, Pryor, and Quint (1998), Bruun (1999), Arifovic (2000), Howitt and Clower (2000), Bruce (2003), Bruun (2003), Dosi, Fagiolo, and Roventini (2005), Howitt (2006), Ashraf, Gershman, and Howitt (2013), Chan and Steiglitz (2008), Canzian (2009), Ashraf, Gershman, and Howitt (2011), Delli Gatti, Desiderio, Gaffeo, Cirillo, and Gallegati (2010), Raberto, Teglio, and Cincotti (2010) and the papers included in the March 2001 issue of the Journal of Economic Dynamics and Control, among others. Early precursors include Eliasson (1977) and the microsimulation-based macro models of Orcutt, Caldwell, and Wertheimer II (1976) and Bennett and Bergmann (1986).

[7] For example, almost every one of the papers cited in footnote 6 above uses a different algorithm. Some eliminate the consumption-saving decision by assuming there are no durable household assets (Albin and Foley (1992), Bruce (2003), Dosi et al. (2005)) or limit its scope by assuming only two period lives (Arifovic (2000)) and the rest have various ways of recognizing that a household's consumption will depend on such factors as income, financial assets and interest rates.

on our theory ensures that consumption behavior

1. has low informational and computational requirements,
2. implies low welfare costs, and
3. generates observable behavior that conforms with empirical evidence.

These criteria are essential in ACE, but also in other macro-modeling frameworks. Criterion (1) is motivated partly by the need to have reasonably fast runtimes in computer simulations with a large number of heterogeneous agents, and partly by the observation (see for example Clark (1997)) that human behavior has evolved under conditions requiring fast reaction to a wide variety of events, such as encounters with potential predators or prey, with little time for processing information. Criterion (2) is likewise motivated by the idea that poorly performing algorithms are unlikely to survive evolution and adaptation. Criterion (3) is self-evidently important.[8] By using our theory and similar methods, ACE practitioners can rid themselves of one of the main critiques they are confronted with.

The paper proceeds as follows: section 2 presents the model and the different consumption rules we use, section 3 describes the adaptive algorithm and its properties, section 4 derives our measures of welfare, section 5 shows the results of simulations and section 6 concludes. All tables and figures are presented in appendix A.

2. The setting

A consumer's lifetime utility is $U = \sum_{t=1}^{\infty} \delta^t u(c_t)$, with an isoelastic flow utility function:

$$u(c_t) = \begin{cases} \frac{c_t^{1-\theta}-1}{1-\theta} & \text{if } \theta \neq 1 \\ \ln c_t & \text{if } \theta = 1 \end{cases} \tag{1}$$

and

$$0 < \delta < 1. \tag{2}$$

She starts each period t with wealth w_t, of which she consumes c_t. She faces a liquidity constraint

$$c_t \leq w_t \tag{3}$$

and her wealth evolves according to the flow budget constraint

$$w_{t+1} = w_t - c_t + y_{t+1}, \tag{4}$$

[8] Descriptive realism of algorithms is not typically sought in ACE models, which often rely on such techniques as neural nets and genetic algorithms that do not map readily into the observable procedures of real people.

where y_{t+1} is next period's income. (We assume for simplicity that the interest rate on saving is zero.) Income is an independently and identically distributed random variable with discrete support $\{y^i\}_{i=1}^n$ where $0 < y^1 < y^2 < \ldots < y^n$ and the probability of each y^i is $p^i > 0$.

The consumer's behavior is determined by a consumption function

$$c_t = c(w_t), \tag{5}$$

which obeys the liquidity constraint (3). We assume that the consumption function is derived from a "notional" consumption function, $\widehat{c}(w_t)$ that ignores the liquidity constraint, so that

$$c(w_t) = \min\{\widehat{c}(w_t), w_t\}. \tag{6}$$

We refer to the notional function $\widehat{c}(\cdot)$ as the consumer's "rule". In what follows, we assume that $\widehat{c}(\cdot)$ is increasing and concave, and satisfies

Assumption A. *There exists $\widetilde{w} > 0$ such that $\widehat{c}(w) > y^n$ for all $w > \widetilde{w}$,*

which guarantees that the consumer's wealth will eventually be bounded above by $\widetilde{w}$. Specifically, theorem 1.2 in Özak (2013, p.7) assures that under the above assumptions there is a unique invariant wealth distribution π, whose support is contained in the interval $[y^1, \widetilde{w}]$.

Assume also that

$$\widehat{c}(0) \geq 0, \text{ and} \tag{7}$$

$$\widehat{c}(\overline{w}) = \overline{w} \text{ for a unique } \overline{w} \in (y^1, y^n) \tag{8}$$

This implies that when wealth surpasses $\overline{w}$ the consumer is no longer liquidity constrained, so that[9]

$$c(w) = \begin{cases} w & \text{if } 0 \leq w \leq \overline{w} \\ \widehat{c}(w) < w & \text{if } w > \overline{w} \end{cases} \tag{9}$$

We refer to $\overline{w}$ as the consumer's "crossover wealth". The assumption that $\overline{w} > y^1$ requires the liquidity constraint to be binding for at least some observable wealth levels in the long-run; otherwise the consumer

[9] *Proof*: (a) If $0 \leq w \leq \overline{w}$, then by concavity $\widehat{c}(w) \geq \frac{\overline{w}-w}{\overline{w}}\widehat{c}(0) + \frac{w}{\overline{w}}\widehat{c}(\overline{w})$, ($\overline{w} > 0$ by (8) and the assumption that $y^1 > 0$). So by (7) and (8) we have $\widehat{c}(w) \geq w$, which implies that $c(w) = \min\{\widehat{c}(w), w\} = w$.
(b) If $w > \overline{w}$, then to show that $c(w) = \widehat{c}(w) < w$ suppose on the contrary that either $c(w) \neq \widehat{c}(w)$ or $\widehat{c}(w) \geq w$. In either case we have $\widehat{c}(w) \geq w$, so by the uniqueness assumed in (8) we have $\widehat{c}(w) > w$. But, by concavity $\widehat{c}(\overline{w}) \geq \frac{w-\overline{w}}{w}\widehat{c}(0) + \frac{\overline{w}}{w}\widehat{c}(w)$ so by (7) $\widehat{c}(\overline{w}) > \overline{w}$, which contradicts (8). ■

would start each period, after the first, with enough wealth that she would never be liquidity constrained.[10] On the other hand, the assumption that $\overline{w} < y^n$ requires the consumer to save with positive probability in the long run; otherwise (given Assumption A) she would in finite time end up with $c_t = w_t = y_t$ for all t. We call a rule $\widehat{c}(\cdot)$ *admissible* if it is increasing, concave and satisfies (7), (8) and Assumption A. The adaptive algorithm we specify below for revising the consumer's rule ensures that the rule always remains admissible. In the next two subsections we give conditions under which the fully optimal consumption rule and a linear rule are admissible.

2.1. Optimal consumption

The optimal consumption function $c^*(w)$ can be derived from the dynamic programming problem:

$$V(w) = \max_{c \leq w} \{u(c) + \delta E_y V(w - c + y)\} \tag{10}$$

where E_y is the expectation with respect to income y. This corresponds to the notional function

$$\widehat{c}^*(w) = \arg\max_c \{u(c) + \delta E_y V(w - c + y)\}, \tag{11}$$

which we refer to as the "fully optimal" rule. For future reference, note that the first-order condition defining $\widehat{c}^*(w)$ is

$$u'(c) = E_y q, \tag{12}$$

where q is the marginal continuation value

$$q = \delta u'\Big(c^*\big(w - c + y\big)\Big) \tag{13}$$

whose value is not known when c is chosen because it depends on next period's income y.

It is known that under our assumptions the fully optimal rule $\widehat{c}^*(\cdot)$ is indeed increasing and concave, that it satisfies (7) and Assumption A, and that for δ close enough to unity (i.e. if the consumer is patient enough) it also satisfies (8) and hence is admissible.[11] From now on we assume that the consumer is indeed patient enough that $\widehat{c}^*(\cdot)$ satisfies (8) and is thus admissible.

[10] This follows directly from the liquidity constraint (3) and the flow budget constraint, which together imply that w_t cannot fall below y^1 for $t \geq 1$.

[11] (See e.g. Carroll, 2004; Carroll and Kimball, 1996; Özak, 2013)

2.2. Linear consumption

Our approach to modelling consumption follows Carroll and Allen (2001) in assuming that each consumer's consumption rule is linear, with coefficients $\gamma = (\alpha, \beta)$:

$$\widehat{c}^{\gamma}(w) = \alpha + \beta w. \tag{14}$$

In order for $\widehat{c}^{\gamma}(\cdot)$ to be increasing and to obey (7) we need $\alpha \geq 0$ and $\beta > 0$. In order to satisfy (8), given that $y^1 > 0$, we also need $\alpha > 0$ and $\beta < 1$. The consumer's crossover wealth is then

$$\overline{w}^{\gamma} = \frac{\alpha}{1-\beta} > 0 \tag{15}$$

and the consumption function can be written as

$$c^{\gamma}(w) = \min\left\{\overline{w}^{\gamma} + \beta\left(w - \overline{w}^{\gamma}\right), w\right\} \tag{16}$$

Assumption (8) also requires

$$(1-\beta)\, y^1 < \alpha < (1-\beta)\, y^n. \tag{17}$$

We say that a coefficient vector γ is admissible if the rule $\widehat{c}^{\gamma}(\cdot)$ is admissible. It is straightforward to demonstrate that:

Proposition 1. *The coefficients* (α, β) *are admissible if and only if* $\alpha > 0$, $\beta < 1$ *and they satisfy (17).*

The set of admissible coefficients is denoted $\mathcal{A}$ and is illustrated by the shaded triangular area in Figure 1 below.

3. Revising the coefficients

At each date t a consumer is free to choose a new linear consumption rule $\widehat{c}^{\gamma_t}(\cdot)$, with coefficients $\gamma_t = (\alpha_t, \beta_t)$. The basic idea behind our proposal is to suppose that in choosing these coefficients she is attempting to solve the Euler equation (12), which can be written as

$$E_y q_{t+1} = u'\left(\alpha + \beta w_t\right), \tag{18}$$

but she doesn't know how to form a rational expectation of $q_{t+1} = \delta u'(c_{t+1})$. Of course, q_{t+1} will depend on the distribution of income, her current choice of coefficients and her current wealth w_t, which together will determine the distribution of w_{t+1}, and on her future choice of coefficients, which together with w_{t+1} will determine c_{t+1}. But she is unable to process all this information.

Instead, we assume she first computes the "error" she made in period $t-1$

$$e_{t-1} = q_t - u'\left(\alpha_{t-1} + \beta_{t-1} w_{t-1}\right) \tag{19}$$

where q_t is the realized marginal continuation value

$$q_t = \delta R u'\left(c_t\right) \tag{20}$$

Clearly if $e_{t-1} = 0$ then if she had known that the marginal continuation value was going to be q_t she would have been happy with her choice of coefficients, because it led her to choose a notional consumption function whose marginal utility was just equal to its marginal cost. In this sense, even with hindsight she did not make a mistake. On the other hand if $e_{t-1} \neq 0$ she made an ex post error of consuming too much (if $e_{t-1} > 0$) or too little (if $e_{t-1} < 0$).

Accordingly, we suppose that her revisions will depend on this error according to the following multi-step procedure. She begins with a symmetric 2×2 "moment" matrix M_{t-1}, and then goes through the following steps:

Algorithm.

1. *Choose a new moment matrix M_t using the formula*

$$M_t = (1-\varepsilon) M_{t-1} + \left[\xi e_{t-1} u'''\left(\alpha_{t-1} + \beta_{t-1} w_{t-1}\right) + u''\left(\alpha_{t-1} + \beta_{t-1} w_{t-1}\right)^2\right] \begin{pmatrix} 1 & w_{t-1} \\ w_{t-1} & w_{t-1}^2 \end{pmatrix} \tag{21}$$

   where $\varepsilon \in [0,1)$ is a constant gain parameter, $\xi \in \{0,1\}$ is a parameter allowing further simplification of the procedure.[12]

2. *If M_t is well conditioned*[13] *choose a provisional coefficient vector $\gamma_{t+1}^p = \left(\alpha_{t+1}^p, \beta_{t+1}^p\right)$ according to*

$$\begin{pmatrix} \alpha_{t+1}^p \\ \beta_{t+1}^p \end{pmatrix} = \begin{pmatrix} \alpha_t^p \\ \beta_t^p \end{pmatrix} + M_t^{-1}\left[e_{t-1} u''\left(\alpha_{t-1} + \beta_{t-1} w_{t-1}\right)\right] \begin{pmatrix} 1 \\ w_{t-1} \end{pmatrix} \tag{22}$$

3. *If M_t is not well conditioned or if the provisional γ_{t+1}^p chosen in step 1 is not admissible, choose the*

[12] If $\varepsilon = 0$ this algorithm can be written as a decreasing gain algorithm. If $\xi = 0$ then what would have been a quasi-Newton method, as explained in the appendix Appendix C, becomes what is known in numerical analysis as a quasi-Gauss-Newton method, which obviates the need for calculating the third derivative u'''.

[13] Specifically, if the condition number $r2\left(M_t\right)$ is less than the conventional limit 10^{10}, indicating that the matrix is reliably nonsingular. See Judd (1998).

nearest vector to γ_t that would have eliminated the most recent error, i.e. let

$$\gamma_{t+1}^p = \arg\min_{\gamma} \|\gamma - \gamma_t\|^2 \text{ subj to } q_t - u'(\alpha + \beta w_{t-1}) = 0 \tag{23}$$

4. *If γ_{t+1}^p is still inadmissible, set $\gamma_{t+1}^p = \gamma_t$.*
5. *Shrink the step size by a factor $\eta \in (0,1]$ and set the new coefficients according to*

$$\gamma_{t+1} = \gamma_t + \eta\left(\gamma_{t+1}^p - \gamma_t\right). \tag{24}$$

Appendix Appendix C gives some motivation of how one can interpret this algorithm.

3.1. Informational and computational requirements

The procedure outlined above requires a certain amount of sophistication, in the sense that moment matrices, condition numbers, and Euler equations are not familiar household names. Still, most of these can be translated into everyday language that most people will be familiar with. Additionally, the consumer must also be sophisticated enough to realize that if her crossover wealth is less than y^1 then her liquidity constraint will never bind and if her crossover wealth is greater than y^n then she will eventually reach a situation in which she never again saves. Moreover she needs to realize that both of these outcomes are suboptimal for someone with her rate of time preference. Although this level of economic sophistication might seem excessive for a boundedly rational consumer, we show below that dispensing with it, i.e. allowing for rules in a superset of the set of admissible rules, does not change our results dramatically.

Nevertheless the procedure makes relatively few informational or computational demands on the consumer, especially in comparison to the demands involved in calculating the optimal consumption function. This is an important consideration for intelligent behavior in a world where information storage capacity and computational time are scarce resources. In particular, all the consumer needs to know is her instantaneous utility function, the gain parameter ε, the shrinkage factor η and the minimal and maximal possible income levels y^1 and y^n. Each period she must remember and update only 9 numbers. In addition to elementary addition and multiplication, she just needs to be able to compute the first three derivatives of her utility function, to determine the conditioning value of a 2×2 matrix and to compute its inverse. Of these, clearly the most stringent assumption are the last two. We dispense with these in appendix Appendix A, where we assume agents do not update nor invert the moment matrix. In particular, we assume agents use a constant 2×2 matrix M, which eliminates both the matrix inversion and the computation of its conditioning number, which were performed in the step 2 of the algorithm. We show in the appendix that removing this additional level of rationality generally affects the speed of convergence to a good rule. Thus, with some luck in her choice of matrix agents can learn a good rule at similar rates as the original version of the algorithm. Reassuringly, although the speed of convergence is affected, the convergence still occurs and at faster rates than have been found in the literature. Thus, in the long-run agents learn a good rule, even in this setting (see also Özak, 2013).

4. Welfare cost

We use two different indices to measure the welfare cost for a consumer of following a specific linear consumption rule rather than the fully optimal rule. Both indices are based on ex-ante equivalent consumption for a consumer following a certain rule, starting from a randomly assigned wealth; but each index uses a different probability distribution for assigning initial wealth. The first index uses the stationary distribution implied by the fully optimal consumption rule, whereas the second one uses the stationary distribution implied by the specific linear rule. In either case, the index measures the percentage difference in certainty equivalent consumption between the fully optimal rule and the linear rule. As will be seen, the two indices produce very similar results.

More specifically, suppose that initial wealth w_0 is assigned randomly according to some distribution λ. The ex ante expected lifetime utility of a consumer using the fully optimal rule is given by

$$EV^* \equiv \int_W V(w)\lambda(dw) \tag{EV^*}$$

where the value function V is defined in section 2 above. Thus we can define the certainty equivalent consumption of the fully optimal rule as

$$CE^* \equiv u^{-1}(EV^* \cdot (1-\delta)) = \Big[1 + (1-\theta)(1-\delta)EV^*\Big]^{\frac{1}{1-\theta}} \tag{CE^*}$$

For any given w_0, the expected life-time utility of a consumer using the specific linear rule $\widehat{c}^{\,\gamma}(w)$ with parameters $\gamma = (\alpha, \beta)$, is given by

$$U^\gamma(w_0) = \sum_{t=0}^{\infty} E_0\Big[\delta^t u\Big(\min\{\alpha + \beta w_t, w_t\}\Big)\Big], \tag{25}$$

where w_t evolves according to the flow budget constraint. So, given the distribution λ of w_0, the ex-ante expected lifetime utility and the certainty equivalent consumption for this specific rule are given by

$$EV^\gamma \equiv \int_W U^\gamma(w)\lambda(dw) \text{ and} \tag{EV^γ}$$

$$CE^\gamma \equiv u^{-1}(EV^\gamma \cdot (1-\delta)) = \Big[1 + (1-\theta)(1-\delta)EV^\gamma\Big]^{\frac{1}{1-\theta}} \tag{CE^γ}$$

If the wealth process generated by $c^*(w)$ satisfies Assumption A, then there exists a unique ergodic invariant distribution over wealth π_*. In this case, let EV_*^*, CE_*^*, EV_*^γ and CE_*^γ be the values implied respectively by

(EV^*), (CE^*), (EV^γ) and (CE^γ) when $\lambda = \pi_*$. Thus our first index of welfare cost for a consumer using the rule $\widehat{c}^\gamma$ is

$$D_1^\gamma = \frac{CE_*^* - CE_*^\gamma}{CE_*^*} * 100. \tag{26}$$

If the linear consumption rule satisfies Assumption A, let π_γ be the unique invariant distribution determined by the rule and let EV_γ^*, CE_γ^*, EV_γ^γ and CE_γ^γ be the values when $\lambda = \pi_\gamma$. Thus our second index of welfare cost for a consumer using the rule $\widehat{c}^\gamma$ is

$$D_2^\gamma = \frac{CE_\gamma^* - CE_\gamma^\gamma}{CE_\gamma^*} * 100. \tag{27}$$

5. Numerical results

5.1. Baseline scenario

In order to study the behavior of our algorithm, we simulate the model for a set of values of the CRRA parameter θ and the discount factor δ, using the same income process starting from different initial conditions. We use the income process studied by Allen and Carroll (2001), which according to these authors "matches (very roughly) the empirical evidence on the amount of transitory variation in annual household income observed in the *Panel Study of Income Dynamics*." The income process is defined by $(y^1, y^2, y^3) = (0.7, 1.0, 1.3)$ with probabilities $(p^1, p^2, p^3) = (0.2, 0.6, 0.2)$ respectively. We take $\delta \in \{0.9, 0.95\}$, $\theta \in \{1.5, 2, 3.0, 3.5, 4\}$, which includes the values Allen and Carroll (2001) assume in their work ($\delta = 0.95$ and $\theta = 3$), $\eta \in \{0, 0.5\}$, $\xi \in \{0, 1\}$ and $\varepsilon \in \{0, 0.2\}$. We evaluate the linear rules in the $[0, 2] \times [0, 2]$ space (a superset of the admissible set) with a grid of 40,000 points, each separated at a distance of 0.01.

5.1.1. Consumption rules and welfare

As a first step, we calculate the optimal consumption function $c^*(w)$ for each parameter configuration, which we show in Figure 2(a). Given that $y^n = 1.3$, Assumption A is satisfied if $c^*(w) = 1.3$ for some w, which the figure shows is true for all our parameters, so there exists a unique distribution π_* for each parameter configuration, allowing us to calculate EV_*^* and CE_*^*. As can be seen in table E.1, EV_*^* is decreasing in both δ and θ, while CE_*^* is decreasing in θ and increasing in δ. In table E.2 we present the optimal linear rule for each set of parameters. As can be seen there, all optimal rules imply that Assumption A holds, so we can also compute EV_*^γ, CE_*^γ, EV_γ^γ, CE_γ^γ, EV_γ^* and CE_γ^*. Our calculations show that the behavior of all EV's and EC's is similar to EV_*^* and CE_*^* with respect to the underlying parameters, i.e. all EV's decrease in both θ and δ, while all EC's decrease in θ and increase in δ.

We find that most linear rules in $[0, 2] \times [0, 2]$ have a low welfare cost, according to both our indices. Figure D.6 presents the welfare loss surface for the case with parameters $\theta = 3.5$ and $\delta = 0.95$, which is qualitatively very similar to the other cases.[14] Furthermore, as figure D.7 shows, the set of consumption rules that achieve a percentage deviation less than or equal to 0.5% under both the optimal and the actual distribution of initial wealth is compact and of positive Lebesgue measure. The optimal linear rule belongs to this set, has a marginal propensity to consume in the range 0.2-0.4 and its costs are in the range 0.2%-0.3% for our set of parameters (see table E.2), which is very low and generalizes the case studied by Allen and Carroll (2001).

Table E.2 also shows the loss incurred when the consumer follows the "consume everything" $\gamma = (0, 1)$, and the expected loss of taking a rule at random from the whole set of parameters and from the admissible set. As can be seen there, if the consumer figures out that she should always have a rule in the admissible set, she can lower her expected loss by almost 40%. On the other hand, if she follows the simplest of all rules, namely the "consume everything" rule, which according to Cochrane's (1989) calculations also produces tiny losses when the agent's income process matches that of quarterly *aggregate income* in the United States, she can (in all but one case) lower her loss even below her expected loss in the admissible set. This is an interesting result, since it can explain why in certain environments people undersave and fail to learn a better rule than simply consuming everything. Experimentation in such environments will generally make them worse off, reinforcing the status quo. (As will be seen below however, the "consume everything" rule does not fare well when we consider different income processes.)

5.1.2. Adaptive behavior

In order to understand the behavior of the algorithm we run 20,000 simulations for each parameter configuration with random initial wealth and a random initial admissible linear rule. Figures 8(a) and 8(b) show the behavior of the distribution of welfare losses $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ across time for the simulations with $\varepsilon = 0$, $\xi = 1$ and $\eta = 0$, $\theta = 3.5$ and $\delta = 0.95$, where $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ are calculated for the rule with parameters $\gamma_t = (\alpha_t, \beta_t)$. The behavior of the distribution is summarized in this figure by the maximum and minimum loss (black lines), the mean (blue), the median (green) and the 25-th, 75-th, 90-th, 95-th, 99-th and 99.9-th percentiles (red lines). Notice that the values of all these measures decrease for the first 100-250 periods and then follow a flatter trajectory, which seems to indicate that within that time frame the algorithm achieves its stationary distribution, which is more concentrated around the mean.

In figures 8(c) and 8(d) we present the behavior of the consumption rule parameters and their distribution. The story here is also similar, showing convergence towards a long-run distribution of the parameters and a higher concentration of the probability of those parameters around their mean in roughly 250 periods or less. In order to better appreciate what is happening, in table E.3 we show for periods 0, 50, 100, 250 and 500,

[14] Lack of space prevents us from including the figures for all cases. For this reason here and below we present only the figures for the particular case with $\theta = 3.5$ and $\delta = 0.95$, since it is also the case with the slowest rates of convergence in the adaptive analysis. The interested reader can see all the figures in the supplemental material to the paper available on the web.

the probability of having a loss less than or equal to 0.5%, i.e. the probability that the consumption rule in that period belongs to the sets identified in figure D.7. As can be seen in this table, the probability rises from almost zero in period 0 for most parameter configurations to values above 30-40% in period 50, to above 40-60% in period 100, 80% in period 250 and above 90% in period 500. Clearly there is much variation in the speed of convergence to these sets, with the cases $(\delta, \theta) = (0.9, 1.5)$ and $(\delta, \theta) = (0.95, 1.5)$ being the ones with the highest rates of convergence and the cases $(\delta, \theta) = (0.9, 4)$ and $(\delta, \theta) = (0.95, 3.5)$ with the slowest convergence rates. This seems to be explained by the effect that both increases in δ and in θ have on the size of the set of parameters that achieve a loss of less than 0.5%. However, it is not clear how changes in the underlying parameters affect the speed of convergence, though this relation appears to be highly non-linear. In part, this seems to be generated by the fact that changes in θ and δ have similar effects on the curvature of c^*, where increases in δ or in θ make the function more concave. So, two functions, one with a high value of δ and a low θ and the other with a low δ and high θ might be very similar on the domain of interest (see e.g. the rational consumption functions for $(\delta, \theta) = (0.9, 4)$ and $(\delta, \theta) = (0.95, 2)$, or, $(\delta, \theta) = (0.9, 3)$ and $(\delta, \theta) = (0.95, 1.5)$), so that their respective rates of convergence are also similar.

For most macroeconomic purposes what matters is the central tendency of consumption demand; it would be too much to ask of a macro model that it account for individual outliers. Tables 6 and 7 show that the average and median welfare losses falls to within about 2% after only 10 years, and to within about 1% after 25 years. In this sense, it takes little time for the central tendency of aggregate consumption under our algorithm to come quite close (in expected utility) to what full optimization would predict, suggesting that an ACE macro model that portrayed a variety of consumers using this algorithm, but differing in random initial conditions, would quickly exhibit the same aggregate consumption as if they were all near-rational agents. At the same time, such a model would allow for considerable heterogeneity across consumers.

We repeated the simulations assuming different values of (ε, η, ξ). Given the overall similarity of the results we do not present them here in detail, but limit ourselves to highlighting the major differences with our previous simulations.[15] Under this new set of parameters, the behavior of the distributions of losses and parameters in terms of convergence to a stationary distribution within 250 periods was similar to before, though the dispersion around the mean increased and the speed of convergence decreased, especially for the "constant gain" cases with $\varepsilon = 0.2$. One striking effect of constant gain is that the probability of losses less than or equal to 0.5% fell, in some cases dramatically, and stayed stationary at that level without any tendency to converge towards 1 as was the case before. Still, this probability is bounded away from zero for all the simulations we realize. In table E.4 we compare the different trajectories for the case $\theta = 3.5$ and $\delta = 0.95$, which is the case with the slowest convergence rate in our baseline simulations. Table 9 shows that again the median welfare losses fall to within

[15] These results can be obtained from the authors by request.

2% after 10 years and to within about 1% after 25 years under all parameter configurations, although Table 8 indicates that average losses remain somewhat higher than this.

5.1.3. Social Learning

We also study the effects of allowing consumers to learn through social interaction. To do so, we set up consumers on circles of 25, 50, 100, 200 individuals and allow each consumer to interact with his left, right or both left and right neighbors. Each consumer can see her neighbor's γ^p_{t+1}. Let i denote the consumer and N_i the set of neighbors of consumer i, $\phi \in [0,1]$, then instead of ending with (Possible Final Step) we end with

$$\gamma_{i,t+1} = \eta\gamma_{i,t} + (1-\eta)\Big(\phi\gamma^p_{i,t+1} + \frac{1-\phi}{|N_i|}\sum_{j\in N_i}\gamma^p_{j,t+1}\Big)$$

This crude form of imitation, in which people (partially) mimick neighbors without trying to distinguish between successful and unsuccessful neighbors, actually speeds up convergence significantly, as can be seen in tables E.5, E.10 and E.11, demonstrating another aspect of the "wisdom of crowds."

5.1.4. Credit constraints

Additionally, we analyze how the credit constraint affects learning by relaxing the credit constraint and allowing consumers to consume at most $w_t + B$ every period, where we take $B \in \{0.1, 0.3\}$. These simulations show that the more the constraint was relaxed the slower consumers seem to learn. This result can be seen as a confirmation of the conjecture proposed by Satz and Ferejohn (1994), who argue that the more constrained consumers are, the more powerful their incentives to behave rationally. Overall we find that the behavior of this new set of simulations is similar to the ones described previously, i.e. the less consumers are constrained the slower they learn, but they *do learn* nonetheless.

5.2. Shocks to income

In all the previous simulations we held the income process fixed and changed the different parameters of the model. In this subsection we hold the parameters fixed at $\theta = 3.5$, $\delta = 0.95$, $B = 0$, $\eta = 0$, $\varepsilon = 0$, $\xi = 1$ and $\phi = 1$, but change the income process. We analyze 4 additional income processes and compare the behavior of the algorithm and the implied welfare costs under these new processes with the one implied by the original process. In order to have comparable results across simulations, we now allow wealth to be in the range $[0, 10]$

and consider a grid of wealth levels 0.0025 apart. The income processes we consider are:

$$Y^1 = (0.7, 1, 1.3) \qquad P^1 = (0.2, 0.6, 0.2)$$

$$Y^2 = (1.4, 2, 2.6) \qquad P^2 = (0.2, 0.6, 0.2)$$

$$Y^3 = (1, 1.4, 2, 4.1) \qquad P^3 = (0.1, 0.2, 0.6, 0.1)$$

$$Y^4 = (0.3, 0.7, 1, 2.1) \qquad P^4 = (0.05, 0.25, 0.6, 0.1)$$

$$Y^5 = (0.1, 0.7, 1, 1.3, 1.9) \qquad P^5 = (0.05, 0.15, 0.6, 0.15, 0.05)$$

For these income processes table E.12 shows some basic statistics, figure 5(b) the optimal consumption rules, and table E.13 presents the losses and optimal rules. As can be seen there, the differences in the income processes generate quite big and striking differences in the expected losses a consumer faces under different rules. For some income processes, e.g. Y^4 and Y^5, very few linear rules have low expected losses;[16] using a random rule, or the consume everything rule, fares badly in these cases because the lowest income is close to zero, generating huge losses for the consumer that ends up being wealth constrained. Still, the optimal linear rule for each process has a relatively low associated loss. Furthermore, the marginal propensity to consume stays in the same range across cases (ca. 0.17-0.25). These results imply that a simple adaptive algorithm, as the one proposed in this paper, can have big welfare effects for boundedly rational consumers, if the algorithm converges to low welfare losses in general.

Table E.14 shows the evolution of the probability of having a rule with a loss lower than 3% when using the algorithm. We increased the range to 3% given that for some of the income processes no rule has a loss less than 0.5% and almost no rule has a loss less than 1%. As can be seen in this table, the qualitative dynamics are similar to those we found in our original setup, so that the rate of convergence is similar to our baseline simulation for most processes. The exception is Y^5, for which convergence requires close to double the time of the other scenarios.

Until now, we have assumed that consumers do not use any other information of the income process except its lowest and highest possible levels. One way in which this might slow convergence to the optimal rule is that the consumer cannot distinguish a situation where the Euler error e_{t-1} is high because of a bad income draw from a situation in which e_{t-1} is high because of a bad rule. Furthermore, the consumer would not notice any changes in the stochastic process determining income, if these do not affect the range of values that can occur. This is especially important in the decreasing gain case ($\varepsilon = 0$), since this would imply that if she had been learning for while, it would be difficult for her to change her behavior. On the other hand, if $\varepsilon > 0$ the

[16] This can be seen in figures E.15-E.16 of the supplemental material

consumer's behavior never settles down, so that she will incorporate any changes in her environment into her behavior, *even if she does not realize there has been a change.*

In order to deal with this issue, we now assume that consumers update their parameters in the same way as before, but change the manner in which they calculate the continuation value q_{t-1} in (Step 2). Instead of using the realized continuation value $\delta u'(c_{t-1})$, a consumer uses the average realization of the continuation value under their most recent parameters, i.e.

$$q_{t-1} = \delta \sum_{i=1}^{n} p^i u'(c^i_{t-1}), \tag{28}$$

where c^i_{t-1} is the amount she would have consumed if she had received income level y^i in period $t-1$, i.e.

$$c^i_{t-1} = \min\left\{\alpha_{t-1} + \beta_{t-1}(w_{t-2} - c_{t-2} + y^i), w_{t-2} - c_{t-2} + y^i\right\}, \quad i = 1, \ldots, n. \tag{29}$$

This change requires the consumer to know the whole distribution of the income process, instead of just knowing y^1 and y^n. Table E.15 shows the evolution of the probability of being at a loss lower than 3%. As expected, the time to convergence is generally lowered, or equivalently, the probability of being close to the optimal rule is increased for any period for all income processes studied.

This modification to our algorithm innoculates it against the Lucas critique. That is, the consumer can now respond immediately to a change in regime by modifying her behavior accordingly. To see how this works, we analyzed the effects of a shock to income in which the income process changes to some other one in the set of processes we have studied, and then 25 years later returns unexpectedly to the original process, assuming that the consumer is informed immediately of each change in the income process. Given that we have assumed that consumers are sophisticated enough to know that the optimal rule lies in the set $\mathcal{A}_Y$ of admissible rules, which depends on the income process, we need an assumption specifying the way in which consumers react to the new information of a change in the income process. One possible assumption is that consumers dismiss all their accumulated experience up to that point and start the process from scratch. This amounts to almost the same exercise we have done in the previous simulations, except that the initial distribution of wealth will be closer to the stationary distribution of the original process. Given the fact that initial wealth conditions do not affect the long-run evolution of the system, we do not need to analyze this scenario again.

Instead, we assume that consumers keep their marginal propensity to consume β, while changing the intercept α in a way that is commensurate with the change in the scale of their income. More specifically, given that the old rule was admissible, the admissibility condition (17) implies that the intercept can be written as a convex combination of the two limiting values in the admissible set given β; i.e. that $\alpha = (1-\beta)\left(\lambda y^1 + \lambda y^n\right)$ for some $\lambda \in (0,1)$. So we assume that the new intercept will be given by the same convex combination of the new limiting values:[17]

$$\alpha' = (1-\beta)\left(\lambda y'^1 + \lambda y'^n\right). \tag{30}$$

Once the consumer starts getting new data on Euler equation errors she will start revising her rule as before.

Figures D.9-D.10 show the dynamics when consumers' income follows Y^1 in periods 1-25 and 51-100, while it follows Y^i, $i \neq 1$ during periods 26-50. As can be seen there, a permanent shock to income has various effects. First, the distribution of consumption, wealth, (α_t, β_t), and losses become more disperse during the shock. Second, welfare losses overshoot, and only slowly decrease towards the minimum loss. Interestingly, the effect on welfare is bigger on the average than on the median. This comes from an asymmetry in the distribution of losses, with very big losses from worse than average rules but small gains from better than average rules. Third, although the effect on consumption comes mostly from the change in the value of the intercept, which jumps discretely at the moment of the shock, it is the distribution of the marginal propensity to consume that changes the most during the shock period.

5.3. Some comments

In order to better appreciate how meaningful these results are, it is useful to compare them with the previous literature. In particular, Allen and Carroll (2001) assume that if an agent's loss is less than 5%, then the agent has learnt the optimal rule. They find that the probability of being at a loss of less than 5% is about 60% only after at least 1,000,000 periods. This contrasts with the speed of convergence for our algorithm, which in less than 250 periods converges with probability close to 1, to a welfare loss of less than 1/2%.[18] This is 3 orders of magnitude faster, even though our rule space is 100 times larger.[19] In a space the size of ours, their algorithm would require 100 times more periods to converge, i.e. about 5 orders of magnitude more! This is a major problem of all algorithms which use a comparison among different rules; the larger the number of rules, the more time is required to analyze their behavior, and also the more computational capacity and memory is required to keep track of the information for all these rules.

A second point that needs to be taken into account, especially when comparing with other stochastic learning schemes, like the ones employed by Lettau and Uhlig (1999) and Yildizoglu, Sénégas, Salle, and Zumpe (2012), is the probability of assigning a good rule to an agent independent of her learning mechanism. Given that in our algorithm agents start with a random rule, the probability that they start with a rule that generates a 1/2% or less loss is 0.2%, while the probability of getting the optimal rule is 1/40401 = 0.0025%. So, in our population of 20000 agents, the probability that no agent has a loss less than 1/2% or the optimal rule is 1.

[17] This scale measure was chosen for its simplicity; other measures such as the average would presumably produce comparable results.

[18] In appendix Appendix B we show that our welfare cutoff is comparable to the cutoff levels based on their welfare measure.

[19] Allen and Carroll (2001) have only 400 possible rules, while our discretized rule space has 40401 rules.

On the other hand, in Yildizoglu et al. (2012) the probability that an agent starts with the optimal rule is $1-(380/400)=5\%$ and the probability that none of them has it is $(380/400)^{20}=36\%$. Thus, in their setting learning is simplified by the high probability that the rule is already in the population. Furthermore, it also explains why their social learning works that well: if an agent was unlucky enough not to have the optimal rule at the start, the probability that she will copy it from one the other agents is still very high. So, although their algorithm converges to the optimum at a speed at most one or two orders of magnitude slower than ours, their setup favors fast convergence and learning, which is not the case in our setting. Social learning in our model increases the speed of learning only through the joint learning process, not by contagion of the good rule.

6. Conclusion

Models of bounded rationality and learning have recently flourished in economics, but 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 we have proposed an adaptive algorithm based on Euler equations and have studied its behavior through time using simulations. The algorithm economizes on information and computation, and we have shown that it allows consumers to have low welfare losses with high probability in a short time. Furthermore, we have generalized our adaptive procedure in order to allow for social learning, relaxation of credit constraints, and changes in the environment. An additional feature of our algorithm is that it allows for considerable individual heterogeneity, while at the same time causing the average or median consumer to behave (nearly) optimally almost all the time. Thus, rational behavior is an emergent property in our model.

All of this is accomplished with people making adaptations only once a year, and with an income process calibrated to annual data. It would be interesting to analyze the model under parameterizations that allowed for updating of the rule to occur on a monthly or quarterly basis, under calibrations that matched the monthly or quarterly time series of individual income. If the time for convergence is still of the same order, this would strengthen the case for using our algorithm for modeling consumption.

If further exploration of the model proves successful, it would be a step in the direction of an ACE approach to macroeconomics, along the lines advocated by Weintraub (1979), Leijonhufvud (1993) and others.[20] This bottom-up approach would endow agents not with decision rules that are always perfectly tailored to their specific environment, but rather with simple all-purpose rules that allow the agent to adapt in a plausibly opportunistic yet imperfect fashion to any given environment. This approach would allow us to ask how an

[20] In particular, Weintraub (1979) argues that "a successful reconciliation of micro and macro might entail a return to Marshallian price theory, or a well worked out statement of individual behavior in a non-optimizing framework" (Weintraub, 1979, p.157).

economic system works to coordinate, for better or worse, the independent decisions of heterogeneous interacting agents; a question that the more conventional top-down approach evades by restricting attention to equilibrium states.

Appendix A. Results under constant M

In the text we argued that using a constant matrix M in the algorithm does not qualitatively change the results.[21] In this section we present results based on the same type of simulations as the main text, but change the algorithm as follows: Instead of step 1, now we fix a 2×2 symmetric matrix M, which has two distinct positive eigenvalues.[22] Agents use now the following modified step 2:

$$\begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix} + \kappa_t M^{-1} \left[q_{t-1} - u' \left(\alpha_{t-2} + \beta_{t-2} w_{t-2} \right) \right] u'' \left(\alpha_{t-2} + \beta_{t-2} w_{t-2} \right) \begin{pmatrix} 1 \\ w_{t-2} \end{pmatrix} \tag{Step 2'}$$

where κ_t is either (a) a decreasing sequence of positive real numbers, such that $\sum \kappa_t = \infty$ and $\sum \kappa_t^2 < \infty$, or (b) equal to some constant κ.

We will focus on simulations where $\theta = 3.5$ and $\delta = 0.95$, which generated the slowest rates of convergence in our previous simulations.[23] Although our choice of matrix M and the sequence $\{\kappa_t\}$ is arbitrary, results shown by Özak (2013) ensure that they hold more generally.

As a first example, let

$$M = 425 \cdot \begin{pmatrix} 1 & 0.975 \\ 0.975 & 1 \end{pmatrix}, \qquad \kappa_t = 0.35.$$

Figure A.1 shows the results of this simulation. Comparison with figure D.8 shows that the results are basically unchanged, even though our agents now have much lower levels of rationality. In particular, notice that again the probability of having a loss lower than 1/2% is close to 90% within 500 periods. Furthermore, the probability of having a loss of welfare lower than 1/2% is higher in this case than when agents had the higher level of rationality necessary to use the original matrix M_t.

Although the previous example performed better than the original analysis, it does not imply that this will always be the case. Özak (2013) shows that the choice of matrix M and of sequence $\{\kappa_t\}$ affect the rate of convergence to the optimal linear rule. To see this, we additionally consider the following cases:

$$M = m * \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}, \qquad m = \begin{cases} 25 \\ 125 \\ 250 \end{cases}, \qquad \kappa_t = \begin{cases} \frac{\sqrt{t}}{t+1} \\ 0.35 \\ 1 \end{cases}. \tag{A.1}$$

[21] See Özak (2013) for the analysis of this and other results.
[22] This ensures that the conditions for convergence shown in Özak (2013) are satisfied.
[23] Similar results can be obtained with other parametrizations.

Figures A.2 and A.3 present the results of these exercises. As can be seen there, the rates of converges are slower than in our previous example and vary across specifications. If $\kappa_t \to 0$ (decreasing gain) the algorithm still converges pretty quickly to the optimal rule. In particular, the probability of being at a loss of less than 1% or 1/2% rises monotonically towards 1. Özak (2013) shows that this is to be expected for a general class of algorithms, of which ours is a particular case. On the other hand, if $\kappa_t = \kappa$ for all $t \geq 0$, then the rate of convergence depends on both the size of κ and the matrix M. This behavior is to be expected and is common in the literature on stochastic approximation (see e.g. Kushner and Yin, 2003), where an optimal κ is usually chosen by the user.[24] These simulations show that using a fixed matrix M might increase the time for convergence in one or two orders of magnitude. Still, the rate of convergence is 3 orders of magnitude smaller than in other algorithms proposed in the literature. Furthermore, if social learning speeds up the rate of convergence as before, the negative effect of having lower levels of rationality can be overcome by socializing the process of learning.

[24] Here we have chosen $\{\kappa_t\}$ constant across periods or convergent to 0. It is often also the case that $\{\kappa_t\}$ is chosen so that it converges to a non-zero constant.

Figure A.1: Distribution of (a) $D_1^{\gamma_t}$, (b) $D_2^{\gamma_t}$, (c) α_t, (d) β_t, and percentage of simulations with (e) $D_1^{\gamma_t}$ and (f) $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 3.5$, $\delta = 0.95$.

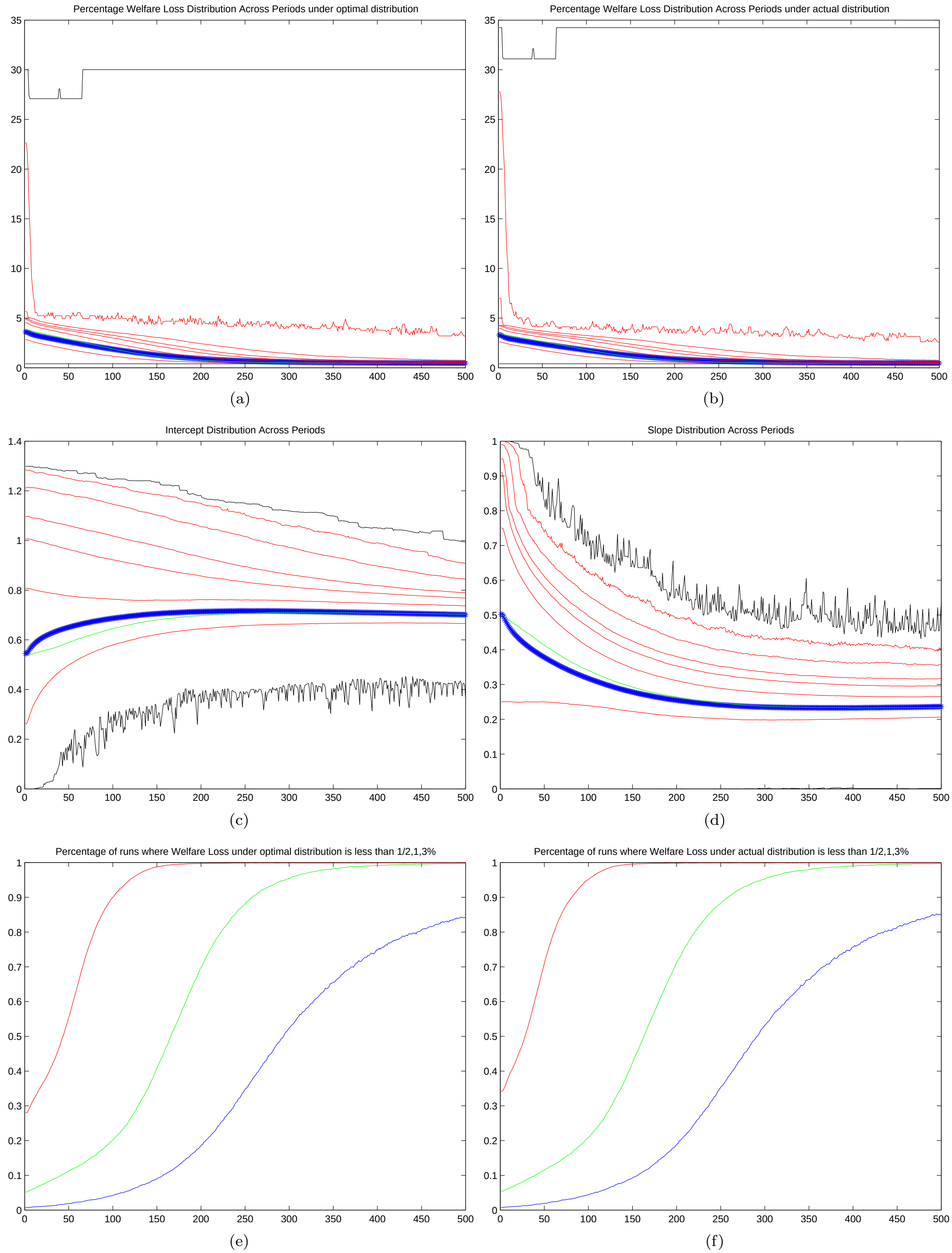


Figure A.2: Simulations with a fixed matrix M.

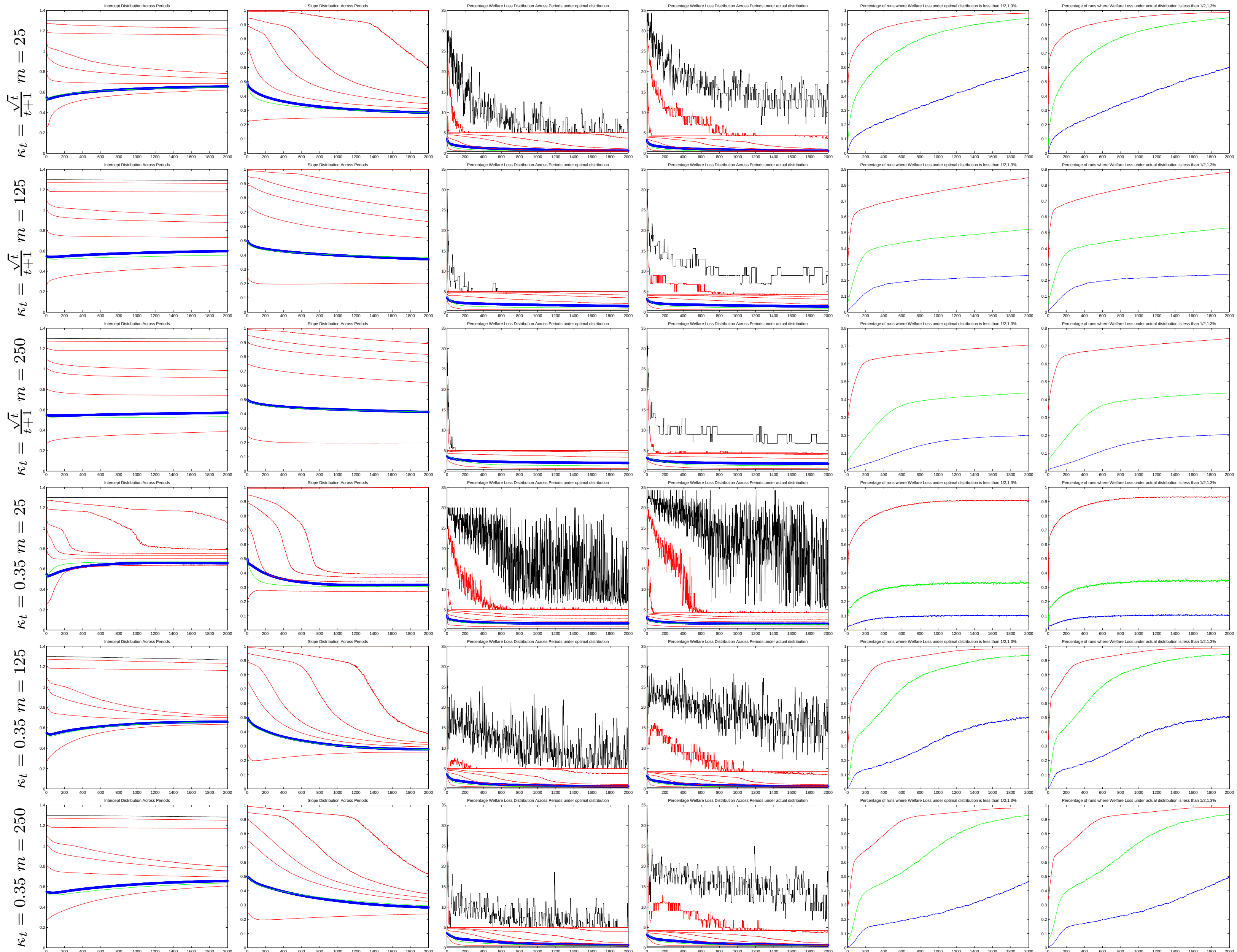
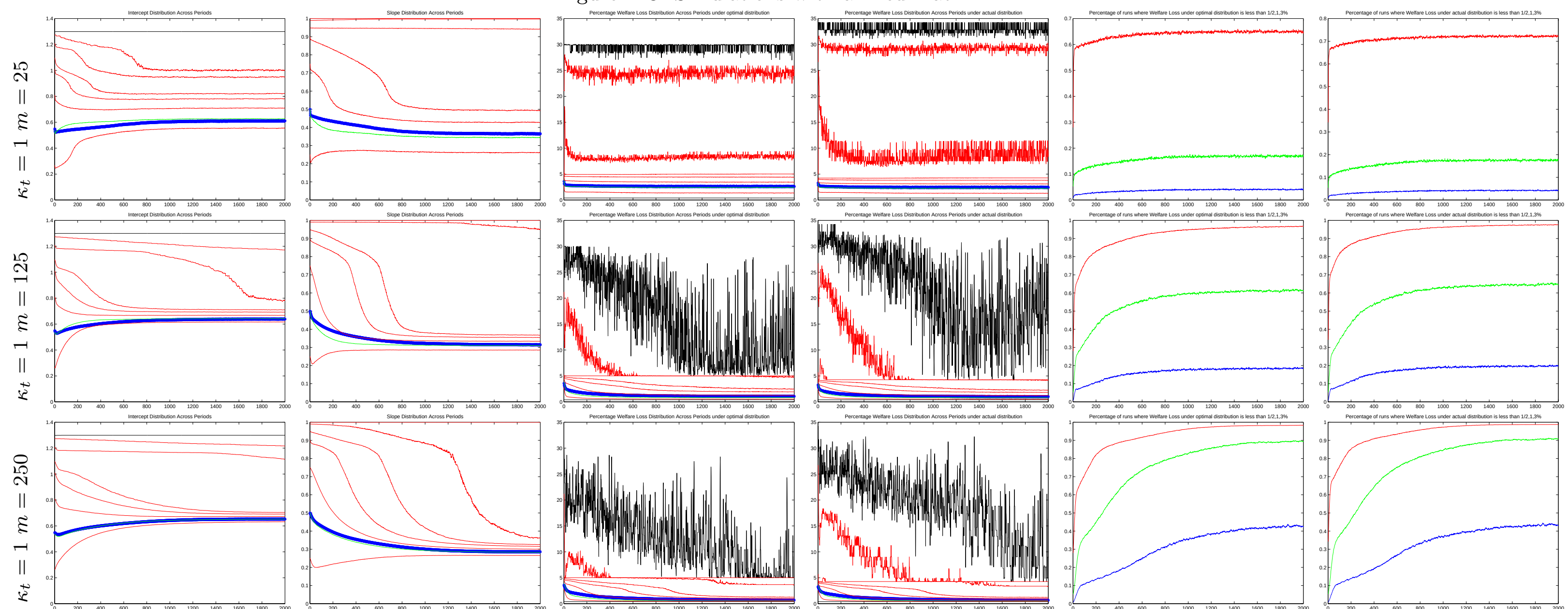


Figure A.3: Simulations with a fixed matrix M.



Appendix B. Interpreting our results

In order to better appreciate our results, it is useful to compare our loss measures with the ones employed by Allen and Carroll (2001). Allen and Carroll (2001) measure the loss of a rule γ as

$$\epsilon^{\gamma} = 1 - \frac{E_W[V^{-1}(U^{\gamma}(w))]}{\bar{w}} \tag{B.1}$$

where E_W is the expected value under the stationary distribution of wealth under the optimal rule, and $\bar{w}$ is the expected wealth level under this distribution.

Let $x = 1 - \frac{CE^{\gamma}}{CE^{*}}$ and $y = \epsilon^{\gamma}$ measure a loss of x and of y according to both measures for a certain rule γ. Then by definition

$$\begin{aligned}
(1-y)\bar{w} =& E_W\left[V^{-1}\left(U^{\gamma}(w)\right)\right] \geq V^{-1}\left(E_W\left[U^{\gamma}(w)\right]\right) \Longrightarrow \\
V((1-y)\bar{w}) \geq& E_W\left[U^{\gamma}(w)\right] \Longrightarrow \\
u^{-1}\left((1-\delta)V((1-y)\bar{w})\right) \geq& u^{-1}\left((1-\delta)E_W\left[U^{\gamma}(w)\right]\right) = (1-x)u^{-1}\left((1-\delta)E_W\left[V(w)\right]\right) \\
&\approx (1-x)u^{-1}\left((1-\delta)E_W\left[V(\bar{w}) + V'(\bar{w})(w-\bar{w})\right]\right) \\
&= (1-x)u^{-1}\left((1-\delta)V(\bar{w})\right),
\end{aligned}$$

which implies

$$\begin{aligned}
x \geq& 1 - \left(\frac{(1-\theta)(1-\delta)V((1-y)\bar{w}) + 1}{(1-\theta)(1-\delta)V(\bar{w}) + 1}\right)^{\frac{1}{1-\theta}} \\
\approx& (1-\theta)\frac{(1-\theta)(1-\delta)V'(\bar{w})\bar{w}}{(1-\theta)(1-\delta)V(\bar{w}) + 1}y \\
\equiv& By
\end{aligned}$$

so that

$$\frac{x}{B} \geq y. \tag{B.2}$$

Thus, as x, the loss in welfare as measured by us, decreases, so does the measure of loss y, the AC measure. For example, for $\delta = 0.95$ and $\theta = 3$, which are the parameters used by Allen and Carroll (2001), $B = 0.095$, so that a 1/2% loss in our measure, generates at most a 5.4% loss in the AC sense. Similarly, a 5% loss according to Allen and Carroll (2001) generates at least a 1/2% loss according to our measure.

Appendix C. Numerical Motivation

In order to understand where such an algorithm might come from, consider the case of a consumer who is an econometrician but cannot process all the information required to find the optimal solution. Suppose instead that she takes as a working hypothesis that $E_y q_{t+1}$ depends only on w_t. More specifically, she posits the existence of a pair (α, β) that solves (18) for all possible values of w_t. Under this working hypothesis, (18) is a bivariate nonlinear regression and her choice of a consumption rule is equivalent to estimating the parameters of that regression.

If she had access to a time series on wealth $(w_0, ..., w_{t-2})$ and the realized continuation values $(q_1, ..., q_{t-1})$, and enough computational resources, then she might use the standard method of nonlinear least squares:

$$(\alpha_t, \beta_t) = \arg\min \sum_{\tau=1}^{\tau=t-1} \left(q_\tau - u'\left(\alpha_t + \beta_t w_{\tau-1}\right)\right)^2, \tag{C.1}$$

solving the minimization problem using an iterative Newton or quasi-Newton method (Davidson and MacKinnon (2004), ch.6; Judd (1998), ch.5). Starting from an initial guess equal to last period's parameters:

$$\begin{pmatrix} \alpha_t^0 \\ \beta_t^0 \end{pmatrix} = \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix}, \tag{C.2}$$

the guess in the k^{th} iteration of the quasi-Newton method would be

$$\begin{aligned} \begin{pmatrix} \alpha_t^k \\ \beta_t^k \end{pmatrix} &= \begin{pmatrix} \alpha_t^{k-1} \\ \beta_t^{k-1} \end{pmatrix} \\ &\quad + \left(M_{t-1}^{k-1}\right)^{-1} \sum_{\tau=1}^{\tau=t-1} \left[q_\tau - u'\left(\alpha_t^{k-1} + \beta_t^{k-1} w_{\tau-1}\right)\right] u''\left(\alpha_t^{k-1} + \beta_t^{k-1} w_{\tau-1}\right) \begin{pmatrix} 1 \\ w_{\tau-1} \end{pmatrix} \end{aligned} \tag{C.3}$$

where

$$M_{t-1}^{k-1} = \sum_{\tau=1}^{\tau=t-1} u''\left(\alpha_t^{k-1} + \beta_t^{k-1} w_{\tau-1}\right)^2 \begin{pmatrix} 1 & w_{\tau-1} \\ w_{\tau-1} & w_{\tau-1}^2 \end{pmatrix}. \tag{C.4}$$

Appendix C.1. The baseline algorithm

What we propose is similar to the simplified variant of (C.3) that does not require storing an ever-expanding time series. First, instead of $M_{\tau-1}^{k-1}$, the consumer uses the historical "moment" matrix:

$$M_{t-1} = \sum_{\tau=1}^{\tau=t-1} u''\left(\alpha_{\tau-1} + \beta_{\tau-1} w_{\tau-1}\right)^2 \begin{pmatrix} 1 & w_{\tau-1} \\ w_{\tau-1} & w_{\tau-1}^2 \end{pmatrix} \tag{C.5}$$

This avoids the need to recompute (C.4) at each iteration with updated coefficients $\left(\alpha_t^{k-1}, \beta_t^{k-1}\right)$. Thus the first step in our proposed algorithm for choosing (α_t, β_t) is to update M_{t-1} recursively:

$$M_{t-1} = M_{t-2} + u''\left(\alpha_{t-2} + \beta_{t-2} w_{t-2}\right)^2 \begin{pmatrix} 1 & w_{t-2} \\ w_{t-2} & w_{t-2}^2 \end{pmatrix} \qquad \text{(Step 1)}$$

which requires the consumer to remember and update only 6 numbers each period, namely α_{t-2}, β_{t-2}, w_{t-2} and the 3 distinct elements of the symmetric matrix M_{t-2}.

The second step is to take a single iteration of the quasi-Newton method, as described above, using the historical moment matrix updated in the first step, and using only a modified version[25] of the last (most recent) term in the sum on the RHS of (C.3)

$$\begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix} + M_{t-1}^{-1}\left[q_{t-1} - u'\left(\alpha_{t-2} + \beta_{t-2} w_{t-2}\right)\right] u''\left(\alpha_{t-2} + \beta_{t-2} w_{t-2}\right) \begin{pmatrix} 1 \\ w_{t-2} \end{pmatrix} \qquad \text{(Step 2)}$$

which requires that she remember three more numbers, namely α_{t-1}, β_{t-1} and c_{t-1}, bringing the total number of variables to be stored and updated equal to 9. (She needs to remember c_{t-1} in order to compute the continuation value $q_{t-1} = \delta u'\left(c_{t-1}\right)$.)

Usually the algorithm requires nothing more than these two steps. Complications arise, however if the moment matrix M_{t-1} is singular, or even close[26] to singular, or if (Step 2) results in an inadmissible set of coefficients. In either of those cases we assume the consumer ignores whatever historical information is in the moment matrix and chooses the parameter pair closest to γ_{t-1} that would have made her most recent error $q_{t-1} - u'\left(\alpha_t + \beta_t w_{t-2}\right)$ equal to zero. This additional step requires no more information than is needed for the first 2 steps. If it still results in an inadmissible rule she simply leaves the coefficients equal to γ_{t-1}.

Appendix C.1.1. Commentary

Our proposed algorithm thus involves having each consumer adjust the parameters of her rule in response to the most recently observed Euler equation error:

$$e_{t-2} = q_{t-1} - u'\left(\alpha_{t-2} + \beta_{t-2} w_{t-2}\right) \qquad \text{(C.6)}$$

[25]The arguments of u' and u'' in (Step 2) are the notional consumption of two periods ago, whereas in the first step of the Newton iteration they would instead be $\alpha_{t-1} + \beta_{t-1} w_{t-2}$.

[26]In our simulations we deemed M_{t-1} to be close to singular whenever the condition number $r2\left(M_t\right)$ was less than the conventional limit 10^{10}. See Judd (1998).

Clearly if $e_{t-2} = 0$ then if she had known that the marginal continuation value was going to be q_{t-1} she would have been happy with her choice of coefficients at $t-2$, because it led her to choose a notional consumption whose marginal utility $u'(\alpha_{t-2} + \beta_{t-2} w_{t-2})$ was just equal to its marginal cost q_{t-1}. In this sense, even with hindsight she did not make a mistake. Accordingly (Step 2), which can be rewritten as

$$\Delta \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} = e_{t-2} u'' (\alpha_{t-2} + \beta_{t-2} w_{t-2}) M_{t-1}^{-1} \begin{pmatrix} 1 \\ w_{t-2} \end{pmatrix}, \tag{C.7}$$

requires the consumer to leave her rule unchanged in this case.[27]

On the other hand, if $e_{t-2} \neq 0$ then she made an ex post error; if $e_{t-2} > 0$ then the marginal cost of her consumption at $t-2$ exceeded its marginal utility, indicating that she consumed too much, whereas if $e_{t-2} < 0$ then she consumed too little. In either case she will adjust her rule in an attempt to learn from this error. The role of the moment matrix M_{t-1} is to help her distribute that adjustment across the two parameters of her rule. Indeed, following a positive error that indicates consumption should have been lower, the algorithm could well lead to an increase in one of the parameters, much as when an econometrician running an OLS regression is led to raise a positive regression coefficient following the addition of a data point below the previous regression line, depending on the nature of the OLS moment matrix.

Of course even a consumer who has chosen c_{t-2} according to the optimal consumption rule $c^*(\cdot)$ will end up making ex post errors when a low draw of y_{t-1} induces her to choose a low c_{t-1} thus raising the continuation value q_{t-1} above what was rationally expected at $t-2$. This rational consumer would have no reason to respond to e_{t-2}. But the rational consumer would also have to know the exact form of the income process, and the exact form of the optimal response, whereas we are trying to specify an algorithm that requires minimal information. To use our algorithm, all the consumer needs to know about the income process are the minimum and maximum values y^1 and y^n, which are needed to determine if a rule is admissible. As we shall see, even with no more awareness than this she will on average come close to the realized utility level of the rational consumer.[28] In section 5.2 below we will consider modifying the process to allow for more awareness of the income process.

Appendix C.2. Some variants

In deriving (Step 1) above we could have started with the original Newton method instead of quasi-Newton. The difference is that in the original Newton method the moment matrix M_{t-1}^{k-1} in (C.3) is replaced by the full Hessian of the term multiplying the inverse moment matrix. (See Davidson and MacKinnon (2004), ch.6 and

[27] Note that, because the error made when choosing α_{t-1} cannot be observed until α_t has been chosen, the consumer is always adjusting α_{t-1} on the basis of errors made when using α_{t-2}.

[28] Likewise, in the canonical stochastic approximation model of Robbins-Monro, a root ϕ^* to the equation $Ef(\phi, x) = 0$ is found iteratively by adjusting the most recent guess ϕ_t in response to the most recent observation $f(\phi_t, x_t)$, where x_t is a realization of x, even though someone who knew the correct value of ϕ^* would know enough not to make any adjustments.

Judd (1998), ch.5 for a comparison of these two methods).

Because the moment matrix grows with each new observation, it will generally be unbounded (and positive definite). Thus Step 2 is like a "decreasing gain" algorithm in stochastic approximation, in which the step sizes will almost certainly fall to zero over time, and thus the consumer will become infinitely slow to adapt to an unobserved change in the distribution of income unless some other procedure is invoked. We discuss in section 5.2 below what will happen if she is aware of the change in regime. But to take make the procedure more responsive to unperceived changes we might follow the literature on macroeconomic learning (Evans and Honkapohja (2001)) by using a "constant gain" method.

These two changes can be allowed for by using the following generalized version of Step 1:

$$M_{t-1} = (1-\varepsilon)\,M_{t-2} + \left[\xi e_{t-2} u'''\left(\alpha_{t-2}+\beta_{t-2}w_{t-2}\right) + u''\left(\alpha_{t-2}+\beta_{t-2}w_{t-2}\right)^2\right]\begin{pmatrix} 1 & w_{t-2} \\ w_{t-2} & w_{t-2}^2 \end{pmatrix}$$

where $\xi = 0$ indicates quasi-Newton, $\xi = 1$ indicates original Newton, and $\varepsilon \in [0,1)$ can be interpreted as the asymptotic gain parameter. As long as $\varepsilon > 0$ the moment matrix will be bounded and the speed of adjustment will not fall to zero.

Finally, a commonly used prudential measure in numerical approximation is to reduce the step size so as to avoid unstable overshooting. Thus we could allow a final step, which is to set

$$\gamma_t = (1-\eta)\,\gamma_{t-1} + \eta\gamma_t^p \qquad \text{(Possible Final Step)}$$

where γ_t^p is the set of coefficients arrived at by all the previous steps and $\eta \in (0,1]$ is the "shrinkage" factor.

References

Akerlof, G. A., Yellen, J. L., 1985a. Can small deviations from rationality make significant differences to economic equilibria? The American Economic Review 75 (4), 708–720. 1

Akerlof, G. A., Yellen, J. L., 1985b. A near-rational model of the business cycle, with wage and price intertia. The Quarterly Journal of Economics 100 (5), 823–38. 1

Albin, P., Foley, D. K., 1992. Decentralized, dispersed exchange without an auctioneer. Journal of Economic Behavior and Organization 18, 27–51. 3

Allen, T. W., Carroll, C. D., 2001. Individual learning about consumption. Macroeconomic Dynamics 5 (02), 255–271. 1, 2, 11, 12, 17, 25

Arifovic, J., 2000. Evolutionary algorithms in macroeconomic models. Macroeconomic Dynamics 4, 373–414. 3

Arthur, W. B., 1994. Inductive reasoning and bounded rationality. American Economic Review 84 (2), 406–411. 1

Ashraf, Q., Gershman, B., Howitt, P., 2011. Banks, market organization and macroeconomic performance: An agent-based computational analysis. Unpublished, Brown University. 3

Ashraf, Q., Gershman, B., Howitt, P., 2013. How inflation affects macroeconomic performance: An agent-based computational investigation. Unpublished, Brown University. 3

Axtell, R. L., 2006. Multi-agent systems macro: A prospectus. In: Colander, D. (Ed.), Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model. Cambridge University Press, New York, pp. 203–220. 3

Basu, N., Pryor, R., Quint, T., 1998. Aspen: A microsimulation model of the economy. Computational Economics 12, 223–241 3

Bennett, R. L., Bergmann, B. R., 1986. A Microsimulated Transactions Model of the United States Economy. The Johns Hopkins University Press, Baltimore. 3

Brown, A. L., Chua, Z. E., Camerer, C. F., 2009. Learning and visceral temptation in dynamic saving experiments. Quarterly Journal of Economics 124 (1), 197–231. 3

Bruce, J., 2003. Learning algorithms in a decentralized general equilibrium model. Ph.D. thesis, McMaster University. 3

Bruun, C., 1999. Agent-based keynesian economics: Simulating a monetary production system bottom-up. Unpublished, Aalborg University. 3

Bruun, C., 2003. The economy as an agent-based whole - simulating schumpeterian dynamics. Industry and Innovation 10 (4), 475–491. 3

Buchanan, M., 2009. Meltdown modeling: Could agent-based computer models prevent another financial crisis? Nature 460 (August 6), 680–682. 3

Canzian, G., 2009. Three essays in agent-based macroeconomics. Ph.D. thesis, University of Trento. 3

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

Carroll, C. D., 2004. Theoretical foundations of buffer stock saving, national Bureau of Economic Research Working Paper No. 10867. 6

Carroll, C. D., Kimball, M. S., 1996. On the concavity of the consumption function. Econometrica 64 (4), 981–992. 6

Chan, C. K., Steiglitz, K., 2008. An agent-based model of a minimal economy. Unpublished, Department of Computer Science, Princeton University. 3

Chen, S.-N., 2003. Agent-based computational macroeconomics: A survey. In: Terano, T., Deguchi, H., Takadama, K. (Eds.), Meeting the Challenge of Social Problems via Agent-Based Simulation. Springer, New York, pp. 141–170. 3

Clark, A., 1997. Being there: putting brain, body, and world together again. MIT Press, Cambridge, Mass. 1, 4

Cochrane, J. H., 1989. The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives. The American Economic Review 79 (3), 319–337. 1, 12

Colander, D., Howitt, P., Kirman, A., Leijonhufvud, A., Mehrling, P., May 2008. Beyond dsge models: Towards an empirically-based macroeconomics. American Economic Review Papers and Proceedings 98 (2), 236–240. 3

Davidson, R., MacKinnon, J. G., 2004. Econometric Theory and Methods. Oxford University Press, New York. 26, 28

Delli Gatti, D., Desiderio, S., Gaffeo, E., Cirillo, P., Gallegati, M., 2010. Macroeconomics from the bottom up. 3

Dosi, G., Fagiolo, G., Roventini, A., 2005. An evolutionalry model of endogenous business cycles. Computational Economics 27, 3–24. 3

Eliasson, G., 1977. Competition and market processes in a simulation model of the swedish economy. American Economic Review Papers and Proceedings 67 (1), 277–281. 3

Evans, G. W., Honkapohja, S., 2001. Learning and expectations in macroeconomics. Princeton University Press, Princeton.
URL `http://www.loc.gov/catdir/bios/prin051/00048320.html` 29

Evans, G. W., McGough, B., 2009. Learning to optimize, unpublished. 2

Farmer, D., Foley, D. K., 2009. The economy needs agent-based modelling. Nature 460 (August 6), 685–686. 3

Feigenbaum, J., Caliendo, F. N., Gahramanov, E., 2011. Optimal irrational behavior. Journal of Economic Behavior & Organization 77 (3), 285–303. 3

Holland, J. H., 1992. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, 1st Edition. MIT Press, Cambridge, Mass. 2

Howitt, P., 1992. Interest rate control and nonconvergence to rational expectations. Journal of Political Economy 100, 776–800 2

Howitt, P., 2006. The microfoundations of the keynesian multiplier process. Journal of Economic Interaction and Coordination 1 (1), 33–44 3

Howitt, P., 2008. Macroeconomics with intelligent autonomous agents. In: Farmer, R. (Ed.), Macroeconomics in the Small and the Large: Essays on Microfoundations, Macroeconomic Applications and Economic History in Honor of Axel Leijonhufvud. Edward Elgar, Cheltenham, pp. 157–177. 3

Howitt, P., Clower, R., 2000. The emergence of economic organization. Journal of Economic Behavior and Organization 41, 55–84 3

Judd, K. L., 1998. Numerical methods in economics. MIT Press, Cambridge, Mass. 8, 26, 27, 29

Kushner, H. J., Yin, G., 2003. Stochastic approximation and recursive algorithms and applications. Vol. 35. Springer, New York.
URL `http://www.loc.gov/catdir/enhancements/fy0813/2003045459-d.html` 2, 21

LeBaron, B., Tesfatsion, L., 2008. Modeling macroeconomics as open-ended dynamic systems of interacting agents. American Economic Review Papers and Proceedings 98, 246–250 3

Leijonhufvud, A., 1993. Towards a not-too-rational macroeconomics. Southern Economic Journal 60 (1), 1–13. 2, 3, 18

Leijonhufvud, A., 2006. Agent-based macro. In: Tesfatsion, L., Judd, K. L. (Eds.), Handbook of Computational Economics. Vol. 2. North-Holland, Amsterdam, pp. 1625–1637. 3

Lettau, M., Uhlig, H., 1999. Rules of thumb versus dynamic programming. American Economic Review 89 (1), 148–174. 1, 2, 17

Ljung, L., 1977. Analysis of recursive stochastic algorithms. IEEE transactions on automatic control 22 (4), 551–575. 2

Orcutt, G., Caldwell, S., Wertheimer II, R., 1976. Policy Exploration through Microanalytic Simulation. The Urban Institute, Washington. 3

Özak, Ö., 2013. Optimal consumption under bounded rationality and liquidity constraints. Working Paper, Southern Methodist University, http://ssrn.com/abstract=2338236. 2, 5, 6, 10, 20, 21

Padula, M., 2010. An approximate consumption function. Journal of Economic Dynamics and Control 34 (3), 404–416. 2

Raberto, M., Teglio, A., Cincotti, S., 2010. Credit money and macroeconomic instability in the agent-based model and simulator eurace. 3

Robbins, H., Monro, S., 1951. A stochastic approximation method. The Annals of Mathematical Statistics, 400–407. 2

Satz, D., Ferejohn, J., 1994. Rational choice and social theory. The Journal of philosophy 91 (2), 71–87. 14

Simon, H. A., 1990. Invariants of human behavior. Annual Reviews in Psychology 41 (1), 1–20. 1

Tesfatsion, L., 2006a. Agent based computational economics: A constructive approach to economic theory. In: Tesfatsion, L., Judd, K. L. (Eds.), Handbook of Computational Economics. Vol. 2. North-Holland, Amsterdam, pp. 831–880. 3

Tesfatsion, L., 2006b. Agent-based computational modeling and macroeconomics. In: Colander, D. (Ed.), Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model. Cambridge University Press, New York, pp. 175–202. 3

Weintraub, E. R., 1979. Microfoundations: the compatibility of microeconomics and macroeconomics. Cambridge surveys of economic literature. Cambridge University Press, Cambridge.
URL `http://www.loc.gov/catdir/description/cam032/78016551.html` 18

Woodford, M., 2003. Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press, Princeton, NJ. 2

Yildizoglu, M., Sénégas, M., Salle, I., Zumpe, M., 2012. Learning the optimal buffer-stock consumption rule of carroll. Available at SSRN. 17, 18

Appendix D. Tables and Figures

Figure D.4: Admissible set of parameters.

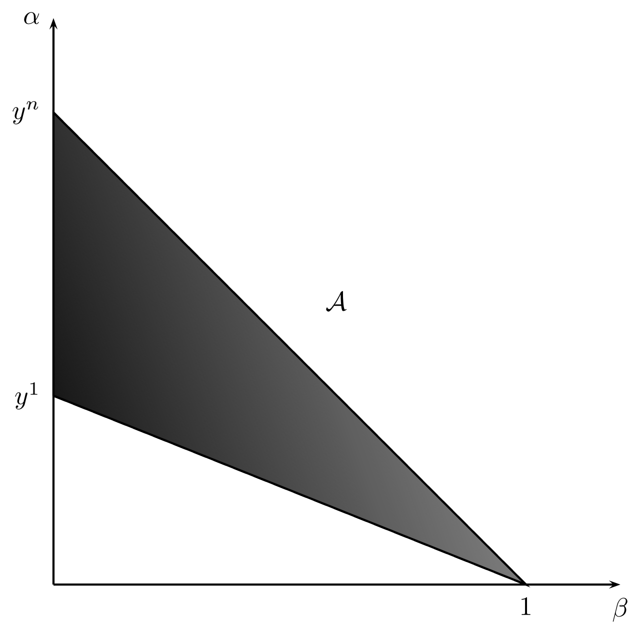
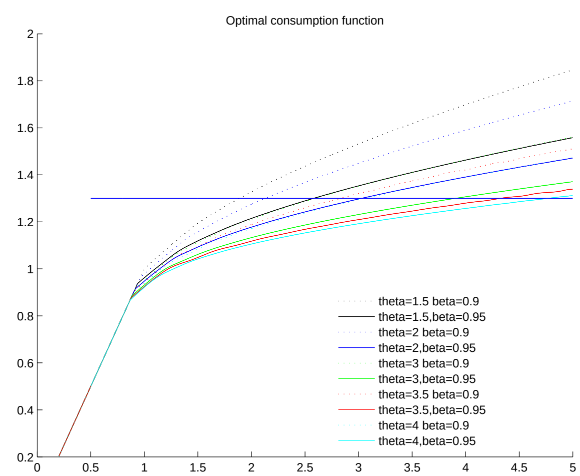
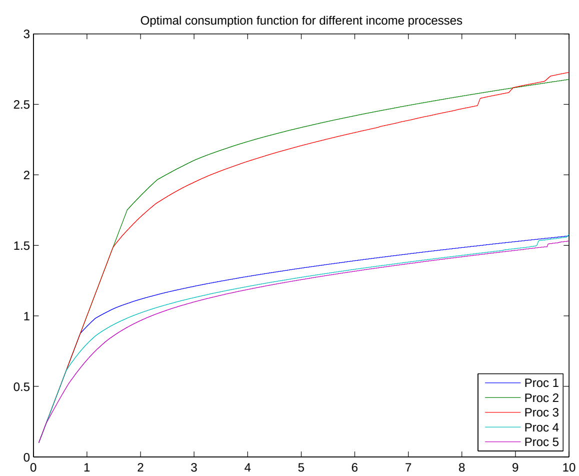


Figure D.5: Optimal consumption functions.

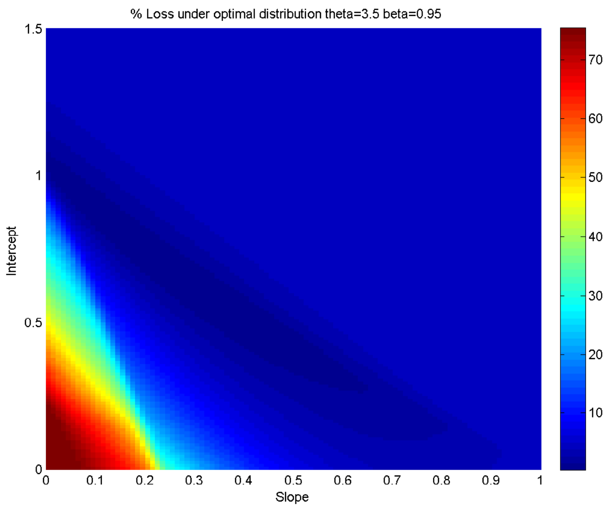


(a) For fixed income process Y^1, $B = 0$, and different values of θ and δ.

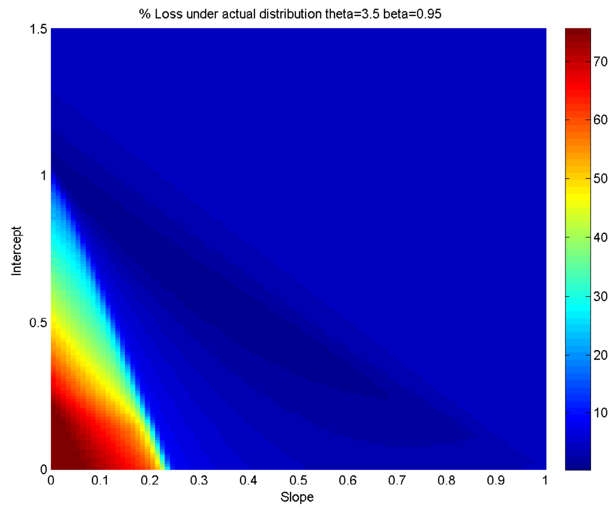


(b) For different income processes for a fixed parametrization with $\theta = 3.5$, $\delta = 0.95$, and $B = 0$.

Figure D.6: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept (vertical axis) and slope (horizontal axis), under (a) π^* and (b) π^γ for $\theta = 3.5$ and $\delta = 0.95$.

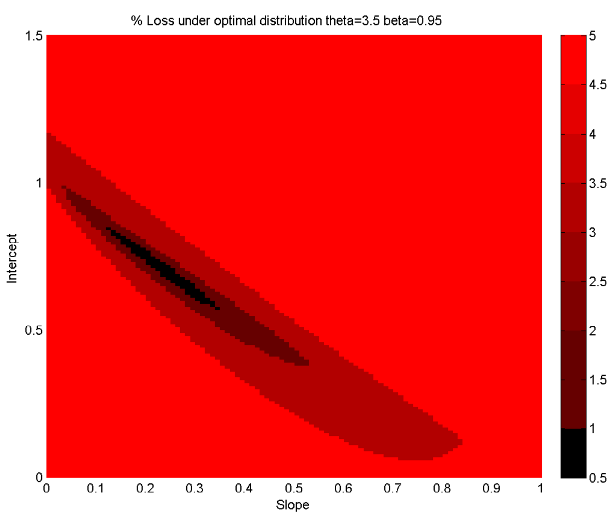


(a)

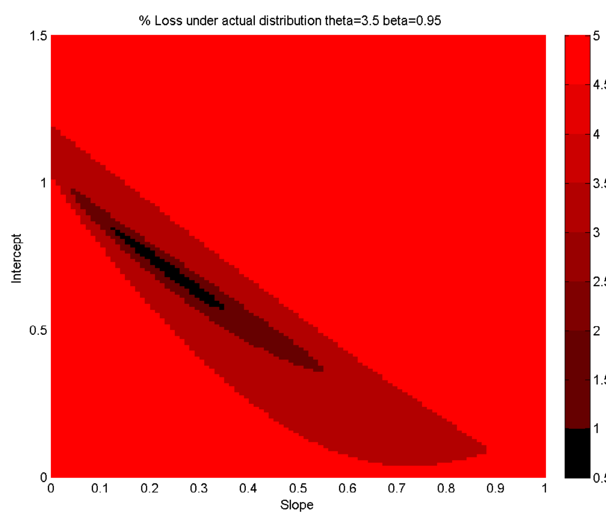


(b)

Figure D.7: Set of consumption rules which have a percentage deviation less than or equal to 0.5% (black), 1% (dark gray), 3% (light gray) and more than 3% (red) from the optimal consumption function for different values of the intercept (vertical axis) and slope (horizontal axis), under (a) π^* and (b) π^γ for $\theta = 3.5$ and $\delta = 0.95$.



(a)



(b)

Figure D.8: Distribution of (a) $D_1^{\gamma t}$, (b) $D_2^{\gamma t}$, (c) α_t, (d) β_t, and percentage of simulations with (e) $D_1^{\gamma t}$ and (f) $D_2^{\gamma t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 3.5$, $\delta = 0.95$.

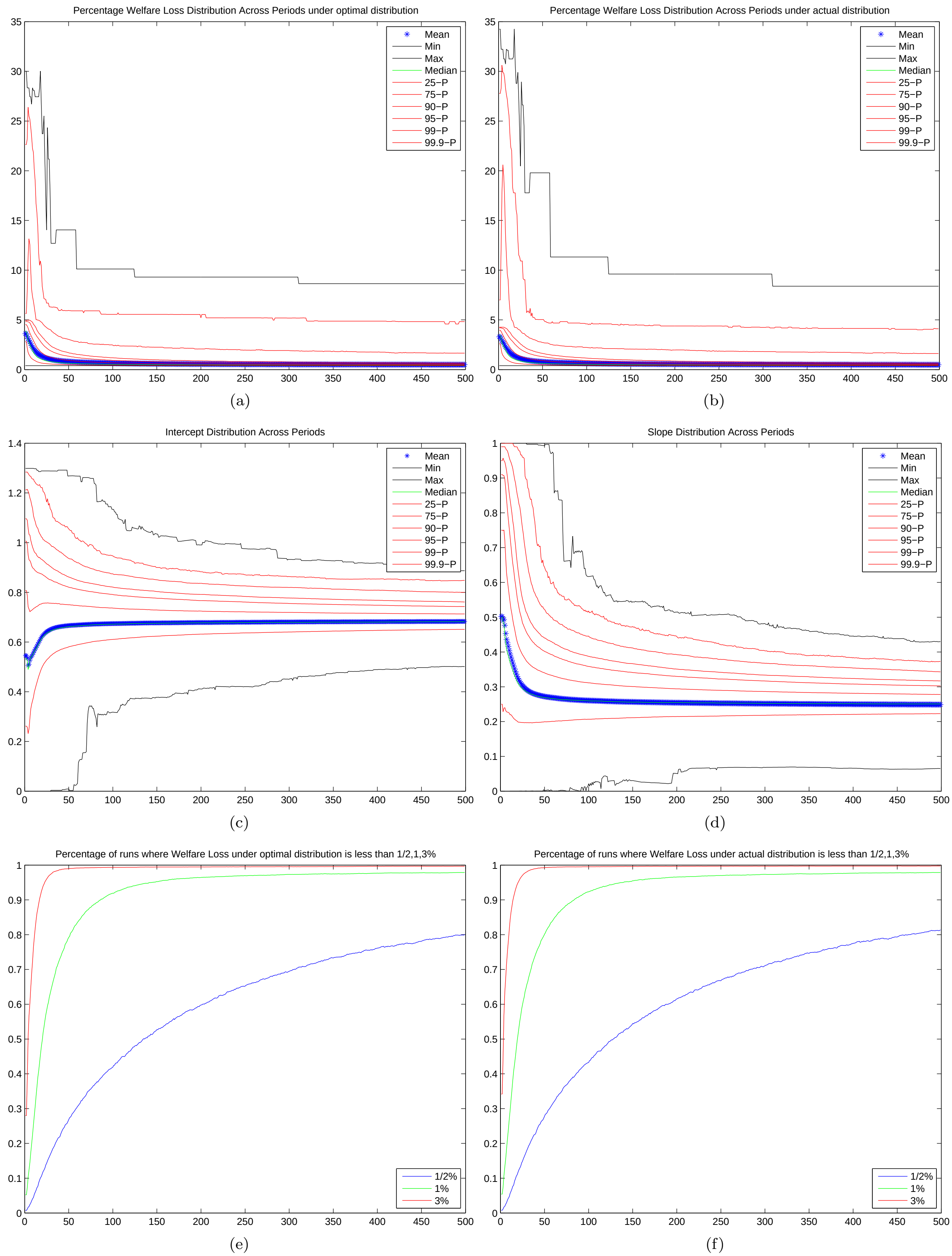


Figure D.9: Dynamics of the distribution of α_t (row 1), β_t (row 2), w_t (row 3), and c_t (row 4) under income shocks. During periods 1-25 and 51-100 agents' income follows Y^1. During periods 26-50 their income follows Y^2, Y^3, Y^4, Y^5.

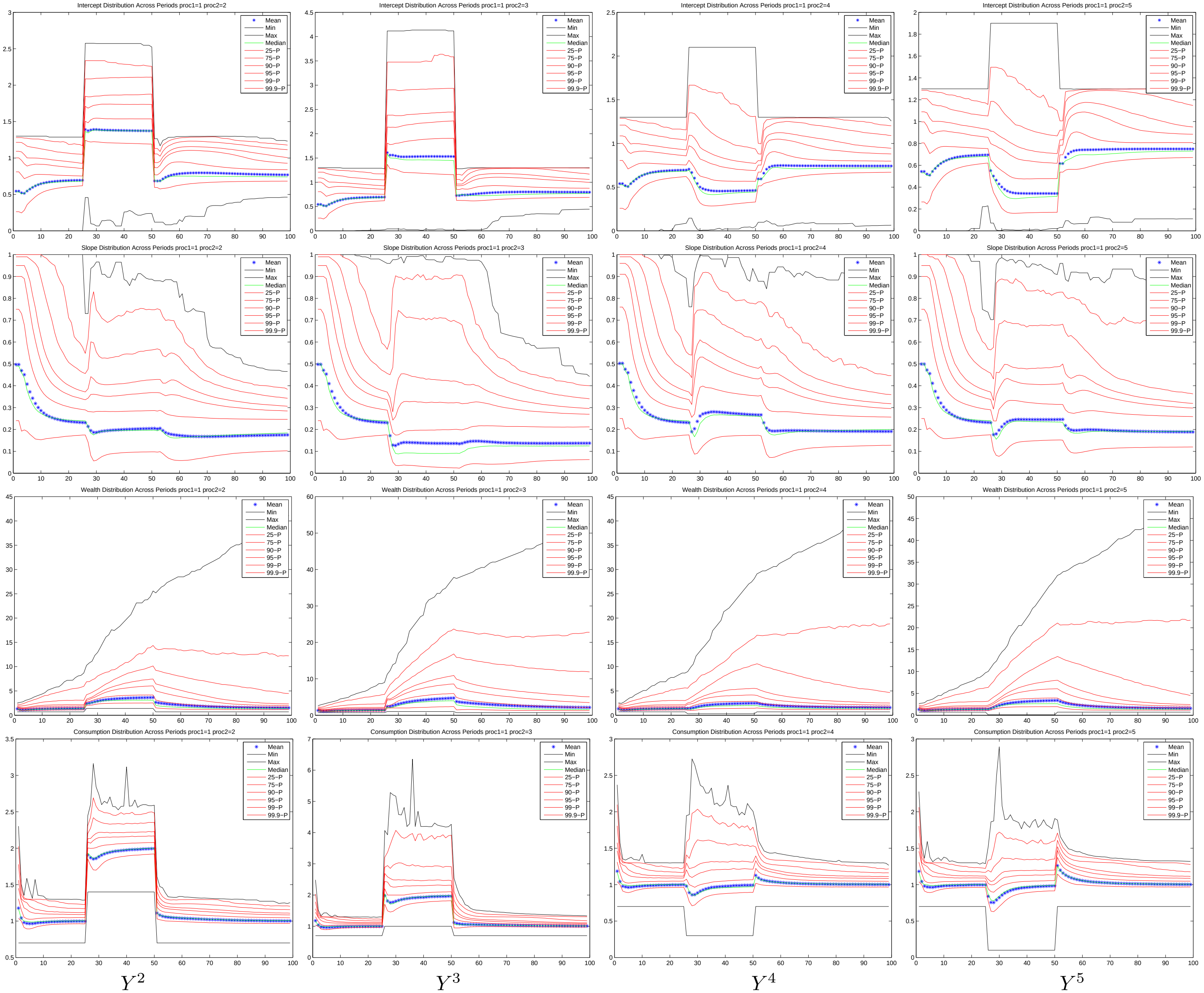
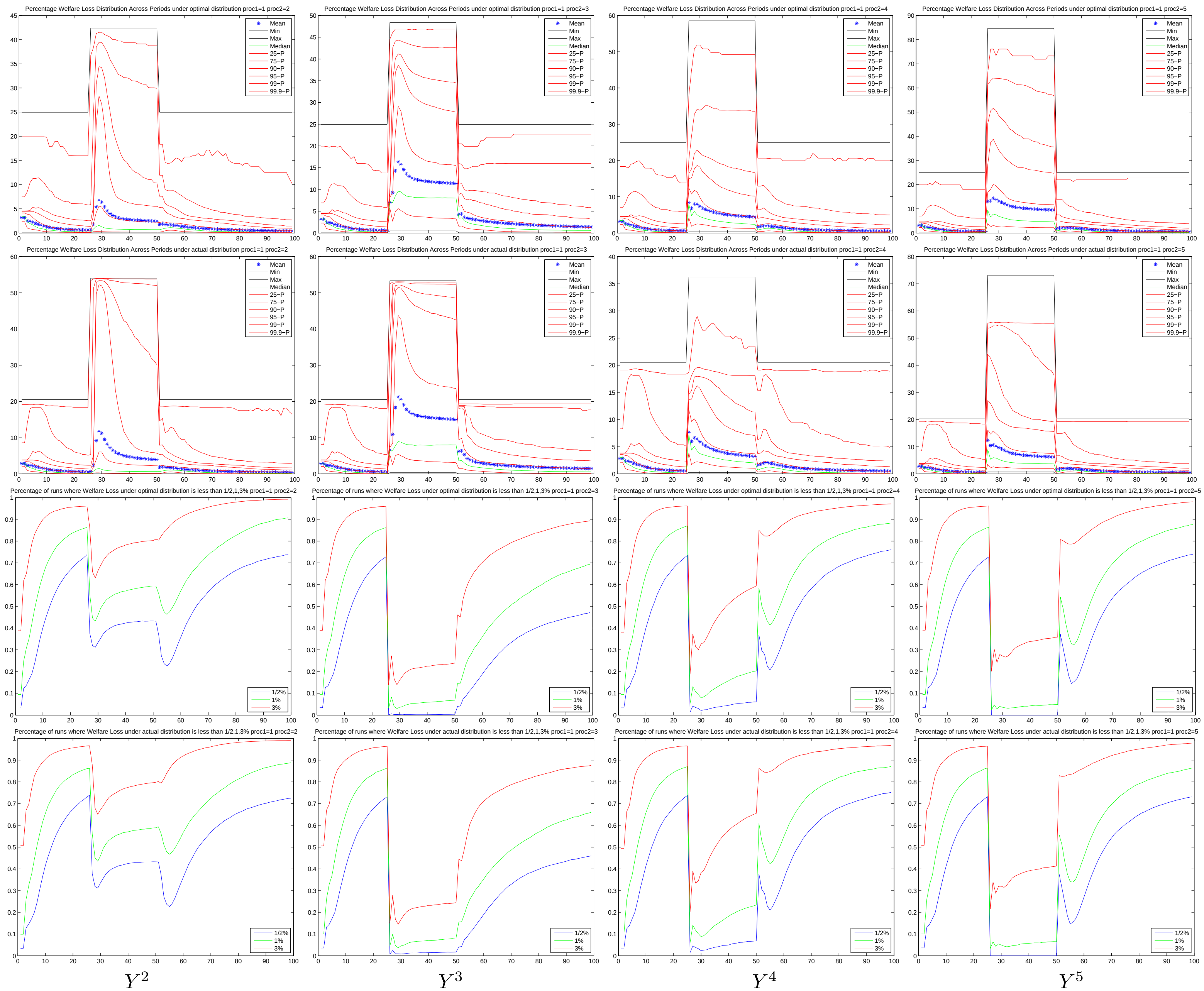


Figure D.10: Dynamics of the distribution of $D_1^{\gamma_t}$ (row 1), $D_2^{\gamma_t}$ (row 2), percentage of runs with $D_1^{\gamma_t}$ (row 3) and $D_2^{\gamma_t}$ (row 4) less than 0.5%, 1% and 3% under income shocks. During periods 1-25 and 51-100 agents' income follows Y^1. During periods 26-50 their income follows Y^2, Y^3, Y^4, Y^5.



Appendix E. Additional Figures

Figure E.11: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept (vertical axis) and slope (horizontal axis), under π^* (first and third rows) and π^γ (second and fourth rows) for $\theta = 1.5, 2, 3, 3.5, 4$, $\delta = 0.9$ (rows 1 and 2) and $\delta = 0.95$ (rows 3 and 4).

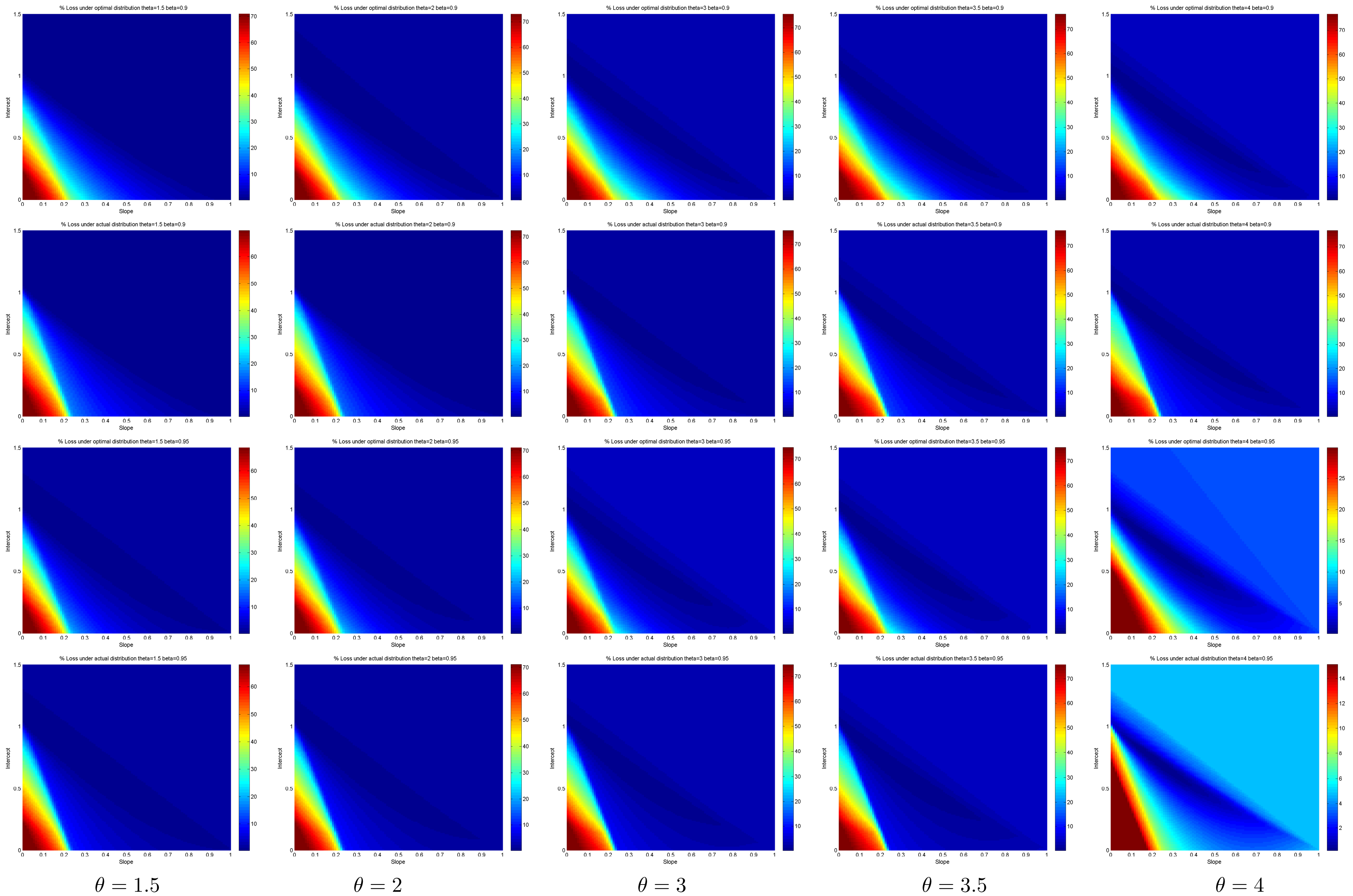


Figure E.12: Set of consumption rules which have a percentage deviation less than or equal to 0.5% (black), 1% (dark gray), 3% (light gray) and more than 3% (red) from the optimal consumption function for different values of the intercept (vertical axis) and slope (horizontal axis), under π^* (first and third rows) and π^γ (second and fourth rows) for $\theta = 1.5, 2, 3, 3.5, 4$, $\delta = 0.9$ (rows 1 and 2) and $\delta = 0.95$ (rows 3 and 4).

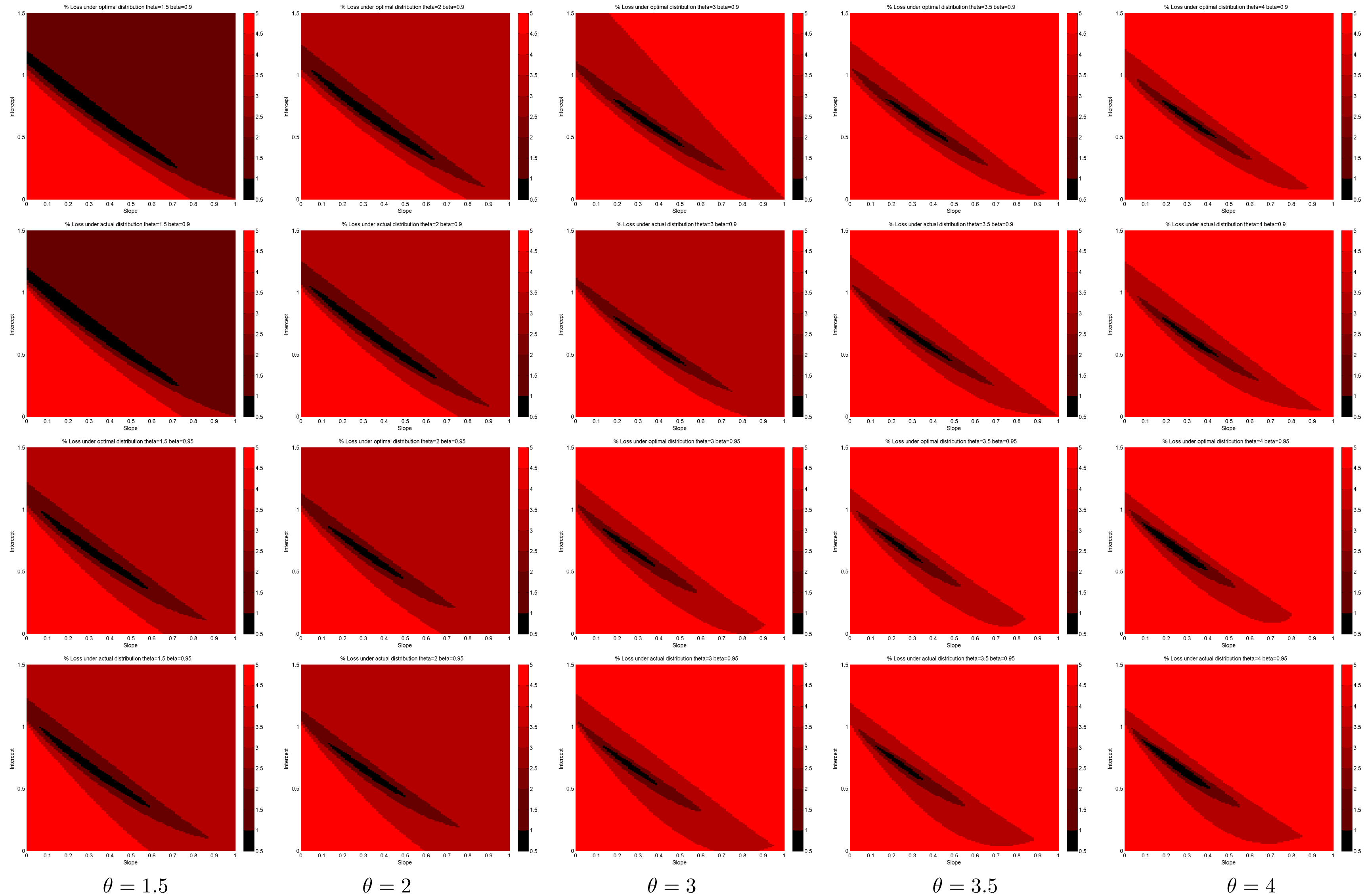


Figure E.13: Distribution of α_t (column 1), β_t (column 2), $D_1^{\gamma t}$ (column 3), $D_2^{\gamma t}$ (column 4) and percentage of simulations with $D_1^{\gamma t}$ (column 5) and $D_2^{\gamma t}$ (column 6) less than $0.5\%, 1\%, 3\%$ for $\theta = 1.5, 2, 3, 3.5, 4$, $\delta = 0.95$.

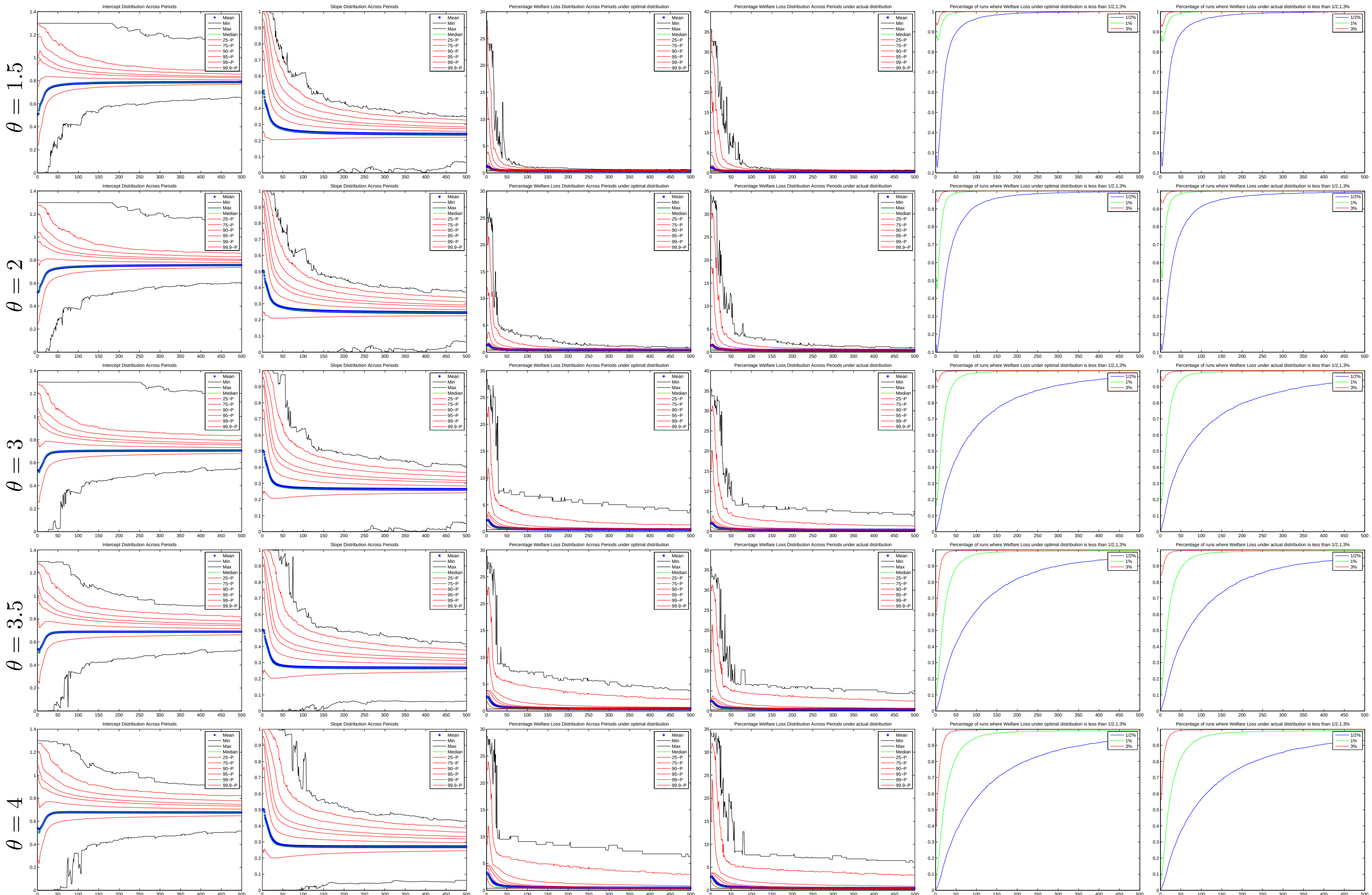


Figure E.14: Distribution of α_t (column 1), β_t (column 2), $D_1^{\gamma t}$ (column 3), $D_2^{\gamma t}$ (column 4) and percentage of simulations with $D_1^{\gamma t}$ (column 5) and $D_2^{\gamma t}$ (column 6) less than $0.5\%, 1\%, 3\%$ for $\theta = 1.5, 2, 3, 3.5, 4$, $\delta = 0.95$.

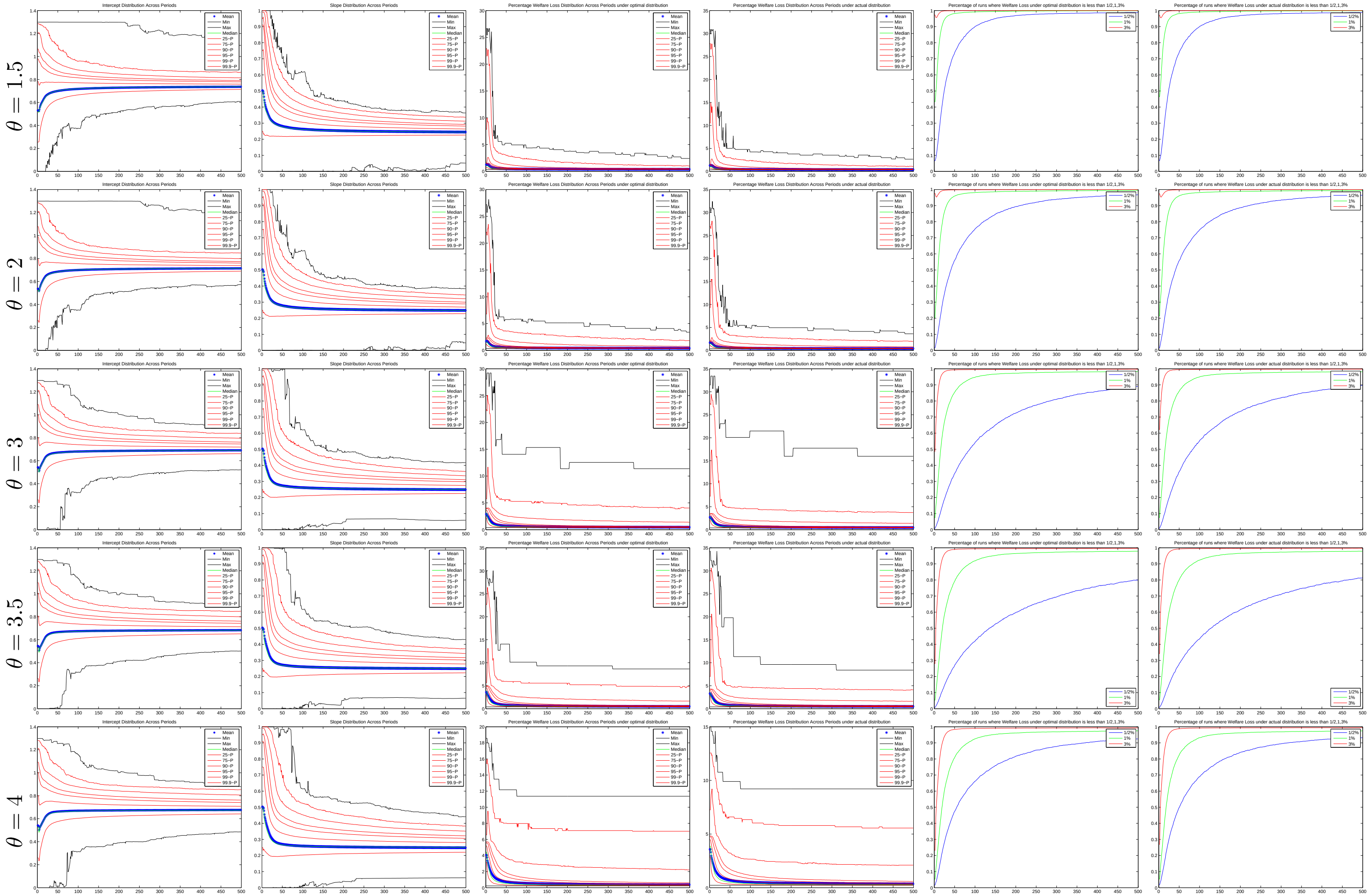
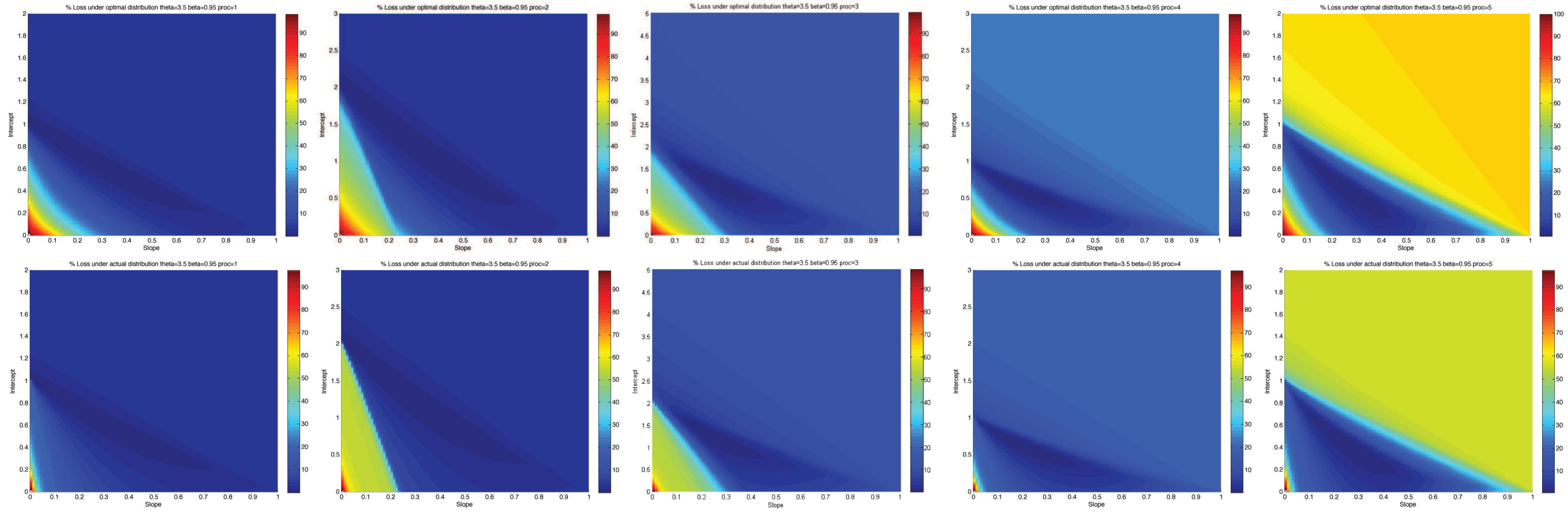
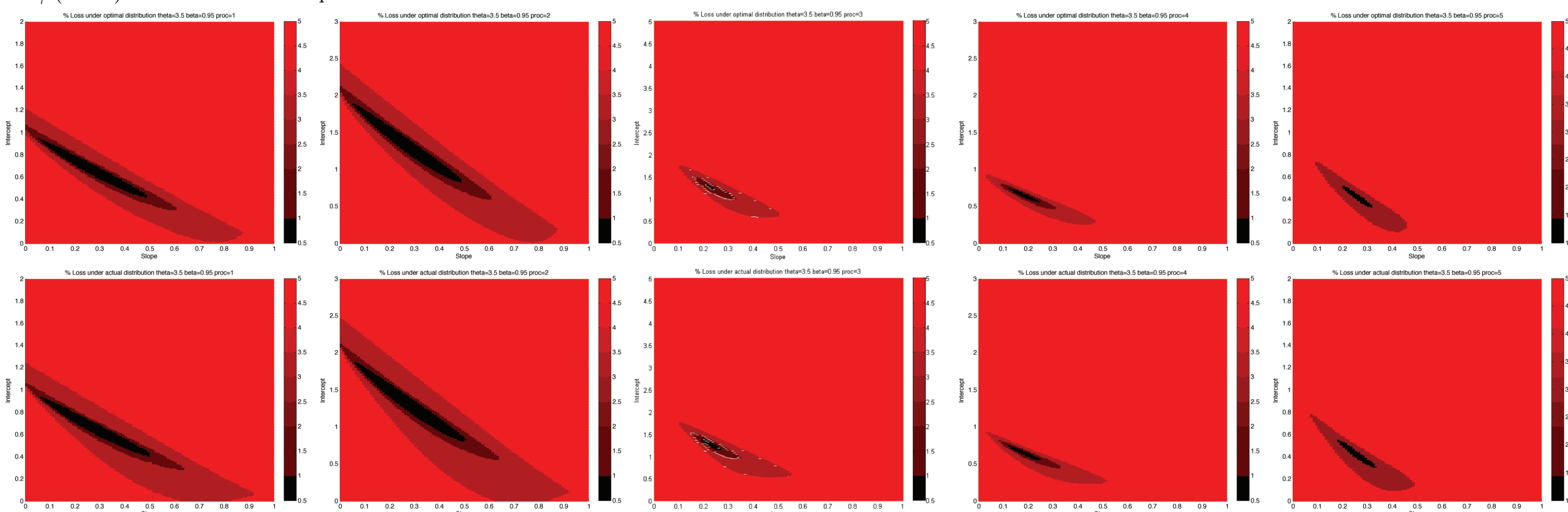


Figure E.15: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept and slope, under π_* (row 1) and π_γ (row 2) for different income processes.



Y^1 Y^2 Y^3 Y^4 Y^5

Figure E.16: Set of consumption rules which have a percentage deviation less than or equal to 0.5% (black), 1% (dark gray), 3% (light gray) and more than 3% (red) under π_* (row 1) and π_γ (row 2) for different income processes.



Y^1 Y^2 Y^3 Y^4 Y^5

Table E.1: EV^*_*, CE^*_*, EV^*_γ and CE^*_γ, EV^γ_*, CE^γ_*, EV^γ_γ and CE^γ_γ under optimal linear rule, for each configuration of δ and θ

δ	θ	EV^*_*	CE^*_*	EV^*_γ	CE^*_γ	EV^γ_*	CE^γ_*	EV^γ_γ	CE^γ_γ
0.9	1.5	-0.1607	0.9841	-0.1650	0.9837	-0.1961	0.9807	-0.2003	0.9803
	2	-0.1731	0.9830	-0.1786	0.9824	-0.2092	0.9795	-0.2148	0.9790
	3	-0.2001	0.9806	-0.2027	0.9803	-0.2382	0.9770	-0.2408	0.9768
	3.5	-0.2128	0.9795	-0.2220	0.9786	-0.2515	0.9759	-0.2607	0.9751
	4	-0.2257	0.9784	-0.2275	0.9782	-0.2662	0.9747	-0.2680	0.9746
0.95	1.5	-0.1962	0.9903	-0.1987	0.9901	-0.2705	0.9866	-0.2730	0.9865
	2	-0.2184	0.9892	-0.2213	0.9891	-0.2943	0.9855	-0.2972	0.9854
	3	-0.2555	0.9875	-0.2554	0.9875	-0.3349	0.9837	-0.3347	0.9837
	3.5	-0.2709	0.9868	-0.2819	0.9862	-0.3525	0.9829	-0.3635	0.9824
	4	-0.2856	0.9861	-0.2848	0.9862	-0.3319	0.9839	-0.3310	0.9840

Table E.2: Percentage difference between between equivalent consumption measures, $D^\gamma_1 = \frac{CE^*_* - CE^\gamma_*}{CE^*_*} * 100$ and $D^\gamma_2 = \frac{CE^*_\gamma - CE^\gamma_\gamma}{CE^*_\gamma} * 100$, and optimal linear rule. Percentage loss under the consume everything rule and expected loss over the whole set of parameters and conditional on rule being in the admissible set.

		Optimal Rule				Consume Everything		Random Rule			
δ	θ	D^γ_1	D^γ_2	α^*	β^*	$D^{(0,1)}_1$	$D^{(0,1)}_2$	$E(D^\gamma_1)$	$E(D^\gamma_1 \mid \mathcal{A})$	$E(D^\gamma_2)$	$E(D^\gamma_2 \mid \mathcal{A})$
0.9	1.5	0.35	0.35	0.61	0.39	0.98	0.93	6.54	3.79	6.34	3.27
	2	0.35	0.35	0.62	0.36	1.54	1.38	7.00	4.19	6.61	3.48
	3	0.36	0.36	0.61	0.34	3.00	2.50	8.16	5.28	7.38	4.16
	3.5	0.37	0.36	0.65	0.30	3.84	3.11	8.80	5.90	7.82	4.55
	4	0.38	0.38	0.65	0.29	4.71	3.74	9.49	6.57	8.28	4.98
0.95	1.5	0.37	0.37	0.63	0.33	1.57	1.45	5.78	3.09	5.64	2.77
	2	0.37	0.37	0.66	0.29	2.34	2.10	6.41	3.65	6.14	3.20
	3	0.38	0.38	0.71	0.23	4.09	3.52	7.84	4.96	7.28	4.21
	3.5	0.39	0.39	0.73	0.21	5.03	4.26	8.60	5.68	7.89	4.75
	4	0.22	0.22	0.72	0.21	5.79	4.83	6.97	5.57	4.94	4.19

Table E.3: Probability of $D^{\gamma t}_1$ or $D^{\gamma t}_2$ below 0.5% at different periods for different parametrizations of the algorithm for $\eta = 0$, $\xi = 1$ and $\varepsilon = 0$.

		$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
δ	θ	$D^{\gamma t}_1$	$D^{\gamma t}_2$	$D^{\gamma t}_1$	$D^{\gamma t}_2$	$D^{\gamma t}_1$	$D^{\gamma t}_2$	$D^{\gamma t}_1$	$D^{\gamma t}_2$	$D^{\gamma t}_1$	$D^{\gamma t}_2$
0.9	1.5	0.2863	0.2950	0.8980	0.8943	0.9650	0.9581	0.9954	0.9927	0.9997	0.9993
	2	0.1410	0.1444	0.7706	0.7748	0.9181	0.9149	0.9847	0.9791	0.9963	0.9945
	3	0.0127	0.0127	0.4573	0.4390	0.6538	0.6232	0.8781	0.8398	0.9644	0.9424
	3.5	0.0104	0.0107	0.4214	0.4204	0.6241	0.6188	0.8678	0.8577	0.9544	0.9490
	4	0.0082	0.0084	0.3706	0.3695	0.5725	0.5667	0.8300	0.8184	0.9405	0.9290
0.95	1.5	0.0659	0.0667	0.7436	0.7461	0.9017	0.9021	0.9772	0.9772	0.9877	0.9877
	2	0.0155	0.0155	0.5636	0.5614	0.7575	0.7516	0.9212	0.9159	0.9675	0.9649
	3	0.0080	0.0086	0.3535	0.3643	0.5453	0.5575	0.7741	0.7827	0.8909	0.8972
	3.5	0.0075	0.0079	0.2739	0.2841	0.4241	0.4376	0.6544	0.6706	0.8003	0.8138
	4	0.0159	0.0176	0.4773	0.5009	0.6622	0.6913	0.8552	0.8736	0.9248	0.9319

Table E.4: Probability of $D^{\gamma t}_1$ or $D^{\gamma t}_2$ below 0.5% at different periods for different parametrizations of the algorithm for $\theta = 3.5$ and $\delta = 0.95$.

			$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
η	ξ	ε	$D^{\gamma t}_1$	$D^{\gamma t}_2$	$D^{\gamma t}_1$	$D^{\gamma t}_2$	$D^{\gamma t}_1$	$D^{\gamma t}_2$	$D^{\gamma t}_1$	$D^{\gamma t}_2$	$D^{\gamma t}_1$	$D^{\gamma t}_2$
0	0	0.2	0.0075	0.0079	0.1355	0.1395	0.1267	0.1303	0.1203	0.1230	0.1216	0.1243
0	0	0	0.0075	0.0079	0.2234	0.2328	0.3355	0.3497	0.5153	0.5344	0.6404	0.6561
0	1	0.2	0.0075	0.0079	0.1273	0.1324	0.1262	0.1309	0.1265	0.1316	0.1279	0.1324
0	1	0	0.0075	0.0079	0.2739	0.2841	0.4241	0.4376	0.6544	0.6706	0.8003	0.8138
0.5	0	0.2	0.0075	0.0079	0.2730	0.2783	0.2497	0.2520	0.2284	0.2307	0.2223	0.2243
0.5	0	0	0.0075	0.0079	0.2222	0.2329	0.2919	0.3057	0.3999	0.4193	0.4970	0.5167
0.5	1	0.2	0.0075	0.0079	0.2549	0.2627	0.2893	0.2981	0.2922	0.3000	0.2933	0.3011
0.5	1	0	0.0075	0.0079	0.1468	0.1519	0.1954	0.2019	0.2799	0.2888	0.3660	0.3750

Table E.5: Probability of $D_1^{\gamma t}$ or $D_2^{\gamma t}$ below 0.5% at different periods for different parametrizations of the algorithm under social learning for $\phi = \frac{1}{3}$, $\theta = 3.5$ and $\delta = 0.95$.

Indiv=200, Neigh=2			$t=0$		$t=50$		$t=100$		$t=250$		$t=500$	
η	ξ	ε	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$
0	0	0.2	0.0075	0.0079	0.4351	0.4356	0.3520	0.3522	0.3308	0.3308	0.3397	0.3397
0	0	0	0.0075	0.0079	0.8995	0.9020	0.9848	0.9849	0.9992	0.9992	1.0000	1.0000
0	1	0.2	0.0075	0.0079	0.4810	0.4875	0.4794	0.4862	0.4850	0.4910	0.4760	0.4826
0	1	0	0.0075	0.0079	0.7925	0.7960	0.9513	0.9513	0.9991	0.9991	1.0000	1.0000
0.5	0	0.2	0.0075	0.0079	0.5216	0.5216	0.2787	0.2787	0.1790	0.1790	0.1841	0.1841
0.5	0	0	0.0075	0.0079	0.6502	0.6620	0.8203	0.8243	0.9606	0.9606	0.9939	0.9939
0.5	1	0.2	0.0075	0.0079	0.5980	0.5997	0.6468	0.6475	0.6531	0.6538	0.6380	0.6389
0.5	1	0	0.0075	0.0079	0.0276	0.0278	0.0549	0.0549	0.1800	0.1800	0.2898	0.2898

Table E.6: Expected loss under optimal distribution, $ED_1^{\gamma t}$, and under linear rule's distribution, $ED_2^{\gamma t}$, at different periods for different parametrizations of the consumer's problem for $\eta = 0$, $\xi = 1$ and $\varepsilon = 0$.

		$t=0$		$t=5$		$t=10$		$t=25$		$t=50$	
δ	θ	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$
0.9	1.5	1.20	1.38	1.09	1.26	0.77	0.85	0.48	0.50	0.42	0.42
	2	1.33	1.43	1.35	1.48	0.96	1.04	0.58	0.59	0.47	0.48
	3	2.12	2.04	1.97	1.97	1.42	1.44	0.81	0.81	0.60	0.60
	3.5	2.65	2.46	2.33	2.26	1.70	1.68	0.94	0.94	0.67	0.66
	4	3.24	2.93	2.72	2.58	2.00	1.94	1.10	1.08	0.75	0.74
0.95	1.5	1.26	1.28	1.25	1.30	0.90	0.93	0.58	0.58	0.49	0.49
	2	1.72	1.69	1.59	1.60	1.12	1.13	0.69	0.69	0.56	0.55
	3	2.91	2.70	2.33	2.25	1.69	1.65	0.95	0.93	0.71	0.70
	3.5	3.61	3.29	2.79	2.65	2.01	1.94	1.12	1.08	0.80	0.78
	4	4.06	3.57	2.96	2.63	2.13	1.94	1.11	1.04	0.73	0.69

Table E.7: Median loss under optimal distribution, $MD_1^{\gamma t}$, and linear rule's distribution, $MD_2^{\gamma t}$, at different periods for different parametrizations of the consumer's problem for $\eta = 0$, $\xi = 1$ and $\varepsilon = 0$.

		$t=0$		$t=5$		$t=10$		$t=25$		$t=50$	
δ	θ	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$
0.9	1.5	0.67	0.65	0.67	0.65	0.55	0.54	0.41	0.42	0.39	0.38
	2	0.98	0.92	0.96	0.91	0.73	0.71	0.49	0.48	0.43	0.43
	3	2.02	1.79	1.68	1.54	1.14	1.10	0.65	0.65	0.51	0.52
	3.5	2.68	2.32	2.10	1.89	1.38	1.30	0.73	0.74	0.54	0.54
	4	3.36	2.86	2.54	2.26	1.65	1.55	0.82	0.81	0.58	0.58
0.95	1.5	1.09	1.04	0.94	0.90	0.69	0.69	0.49	0.49	0.44	0.44
	2	1.66	1.55	1.27	1.22	0.89	0.87	0.57	0.57	0.48	0.48
	3	3.04	2.74	2.06	1.92	1.33	1.28	0.75	0.73	0.57	0.57
	3.5	3.84	3.42	2.52	2.32	1.59	1.50	0.85	0.83	0.63	0.62
	4	4.42	3.88	2.79	2.52	1.69	1.57	0.78	0.74	0.52	0.50

Table E.8: Expected loss under optimal distribution, $ED_1^{\gamma t}$, and under linear rule's distribution, $ED_2^{\gamma t}$, at different periods for different parametrizations of the algorithm for $\theta = 3.5$ and $\delta = 0.95$.

			$t=0$		$t=5$		$t=10$		$t=25$		$t=50$	
η	ξ	ε	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$
0	0	0.2	3.61	3.29	2.79	2.82	2.67	2.86	2.01	2.07	1.95	2.05
0	0	0	3.61	3.29	2.79	2.81	2.38	2.51	1.51	1.49	1.28	1.26
0	1	0.2	3.61	3.29	2.79	2.65	2.17	2.13	1.40	1.42	1.24	1.25
0	1	0	3.61	3.29	2.79	2.65	2.01	1.94	1.12	1.08	0.80	0.78
0.5	0	0.2	3.61	3.29	2.51	2.30	1.74	1.63	1.05	1.01	0.94	0.93
0.5	0	0	3.61	3.29	2.49	2.30	1.71	1.60	1.12	1.07	0.92	0.89
0.5	1	0.2	3.61	3.29	2.73	2.47	2.02	1.85	1.14	1.09	0.81	0.80
0.5	1	0	3.61	3.29	2.75	2.49	2.13	1.95	1.47	1.38	1.12	1.07

Table E.9: Median loss under optimal distribution, $MD_1^{\gamma t}$, and under linear rule's distribution, $MD_2^{\gamma t}$, at different periods for different parametrizations of the algorithm for $\theta = 3.5$ and $\delta = 0.95$.

			$t = 0$		$t = 5$		$t = 10$		$t = 25$		$t = 50$	
η	ξ	ε	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$
0	0	0.2	3.84	3.42	2.13	1.98	1.69	1.61	1.10	1.08	1.00	1.00
0	0	0	3.84	3.42	2.13	2.01	1.52	1.45	0.91	0.88	0.70	0.68
0	1	0.2	3.84	3.42	2.52	2.32	1.72	1.62	1.06	1.03	0.98	0.96
0	1	0	3.84	3.42	2.52	2.32	1.59	1.50	0.85	0.83	0.63	0.62
0.5	0	0.2	3.84	3.42	2.46	2.28	1.39	1.33	0.75	0.74	0.67	0.66
0.5	0	0	3.84	3.42	2.46	2.28	1.31	1.26	0.82	0.79	0.69	0.67
0.5	1	0.2	3.84	3.42	2.84	2.59	1.82	1.69	0.85	0.83	0.66	0.65
0.5	1	0	3.84	3.42	2.87	2.62	1.97	1.84	1.21	1.16	0.92	0.89

Table E.10: Expected loss under optimal distribution, $ED_1^{\gamma t}$, and under linear rule's distribution, $ED_2^{\gamma t}$, at different periods for different parametrizations of the algorithm under social learning for $\phi = \frac{1}{3}$, $\theta = 3.5$ and $\delta = 0.95$.

Indiv=200, Neigh=2			$t = 0$		$t = 5$		$t = 10$		$t = 25$		$t = 50$	
η	ξ	ε	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$	$ED_1^{\gamma t}$	$ED_2^{\gamma t}$
0	0	0.2	3.61	3.29	1.65	1.53	0.97	0.93	0.64	0.63	0.71	0.71
0	0	0	3.61	3.29	1.66	1.54	0.88	0.85	0.50	0.49	0.44	0.44
0	1	0.2	3.61	3.29	1.69	1.56	0.99	0.95	0.61	0.60	0.58	0.58
0	1	0	3.61	3.29	1.71	1.57	0.92	0.88	0.54	0.53	0.46	0.46
0.5	0	0.2	3.61	3.29	2.30	2.11	1.09	1.03	0.54	0.54	0.58	0.59
0.5	0	0	3.61	3.29	2.31	2.11	1.09	1.03	0.57	0.56	0.49	0.48
0.5	1	0.2	3.61	3.29	2.44	2.22	1.52	1.42	0.69	0.68	0.52	0.52
0.5	1	0	3.61	3.29	2.45	2.23	1.63	1.52	1.00	0.97	0.77	0.75

Table E.11: Median loss under optimal distribution, $MD_1^{\gamma t}$, and under linear rule's distribution, $MD_2^{\gamma t}$, at different periods for different parametrizations of the algorithm under social learning for $\phi = \frac{1}{3}$, $\theta = 3.5$ and $\delta = 0.95$.

Indiv=200, Neigh=2			$t = 0$		$t = 5$		$t = 10$		$t = 25$		$t = 50$	
η	ξ	ε	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$
0	0	0.2	3.84	3.42	1.52	1.41	0.74	0.71	0.50	0.49	0.52	0.53
0	0	0	3.84	3.42	1.52	1.42	0.66	0.65	0.46	0.45	0.42	0.42
0	1	0.2	3.84	3.42	1.57	1.44	0.81	0.78	0.53	0.53	0.51	0.51
0	1	0	3.84	3.42	1.58	1.45	0.75	0.73	0.50	0.49	0.45	0.45
0.5	0	0.2	3.84	3.42	2.39	2.21	0.91	0.87	0.46	0.46	0.48	0.48
0.5	0	0	3.84	3.42	2.39	2.22	0.92	0.87	0.51	0.51	0.47	0.47
0.5	1	0.2	3.84	3.42	2.54	2.32	1.46	1.37	0.63	0.62	0.47	0.47
0.5	1	0	3.84	3.42	2.55	2.34	1.59	1.49	0.96	0.93	0.75	0.73

Table E.12: Statistics of the various income processes.

Income Process	Mean	Std	Kurtosis	Skewness
Y^1	1.00	0.19	-0.51	-0.00
Y^2	2.00	0.38	-0.50	0.00
Y^3	1.99	0.78	2.64	1.68
Y^4	1.00	0.41	2.74	1.59
Y^5	0.95	0.26	3.84	-1.64

Table E.13: Percentage difference between between equivalent consumption measures, $D_1^{\gamma} = \frac{CE_*^* - CE_*^{\gamma}}{CE_*^*} * 100$ and $D_2^{\gamma} = \frac{CE_{\gamma}^* - CE_{\gamma}^{\gamma}}{CE_{\gamma}^*} * 100$, and optimal linear rule. Percentage loss under the consume everything rule and expected loss over the whole set of parameters and conditional on rule being in the admissible set for $\theta = 3.5$, $\delta = 0.95$, $B = 0$.

	Optimal Rule				Consume Everything		Random Rule			
Process	D_1^{γ}	D_2^{γ}	α^*	β^*	$D_1^{(0,1)}$	$D_2^{(0,1)}$	$E(D_1^{\gamma})$	$E(D_1^{\gamma} \mid \mathcal{A})$	$E(D_2^{\gamma})$	$E(D_2^{\gamma} \mid \mathcal{A})$
1	0.02	0.02	0.72	0.22	4.62	3.88	6.40	3.27	4.46	2.80
2	0.02	0.02	1.46	0.21	4.62	3.88	7.91	6.94	7.50	5.47
3	0.47	0.36	1.26	0.22	14.64	11.76	14.50	13.97	12.64	12.32
4	0.29	0.28	0.66	0.17	23.79	19.51	21.21	18.98	17.17	15.41
5	0.80	0.76	0.44	0.24	65.99	55.90	53.78	52.53	45.53	44.87

Table E.14: Probability of $D_1^{\gamma t}$ or $D_2^{\gamma t}$ below 3% at different periods for fixed parametrization of the algorithm under different income processes for $\theta = 3.5$, $\delta = 0.95$, $B = 0$, $\eta = 0$, $\epsilon = 0$, $\xi = 1$ and $\phi = 1$.

	$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
Process	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$
1	0.3846	0.4980	0.8952	0.8866	0.9064	0.9011	0.9194	0.9141	0.9291	0.9224
2	0.3860	0.4989	0.9054	0.8987	0.9156	0.9124	0.9290	0.9254	0.9373	0.9318
3	0.0252	0.0299	0.3394	0.3441	0.4497	0.4460	0.5427	0.5289	0.5885	0.5701
4	0.0120	0.0129	0.3537	0.3420	0.5006	0.4865	0.6472	0.6335	0.7040	0.6961
5	0.0031	0.0038	0.0462	0.0585	0.0785	0.0958	0.1423	0.1651	0.2142	0.2415

Table E.15: Probability of $D_1^{\gamma t}$ or $D_2^{\gamma t}$ below 3% at different periods for fixed parametrization of the algorithm under different income processes under ID version of the algorithm for $\theta = 3.5$, $\delta = 0.95$, $B = 0$, $\eta = 0$, $\epsilon = 0$, $\xi = 1$ and $\phi = 1$.

	$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
Process	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$
1	0.3846	0.4980	0.9680	0.9742	0.9735	0.9785	0.9804	0.9833	0.9847	0.9862
2	0.3860	0.4989	0.9678	0.9742	0.9728	0.9783	0.9790	0.9821	0.9832	0.9847
3	0.0252	0.0299	0.8103	0.8150	0.8495	0.8468	0.8707	0.8650	0.8840	0.8762
4	0.0120	0.0129	0.6666	0.7393	0.7615	0.8062	0.8317	0.8538	0.8590	0.8707
5	0.0031	0.0038	0.2321	0.2829	0.2652	0.3251	0.3024	0.3659	0.3109	0.3773