Expectations with No Regrets

Omer Edhan‡

November 7, 2019

Abstract

I develop a general learning framework in order to answer the question “how do expectations evolve?” I consider a recursive general-equilibrium framework that nests a large class of macroeconomic models due to the agents’ uncertainty of the economy’s law of motion. I impose a learning framework based on the idea that agents try to minimise losses in their future expected payoff occurring due to such uncertainties. Unlike the vast majority of the learning literature in macroeconomics, the entire learning process is phrased in terms of the structural model and a reduced form, thus I offer a solution to the critique of Williams (2003). I give conditions under which the cumulative payoff loss, or ‘regret’, is small compared to the total payoff that the agent might have had had she adhered to the model supplying the optimal payoff in hindsight. An immediate outcome of the theory is that an agent’s expectations are fundamentally both forward-looking and backward looking, which is consistent with a very large body of empirical evidence from recent years. I examine some of the economic implications in a simple permanent income model with linear marginal utility. No-regret learning results in generic violation of the “random walk hypothesis” studied ever since Hall’s (1978) classical paper. Namely – contrary to rational expectations prediction, under no-regret learning consumption would typically not follow a random walk nor would it be a martingale. Nevertheless, under some conditions on the convergence of the learning process and its rate of convergence, consumption has a central limit theorem.

∗Department of Economics, University of Manchester; omer.edhan@gmail.com
†I would like to thank the following people for the helpful discussions: Abraham Neyman, David Levine, Igor Evstigneev, Eugene Feinberg, Ziv Hellman, James Banks, Raffaele Rossi, Agnes Kovacs, and Michele Berardi.
1 Introduction

Modern macroeconomic modelling heavily relies on the rational-expectations assumption which makes expectations model-consistent and predicts their behaviour under changes in policy. It is nevertheless unrealistic regarding its information requirement, as agents are supposed to have complete information of the economic environment as well as perfect foresight. In practice, rational expectations is challenged by a very large body of empirical literature (e.g., [4], [5], [6], [10], [11], [18], [19], [20], [23], [30]). Many of the proposed alternatives, such as rational inattention (Sims [33]) or sticky information (Mankiw and Reis [25]) impose costs on optimisation, while others explicitly or implicitly assume that agents beliefs may indeed be able to capture the actual complexity of the economic reality (e.g., Marcet and Sargent [26], [27], Hansen and Sargent [17]). Nevertheless, it is unclear how an agent with no optimisation costs should behave when she is facing the uncertainty that her beliefs may fail to capture the actual complexity of the economic reality.

This paper solves this problem using a general framework which is based in the notion of regret minimisation (e.g. [8]). It considers a recursive general-equilibrium framework in which agents face uncertainty regarding the dynamic economic environment. Agents are endowed with a set of models for the economy, use them to learn about the environment, and update their beliefs to minimise ‘regret’ - the loss in their future expected payoff occurring due to their uncertainty, including the possibility that none of their models capture the economy’s actual evolution.

The paper makes the following contributions:

1. It proposes a unified coherent deviation from rational expectations, in which agents wish to insure themselves against model uncertainty that includes the possibility that their models do not capture the economy’s complexity.

2. It proposes a general principle that allows to model learning in a general family of structural models\textsuperscript{1}.

3. It studies implications of regret minimisation in a simple permanent income model in which agents do not fully observe the relevant states contributing to their income.

In the first part of the paper I develop a framework of counterfactual dynamic regret within a dynamic general-equilibrium economy with a representative agent. A model’s

\textsuperscript{1}Up to date, the majority of works on learning in macroeconomics have focused on ‘learning’ in the reduced form. The critique of Williams [34] on this matter is still relevant, and to the best of my knowledge this is the first work that gives a general and coherent principle that resolves this matter.
counterfactual value is formed as the agent considers the counterfactual stating “what if I would have believed a particular model today and returned to my original belief from tomorrow and onwards?” The agent can then consider the difference between the counterfactual value of each model and her payoff given her beliefs. A model’s gain is the difference between its counterfactual value and the agent’s payoff given her belief. The counterfactual dynamic regret of a model at a certain time is the totality of its gains until that time. In other words, it is the amount of payoff the agent may have gained had she adhered to this particular model. An optimal model in hindsight is a model producing the highest level of regret. In my framework, the agent would thus wish to make choices and update her beliefs so her regret is minimised, while considering the possibility that none of her models may capture the actual dynamics of the economy.

I prove that the agent can make her average-per-period regret diminish to zero. Namely, the experienced payoff loss due to model uncertainty can be made quite insignificant in the long run. Furthermore, I show that the agent may update her beliefs in such a manner that any of her policy choices would lead to diminishing average-per-period regret. The regret minimising belief update turns up to be non-Bayesian, and depends on the agent’s expected payoffs at every stage of the game. Due to the recursive structure of the optimisation problem, the dependency of the belief updates on the expected payoff makes the belief updates both backward-looking and forward-looking, which is in agreement with survey evidence (e.g., Coibion et al. [11]).

As an example of the economic implication of my framework, I consider a simple permanent income model with income uncertainty following Hall’s [15] classical paper. In this model an agent wishes to optimise her consumption from a quadratic utility subject to budget constraints and income uncertainties. Contrary to rational expectations prediction (e.g., Hall [15]), under regret minimisation the consumption is typically not a martingale. I show that the size of the deviation from the rational expectations martingale property is pinned down by the income expectations gap - the difference in permanent income expectations between two consecutive time periods. If the income expectations gap diminishes to zero quickly enough, some of the classical rational expectations outcomes can be salvaged. In such a case consumption may still fail to be a martingale, but it would nevertheless have a central limit theorem.

My work is related and contributes to several avenues in the Literature. The literature on robustness concerns (as in Hansen and Sargent [17]) discusses agents who entertain models and wish to insure themselves against model uncertainty. Nevertheless, the approach in that literature is that one of the underlying models of the economy captures the true economic environment, while other models are misspecified. In my work the agent’s uncertainty includes the option that none of the models represent the true economic dy-
dynamics. My framework thus captures a prevalent economic reality in which models are formed using historical, and at times incomplete data. A similar difference exists between my work and other works concerning learning such as Marcet and Sargent [26]-[27], Sargent [31], and Marcet and Nicolini [28].

A seemingly separate avenue of the literature is the one concerned with optimisation restrictions due to various ‘information costs’, such as the rational inattention model that was introduced by Sims [33] and the sticky information model that was introduced by Mankiw and Reis [25]. In both models constraints are imposed using various cost functions on either processing or acquisition of information. In my framework, no such costs are imposed. The restriction on the agent relates to the richness of her set of models and the fundamental uncertainty in their ability to capture the economic environment complexity. In relation to that, one of the most striking results of my analysis is that an attempt mitigate the uncertainty by entertaining a more complex set of models will result in higher regret. It is certainly of interest to examine possible connections between these two seemingly separate lines of research, but this is left as a matter for future investigation.

2 Framework

2.1 The Macroeconomic Agent

An infinitely lived macroeconomic agent has\(^2\) subjective expectation \(E_t\) regarding the evolution of the economy. By time \(t\) the agent has already chosen a sequence \(x_1, \ldots, x_{t-1} \in X\) of choice variables and wishes to choose a continuation sequence \(\{x_s\}_{s=t}^\infty\) of choice variables to maximise the expected present-discounted value of her payoff:

\[
\max_{\{x_s\}_{s=t}^\infty} E_t \left[ \sum_{s=t}^\infty \beta^{s-t} u(x_{s-1}, x_s, y_s) \right],
\]

(1)

where \(x_{t-1}\) is given and subject to constraints

\[
x_s \in \Gamma(x_{s-1}, y_s), \quad s = t, t + 1, \ldots
\]

(2)

Here \(\beta \in (0, 1)\) is the discount factor, \(u\) is the per-period utility function, and \(\Gamma\) is a correspondence that determines the agent’s choice set as a function of the choice \(x_{s-1}\) made by the agent in the previous period and an observable \(y_s \in Y\) of the economy at

\(^2\)Not to be confused with the Rational Expectations operator. The nature of this operator and the way in which it is chosen will be subsequently explained.
time $s$. Faithful to its name, the observable $y_t$ is assumed to be observed by the agent. It can affect her optimal choices by shifting her payoffs or changing her choice set, and is out of the agent’s control. The observable evolves in manner unknown to the agent, yet the agent has a subjective set of models which she believes, and may update her beliefs from one period to the other. More specifically, the agent entertains a set of models $\Theta$, s.t. each model $\theta \in \Theta$ corresponds to a transition probability $Q_\theta$. I assume that $y_{t+1} \sim Q_\theta(\cdot | y_t)$, namely the transition depends on the identity of the previous observable.

Problem (1)-(2) is identical to the canonical dynamic macroeconomic decision problem studied in Chapter 9 of Lucas and Stokey [24] except for two important differences. First, the operator $E_t$ represents the agent’s subjective expectation about the path $\{y_t\}_{t=1}^\infty$ of the observables. It can (and will) in general be different than the model-implied (rational) expectation. I develop the modeling and analysis of the evolution of the agent’s subjective expectation in the theory part of this paper. My analysis deviates from the dogma of Bayesian updating common in the literature (e.g., [22], [29], [31]) as I will assume an agent that wishes to minimise her long run regrets, perhaps even against sequence of observables that are chosen adversely in a manner that shifts the sequence away from being asymptotically mean stationary\(^3\).

The second difference concerns the determination of the observable $y_t$, and will only become relevant in applications. In applications, my approach generalises the textbook dynamic decision problem by allowing the observable process $(y_t)_{t \geq 0}$ to have any form, which accommodates the option that it is a projection of Markov process $(y_t, z_t)_{t \geq 0}$, in which the process $z_t$ is unobservable to the agent. In other words, the agent may entertain hidden factor models.

### 2.2 The Recursive Framework

Reformulating problems (1)-(2) in recursive terms, and assuming that Bellman’s principle of optimality holds\(^4\), given a model $\theta \in \Theta$ the agent’s Bellman equation at time $t$ is

$$ V(x_-, y, \theta) = \max_{x \in \Gamma(x_- | y)} \left[ u(x_-, x, y) + \beta \int V(x, y_+, \theta) Q_\theta(dy_+ | y) \right]. $$

\(^3\)Not being asymptotically mean stationary distinguishes my work from the “robust optimisation” framework of Hansen and Sargent [17].

\(^4\)The Bellman equation is a Theorem. It may fail to hold. I develop my theory conditionally on it being satisfied. The question of whether or not it is satisfied will only become important in applications.

---

5
The agent’s *policy correspondence* is

\[
x(x_-, y, \theta) \equiv \arg \max_{x \in \Gamma(x_-, y)} \left( u(x_-, x, y) + \beta \int \int_V (x, y_+, \theta) Q_t(dy_+|y) \lambda(d\theta) \right).
\]  

(4)

A *policy* is a map \( \hat{x} : X \times Y \rightarrow X \) mapping the previous choice \( x_- \) and the observable \( y \) to a new choice \( \hat{x}(x_-, y) \in \Gamma(x_-, y) \). It is an *optimal policy* iff \( \hat{x}(x_-, y) \in x(x_-, y, \theta) \).

In my framework, the agent may have a belief over the models, namely, a measure \( \lambda \) on \( \Theta \). I would like to allow the agent to update her beliefs. Unlike other works in macroeconomics (e.g., Sargent [31], Hansen and Sargent ([17]) I will not assume that the agent is a Bayesian updater but rather allow for general belief-updating which I study under an imposed regret minimisation assumption. The existence of a “no-regret” belief updating is one of my main results and will be discussed in Section 3.

I now introduce a straightforward generalization that incorporates model belief updating into my framework. In the spirit of anticipated utility (see Kreps [21], Sargent [31]), I assume the agent doesn’t anticipate belief updating until a new observable is realised. At each time period the agent thus optimises her expected discounted payoff subject to her belief. As beliefs remain un-updated throughout this optimisation process due to the anticipated utility assumption, if the agent’s belief in the current period is given by \( \lambda \in \Delta(\Theta) \) then the problem’s recursive nature is captured by the following Bellman equation:

\[
V(x_-, y, \lambda) = \max_{x \in \Gamma(x_-, y)} \left( u(x_-, x, y) + \beta \int \int_Y (x, y_+, \lambda) Q_t(dy_+|y) \lambda(d\theta) \right).
\]  

(5)

The agent’s *policy correspondence* is

\[
x(x_-, y, \lambda) \equiv \arg \max_{x \in \Gamma(x_-, y)} \left( u(x_-, x, y) + \beta \int \int_V (x, y_+, \theta) Q_t(dy_+|y) \lambda(d\theta) \right),
\]  

(6)

and the definitions of a policy and an optimal policy can be adjusted accordingly.

It will be of use to consider the *counterfactual value* \( U_\theta \) of a model \( \theta \in \Theta \) w.r.t. a belief \( \lambda \) given by

\[
U_\theta(x_-, x, y, \lambda) \equiv \int_Y (u(x_-, x, y) + \beta V(x, y_+, \lambda)) Q_\theta(dy_+|y).
\]  

(7)

\[
^5\text{Again, this is an assumption of my theory. The relevant conditions for the Bellman equation to hold should and will be discussed in each application of the theory.}
\]
I interpret $U_\theta$ as the current and the continuation payoff that the agent would have obtained had she suspended her belief $\lambda$ for until tomorrow, replacing it with the belief that observable $y$ evolved according to the model $\theta$ today. Thus, the counterfactual value represents a decision problem in which today’s transition is assumed by the agent to be properly modelled by $\theta \in \Theta$, but thereafter the belief over the models returns to be $\lambda$. Notice that Bellman’s equation (5) may be rewritten as

$$V(x_-, y, \lambda) = \max_{x \in \Gamma(x_-, y)} \int_\Theta U_\theta(x_-, x, y_+, \lambda) \lambda(d\theta).$$  \hfill (8)

### 2.3 Regret

I postulate that a learning macroeconomic agent would wish to minimise losses incurred by her value due to model uncertainties. In this Section I will make this statement exact. An agent holding a belief $\lambda$ over models can consider counterfactuals. At any time $t$, the agent can ask herself “what if I have believed model $\theta$ today and then returned to my belief $\lambda$ for all eternity tomorrow?”. This is certainly not the only counterfactual an agent can consider, but it is a counterfactual whose implications the agent can actually assess given her information. A model that has better predictive capabilities than other models would tend to yield a higher counterfactual value more often than other models. Thus, the agent can assess her total loss in payoff terms compared to an optimal in-hindsight model at time $t$ - a model\(^6\) that by time period $t$ gives the agent the highest totality of counterfactual values. As the agent does not know the exact environment, she may not behave optimally, and thus suffer regret compared to the optimal in-hindsight model. I shall refer to the gap between the agent’s payoff and the counterfactual payoff she might have had as counterfactual dynamic regret. I will then consider a learning game between the agent and Nature that will show that the agent may update her beliefs so the regret becomes asymptotically small w.r.t. the number of time periods, which will be interpreted as the agent’s ability to adopt a belief that will bring his payoff near the one that can be obtained by the optimal in-hindsight model.

---

\(^6\)There is a non-empty set of such models from which a representative may be chosen.
2.3.1 Counterfactual Dynamic Regret

Given the agent’s beliefs $\lambda_1, \ldots, \lambda_s$, the observables $y_1, \ldots, y_t$, and her choices $x_s^* \in x(x_{s-1}^*, y_s, \lambda_s)$, I define the counterfactual dynamic regret $R^t_C$ at time $t$ to be the following difference

$$R^t_C = \sup_{\theta^* \in \Theta} \sum_{s=1}^{t} \max_{x \in \Gamma(x_{s-1}^*, y_s)} U_{\theta^*}(x_{s-1}^*, x, y_s, \lambda_s) - \sum_{s=1}^{t} \max_{x \in \Gamma(x_{s-1}^*, y_s)} \int_{x_{s-1}}^{x_s} U_{\theta}(x_{s-1}^*, x, y_s, \lambda_s) \lambda_s(d\theta)$$

$$= \sup_{\theta^* \in \Theta} \sum_{s=1}^{t} \max_{x \in \Gamma(x_{s-1}^*, y_s)} U_{\theta^*}(x_{s-1}^*, x, y_s, \lambda_s) - \sum_{s=1}^{t} V(x_s^*, y_s, \lambda_s)$$

The counterfactual dynamic regret can be interpreted as follows. Consider an agent with belief $\lambda_s$ at time $s$. The agent would like to know which model gives the best predictions. In order to do that she considers the change in the value of her dynamic optimisation problem had she believed model $\theta$ to hold at time $s$ and then returned to her belief $\lambda_s$ from time $s + 1$ onwards. To examine this more closely, I notice that at time $s$, the agent has already made the decision $x_{s-1}^*$, hence she faces a decision set $\Gamma(x_{s-1}^*, \cdot)$ that cannot be changed, and her choice variable at time $s$ would be chosen from this set regardless of her beliefs. Thus, the agent now has a benchmark for comparing the performance of model $\theta$ with that of her belief $\lambda_s$ as I have described. If the counterfactual dynamic value produced by model $\theta$ w.r.t. $\lambda_s$ at time $s$ surpasses the one produced by holding to the belief $\lambda_s$ then model $\theta$ is advantageous to the agent’s belief. At any time $t$, the most advantageous-in-hindsight model would be the one that has produced the highest total gap between its counterfactual values and the agent’s values due to her beliefs. The agent would thus strive to minimise this gap.

2.3.2 Forward and Backward Looking Regret

I have defined the counterfactual dynamic regret using the agent’s values and counterfactual values. Thus it may seem that the counterfactual regret is an inherently forward-looking object. In this Subsection I will show that the recursive structure of the optimisation problem makes it also backward-looking.

I consider the correspondence of counterfactually-optimal choice for model $\theta$ given the belief $\lambda$, which is given by

$$x(x_, y, \lambda|\theta) = \arg \max_{x \in \Gamma(x_, y, \lambda)} U_{\theta}(x, y, \lambda).$$

(11)
For sequences of optimal choice variables \( x^*_s(\lambda_s) \in \mathbf{x}(x_-, y, \lambda_s) \) and counterfactually-optimal choice variables \( x^*_s(\lambda_s|\theta) \in \mathbf{x}(x_-, y, \lambda_s|\theta) \), with \( 1 \leq s \leq t \), I can use Bellman’s Equation (5) to rewrite the average counterfactual regret at time \( t \) as

\[
R^t = \sup_{\theta^* \in \Theta} \sum_{s=1}^{t} \left[ \left( u(x^*_{s-1}(\lambda_{s-1}), x^*_{s}(\lambda_s|\theta^*), y_s) - u(x^*_{s-1}(\lambda_{s-1}), x^*_{s}(\lambda_s), y_s) \right) \right]
+ \beta \left. \left( y_{s+1} \sim Q_{\theta^*}(\cdot|y_s) \right) \left[ V(x^*_{s}(\lambda_s|\theta^*), y_{s+1}, \lambda_s) \right] - \int_{T} E_{y_{s+1} \sim Q_{\theta}(\cdot|y_s)} \left[ V(x^*_{s}(\lambda_s), y_{s+1}, \lambda_s) \right] \lambda_s(d\theta) \right] .
\]

This representation brings forward an interesting interpretation of the counterfactual regret. For simplicity, suppose that the supremum was indeed attained at some \( \theta^* \in \Theta \). If it were the case, then expression (1), can be viewed as the gap at time \( s \) between the utility produced by the counterfactually optimal choice for the model \( \theta^* \) and the utility of the optimal choice under the belief \( \lambda_s \). Expression (2), can be similarly viewed as the difference in the continuation payoffs had the agent chosen a counterfactually-optimal choice instead of the choice prescribed by her beliefs. Thus, the counterfactual regret can be viewed as the loss in-hindsight in both current and continuation payoff had the agent behaved counterfactually-optimal. Thus, minimising regret inherently contains both forward-looking and backward-looking components.

In this paper I take the approach that the agent wishes to make her choice as robust as possible to modelling failure. To achieve a high level of robustness, I will try to obtain robustness even against the possibility that Nature ‘chooses’ the sequence of observables in a manner that would lead to the highest counterfactual dynamic regret for the agent. More precisely, if the agent employs policy \( \hat{x}^*_{s} \) at time \( s \), and \( x^*_s = \hat{x}^*_{s}(x^*_{s-1}, y_s) \), then Nature would choose the observables s.t. the dynamic counterfactual regret \( R^t_C \) is maximised. This idea can be rephrased in terms of a game. The agent and Nature are involved in a zero-sum game in which the agent pays Nature \( R^t_C \) at time \( t \). Nature chooses observables \( y_1, ..., y_t, ... \) and the agent chooses policies \( \hat{x}^*_{1}, ..., \hat{x}^*_{t}, ... \) and beliefs \( \lambda_1, ..., \lambda_t, ... \) I would therefore like to consider the long run behaviour of

\[
\inf_{x^*_1, ..., x^*_t} \sup_{\lambda_1, ..., \lambda_t} R^t_C .
\]

9
2.3.3 The Auxiliary Learning Game

I will be interested in showing that the agent can behave in a manner that diminishes the average per-stage counterfactual dynamic regret \( \theta \) regardless of Nature’s choices. More precisely, I would like to show that the expression in Line (13) is asymptotically sublinear\(^7\) in the time \( t \). The setup considered in the previous subsection states the problem but may leave some readers who are familiar with either the reality of dynamic programming or that of regret minimisation a bit puzzled. Dynamic programming problems may be quite complex in their own right, and making their solutions robust to such adversarial behaviour from the side of Nature may result in an even much more complex beast. Any attempt to solve the problem should now not only consider a good choice of states but also a severe robustness requirement for the agent’s policies. As far as regret minimisation is concerned the dynamical features of the problem distinguishes it from other works in the field due to its dynamic structure in which agents actions are policies whose choice depends on future expectations. Furthermore, the beliefs are formed over a set \( \Theta \) that may fail to be finite in many applications.

I will now consider a somewhat different setting that will resolve the robustness problem and will also make the problem closer in its nature to the framework of decisions with experts’ advice that was studied by Freund and Schapire [12] (see, as well, Cesa-Bianchi and Lugosi [8]). In essence, I will resolve the aforementioned issues by moving the choice of correspondences \( \hat{x}^s \) from the agent to Nature. As Nature also chooses the observable \( y \) it knows the value of \( x^* = \hat{x}^s(x^*, y) \), which will lead to a great simplification of the problem.

I will now describe the game’s protocol. To do so, I define the agent’s gain w.r.t. previous choice \( x_- \), current choice \( x \), observable \( y \), and belief \( \lambda \) to be

\[
G = \int_\Theta U_\theta(x_-, x, y, \lambda) \lambda(d\theta), \tag{14}
\]

and the gain attributed to the model \( \theta \in \Theta \) to be

\[
G_\theta = U_\theta(x_-, x, y, \lambda). \tag{15}
\]

In each time period \( s = 1, 2, \ldots \)

(1) The agent chooses a belief \( \lambda_s \);

\( ^7\)Namely, divided by \( t \) it diminishes to 0 as \( t \to \infty \).
(2) Nature chooses a choice variable and an observable \((x_s^*, y_s) \in C_s = \{(x, y) : y \in Y, x \in x(x_s^{*-1}, y, \lambda_s)\}\). I assume Nature doesn’t know the structure of the correspondence \(x\), and only knows her choice set \(C_s\) and the payoff functions \(G_s, G_{\theta,s} : C_s \to \mathbb{R}\), given by \(G_s(x, y) = G(x_{s-1}^*, x, y, \lambda_s)\), and \(G_{\theta,s}(x, y) = G_\theta(x_{s-1}^*, x, y, \lambda_s)\);

(3) The choices \(x_s^*, y_s, \lambda_s\) are revealed. The agent gains \(G_s\), and views every model as generating a gain \(G_{\theta,s}\).

The agent’s goal is to keep as small as possible the cumulative regret:

\[
R^t_L = \sup_{\theta^* \in \Theta} \sum_{s=1}^{t} G_{\theta^*,s} - \sum_{s=1}^{t} G_s
\]

regardless of the choice made by Nature. In other words, she wishes to minimise the regret even in the face of Nature adversely choosing the sequence of choice variables and observables. I interpret this assumption as the agent seeking her decision to be as robust as possible, insuring her against possible failures of her entire family of models. This idea has some connections to Hansen and Sargent’s agenda on robustness (see [17] for a thorough introduction). The main difference between the frameworks is that underlying Hansen and Sargent’s framework are the assumptions that the sequences of observables can be modelled by as an asymptotically mean-stationary process, that Nature chooses the worse model in a given set of such models, and that the set of models is known to the agent. My approach allows the agent to assume that none of her models may captures the observables dynamics. As I shall subsequently show, these seemingly small differences lead to a completely different behaviour for the agent.

**Remark 1.** As in the case of the counterfactual regret, I will represent the learning game’s regret in a way that will result in an interesting interpretation. Substituting Equation (7) into Equation (16) gives

\[
R^t_L = \beta \sup_{\theta^* \in \Theta} \sum_{s=1}^{t} \left( E_{y_s \sim Q_{\theta^*}(\cdot|y_s)}[V(x_s^*, y_s, \lambda_s)] - \int_{\Theta} E_{y_s \sim Q_{\theta}(\cdot|y_s)}[V(x_s^*, y_s, \lambda_s)] \lambda_s(d\theta) \right).
\]

Assume, for simplicity, that the supremum is attained by some model \(\theta^*\). Then the regret in the learning game is proportional to the totality of the gap between the expected continuation payoff had the observables followed the model \(\theta\), and the expected continuation payoff had the observables followed the agent’s beliefs\(^8\). Thus, the learning game is, in

\(^8\)Recall that here, the agent has no influence on the choice of \(x_s^*\).
essence, a game in which the agent learn the “right expectations” regardless of the choice variables $x_1^*, \ldots, x_t^*, \ldots$. The need to make the choices $x^*$ robust is made moot as Nature has been made ‘responsible’ of it, and the matter of robustness has been delegated to the choice of beliefs.

2.4 Technical Assumptions

I will now state some weak technical assumptions that are maintained throughout the paper. More substantive assumptions are explicitly stated in the paper in tandem to the results that use them.

The sets $X$ and $Y$ are nonempty subsets of metric spaces with the corresponding Borel sigma-algebras denoted by $\mathcal{X}$, $\mathcal{Y}$. The function $u : X \times X \times Y \to \mathbb{R}$ and the correspondence $\Gamma : X \times Y \Rightarrow X$ are measurable. The set $\Theta$ is a nonempty subset of a metric space, and it has a compact closure. The set $\Delta(\Theta)$ of probability distributions over $\Theta$ is endowed with the topology of weak convergence and the corresponding Borel sigma-algebra $\mathcal{B}(\Delta(\Theta))$. For any $\theta \in \Theta$, the function $Q_\theta : \mathcal{Y} \times Y \to [0, 1]$ is a transition probability from $Y$ to itself.

When it comes to the agent’s beliefs, I will assume that there is an underlying prior $\xi \in \Delta(\Theta)$ and that each belief $\lambda$ has a density w.r.t. $\xi$. The family of functions

$$\{\theta \to U_\theta(x, y, x, y, \lambda) : y \in Y, x, y \in X, x \in \Gamma(x, y), \lambda \in L^1(\xi)\}$$

is assumed to be equicontinuous with uniform modulus of continuity $w$, which will be assumed to be increasing and differentiable on $(0, \infty)$. I further impose that the modulus of continuity satisfies

$$w\left(e^{-o(\sqrt{t})}\right) = O\left(t^{-1/2}\right)$$

as $t \to \infty$.

I will always assume that the value function is a solution of the Bellman equation (5) and vice versa - that the Equation has a unique solution which is the value function. These are not a trivial assumptions, and they require verification with every example. Nevertheless, the focus of my theory is on no-regret learning in examples where the recursive

\textsuperscript{9}Namely, that there is a $w \in L^1(\xi)$ s.t. $\lambda(B) = \int_B w(\theta) d\xi(\theta)$.

\textsuperscript{10}There is no lose of generality here.

\textsuperscript{11}Note that this assumption is satisfied by any Holder-continuous function, for example, thus I shall be able to consider very general statistical models.
framework in macroeconomics is known to hold, hence such an assumption is reasonable in order to develop a general theory.

Finally, I shall make some assumptions on the “complexity” of the set of models $\Theta$. An $\epsilon$-cover of $\Theta$ is a cover of $\Theta$ by open balls of radius $\epsilon$. The $\epsilon$-covering number $N(\epsilon)$ of $\Theta$ is the infimum over the size of $\epsilon$-covers. My technical assumption that the closure of $\Theta$ is compact ensures that the $\epsilon$-covering number is finite. The metric entropy of $\Theta$ is the function $h(\epsilon) = \ln N(\epsilon)$. The set of models is of bounded metric dimension if there is some $C > 0$ s.t.

$$h(\epsilon) \leq -C \ln \epsilon.$$  

as $\epsilon \to 0$. I will only consider sets of models with bounded metric dimension.

**Example 1.** Every compact subset $\Theta \subseteq \mathbb{R}^d$ has bounded topological dimension. If $B'$ is a ball containing $\Theta$ and $B$ denotes the Euclidean unit ball then $N(\epsilon) \leq \frac{\text{vol}(B')}{\text{vol}(B)} \left( \frac{1}{\epsilon} \right)^d$. Thus, my framework can capture almost any modelling framework appearing in macroeconomic applications.

## 3 Expectations with No Regrets

I am now ready to present my results on the counterfactual dynamic regret. My first result assures that the agent can update her beliefs so that her counterfactual dynamic regret is asymptotically sublinear in the number of periods $t$ regardless of Nature’s choices:

**Theorem 1.** The agent can choose her optimal policies and update her beliefs so she may assure a counterfactual dynamic regret that is asymptotically sublinear in the number of time periods, and this may be guarantied to hold uniformly w.r.t. Nature’s choice of the sequence of observables. In other words, the agent can choose policies $\hat{x}^*_t(\cdot), \ldots, \hat{x}^*_t(\cdot), \ldots$, and beliefs $\lambda_1, \ldots, \lambda_t, \ldots$ s.t.

$$\sup_{y_1, \ldots, y_t} R^t_C = o(t)$$

as $t \to \infty$.

The dynamic counterfactual regret has close relationship with the cumulative regret in the learning game despite the clear differences in their definitions. This relationship is captured by the following result:
Proposition 1. For every choice $\lambda_1, \ldots, \lambda_t$ of the agents belief, and $\hat{x}_1^*, \ldots, \hat{x}_t^*$ of optimal policies

$$
\sup_{y_1, \ldots, y_t} R_C^t \leq \sup_{x_s^* \in \Gamma(x_{s-1}^*, y_s), s=1, \ldots, t} R_L^t
$$

Proposition 1 makes it clear that Theorem 1 follows as a consequence of the following Theorem:

Theorem 2. In the auxiliary learning game the agent can update her beliefs so she may assure a cumulative regret that is asymptotically sublinear in the number of time periods, and this may be guarantied to hold uniformly w.r.t. Nature’s choice of the sequence of observables and choice variables. In other words the agent may choose her beliefs $\lambda_1, \ldots \lambda_t, \ldots$ s.t.

$$
\sup_{y_1, \ldots, y_t} \frac{R_L^t}{x_s^* \in \Gamma(x_{s-1}^*, y_s)} = o(t)
$$

as $t \to \infty$.

It is worthwhile to remark that Theorem 2 not only contributes to the current macroeconomic framework but also to the literature on no-regret strategies, as the set of ‘experts’ here may be a continuum. The results will be proved in Appendix A. I will in fact prove the following much stronger bound for the cumulative regret:

Theorem 3. Under the assumptions made in the technical assumptions Section 2.4, there exist positive constants $C_1(\Theta)$, $C_2(u)$, $C_3(u)$ s.t.

$$
R_L^t \leq \sqrt{t} \left( -C_1(\Theta) \ln \left( w^{-1} \left( O \left( t^{-1/2} \right) \right) \right) + 1 + 6 \log_2 t + C_2(u) \right) + C_3(u) \left( \frac{\log_2 t}{\sqrt{t}} \right)^3.
$$

Theorem 3 may seem quite technical but it is not at all so. It actually unravels the relationship between the regret and the involved economic components of the model. It is thus of interest to discuss the given bound as it uncovers several insights regarding the factors involved in controlling the levels of regret. I will now follow how each component emerges, in a non-technical review of the proof. The interested reader is referred to the proof in Appendix A.
Cost of Complexity (COC)

Component (COC) accounts for the “cost of complexity” in the regret bound. This component has two factors, each corresponding for different source of complexity. The first factor, $C_1(\Theta)$, relates to the ‘size’ of the set of models $\Theta$. It is essentially the topological dimension\(^{12}\) measuring the size of a finite subset $A$ of $\Theta$ s.t. every model in $\Theta$ is sufficiently close to one of the models in $A$. The first step in the proof is to approximate the set of models $\Theta$ with a growing sequence of such approximating sets $(A_t)_{t \geq 1}$. Every model in $\Theta$ is associated with a nearby “approximated model” in $A_t$ that guaranties a required approximation in terms of payoff. This leads to the second type of complexity which is the model complexity, measured by the modulus of continuity $w$. The modulus of continuity $w$ maps the approximation made in terms of differences between models into differences in terms of payoffs.

As my results rely on the uncertainty that models capture the economy’s complexity, one might be tempted to mitigate that by considering a richer set of models, or include models that have higher level of complexity. Nevertheless, Theorem 3 implies that this will result in either increasing $C(\Theta)$ or the modulus of continuity $w$, thus resulting in higher counterfactual dynamic regret.

Cost of Information (COI)

Component (COI) accounts for the “cost of information” in the regret bound. This is the cost of carrying forward the information contained in the prior beliefs following a change in the approximating set $A_t$. A naive approach to the problem would have been to consider time blocks of increasing length in which the problem is initialised at the beginning of every time block, namely, a uniform prior is taken over the new set of approximating models. This would not have mattered much\(^{13}\) for the worst case regret, but if the underlying dynamics was such that enables learning, I would have resulted in the loss of all the information already encrypted in the prior belief.

Cost of Fluctuations (COF)

Component (COF) accounts for the “cost of fluctuations” in the regret bound. Models may have different payoff distributions around the mean. If the agent doesn’t wish to overshoot with their belief updating then they should account for the possibility that some models are less or more advantageous than they may seem. This term accounts for that. It arrives from

\(^{12}\)See the technical assumptions Section 2.4 for details

\(^{13}\)The regret bound I obtain here for the worst case is still better but not by much.
second order contributions of payoff to the regret, which may be thought of as fluctuations around the mean payoff regret. What is shown here is that with the right choice of belief update, the cost of such fluctuations becomes insignificant in the long run compared to the other factors I have mentioned.

4 Permanent Income with No Regrets

I will now consider an application of the theory in studying consumption in a permanent income model when consumers are regret minimisers. My main focus here would be to understand how a regret minimising framework fairs w.r.t. the benchmark rational expectations model presented in Hall [15]. I will thus concentrate on a simple quadratic utility case. In such a case, full information rational expectations (FIRE) predicts that consumption is a martingale. This type of predictions\textsuperscript{14} have been the cornerstone of the empirical literature on such matters (see Attanasio and Low [1], and Attanasio and Weber [2] for extensive reviews). I would thus wish to examine this prediction and the ways it may change due to no-regret learning.

My findings here are simple - for such a prediction to hold it is necessary that under no-regret learning beliefs resemble FIRE outcomes in the long run in the sense that the gap between the expected permanent income today and tomorrow diminishes to zero. More generally, the difference between the expected consumption tomorrow and the consumption today is fully characterised by the rate in which this gap diminishes to zero. I will characterise cases in which 1. the gap diminishes to 0, and 2. the convergence is sufficiently fast, so that consumption attains a central limit theorem, which salvages some of the structure of the classical FIRE results and gives a falsifiable empirical test.

4.1 The Framework

The agent wishes to maximise the expected discounted sum of distances between her actual consumption $c$ and her preferred consumption $\bar{c}$:

$$\max E_t \left[ -\sum_{s=t}^{\infty} \beta^{s-t} (c_s - \bar{c})^2 \right].$$  

\textsuperscript{14}And more generally, certain properties of Euler equations.
I assume she does so by choosing, at each time \( t \), a borrowing plan \( b_{t+1} \) and has an observable exogenous income process \( y_t \), s.t. the following budget constraint is satisfied

\[
c_t + b_t = \frac{1}{1 + r} b_{t+1} + y_t,
\]

where \( r > 0 \) is a time-invariant risk-free net interest rate constant, and the variable \( b_t \) may be interpreted as one period risk-free bond maturing at period \( t \). To maintain simplicity, I adopt Hall’s [15] convention of assuming \( \beta = \frac{1}{1+r} \).

I now note that even before special modelling considerations are made, several general observations can be given. My first observation is that due to the budget constraint (21), the Euler equation is

\[
c_s = E_t^t[c_{s+1}]
\]

for \( s \geq t \), where \( E_t^s \) is the expectation operator resulting from the belief \( \lambda_t \) in time period \( s \). The operator \( E_t^t \) thus coincides with \( E_t \).

**Remark 2.** Equation (22) does not imply that the actual consumption \( c_t \) is a martingale. Indeed, the consumptions \( c_s \) and \( c_{s+1} \) in Equation (22) are those derived under the belief \( \lambda_t \), namely they are \( c_s(\lambda_t) \) and \( c_{s+1}(\lambda_t) \). Nevertheless, the actual consumption in period \( t \) is \( c_t(\lambda_t) \), hence in period \( t + 1 \) it is \( c_{t+1}(\lambda_{t+1}) \) and not \( c_{t+1}(\lambda_t) \). The consumption \( c_{t+1}(\lambda_t) \) is thus a “virtual consumption”, which might have coincided with the actual one had the belief in period \( t + 1 \) remained \( \lambda_t \). As I shall subsequently show in a specific example, the actual consumption would typically not be a martingale.

My second observation is that Equation (21) can be solved forward to give

\[
b_t = \sum_{j=0}^{\infty} \beta^j(E_t[y_{t+j}] - E_t[c_{t+j}]). \tag{23}
\]

I will now move to describe some specialised cases of this model and their solutions.

### 4.2 Income model uncertainty

I now turn to consider model income uncertainty. To simplify some of the analysis, I assume that the income process follows\(^{15}\)

\(^{15}\)I can entertain other processes as well. My point in this discussion is, nevertheless, to simplify the analysis as much as possible.
\begin{align*}
y_t &= U u_t, \quad (24) \\
u_{t+1} &= A u_t + C \epsilon_{t+1}. \quad (25)
\end{align*}

I will assume, to maintain simplicity, that the sequence of vectors \((u_t)_{t \geq 1}\) is unobservable\(^{16}\). The matrices \(A\) and \(C\) are unknown, \(U\) is an unknown selection vector, and \((\epsilon_t)\) is a standard bounded\(^{17}\) i.i.d. noise process, namely, \(\epsilon_t\) is bounded \(E \epsilon_t = 0\) and \(E[\epsilon_t^2] = 1\). In other words, the agent entertains a finite dimensional hidden-factor model. I shall make the standard assumption that the spectral radius of \(A\) is smaller than \(\beta^{-1/2}\). Nevertheless, in my example the agent’s belief \(\lambda_t\) at time \(t\) is defined over models \(\theta = (u, U, A, C)\) where \(u\) denotes the value of \(u_t\). To maintain the compactness of the model space, I will assume \(||u|| \leq M_0\) for some large \(M_0 > 0\), and I will further assume that \(||A|| + ||C|| ||\epsilon|| / M_0 \leq 1\) for any choice of \(A, C\), and every possible value of \(\epsilon\). Moreover, I will also assume that \(||U||\) is bounded. Thus, \(y_t\) takes values in a compact set which implies that the respective Bellman equation admits a unique solution, and the theory I have developed can thus be employed.

As Euler’s Equation (22) states \(E_t[c_{s+1}] = c_s\), then combined with (23) it gives the standard relation

\[c_t = (1 - \beta) \left( \sum_{j=0}^{\infty} E_t[y_{t+j}] - b_t \right). \quad (26)\]

I turn to evaluate \(E_t[y_{t+j}]\). I notice \(E_t[X] = E_{\lambda_t}[E_{\theta^t}[X|\theta^t]]\). By the assumption of anticipated utility, the operator \(E_t\) doesn’t include virtual updating. Hence I obtain

\[E_t[y_{t+j}] = E_{\lambda_t} E_{\theta^t} \left[ E_{\epsilon_{t+j}} [U A u_{t+j-1} + U C \epsilon_{t+j-1}] \right] = E_{\lambda_t} E_{\theta^t} [U A u_{t+j-1} \theta^t] \]

\[= E_{\lambda_t} E_{\theta^t} [U A^2 u_{t+j-2} \theta^t] = \cdots = E_{\lambda_t} [U A^j u], \]

and therefore

\[^{16}\text{Discussing more general example, though interesting, may create unnecessary complications, and may result in obscuring the role of learning in such a model.}\]

\[^{17}\text{I am making these assumption in order to meet the assumptions of Theorem 1. For that, I will need the model space } \Theta \text{ to be compact. In application, there is no loss of generality here, as I may consider values of } u, A, C, U \text{ and } \epsilon \text{ leading to income } y \text{ that is far greater than any real income.}\]
\[ c_t = (1 - \beta) \left( E_{\lambda_t} \left[ u \right] \right) - b_t \]

\[ = (1 - \beta) \left( E_{\lambda_t} \left[ U(I - \beta A)^{-1} u \right] - b_t \right) \]

\[ = (1 - \beta) \left( E_{\lambda_t} \left[ U(I - \beta A)^{-1} u \right] - b_t \right) \]

Note that if the belief update is performed using Algorithm 1, then the operator \( \lambda_t \) depends on past states, namely, it depends on realised past income and hence realised past consumption inter alia. As I have remarked in Remark 2, the actual consumption may fail to be a martingale w.r.t. the expectation at time \( t \). As a preliminary stage in making this statement exact, I compute the difference in actual consumption

\[ c_{t+1}(\lambda_{t+1}) - c_t(\lambda_t) \]

\[ = (1 - \beta) \left( E_{\lambda_{t+1}} \left[ U(I - \beta A)^{-1} u \right] - E_{\lambda_t} \left[ U(I - \beta A)^{-1} u \right] - (b_{t+1} - b_t) \right). \]  

(28)

To ease the algebra, I write \( \beta = 1 + r \) and obtain from the budget constraint

\[ b_{t+1} \]

\[ = (1 + r)b_t + r \left( E_{\lambda_t} \left[ U(I - \beta A)^{-1} u \right] - b_t \right) - (1 + r)E_{\lambda_t}[U u] \]

\[ = b_t + E_{\lambda_t} \left[ U(r(I - \beta A)^{-1} - (1 + r)I) u \right] \]

\[ = b_t + E_{\lambda_t} \left[ U(I - \beta A)^{-1} (A - I) u \right], \]  

(29)

where the equality in Line (29) follows as\(^{19} \) \( r(I - \beta A)^{-1} - (1 + r)I = (I - \beta A)^{-1}(A - I) \). Substituting Equation (29) into Equation (28), I obtain

\[ c_{t+1}(\lambda_{t+1}) - c_t(\lambda_t) \]

\[ = (1 - \beta) \left( E_{\lambda_{t+1}} \left[ U(I - \beta A)^{-1} u \right] - E_{\lambda_t} \left[ U(I - \beta A)^{-1} u \right] \right). \]  

(30)

\(^{18}\)Namely, the sequence of consumptions of the form \( c_t(\lambda_t) \) - the consumption at time \( t \) that is prescribed by the belief \( \lambda_t \).

\(^{19}\)Indeed, this is true iff \( rI - (1 + r)(I - \beta A) = A - I \). Rearranging gives that this is equivalent to \( A = (1 + r)\beta A \), which holds as \( (1 + r)\beta = 1 \).
4.3 The Random Walk Hypothesis

The relationship described in Equation (30) may contribute the understanding of the implication of no-regret learning on the “random walk hypothesis” (RWH\textsuperscript{20}).

To explore that a little bit further, Suppose that income would have followed an ‘objective’ random process \( (Y_t) \) that is unknown to the agent, namely, the observed income \( y_t \) at time period \( t \) is a realisation of the random variable \( Y_t \). Denote by \( E_t \) the time \( t \) expectation of the process. In that case, an econometrician knowing the agent’s consumption at time \( t \) would be facing a random belief \( \Lambda_t \) at time \( t \), where \( \Lambda_{t+1} \) is the random belief obtained from \( \Lambda_t \) using the Hedge with Procreating Experts updating Algorithm 1 using the income generated from the random variable \( Y_{t+1} \). The econometrician would thus set to estimate equation\textsuperscript{21}

\[
\mathcal{E}_t [c_{t+1}(\Lambda_{t+1})] - c_t(\Lambda_t) = (1 - \beta) \mathcal{E}_t \left( E_{\Lambda_{t+1}} \left[ U(I - \beta A)^{-1}u \right] - E_{\Lambda_t} \left[ U(I - \beta A)^{-1}Au \right] \right). \tag{31}
\]

Suppose now that the agent is attempting to minimise regret and that she applies Algorithm 1 in the process. Recall that if \( j \geq 0 \), then at time \( 2^j \leq t < 2^{j+1} - 1 \) Algorithm 1 uses a partition \( \Pi_j \) to give a finite approximation of the space \( \Theta \) represented by models \( \theta_j(a) \in a \) for each atom \( a \in \Pi_j \). By Proposition 5 in Appendix B for \( 2^j \leq t < 2^{j+1} - 1 \) there is a bounded measurable function \( G_t : \Delta(\Theta) \times [0, 1] \times \Theta \) s.t. if \( W_t \) is the random\textsuperscript{22} density of \( \Lambda_t \) then the difference between the densities \( W_{t+1} \) and \( W_t \) has the following form

\[
W_{t+1}(\theta) - W_t(\theta) = \eta_j W_t(\theta) G(\Lambda_t, \eta_j, \theta). \tag{32}
\]

Substituting that back into Equation (31) I obtain\textsuperscript{23}

\[
\mathcal{E}_t [c_{t+1}(\Lambda_{t+1})] - c_t(\Lambda_t) = (1 - \beta) \mathcal{E}_t \left( E_{\Lambda_{t+1}} \left[ U(I - \beta A)^{-1}(I - A)u \right] \right)
+ \eta_j (1 - \beta) \int_\Theta \mathcal{E}_t \left[ U(I - \beta A)^{-1}uW_t(\theta)G_t(\Lambda_t, \eta_j, \theta) \right] d\xi(\theta). \tag{33}
\]

\textsuperscript{20}The RWH states that in the case of linear marginal utility that I am studying here, \( E_t[c_{t+1}] = c_t \). A more general theory states that \( E_t[u'(c_{t+1})] = u'(c_t) \), so the only information required to know the expected marginal utility of consumption tomorrow is the marginal utility today. I will not study the implication in such a general case here.

\textsuperscript{21}The left hand side is of the form \( E_t[c_{t+1}] - c_t \) which under FIRE equals 0. The left hand side in a residual that is due to the learning process.

\textsuperscript{22}Namely the density of \( \Lambda_t(Y_t) \).

\textsuperscript{23}Recall that \( \theta = (u, U, A, C) \).
In order for this difference to be a martingale it must hold that the expectation w.r.t. $\mathcal{E}_t$ equals zero, namely
\begin{align}
(1 - \beta)E_{\Lambda_t} \left[ U(I - \beta A)^{-1}(I - A)u \right] = -\eta_j (1 - \beta) \int_{\Omega} \mathcal{E}_t \left[ U(I - \beta A)^{-1}uW_t(\theta)G_t(\Lambda_t, \eta_j, \theta) \right] d\xi(\theta).
\end{align}

I denote the sequence of values of $t$ excluding the values $t_{j+1} - 1$ for $j \geq 0$ by $\mathcal{J}$. Taking the limit over this sequence I also have $\eta_j \to 0$ hence, as the expectation on the right hand side is bounded, I have shown that

**Proposition 2.** For the sequence $(c_t(\Lambda_t))_{t \in \mathcal{J}}$ to be a martingale it is necessary that
\begin{equation}
\lim_{t \in \mathcal{J}, t \to \infty} E_{\Lambda_t} \left[ U(I - \beta A)^{-1}(I - A)u \right] = 0 \ a.s.
\end{equation}

Proposition 2 considers the consumption but excludes its values in times in which the learning rate $\eta$ changes. The number of learning rate changes up to time $t$ behave asymptotically as $\log t$, thus the share of time points to be excluded from the sample is small.

Suppose that instead of wishing the sequence $(\mathcal{E}_t[c_{t+1}(\Lambda_{t+1})] - c_t(\Lambda_t))_{t \in \mathcal{J}}$ to be a martingale, I would have only wished it to diminish to 0 in the long run. Then the same necessary condition must hold. On the other hand, if it doesn’t hold, then $\mathcal{E}_t[c_{t+1}(\Lambda_{t+1})] - c_t(\Lambda_t)$ doesn’t diminish to 0. It follows that

**Proposition 3.** The sequence $(\mathcal{E}_t[c_{t+1}(\Lambda_{t+1})] - c_t(\Lambda_t))_{t \in \mathcal{J}}$ diminishes to 0 as $t \to \infty$ a.s., iff
\begin{equation}
\lim_{t \in \mathcal{J}, t \to \infty} E_{\Lambda_t} \left[ U(I - \beta A)^{-1}(I - A)u \right] = 0 \ a.s.
\end{equation}

Proposition 3 essentially states that for consumption to come anywhere near a “random walk” property, a very demanding condition needs to be satisfied. In the following Subsections I will discuss this condition in greater detail.

**4.4 Income Processes Leading to Learning**

I say that the income process $(Y_t)_{t \geq 0}$ leads to learning if and only if
\begin{equation}
\lim_{t \in \mathcal{J}, t \to \infty} E_{\Lambda_t} \left[ U(I - \beta A)^{-1}(I - A)u \right] = 0
\end{equation}
almost-surely.
Remark 3. To understand the idea behind this condition, I recall that
\[
\sum_{j=0}^{\infty} \beta^j E_t[y_{t+j}] = E_{t_1} \left[ U(I - \beta A)^{-1}u \right] 
\]
which is the agent’s expected permanent income at time \( t \). Furthermore, a similar computations lead to
\[
\sum_{j=0}^{\infty} \beta^j E_t[y_{t+j+1}] = E_{t_1} \left[ U(I - \beta A)^{-1}Au \right] 
\]
which is the agent’s time \( t \) expectation of her permanent income from time \( t + 1 \) onwards. Thus the expression in Equation (35) is the expected difference, computed at time \( t \), between today’s and tomorrow’s permanent incomes. Thus, the condition that the process \( (Y_t)_{t \geq 0} \) leads to learning means that, excluding times in which the agent’s learning rate changes, this expected difference diminishes to 0 a.s.

In light of Remark 3, Proposition 3 can now be given an interesting interpretation. It pins down the gap between the results of Hall’s [15] classical theory and my theory to the question of whether or not the process \( (Y_t) \) leads to learning, namely - if the expected gap between today’s and tomorrow’s permanent income diminishes to zero. In such a case, writing \( \Delta c_s = c_s(\Lambda_s) - c_{s-1}(\Lambda_s) \), \( X_s = \Delta c_s - E_{s-1}[\Delta c_s] \), \( M_t = \sum_{s=1}^{t} X_s \), and \( R_t = \sum_{s=1}^{t} E_{s-1}[\Delta c_s] \) implies that \( M_t \) is a martingale and that
\[
c_t(\Lambda_t) - c_0 = M_t + R_t,
\]
which gives a decomposition of the consumption difference \( c_t(\Lambda_t) - c_0 \) into a martingale \( M_t \) and a remainder term \( R_t \). Thus, if the process \( Y_t \) leads to learning and the remainder is “sufficiently small” compared to \( t \) then the consumption is “almost” a martingale. I will now make this statement exact, and study its empirical and theoretical implications.

I say that the process \( Y_t \) leads to tempered learning iff it leads to learning and \( E_{t-1}[\Delta c_t] = o \left( t^{-1/2} \right) \) for every \( t \in \mathcal{J} \). Thus tempered learning essentially\(^{24}\) means that the expected gap in today’s and tomorrow’s permanent incomes diminishes to 0 as well as that the ex-

\(^{24}\)I use the word essentially here, as their may be exceptions for that, but they occur very rarely - only for \( t = t_j - 1 \).
pected gaps between today’s and tomorrow’s consumption diminish faster to 0 than \( \frac{1}{\sqrt{t}} \).

Under such a condition it is clear\(^{25}\) that \(|R_t| = o(\sqrt{t})\) as \( t \to \infty \).

A decomposition of a process into the form \( M_t + R_t \) s.t. \( M_t \) is a martingale and \(|R_t| = o(\sqrt{t})\) is known in the mathematical literature as **martingale approximation** (see Gordin [14], Zhao and Woodroffe [36]). The theory of martingale approximation now leads me to a result whose empirical implications extend those of the RWH. Consider the distribution of the normalised consumption \( c_t^* = \frac{c_t}{\sqrt{t}} \) conditional on the value of \( c_0 \), namely

\[
F_t(z) = \Pr(c_t^* \leq z|c_0)
\]

Then the following Theorem is an immediate consequence of [36, Section 6]:

**Theorem 4.** If the process \((Y_t)_{t \geq 0}\) leads to tempered learning then the consumption \( c_t \) has a conditional central limit theorem, namely there is some\(^{26}\) \( \sigma \geq 0 \) s.t. \( \frac{\mathbb{E}[c_t^2]}{t} \to \sigma^2 \) and \( c_t^* \to \mathcal{N}(0, \sigma^2) \) in distribution as \( t \to \infty \).

Theorem 4 gives an empirically verifiable prediction. Thus, although the econometrician may lack the ability to look at the agent’s mind, she may still be able refute the claim that “the agent employs a no-regret learning algorithm which leads to tempered learning”.

**Remark 4.** As the RWH is a private case of a more general Euler equation, and as Euler equations play an important role in the estimation of dynamic models (see Attanasio and Low [1]), the current exercise can be viewed as a preliminary step in the direction of micro-founded theory from estimating Euler equations without the rational expectations assumption\(^{27}\).

### References


\(^{25}\)If I sum over all \( 1 \leq s \leq t \), then for a very large value of \( s \) the only elements not behaving as \( o(s^{-1/2}) \) are those for which \( s = t_j - 1 \) for some \( j \geq 1 \), of which there are at most \( \log t + 1 \), each contributing a bounded element to the sum, which in total = \( o(\sqrt{t}) \).

\(^{26}\)If \( \sigma = 0 \) then \( \mathcal{N}(0, \sigma^2) \) is defined to be the unit mass at 0.

\(^{27}\)Attansio et al [3] have estimated Euler equations subject to “expectation errors”. Nevertheless, they do not supply any micro-foundation for these ‘errors’ nor the way they are formed, but rather exogenously impose them on the Euler equation.


24


In this Appendix I will prove Theorem 2. I will in fact prove Theorem 3, which will imply Theorem 2, and in turn Theorem 1 via Proposition 1.

A. Proof of Results in Section 3

The proof has three moves, each contained in a separate Subsection. In the first move, I will introduce a finite approximation of this game. In the second move I will introduce the
updating algorithm for the agent’s belief and will show that it can be thought of, most of the time, as a version of the Hedge Algorithm (e.g., [12]). In the third move I will use this observation to obtain a regret bound for the game.

A.1.1 Finite Approximation for Models

I will now construct an approximation of the problem consisting a finite set of models. The number of models will increases with time. I will later use this approximation to define the belief updating and then will show that it assures sublinear regret.

In my construction I shall use an integer valued counter \( j \in \mathbb{N} \). Set \( t_j \) to be an increasing sequence of time periods. I will choose \( t_j \) optimally once the regret bounds have been found. For every \( j \in \mathbb{N} \), let \( \Delta_j > 0 \) be s.t. for every observable \( y \in Y \), choices \( x_- \in X \) and \( x \in \Gamma(x_-, y) \), and belief \( \lambda \) if \(||\theta - \theta'|| < \Delta_j\) then

\[
|U_\theta(x_-, x, y, \lambda) - U_{\theta'}(x_-, x, y, \lambda)| < \frac{1}{2\sqrt{t_j}}. \tag{39}
\]

The existence of such \( \Delta_j > 0 \) is assured by the equicontinuity of the family of mappings

\[
\{\theta \to U_\theta(x_-, x, y, \lambda) : y \in Y, x_- \in X, x \in \Gamma(x_-, y), \lambda \in L^1(\xi)\} \tag{40}
\]

which as been assumed in the Technical Assumptions Section 2.4. An admissible sequence of partitions \((\Pi_j)_{j=0}^\infty\) of \( \Theta \) consists of finite partition \( \Pi_j \) of \( \Theta \) for each \( j \) s.t. \( \Pi_0 = \{\Theta\} \), \( \Pi_{j+1} \) is a refinement of \( \Pi_j \), namely each atom \( a \in \Pi_j \) is a union of atoms of \( \Pi_{j+1} \), and each atom \( a \in \Pi_j \) has a diameter \( \leq \Delta_j \).

I consider an admissible sequence of partitions \((\Pi_j)_{j=0}^\infty\). For every \( j \geq 1 \) and every atom \( a \in \Pi_j \) I denote by \( p_j(a) \) the parent of \( a \) - the unique element of \( \Pi_{j-1} \) containing \( a \), namely \( a \subseteq p_j(a) \). I will denote it, by abuse of notation, \( p(a) \) whenever the value of \( j \) is clear from the context, and \( p^k(a) \) the composition \( p_{j-k+1} \circ \cdots \circ p_j(a) \) whenever it exists and the value of \( j \) is clear from the context. Every \( a \in \Pi_{j+1} \) is said to be the child of \( p(a) \), and the number of children of \( a \in \Pi_j \) is denoted \( n(a) \). For every atom \( a \in \Pi_j \) I fix a point \( \theta_j(a) \in a \).

For \( t_j \leq t < t_{j+1} \) define \( \hat{g}_t : \Pi_j \to \mathbb{R} \) by \( \hat{g}_t(a) = G_{\theta_j(a), t} \). Notice that if \( a \in \Pi_j \) then for every model \( \theta \in a \) I have

\[
|\hat{g}_t(a) - G_{\theta, t}| < \frac{1}{8\sqrt{t_j}}. \tag{41}
\]

Furthermore, for \( a \in \Pi_j \) let \( \hat{\lambda}_t(a) = \lambda_t(a) \).

27
A.1.2 The Algorithm

I will now present the Algorithm:

1 **Parameters:** Counter $j$, Interval times $t_j$, Learning Rates $\eta_j$, Initial density function $f_0$, Partitions $\Pi_j$, density $f_0 = \frac{d\lambda}{d\xi}$, weights $w_t$;

2 **Initialisation:** $j = 0$, values of $t_j$, $t_0 = 1$, initial weight function $w_0(\theta) = f_0(\theta)$;

3 while $j \geq 0$ do

4 For each $\theta$ compute $a(\theta) \in \Pi_j$ s.t. $\theta \in a(\theta)$;

5 for $t \leftarrow t_j$ to $t_{j+1}$ do

6 Observe payoff function $\hat{g}_t$;

7 Update $w_{t+1}(\theta) = w_t(\theta)e^{-\eta_j \hat{g}_t(a(\theta))}$;

8 end

9 For each $a \in \Pi_{j+1}$ and $\theta \in a$ set $w_{t+1}(\theta) = \frac{1}{n_p(a)^{\prime}} w_t(\theta)$;

10 $j \leftarrow j + 1$;

end

Algorithm 1: Hedge with Reproducing Experts

In words, the algorithm attaches weights for each model $\theta \in T$, and within the $j$-th period, namely times $t_j \leq t \leq t_{j+1} - 1$, it updates the weights according to the exponential multiplicative weights algorithm with rate $\eta_j$ and weights given by the relevant approximated payoff $\hat{g}_t$. At the end of the $j$-th period I replace the approximating partition with the partition containing its “children” attaching equal weights to each element’s ‘children’, hence the name “Hedge with Reproducing Experts”.

I now turn to analyse the algorithm and prove that it leads to sublinear regret. Let $\lambda_t$ be the measure obtained by normalisation from the weight function $w_t$. I define a probability distribution $\hat{\lambda}_t$ over $\Pi$ by $\hat{\lambda}_t(a) = \lambda_t(a)$. Then for every time period $t_j \leq t \leq t_{j+1} - 1$

\[
\hat{\lambda}_{t+1}(a) = \lambda_{t+1}(a) = \frac{\int_a w_t(\theta)e^{-\eta_j \hat{g}_t(a)}\xi(d\theta)}{\int_{\Theta} w_{t+1}(\theta)\xi(d\theta)}
= \frac{\int_a w_t(\theta)\xi(d\theta)}{\sum_{a' \in \Pi_j} e^{-\eta_j \hat{g}_t(a')} \int_{a'} w_t(\theta)\xi(d\theta)}} \frac{e^{-\eta_j \hat{g}_t(a)}}{\sum_{a' \in \Pi_j} e^{-\eta_j \hat{g}_t(a')} \hat{\lambda}_t(a')},
\]

(42)

Thus the update we have defined for $\lambda_t$ results in the exponential weights algorithm for the approximate gain $\hat{g}_t$ and the respective normalised weights $\hat{\lambda}_t$. Recall that I denote the
regret in the learning game by $R_t^L$. For the rest of the proof I will use the notation $R_t = R_t^L$ in order to keep the rest of the notation system as simple as possible. I thus obtain

**Lemma 1.** Set $t_j = 2^j$. If $t_j \leq t < t_{j+1}$, $\theta^t \in \Theta$ is the optimal model in-hindsight at stage $t$, and $\theta^t \in a^t \in \Pi_j$ then

$$R_t < \sum_{k=1}^{j} \sum_{s=t_k-1}^{t} \hat{g}_s(p^{k-1}(a')) - \sum_{s=1}^{t} E_{a' \sim \tilde{\lambda}_s} [\hat{g}_s(a')] + \sqrt{t}. \quad (43)$$

**Proof.** By my construction I have

$$R_t < \sum_{k=0}^{j} \left( \sum_{s=t_k}^{t} \hat{g}_s(p^k(a')) - \sum_{s=1}^{t} E_{a' \sim \tilde{\lambda}_s} [\hat{g}_s(a')] \right)$$

+ \sum_{k=0}^{j} \sum_{s=t_k}^{t} \frac{1}{4\sqrt{t_k}} \quad (44)

Notice that for the approximation term $(AT)$ above it holds that

$$(AT) = \frac{1}{4} \sum_{k=1}^{j} \sum_{s=2^k}^{\min\{t, 2^{k+1} - 1\}} 2^{-k/2}$$

$$\leq \frac{1}{4} \sum_{k=1}^{\lceil \log t \rceil} 2^{k/2} \leq \frac{1}{4} \frac{\sqrt{2}}{\sqrt{2} - 1} 2^{(\log t + 1)/2} < \sqrt{t}. \quad (45)$$

**A.1.3 Regret Bounds**

I now turn to bound the regret on the RHS of Equation (43). To do so, I denote $W_t(a) = \int_a w_t(\theta) \xi(d\theta)$ for every $a \in \Pi_j$ and $\Phi_t = \sum_{a' \in \Pi_j} W_t(a')$ whenever $t_j \leq t \leq t_{j+1} - 1$. By definition $W_0(\Theta) = 1$ and recall that $\tilde{\lambda}_t(a) = \frac{W_t(a)}{\Phi_t}$ are the normalised weights. The following Lemmata are helpful
Lemma 2. For every \( j \geq 0 \)
\[
\log(\Phi^{t_{j+1}}) \leq \log(\Phi^{t_j}) - \sum_{t=t_j}^{t_{j+1}-1} (\eta_j E_{a' \sim \tilde{\lambda}_t} [\tilde{g}_t(a')] + \eta_j^2 E_{a' \sim \tilde{\lambda}_t} [\tilde{g}_t(a')^2]).
\]  

Proof. For each \( t_j + 1 \leq t \leq t_{j+1} - 1 \) I use the following standard bounds (see, e.g., Cesa-Bianchi and Lugosi [8])
\[
\Phi^t = \sum_{a' \in \Pi_j} W_t(a') = \sum_{a' \in \Pi_j} W_{t-1}(a') e^{-\eta_j \tilde{g}_{t-1}(a')} 
= \Phi_{t-1} \sum_{a' \in \Pi_j} \tilde{\lambda}_{t-1}(a') e^{-\eta_j \tilde{g}_{t-1}(a')}
\leq \Phi_{t-1} \sum_{a' \in \Pi_j} \tilde{\lambda}_{t-1}(a') (1 - \eta_j \tilde{g}_{t-1}(a') + \eta_j^2 \tilde{g}_{t-1}(a')^2)
= \Phi^t \left(1 - \eta_j E_{a' \sim \tilde{\lambda}_{t-1}} [\tilde{g}_{t-1}(a')] + \eta_j^2 E_{a' \sim \tilde{\lambda}_{t-1}} [\tilde{g}_{t-1}(a')^2]\right)
\leq \Phi^t \exp \left(-\eta_j E_{a' \sim \tilde{\lambda}_{t-1}} [\tilde{g}_{t-1}(a')] + \eta_j^2 E_{a' \sim \tilde{\lambda}_{t-1}} [\tilde{g}_{t-1}(a')^2]\right).
\]  

Applying recursively I obtain
\[
\Phi^{t_{j+1}-1} \leq \Phi^{t_j} \exp \left(-\sum_{t=t_j}^{t_{j+1}-2} (\eta_j E_{a' \sim \tilde{\lambda}_t} [\tilde{g}_t(a')] + \eta_j^2 E_{a' \sim \tilde{\lambda}_t} [\tilde{g}_t(a')^2])\right).
\]  

Moving from time \( t_{j+1} - 1 \) to time \( t_{j+1} \) the algorithm not only multiplies by the weight but also splits the priors, so \( W_{t_{j+1}}(a') = \frac{W_{t_{j+1}-1}(p(a'))}{n(p(a'))} e^{-\eta_j \tilde{g}_{t-1}(a')} \) for every atom \( a' \in \Pi_{j+1}. \) Applying this one more step as before I obtain
\[
\Phi^{t_{j+1}} \leq \Phi^{t_j} \exp \left(-\sum_{t=t_j}^{t_{j+1}-1} (\eta_j E_{a' \sim \tilde{\lambda}_t} [\tilde{g}_t(a')] + \eta_j^2 E_{a' \sim \tilde{\lambda}_t} [\tilde{g}_t(a')^2])\right).
\]  

By taking logarithms from both sides of the last inequality the Lemma follows. \( \square \)

Lemma 3. For every \( j \) and \( a \in \Pi_j \)
\[
\log \Phi^{t_{j+1}-1} \geq \log(W_{t_{j+1}-1}(a)) = \log(W_{t_j}(a)) - \sum_{t=t_j}^{t_{j+1}-2} \eta_j g_t(a).
\]
Proof. As Φ_{t+1} ≥ W_{t+1}(a) I obtain by the definition of the weight and the multiplicative algorithm

$$\Phi_{t+1} ≥ W_{t+1}(a) = W_t(a) \cdot \exp \left( - \sum_{t=t_j}^{t_j+2} \eta_j g_t(a) \right), \quad (56)$$

and the Lemma follows by taking logarithms.

Consider the regret

$$\hat{R}_i^j(a) = \sum_{s=t_j}^t \hat{g}_s(a) - \sum_{s=t_j}^t E_{a' \sim \tilde{\lambda}_t} \hat{g}_s(a'). \quad (57)$$

By Cesa-Bianchi and Lugosi [8, Corollary 3.1] it holds for every $j$ that

$$\eta_j \hat{R}_{t+1}^j(a) \leq -\log(\pi_j(a)), \quad (58)$$

where $\pi_j$ is the prior distribution at the beginning of the $j$-th period, namely the ratio $\frac{W_{t_j}(a)}{\Phi_{t_j}}$. Recalling that $W_t(a) = \frac{W_{t-1}(a)}{n(p(a))} e^{-\eta_j \hat{g}_{t-1}(p(a))}$, I obtain:

$$\eta_j \hat{R}_{t+1}^j(a) \leq \log(\Phi_{t_j}) - \log(W_{t_j-1}(p(a))) + \log(n(p(a))) - \eta_j \hat{g}_{t-1}(p(a)) \quad (59)$$

By first applying Lemmata 2 and 3 to Inequality (59), I obtain

$$\eta_j \hat{R}_{t+1}^j(a) \leq \log(\Phi_{t_j}) - \sum_{t=t_j}^{t_j-1} (\eta_{t-1} E_{a' \sim \tilde{\lambda}_t} \hat{g}_{t}(a') - \eta_{t-1}^2 E_{a' \sim \tilde{\lambda}_t} \hat{g}_{t}(a')^2)) +$$

$$- \log(W_{t_j-1}(p(a))) + \sum_{t=t_j}^{t_j-1} \eta_{t-1} \hat{g}_{t}(p(a)) + \log(n(p(a))). \quad (60)$$

After some rearrangement, collecting terms of $\eta_{t-1}$ and recalling that the ratio $\frac{W_{t_j}(a)}{\Phi_{t_j}}$ equals the prior $\pi_{t_j}(p(a))$, it follows that

$$\eta_j \hat{R}_{t+1}^j(a) \leq \eta_{t-1} \hat{R}_{t_j}^j(p(a)) + \log(n(p(a))) + \eta_{t-1}^2 \sum_{t=t_j}^{t_j-1} E_{a' \sim \tilde{\lambda}_t} \hat{g}_{t}(a')^2 - \log(\pi_{t_j-1}(p(a))) \quad (61)$$
I shall recursively apply the same logic that followed Equation (58) to summand ("$R_{j-1}$"). Recalling that $W_0(\Theta) = 1$, I obtain the following recursive equation

$$\eta_j \hat{R}_{t_{j+1}}^j \leq \sum_{k=0}^{j-1} \eta_k \hat{R}_{t_{k+1}}^k (p^{j-k}(a)) + \sum_{k=0}^{j-1} \log n(p^{j-k}(a)) + \sum_{k=0}^{j-1} \sum_{t=t_k}^{t_{k+1}-1} E_{a' \sim \hat{\lambda}_t} [\hat{g}_t(a')^2].$$

By recursively applying Inequality (62) I obtain

$$\hat{R}_{t_{j-1}}^j \leq \eta_0 R_0(\Theta)(j-1)\eta_j^{-1} + \eta_j^{-1} \sum_{s=0}^{j-1} (j-s) \log n(p^{j-s-1}(a)) + \sum_{s=0}^{j-1} (j-s)\eta_s^2 \sum_{t=t_s}^{t_{s+1}-1} E_{a' \sim \hat{\lambda}_t} [\hat{g}_t(a')^2].$$

By combining Lemma (1) and Inequality (63), and denoting $k_t = \lceil \log t \rceil$ I have thus proved

**Proposition 4.**

$$R_t < R_0(\Theta)\eta_0 \sum_{j=1}^{k_t} (j-1)\eta_j^{-1} + \sum_{j=1}^{k_t} \sum_{s=0}^{j-1} \eta_j^{-1} (j-s) \log n(p^{j-s-1}(a')) \quad (S1)$$

$$+ \sum_{j=1}^{k_t} \sum_{s=0}^{j-1} (j-s)\eta_s^2 \sum_{t=t_s}^{t_{s+1}-1} E_{a' \sim \hat{\lambda}_t} [\hat{g}_t(a')^2] + \sqrt{t}. \quad (S3)$$

To complete the proof of Theorem 3, I will now bound summands $(S1) - (S3)$ in Inequality (64).
**Summand (S1):** As I have chosen \( \eta_j = \frac{1}{\sqrt{t_j}} = 2^{-j/2} \), it gives the bound

\[
(S1) = \sum_{j=1}^{k_t} (j - 1)2^{j/2} \leq k_t \sum_{j=1}^{k_t} 2^{j/2}
\]

\[
\leq 3k_t(2^{-k_t/2} - 1) < 3(\log t + 1)2^{(\log t + 1)/2} < 6\sqrt{t} \log t.
\]

**Summand (S2):** It holds that for every \( a \in \Pi_{k_t} \)

\[
(S2) \leq (k_t)^22^{k_t/2} \left( \prod_{s=0}^{k_t-1} n(p^{k_t-s-1}(a)) \right) - \sqrt{t} \log |\Pi_{k_t}|
\]

\[
\leq 2\sqrt{t} \log |\Pi_{k_t}| - \sqrt{t} \log |\Pi_{k_t}| = \sqrt{t} \log |\Pi_{k_t}|.
\]

I will now show that the partitions can be chosen in advance s.t. the size of \( k_t \)-th partition \( \Pi_{k_t} \) will have a small contribution to the regret bound (S2). Indeed, let \( C \) be an Euclidean cube containing \( \Theta \). Partition \( C \) recursively, setting \( \hat{\Pi}_0 = \{C\} \) and \( \hat{\Pi}_j \) is obtained from \( \hat{\Pi}_{j-1} \) by partitioning every element into cubes of equal edge length \( r_j \), where \( r_j \) is chosen to be the minimal s.t. each cube contains a Euclidean ball of radius \( \Delta_j \). I define the partition \( \Pi_j = \{a \cap \Theta : a \in \hat{\Pi}_j\} \). Notice that if \( M(\Delta_j) \) is the \( \Delta_j \)-packing number of \( \Theta \), namely, the infimum over the number of points in \( \Theta \) s.t. each two points have distance \( \geq \Delta_j \) from each other, then \( |\Pi_j| \leq M(\Delta_j) \), and as \( M(\Delta_j) \leq N(\Delta_j/2) \) I have for some constant \( C(\Theta) > 0 \)

\[
\ln |\Pi_j| \leq \ln N(\Delta_j/2) \leq -C(\Theta) \ln \Delta_j
\]

where the last inequality follows as \( \Theta \) is assumed to have a bounded metric dimension.

Returning to summand (S2), I recall that \( w^{-1} \) is the inverse of the uniform modulus of continuity of \( g \). As \( \Delta_j = w^{-1} \left( O \left( t_j^{-1} \right) \right) \), it follows from Inequality (67) that for \( a \in \Pi_{k_t} \)

\[
(S2) \leq -C(\Theta)\sqrt{t} \ln \left( w^{-1} \left( O \left( t^{-1/2} \right) \right) \right),
\]

**Summand (S3):** Denote by \( M \) the bound of \( \tilde{g} \). I thus obtain for any \( t \geq 2 \)

\[
(S3) \leq \sum_{j=1}^{k_t} \sum_{s=0}^{j-1} (j - s)M^2 \leq M^2k_t^3 \leq 8M^2 (\log t)^3
\]
A.2 Proof of Proposition 1

Recall that given the agent’s beliefs \( \lambda^t = (\lambda_1, ..., \lambda_t) \) the observables \( y_t = (y_1, ..., y_t) \), and her choices \( x^*_t = (x^*_1, ..., x^*_t) \) s.t. \( x^*_s \in x(x^*_{s-1}, y_s, \lambda_s) \) for each \( 1 \leq s \leq t \), I have defined the counterfactual dynamic regret at time \( t \) is to be

\[
R^t_C(x_t, y_t, \lambda^t) = \sup_{\theta^* \in \Theta} \sum_{s=1}^t \max_{x \in \Gamma(x^*_{s-1}, y_s)} \left( \sum_{t=1}^s U_{\theta^*}(x^*_{s-1}, x, y_s, \lambda_s) - \sum_{s=1}^t \max_{x \in \Gamma(x^*_{s-1}, y_s)} \int_T U_{\theta}(x^*_{s-1}, x, y_s, \lambda_s) l_s(d\theta) \right). \tag{72}
\]

Let \( \epsilon > 0 \). For fixed vectors \( y_t, x_t, \) and \( \lambda^t \) I choose a model \( \theta^t \in \Theta \) s.t.

\[
\sup_{\theta^* \in \Theta} \sum_{s=1}^t \max_{x \in \Gamma(x^*_{s-1}, y_s)} U_{\theta^*}(x^*_{s-1}, x, y_s, \lambda_s) < \sum_{s=1}^t \max_{x \in \Gamma(x^*_{s-1}, y_s)} U_{\theta^t}(x^*_{s-1}, x, y_s, \lambda_s) + \epsilon. \tag{73}
\]

Suppose the supremum for \( R^t_C \) is obtained at and that the optimal in-hindsight model in period \( t \) is \( \theta^t \). Then

\[
R^t_C(x^*_t, y_t, \lambda^t) - \epsilon \tag{74}
\]

\[
< \sum_{s=1}^t \max_{x \in \Gamma(x^*_{s-1}, y_s)} U_{\theta^t}(x^*_{s-1}, x, y_s, \lambda_s) - \sum_{s=1}^t \max_{x \in \Gamma(x^*_{s-1}, y_s)} \left( \int U_{\theta}(x^*_{s-1}, x, y_s, \lambda_s) d\lambda_s(\theta) \right) \tag{75}
\]

\[
\leq \max_{x \in \Gamma(x^*_{s-1}, y_s)} \left( \sum_{s=1}^t \max_{x \in \Gamma(x^*_{s-1}, y_s)} U_{\theta^t}(x^*_{s-1}, x, y_s, \lambda_s) - \int U_{\theta}(x^*_{s-1}, x, y_s, \lambda_s) d\lambda_s(\theta) \right) \tag{76}
\]

where the inequality in line (75) follows from the fact that for every choice of real valued functions \( f_1, ..., f_t \) and \( g_1, ..., g_t \) over common sets \( W_1, ..., W_t \) respectively we have

\[
\sum_{s=1}^t \left( \max_{w \in W_s} f_s(w) - \max_{w \in W_s} g_s(w) \right) \leq \max_{w \in W_s} \sum_{s=1}^t (f_s(w) - g_s(w)).
\]

34
Indeed, if \( w^*_s \in \text{arg max}\{f_s(w) : w \in W_s\} \) then

\[
\sum_{s=1}^t \left( \max_{w \in W_s} f_s(w) - \max_{w \in W_s} g_s(w) \right) = \sum_{s=1}^t \left( f_s(w^*_s) - \max_{w \in W_s} g_s(w) \right)
\]

\[
\leq \sum_{s=1}^t (f_s(w^*_s) - g_s(w^*_s)) \leq \max_{w_s \in W_s} \sum_{s=1}^t (f_s(w_s) - g_s(w_s)).
\]

Taking supremum over \( y \) in Inequality (74), I obtain for every \( \epsilon > 0 \)

\[
\sup_{y_1, \ldots, y_t} R^t_C(x^*_t, y_t, \lambda^t) - \epsilon \leq \sup_{y_1, \ldots, y_t} R^t_L(x_t, y_t, \lambda^t),
\]

and the Proposition follows.

### A.3 Proof of Theorem 1

Combining Proposition 1 and Theorem 3 I obtain that the agent may choose her policies \( x^*_1, \ldots, x^*_t, \ldots \) and beliefs \( \lambda_1, \ldots, \lambda_t, \ldots \) s.t. as \( t \to \infty \)

\[
\sup_{y_1, \ldots, y_t} R^t_C \leq -\sqrt{t}C_1(\Theta) \ln \left( w^{-1} \left( O \left( t^{-1/2} \right) \right) \right) + o(t).
\]

A I have assumed\(^{28}\) that \( w \left( e^{-O(t)} \right) = O \left( t^{-1/2} \right) \) I obtain that \( -\ln \left( w^{-1} \left( O \left( t^{-1/2} \right) \right) \right) = o(t) \) hence the agent can choose her policies \( x^*_1, \ldots, x^*_t, \ldots \) and beliefs \( \lambda_1, \ldots, \lambda_t, \ldots \) s.t. as \( t \to \infty \)

\[
\sup_{y_1, \ldots, y_t} R^t_C = o(t),
\]

and the Theorem follows.

### B Approximation of the Belief Updating Operator

Here I would like to study how the belief densities \( w_t \) evolve over time and the relationship of this evolution with the learning rate \( \eta \). For convenience, I will omit time, and write \( w \)

\(^{28}\)See Equation (18) in the Technical Assumptions Section 2.4
for the density “today” and $w_+$ for the density “tomorrow”, and the same for the belief $\lambda$ and any other parameter. I will fix a partition $\Pi = \Pi_j$ for some $j \geq 0$.

First assume that $t_j \leq t < t_{j+1} - 1$. Then for each atom $a \in \Pi$ and model $\theta \in a$

$$w_+(\theta) = \frac{w(\theta) e^{-\eta \widehat{g}(a)}}{\sum_{a' \in \Pi} \lambda(a') e^{-\eta \widehat{g}(a')}} = \frac{w(\theta)}{\sum_{a' \in \Pi} \lambda(a') e^{-\eta (\widehat{g}(a') - \widehat{g}(a))}}$$

Due to the continuity and monotonicity of the exponential function and the fact that $\widehat{g}$ is bounded by some $M > 0$, for every $\lambda$, $\eta$, and $a \in \Pi$ there is a number $Z(\lambda, \eta, a)$ s.t.

$$|Z(\lambda, \eta, a)| \leq M$$

and

$$\sum_{a' \in \Pi} \lambda(a') e^{-\eta (\widehat{g}(a') - \widehat{g}(a))} = e^{-\eta Z(\lambda, \eta, a)}.$$  

Thus for $\theta \in a$

$$w_+(\theta) - w(\theta) = w(\theta) \left( 1 - e^{\eta Z(\lambda, \eta, a)} \right).$$

Expanding the exponential as an infinite series I obtain, as $|Z(\lambda, \eta, a)| \leq M$ is bounded

$$w_+(\theta) - w(\theta) = \eta w(\theta) (Z(\lambda, \eta, a) + O(\eta)).$$

By abuse of notation, I define a function $Z(\lambda, \eta, \cdot) : \Theta \to \mathbb{R}$ via $Z(\lambda, \eta, \theta) = Z(\lambda, \eta, a)$ iff $\theta \in a$. I have thus proved

**Proposition 5.** There is a bounded measurable function $G : \Delta(\Theta) \times [0, 1] \times \Theta \to \mathbb{R}_+$ s.t.

$$w_+(\theta) - w(\theta) = \eta w(\theta) G(\lambda, \eta, a).$$