On Incomplete Learning and Certainty-Equivalence Control

N. Bora Keskin* Duke University

Assaf Zeevi† Columbia University

This version: November 17, 2017

Abstract

We consider a dynamic learning problem where a decision maker sequentially selects a control and observes a response variable that depends on chosen control and an unknown sensitivity parameter. After every observation, the decision maker updates her/his estimate of the unknown parameter and uses a certainty-equivalence decision rule to determine subsequent controls based on this estimate. We show that under this certainty-equivalence learning policy the parameter estimates converge with positive probability to an *uninformative* fixed point that can differ from the true value of the unknown parameter; a phenomenon that will be referred to as *incomplete learning*. In stark contrast, it will be shown that this certainty-equivalence policy may avoid incomplete learning if the parameter value of interest "drifts away" from the uninformative fixed point at a critical rate. Finally, we prove that one can adaptively limit the learning *memory* to improve the accuracy of the certainty-equivalence policy in both static (estimation), as well as slowly varying (tracking) environments, without relying on forced exploration.

Keywords: Dynamic control, sequential estimation, certainty equivalence, incomplete learning.

1 Introduction

1.1 Background and overview of contribution

Background and motivation. Dynamic decision making under uncertainty arises in many application domains. For example, consider a seller who is uncertain about the price-elasticity of the demand for its product and can dynamically adjust prices to learn about the elasticity of demand, or a physician who is uncertain about how a drug's dosage will treat a medical condition and makes sequential observations on patient outcomes to learn about that effect; see §3 for detailed models of these and other applications. A common strategy in this context is to first estimate the unknown effect (e.g., the price-elasticity of demand, or the drug's treatment effect) and then make a decision that optimizes an objective function that is parameterized by the estimate. The repetitive use of this estimate-and-optimize routine at every decision epoch provides a *dynamic learning policy*. A salient feature of this type of policy is that it optimizes as if there is no estimation, and estimates as if there is no optimization; a principle often referred to as *certainty-equivalence*. This paper is concerned with the question whether (and when) learning "takes care of itself," as implicitly stipulated by this class of policies.

*Fuqua School of Business, Durham, NC 27708, e-mail: bora.keskin@duke.edu

†Graduate School of Business, New York, NY 10027, e-mail: assaf@gsb.columbia.edu

More formally, we consider a dynamic control problem where the response structure is a function of the decision maker's controls and an unknown sensitivity parameter. We analyze learning policies that iteratively estimate the unknown parameter and on the basis of this choose controls via a certainty-equivalence decision rule, which would have been optimal if the unknown parameter were equal to its estimate with certainty. A natural question in this context is whether the resulting sequence of parameter estimates eventually, as more and more observations are collected, reveals the true nature of the response function parameter. Failure to do so is usually referred to as *inconsistency* of the estimates, and our primary focus is on an extreme form of inconsistency called *incomplete learning*, which occurs if the parameter estimates not only fail to converge to the true value of the unknown parameter but in fact converge to an incorrect value. In this paper, we study environments in which the incomplete learning phenomenon is observed, and elucidate when and how certainty-equivalence decision making can avoid incomplete learning. Moreover, when incomplete learning can be avoided, we are interested in the asymptotic accuracy and tracking performance of certainty-equivalence estimates.

As will be discussed in detail below, antecedent literature related to incomplete learning has almost exclusively focused on avoiding this phenomenon via *forced exploration* (i.e., carefully substituting information collection for inference purposes, for the decisions that would be otherwise prescribed by the learning policy). Roughly speaking, forced exploration judiciously turns "on and off" a given decision rule to improve inference. In contrast to the literature on forced exploration, the question we are interested in is how the iterative and uninterrupted use of a certainty-equivalence decision rule, which is a *passive* learning approach that does not rely on forced exploration, can avoid the incomplete learning phenomenon.

Overview of main results. Our study makes two contributions to the literature on dynamic decision making under uncertainty.

The first concerns the nature of incomplete learning. We prove that in a *static* environment the estimates of a certainty-equivalence learning policy can *fail* to converge to the true value of the unknown model parameter with positive probability (see Example 1 and Theorem 1). Roughly speaking, a certainty-equivalence learning policy can stop learning prematurely. This type of observation is not new in and of itself. Lai and Robbins (1982) were the first to show that the controls of an iterated least squares policy can converge to the boundary of the feasible set of controls, disproving a conjecture of Anderson and Taylor (1976); see Prescott (1972) for an earlier reference, and den Boer and Zwart (2014) for a more recent one, as well as other follow-up work discussed in §1.2. However, the analysis of incomplete learning in these papers suggests that it is a consequence of the possibly problematic boundaries of feasible control sets. The setting considered in this paper shows that incomplete learning occurs due to the controls and estimates of a certainty-equivalence policy converging to an *uninformative equilibrium*, which has nothing to do with boundaries, but

rather is a fixed point (attractor) of the dynamical system induced by the response function and the certainty-equivalence rule.

As incomplete learning is identified with a fixed point of a dynamical system, an obvious question is whether this is a *stable* equilibrium point (i.e., does perturbing this point result in the dynamical system being "attracted" back to it, or diverging from it). In the context of dynamic estimation and control, it has been established that if the cumulative "variation" of the controls is *forced* to grow over time at a judiciously selected rate then the corresponding estimate sequence would be consistent (this is a classical observation that pertains to forced exploration, further discussed in §1.2). Along similar lines of thinking, one might intuitively expect that if the unknown parameter of the response function is changing over time such that its cumulative variation grows at a suitable rate then incomplete learning should not happen. However, our analysis reveals that this intuition is incorrect (see Example 2 and Theorem 2). Thus, variation in controls and variation in unknown parameters have distinct impacts on learning; small perturbations to the control sequence are effective in mitigating incomplete learning, whereas similar fluctuations in the unknown parameter sequence do not rule out incomplete learning (see also the discussion following Theorem 2 for further details). Expanding on this result, we also investigate the question whether there exists a changing environment in which incomplete learning can be avoided without using any forced exploration. For example, what happens if the unknown parameter varies over time in a manner that can "push" the trajectory of estimates and controls "away" from the attractor discussed above. To that end, we identify the following phenomenon: if the parameter drifts away from the uninformative equilibrium faster than some critical rate, then incomplete learning is eliminated in a suitable sense (see Example 4 and Theorem 3). In this setting, the changes in the "environment" facilitate dynamic learning.

Motivated by these observations, we propose a general adaptive scheme that can mitigate incomplete learning in *both* static as well as "slowly varying" environments. For that purpose, we limit the memory of certainty-equivalence learning by adaptively choosing a sequence of estimation windows. In Theorem 5, we prove that such a policy avoids incomplete learning in a fairly general class of static and changing environments, without relying on any forced exploration. Moreover, we show that limiting estimation memory achieves asymptotic accuracy in static environments (see Theorem 6), and exhibits good tracking performance in slowly changing environments (see Theorem 7).

Exposition, conventions, and organization of the paper. Throughout the sequel, we will use some modeling elements primarily for illustrative purposes. For example, we will focus on a linear-Gaussian response model that will greatly facilitate development of basic ideas and intuition and allow us to study both static and drifting parameter sequences, deferring treatment of a general response model to §6. We will also employ nonlinear least squares estimation, and provide an extension to more general estimation techniques in §7. The remainder of this paper is organized as

follows. This section concludes with a review of related literature. Section 2 describes our model and the main salient features of the problem studied in this paper, and §3 presents illustrative examples of the model. Our main results are presented in §§4-6. In §4, we show several negative and positive outcomes driven by certainty-equivalence learning policies in static and changing environments in the context of a linear-Gaussian model, and in §5, we extend our analysis of incomplete learning in static environments to a family of nonlinear models. In §6, we study certainty-equivalence learning with limited memory as a general method for eliminating the negative outcomes and guaranteeing "good" performance in static and in slowly changing environments. We provide our concluding remarks in §7. All proofs are in appendices.

1.2 Origins of certainty-equivalence and related literature

There is a rich academic literature on multiperiod control and sequential estimation problems, especially in the area of adaptive control (see, e.g., Åström and Wittenmark 2013), stochastic approximation (see, e.g., the survey paper by Lai 2003), and reinforcement learning (see, e.g., Kaelbling, Littman and Moore 1996, Gosavi 2009, for comprehensive surveys): to avoid exhaustively surveying said literature, we will focus on work which is closely related to, and serves best to motivate, the problems studied in this paper.

The principle of certainty-equivalence is a widely used heuristic in the design of adaptive control policies. It can be viewed as an "extreme point" in the space of dynamic programming-based policies. The significant computational challenge there, primarily due to the curse of dimensionality, is further exacerbated in problems with parameter uncertainty. One approach to deal with this is *model predictive control*, which uses a limited rolling horizon to account for the evolution of controls and estimates (see the survey paper by Garcia, Prett and Morari 1989). A particular form of model predictive control is the restriction of policy space to what is known as *limited lookahead* policies, which reduce the computational burden by solving the dynamic programming recursion for a shorter time horizon, leading to a smaller-scale problem. For instance, a simple and commonly used policy within this family is the one-step lookahead policy that needs to iterate the dynamic programming recursion only once. An even more extreme policy is the *certainty-equivalence control*, which is a myopic policy that does not look ahead at all, but instead focuses only on optimizing immediate rewards. To be precise, the certainty-equivalence control operates under the assumption that the decision maker's beliefs or estimates on an unobservable system state will remain the same in the future, as if these beliefs or estimates are *certain* values rather than random variables. Early examples of estimation methods in this context typically involve least squares estimation of the parameters of a linear dynamical system, referred to as linear quadratic estimation, or more generally as the Kalman-Bucy filter (see Kalman and Bucy 1961). As explained above, the certainty-equivalence control separates the dual goals of estimation and optimization, and is known to perform well in some of the fundamental dynamic control problems such as the

linear quadratic Gaussian (LQG) control problem (see Åström and Wittenmark 2013, chap. 4). These appealing features have brought forth certainty-equivalence control as a viable heuristic in the broader context of dynamic learning problems.

A prototypical and widely studied example in this context is the multiarmed bandit problem, in which a decision maker attempts to find the best option within a finite feasible action set by sequentially sampling and obtaining noisy observations on the expected rewards of sampled options, also referred to as "arms"; see Thompson (1933) and Robbins (1952) for the origin of this literature. In this context it is clear that if one employs certainty-equivalence, the policy would sample the arm with the highest empirical mean. Because sampling an arm does not provide information about other arms, it is not difficult to see that in most settings this policy will get stuck on an inferior arm with positive probability. Robbins (1952) identified this issue and proposed the use of *forced exploration*, defined as departure from the certainty-equivalence decision rule on a pre-scheduled sequence of experiments. Lai and Robbins (1985) refined this proposal by introducing an adaptive version of forced exploration based on upper confidence bounds (UCB), which does not pre-schedule experimentation; see also Auer, Cesa-Bianchi and Fischer (2002) for further study of these UCB policies as well as randomization-based alternatives. Rothschild (1974) asked a slightly different question in this context. If one were to study the multiarmed bandit problem within a Bayesian infinite horizon discounted formulation, is the optimal policy going to sample the best arm infinitely often? While this is a property that seems natural to expect, it turns out that this need not hold, and the optimal action is not identified with positive probability. Rothschild (1974) called this phenomenon "incomplete learning" (see also Brezzi and Lai 2000, 2002, McLennan 1984), and we use this term in our paper, with slight abuse of terminology, to describe the inability of certainty-equivalence to identify the underlying parameter (and optimal action).

Another research stream related to incomplete learning focuses on the consistency of iterated least squares in multiperiod control and estimation. As mentioned in §1.1, an early study by Anderson and Taylor (1976) provided simulation results that demonstrate the consistency of iterated least squares in a multiperiod control problem, and following this, Lai and Robbins (1982) derived a counterexample where a control sequence based on iterated least squares can incorrectly converge to the boundary of the feasible control set. More recently, den Boer and Zwart (2014) proved a similar incomplete learning result in a dynamic pricing context. In a sequence of papers, Lai and Robbins (1979, 1981, 1982) derived conditions that ensure the consistency of iterated least squares and stochastic approximation based schemes in similar settings.

In the context of adaptive control, Borkar and Varaiya (1979, 1982) studied the control of discrete-state-space Markov chains whose transition probabilities depend on an unknown parameter. They derived conditions for identifiability, and showed that adaptive control rules may not necessarily identify the unknown parameter that governs the Markov chain transition probability. The broader

domain of dynamic learning and adaptive control also includes variants of certainty-equivalence policies that use different forms of forced exploration. A prominent example is ϵ-greedy exploration, which prescribes choosing a random control with probability ϵ at every decision opportunity and using certainty-equivalence control otherwise (see Sutton and Barto 1998, chap. 5). Another approach is to employ extensions of the aforementioned UCB policies when the feasible control set is continuous rather than discrete. One obvious approach is to quantize the feasible control set and treat each as an "arm" within a multiarmed bandit problem (see, e.g., Auer, Ortner and Szepesvári 2007); Thompson sampling (Thompson 1933) has recently received a lot of attention as a Bayesian-based UCB alternative (see, e.g., Agrawal and Goyal 2012). In control theory, dithering signals is used for maintaining system stability by adding random perturbations on top of the certainty-equivalence control sequence (see Åström and Wittenmark 2013, chap. 10). There has also been a flurry of recent work in revenue management that considers dynamic pricing policies that might be described as semi-myopic yet focuses on avoiding incomplete learning via repetitive use of forced exploration (see, e.g., Lobo and Boyd 2003, Harrison, Keskin and Zeevi 2012, Broder and Rusmevichientong 2012, den Boer and Zwart 2014, Keskin and Zeevi 2014, den Boer 2014, Besbes and Zeevi 2015, Cheung, Simchi-Levi and Wang 2017).

In terms of formulation, our paper has several distinguishing features: (i) the dynamical system we analyze has a continuous and unbounded state space; (ii) there is a (possibly unbounded) continuum of feasible controls; (iii) the unknown parameter that governs the system evolution can be static or changing over time; and (iv) we introduce and study adaptive and non-stationary control policies (e.g., adaptively limiting the memory in estimation) as a way to mitigate incomplete learning. In that way our work sheds light on the boundary of environments in which passive learning (i.e., absent forced exploration) works well. As alluded to earlier, the statistical inference methods we employ in this paper are related to nonlinear least squares that is first developed and analyzed in Marquardt (1963) and Jennrich (1969), and studied in detail by Wu (1981) and Lai (1994).

2 Problem Formulation

2.1 The model and preliminaries

The observation process and certainty-equivalence control. Consider a dynamic control problem in which a decision maker chooses controls $x_1, x_2, \ldots$ from a set $\mathcal{X} \subseteq \mathbb{R}$ over a discrete time horizon. In response to the controls, s/he observes outputs $y_1, y_2, \ldots$ generated according to the following response model:

$$y_t \;=\; f(x_t, \theta) + \epsilon_t \quad \text{ for } t = 1, 2, \ldots, \tag{2.1}$$

where θ is an unknown model parameter that can take values in a set $\Theta \subseteq \mathbb{R}$, $f : \mathcal{X} \times \Theta \to \mathbb{R}$ is a continuously differentiable function, and $\{\epsilon_t,\, t = 1, 2, \ldots\}$ are unobservable noise terms, which

are independent and identically distributed random variables with a density $h_\epsilon(\cdot)$ and support $\mathbb{R}$. We assume that the mean and the variance of these noise terms are zero and σ^2 respectively. The unknown parameter θ represents the sensitivity of the responses to controls. To accommodate the largest possible set of values for θ, we will assume that $\Theta = \mathbb{R}$ unless otherwise stated (see §7 for a discussion of the case where Θ is a strict subset of $\mathbb{R}$).

In the first period, the decision maker deterministically chooses the value of x_1 to generate an initial observation. (The case where several initial observations are taken at different points $x_1, x_2, \ldots$ can be treated similarly.) After that, at the end of every period $t \geq 1$, the decision maker aims to compute the least squares estimate $\hat{\theta}_{t+1}$ that minimizes $S_t(\theta) = \sum_{s=1}^{t} \big(y_s - f(x_s, \theta)\big)^2$. In general, there need not be a closed-form solution to this optimization problem, and we stipulate that $\hat{\theta}_{t+1}$ is computed by solving the first-order optimality condition:

$$\frac{\partial S_t(\hat{\theta}_{t+1})}{\partial \theta} \;=\; 0, \qquad\qquad \text{[estimation]} \tag{2.2}$$

where $\partial S_t(\theta)/\partial\theta = -2\sum_{s=1}^{t} \big(y_s - f(x_s,\theta)\big)\, f_\theta(x_s,\theta)$, and $f_\theta(x,\theta) = \partial f(x,\theta)/\partial\theta$. We assume the existence of a unique solution to (2.2). (If Θ is a strict subset of $\mathbb{R}$, then $\hat{\theta}_{t+1}$ is computed by projecting the solution to (2.2) onto Θ.)

Remark 1 The use of least squares estimation in the computation of $\hat{\theta}_{t+1}$ is to make the exposition concrete. The analysis in §6 is valid for any M-estimator, with $\phi : \mathbb{R}^{2t} \to \Theta$ such that $\hat{\theta}_{t+1} = \phi(x_1, y_1, \ldots, x_t, y_t) = \arg\max_\theta \sum_{s=1}^{t} \lambda\big(y_s - f(x_s,\theta)\big)$, and $\lambda(\cdot)$ is a suitably chosen score function. See §7 for a detailed discussion of the extension from least squares to M-estimation.

Following the estimation in period t, the decision maker chooses the control in period $t+1$ as follows:

$$x_{t+1} \;=\; \psi(\hat{\theta}_{t+1}), \qquad\qquad \text{[control]} \tag{2.3}$$

where $\psi : \Theta \to \mathcal{X}$ is a control function that satisfies the following properties.

Definition (admissible control functions) *A function $\psi : \Theta \to \mathcal{X}$ is said to be an admissible control function if $\psi(\cdot)$ is differentiable and monotone, and satisfies $\ell \leq |\psi'(\theta)| \leq L$ for all $\theta \in \Theta$, where $0 < \ell \leq L < \infty$. The set of all admissible control functions is denoted by Ψ.*

The value of $\psi(\theta)$ is interpreted as the best control the decision maker could have chosen in period $t+1$ if s/he had perfect knowledge of θ. However, in the absence of this information the mapping to action space replaces θ with the estimate $\hat{\theta}_{t+1}$ in (2.3). The monotonicity of $\psi(\cdot)$ implies that the control is always sensitive to the unknown model parameter, and the decision maker reacts to more responsive systems in a particular direction, by either increasing or decreasing controls (see the applications in §3 for a more detailed explanation of how such monotonicity conditions naturally arise in practice). Unless otherwise noted, we assume without loss of generality that $\psi(\cdot)$ is increasing, as the analysis for case where $\psi(\cdot)$ is decreasing follows by symmetry. Because $\psi(\cdot)$ is monotone it is invertible, and we denote by $\psi^{-1}(\cdot)$ its inverse.

The iterative use of equations (2.2) and (2.3), which interlace estimation and control, describes a dynamical system that induces a family of probability measures on the sample space of response sequences $\{y_t,\, t = 1, 2, \ldots\}$. Given $\theta \in \Theta$, let $\mathbb{P}_\theta$ be a probability measure with density

$$h_\theta(y_1, \ldots, y_t) \;=\; \prod_{s=1}^{t} h_\epsilon\big(y_s - f(x_s, \theta)\big) \quad \text{for } y_1, \ldots, y_t \in \mathbb{R}, \tag{2.4}$$

where $h_\epsilon(\cdot)$ is the density of the random variables ϵ_t, and $\{x_t,\, t = 1, 2, \ldots\}$ is the control sequence formed under the decision rule (2.3) and responses $y_1, y_2, \ldots$

Performance metric and formulation for drifting parameter sequences. In the subsequent sections, we will also consider a more general time-varying version of the response model (2.1), which is expressed as follows:

$$y_t \;=\; f(x_t, \theta_t) + \epsilon_t \quad \text{for } t = 1, 2, \ldots, \tag{2.5}$$

where $\boldsymbol{\theta} = \{\theta_t,\, t = 1, 2, \ldots\}$ is a sequence of unknown model parameters taking values in $\Theta \subseteq \mathbb{R}$. Replacing θ with $\{\theta_t\}$ in all preceding response equations, one obtains the time-varying counterparts of our learning problem in static environments. In these time-varying environments, we use the probability measure $\mathbb{P}_{\boldsymbol{\theta}}$ with density $h_{\boldsymbol{\theta}}(y_1, \ldots, y_t) = \prod_{s=1}^{t} h_\epsilon\big(y_s - f(x_s, \theta_s)\big)$ for $y_1, \ldots, y_t \in \mathbb{R}$. We measure the inaccuracy of the estimates $\hat{\theta}_t$ as normalized deviations from unity,

$$\Delta_t \;:=\; \left| 1 - \frac{\hat{\theta}_t}{\theta_t} \right|, \tag{2.6}$$

where $\theta_t \neq 0$ for all t. In settings where $\{\theta_t\}$ is static, the convergence of $\{\hat{\theta}_t\}$ to the true value of the unknown parameter θ (in which case we say $\hat{\theta}_t$ is *consistent*) is tantamount to $\{\Delta_t \to 0\}$. Given $\varepsilon > 0$, we say that the estimate $\hat{\theta}_t$ is *ε-accurate* if

$$\Delta_t \;\leq\; \varepsilon, \tag{2.7}$$

and the estimate sequence $\{\hat{\theta}_t\}$ is *asymptotically ε-accurate* if

$$\mathbb{P}_{\boldsymbol{\theta}}\big\{\hat{\theta}_t \text{ is } \varepsilon\text{-accurate eventually}\big\} \;=\; \mathbb{P}_{\boldsymbol{\theta}}\big\{ \textstyle\bigcup_{n=1}^{\infty} \bigcap_{t=n}^{\infty} \{\Delta_t \leq \varepsilon\}\big\} \;\geq\; 1 - \varepsilon. \tag{2.8}$$

The preceding definition of asymptotic accuracy is a basic requirement for any consistent estimator in a static environment, and reflects our focus on the pathwise properties of said estimates. As will be shown below, there exist several different examples in which $\{\Delta_t\}$ fails to converge to zero.

2.2 Incomplete learning and certainty-equivalence

The dynamical system in (2.2-2.3) is induced by an iterative process of estimation and optimization. But, the estimation and optimization steps of this process are executed in isolation, i.e., we estimate the unknown parameter as if there were no optimization of controls and we choose the controls as if there were no estimation. For brevity, we call the dynamical system in (2.2-2.3) the *certainty-equivalence learning policy* and denote it by $\mathcal{C}$. A fundamental question concerning this policy is whether learning "takes care of itself" if we carry out estimation and optimization in isolation. To that end, consider the following illustrative example where the unknown parameter is fixed over time.

Example 1: A static environment. *Assume that* $f(x, \theta) = \theta x$ *for all* $x \in \mathcal{X} = \mathbb{R}$ *and* $\theta \in \Theta = \mathbb{R}$, *and that* $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$ *with* $\sigma = 3$. *Let* $\{\theta_t,\, t = 1, 2, \ldots\}$ *be a constant sequence with* $\theta_t = 2.5$ *for all* t. *The decision maker sets the initial control as* $x_1 = 1$, *and subsequently uses the control function* $\psi(\theta) = -1 + \theta$.

Figure 1 displays sample paths of $\{\hat{\theta}_t\}$ under the certainty-equivalence learning policy $\mathcal{C}$ in Example 1. Interestingly, a substantial portion of the sample paths converge to a parameter value that is different from the true value of the unknown parameter.

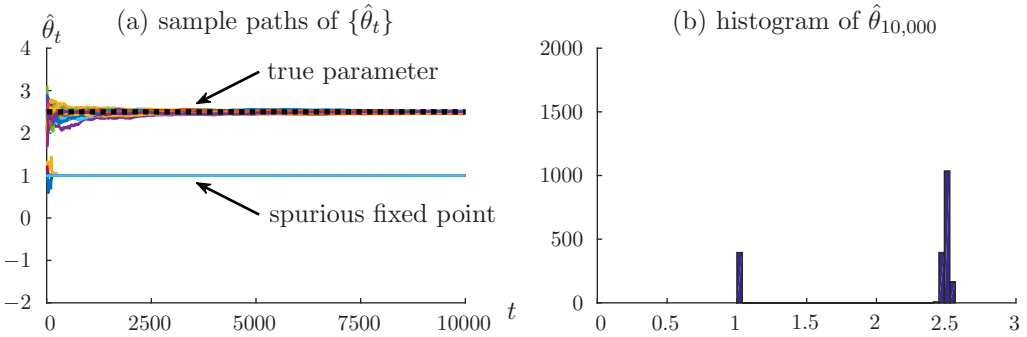

Figure 1: **Certainty-equivalence estimates in another static environment.** Panels (a) and (b) depict sample paths of the estimate sequence $\{\hat{\theta}_t\}$, and the histogram of the estimate in period 10,000, respectively, generated under the certainty-equivalence learning policy $\mathcal{C}$ in Example 1. There are 2,000 sample paths in total, and on 20% of the sample paths, $\{\hat{\theta}_t\}$ converges to 1. The separation in the distribution of estimates is stable after period 10,000; when we extend the graph in panel (a) up to t = 30,000, we observe that the same 20% of the sample paths remain around a small neighborhood of 1.

The behavior in Figure 1 suggests that certainty-equivalence control can "stop learning" prematurely with positive probability. This phenomenon, which we call *incomplete learning*, is an extreme form of asymptotic inaccuracy. To formally define the incomplete learning phenomenon, let us consider how information is collected in the linear-Gaussian setting of Example 1. The choice of the control x_t determines how fast the information accumulates. If $x_t = 0$, then $f_\theta(x_t, \theta) = x_t = 0$ and the estimation equation (2.2) implies that $\hat{\theta}_{t+1} = \hat{\theta}_t$; i.e., the estimate stays the same in the following period. Letting $\zeta = \psi^{-1}(0)$, we note that if $\hat{\theta}_t = \zeta$ at some period t then $x_t = \psi(\hat{\theta}_t) = \psi(\zeta) = 0$, implying that $\hat{\theta}_{t+1} = \hat{\theta}_t$. Repeating this argument we deduce that whenever $\hat{\theta}_t = \zeta$ for some t we have $\hat{\theta}_s = \zeta$ for all $s > t$; that is, the estimate sequence $\{\hat{\theta}_t\}$ becomes permanently "stuck" at the fixed point ζ of the dynamical system of estimation and control. In light of this, we hereafter call ζ the *uninformative estimate*. Accordingly, we refer to $\psi(\zeta) = 0$ as the *uninformative control*. (This definition of the uninformative estimate ζ extends to a fixed point of the general dynamical system in (2.2-2.3); see §6.1 for details). We define incomplete learning as the convergence of the estimate sequence $\{\hat{\theta}_t\}$ to the uninformative estimate ζ.

Definition (incomplete learning) *The sequence of estimates* $\{\hat{\theta}_t\}$ *is said to exhibit incomplete learning if* $\hat{\theta}_t \to \zeta$ *with positive probability and* $\{\theta_t\}$ *does not converge to* ζ *as* $t \to \infty$.

Remark 2 (static case) In the case where θ_t is constant and equal to $\theta \neq \zeta$, the above definition of incomplete learning implies that the sequence of estimates $\{\hat{\theta}_t\}$ fails to converge to θ; in other words, the estimator sequence is not consistent.

The *reachability* of ζ by the estimates of the certainty-equivalence learning policy plays a key role in incomplete learning and asymptotic accuracy.

Definition (reachability of the uninformative estimate) *Let $\delta > 0$ and $\varepsilon \in [0,1]$. The uninformative estimate ζ is said to be δ-reachable by $\{\hat{\theta}_t\}$ with probability ε if*

$$\mathbb{P}_{\boldsymbol{\theta}}\big\{|\hat{\theta}_t - \zeta| \leq \delta \text{ for some } t = 1, 2, \dots\big\} \;=\; \mathbb{P}_{\boldsymbol{\theta}}\big\{\textstyle\bigcup_{t=1}^{\infty}\{|\hat{\theta}_t - \zeta| \leq \delta\}\big\} \;\geq\; \varepsilon. \tag{2.9}$$

Note that despite the certainty-equivalence learning policy $\mathcal{C}$ being designed primarily for static environments, it can be also used in changing environments. Motivated by the case of a decision maker who is oblivious to changes in $\{\theta_t\}$, we are also interested in how $\mathcal{C}$ would perform if $\{\theta_t\}$ can change over time.

Remark 3 We would like to note that inconsistency of parameter estimates can arise also due to the empirical objective function (e.g., the sum of squared residuals) being multimodal. In this paper we do not consider this potential source of incomplete learning and hence restrict attention to settings where the empirical objective defining the estimator has a unique optimizer given by (2.2) and the estimates of the unknown parameter can be uniquely computed in every period of the problem. (See also §7 for a discussion that extends least squares to general M-estimation.)

3 Illustrative Examples of the Model

In this section, we present examples of the model presented in §2, with explicit forms of the response function $f(\cdot\,,\cdot)$ and the control function $\psi(\cdot)$. As will be explained below, the antecedent work on these examples have almost exclusively focused on static environments; time-varying generalizations of such examples can be constructed by modifying the response as in (2.5). In all of these examples, the asymptotic estimation accuracy plays an important role in determining whether the decision maker can ultimately identify $\psi(\theta)$, the ideal control under perfect information on the unknown parameter θ. To be precise, if the decision maker's estimate sequence $\{\hat{\theta}_t\}$, which is computed via (2.2), does not converge to θ then the control sequence $\{x_t = \psi(\hat{\theta}_t),\, t = 1, 2, \dots\}$ would fail to converge to $\psi(\theta)$. This demonstrates that the inaccuracy of estimates defined in (2.6) is a relevant performance metric in all the examples below. Consequently, as an extreme form of asymptotic inaccuracy, incomplete learning is pertinent to these examples. The results in §6 provide a method that eliminates any possibility of incomplete learning under the certainty-equivalence learning policy $\mathcal{C}$ in these settings.

Dynamic control for eliciting a target response. Let $y_1, y_2, \dots$ be a sequence of response variables satisfying $y_t = \theta x_t + \epsilon_t$ for $t = 1, 2, \dots$, where: $\theta \in \Theta = [\theta_{\min}, \theta_{\max}]$ is an unknown

parameter, $0 < \theta_{\min} < \theta_{\max} < \infty$, $\{\epsilon_t\}$ are independent and identically distributed random variables with zero mean and variance σ^2, and $x_t \in \mathcal{X} = \mathbb{R}$. The decision maker sequentially chooses $x_1, x_2, \ldots$ to bring $y_1, y_2, \ldots$ as close as possible to some target value y^*. Here the estimation (2.2) is the projection onto Θ of the ordinary least squares estimate, and the control function is given by $\psi(\theta) = y^*/\theta$. In period t, given the projected least squares estimate $\hat{\theta}_t$, the control is $x_t = y^*/\hat{\theta}_t$. On paths where $\Delta_t = \varepsilon > 0$, x_t is either $y^*/\big((1-\varepsilon)\theta\big)$ or $y^*/\big((1+\varepsilon)\theta\big)$, and consequently $|y^* - \theta x_t|$ would equal either $y^*\varepsilon/(1-\varepsilon)$ or $y^*\varepsilon/(1+\varepsilon)$. Thus, smaller inaccuracy makes the mean response θx_t closer to y^*. See Prescott (1972), Anderson and Taylor (1976) and Lai and Robbins (1982) for some examples of studies that consider variants of this problem, with the latter study focusing on incomplete learning in this context. The treatment in §4 will further illuminate the incomplete learning phenomenon in such settings.

Stochastic optimization of a quadratic function. Consider a decision maker who observes a sequence of responses $y_1, y_2, \ldots$ such that $y_t = (\theta - a x_t)^2 + \epsilon_t$ for $t = 1, 2, \ldots$, where $a > 0$ is a known constant, $\theta \in \Theta$ is an unknown parameter, and $\{\epsilon_t\}$ are independent and identically distributed random variables with zero mean and variance σ^2, and the control $x_t \in \mathcal{X}$. The decision maker aims to minimize $(\theta - a x_t)^2$ by choosing certainty-equivalence controls $x_1, x_2, \ldots$ in a sequential fashion. Specifically, the estimation (2.2) is given by $\sum_{s=1}^{t} \big(y_s - (\hat{\theta}_{t+1} - a x_s)^2\big)(\hat{\theta}_{t+1} - a x_s) = 0$, and the control function in (2.3) is $\psi(\theta) = \theta/a$, hence in period t the control is $x_t = \psi(\hat{\theta}_t) = \hat{\theta}_t/a$. Note that on paths for which $\Delta_t = \varepsilon > 0$, x_t equals either $(1+\varepsilon)\theta/a$ or $(1-\varepsilon)\theta/a$. In either case, we have $(\theta - a x_t)^2 = \theta^2\varepsilon^2$, meaning that smaller values of inaccuracy Δ_t help the decision-maker achieve her/his goal of minimizing $(\theta - a x_t)^2$. Several variants of the above setting have been studied in the literature, starting with an early paper by Kiefer and Wolfowitz (1952). The examples in §4 and §5 indicate that this procedure is possibly subject to incomplete learning, and as mentioned above, §6 presents a general method of avoiding incomplete learning in this setting.

Dynamic pricing with demand learning. Consider a price-setting monopolist facing an isoelastic demand curve $D(p, \theta) = k p^{-\theta}$. The demand in period t is given by

$$d_t \;=\; D(p_t, \theta)\, e_t \;=\; k p_t^{-\theta}\, e_t \quad \text{for } t = 1, 2, \ldots, \tag{3.1}$$

where: $k > 0$ is a known constant, $p_t > 0$ is the price charged in period t, $\theta \in \Theta = [\theta_{\min}, \theta_{\max}]$ is the price-elasticity of demand, $1 < \theta_{\min} < \theta_{\max} < \infty$, and $e_t \overset{\text{iid}}{\sim} \text{Lognormal}(0, \sigma^2)$ are unobservable multiplicative demand shocks. Taking the logarithm on both sides of (3.1), we obtain the following response model:

$$y_t \;=\; a - \theta x_t + \epsilon_t, \quad \text{for } t = 1, 2, \ldots,$$

where $y_t = \log d_t$, $a = \log k$, $x_t = \log p_t \in \mathcal{X} = \mathbb{R}$, and $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$. Note that the above model is a special case of the general response model (2.1). The monopolist's expected profit can

be expressed as a function of log-price x and elasticity θ as follows:

$$\pi(x,\theta) \;=\; \big(p(x)-c\big)Kp(x)^{-\theta} \;=\; K(e^x-c)\,e^{-\theta x},$$

where $p(x)=e^x$, $K=ke^{\sigma^2/2}>0$, and $c>0$ is the marginal cost of production. In the above setup the estimation (2.2) is the projection onto Θ of the least squares estimate $\sum_{s=1}^t x_s(a-y_s)/\sum_{s=1}^t x_s^2$, and the control (2.3) is given by the profit-maximizing decision $\psi(\theta)=\arg\max_{x\in\mathcal{X}}\{\pi(x,\theta)\}=\log c-\log(1-1/\theta)$. Note that $\psi(\theta)$ is monotone decreasing in θ. Intuitively, this means that if the demand is more price-elastic then the monopolist would charge a lower price, as this would increase profits. To see the impact of estimation inaccuracy on profits, note that there exists a positive constant z_0 such that for all x satisfying $|x-\psi(\theta)|\le z_0$, $\pi(\psi(\theta),\theta)-\pi(x,\theta)\ge a_\theta\big(x-\psi(\theta)\big)^2$, where $a_\theta=\frac{1}{4}Kc^{1-\theta}\theta^{1-\theta}(\theta-1)^\theta$. This implies that if $\Delta_t=\varepsilon\in(0,b_\theta)$ then $\pi(\psi(\theta),\theta)-\pi(\psi(\hat\theta_t),\theta)\ge\tilde a_\theta\varepsilon^2$, where $\tilde a_\theta=a_\theta/\big(2\theta(\theta-1)\big)$ and $b_\theta=\theta(\theta-1)z_0/2$. Thus, to get closer to the maximal profit $\pi(\psi(\theta),\theta)$, the monopolist needs to reduce Δ_t. For an illustration of incomplete learning in a related dynamic pricing setting, see den Boer and Zwart (2014).

Dynamic medical treatment. Consider a physician who sequentially decides on medical treatment levels (e.g., dosage of a drug) for patients. Viewing the response model (2.1) in this healthcare context, the outputs $\{y_t\}$ are sequential responses that reflect the patients' medical condition, the controls $\{x_t\}$ are the treatment levels, the unknown model parameter θ represents the patients' responsiveness to treatment, and $\{\epsilon_t\}$ are temporal shocks that depend on unobservable factors. The treatment levels are chosen from a set $\mathcal{X}=[x_{\min},\infty)$, where $x_{\min}>0$. Suppose that there exists a current medical practice that prescribes a treatment level $x_0\in\mathcal{X}$, and the physician knows the expected response to x_0. In this setting, a simple example for the response curve is $f(x,\theta)=\theta(x-x_0)$. Alternatively, one can consider nonlinear response curves such as $f(x,\theta)=k_1e^{\theta(x-x_0)}+k_2\theta^2(x-x_0)$, where k_1 and k_2 are known constants. To determine the treatment sequence, suppose that the physician uses the estimation (2.2) in conjunction with the control function $\psi(\theta)=x_{\min}+\alpha(\theta-\theta_{\min})$, where $\theta\in\Theta=[\theta_{\min},\infty)$, $\alpha>0$, and $\theta_{\min}\in\mathbb{R}$. This control function prescribes linearly adjusting the treatment level for more responsive patients, where the policy parameter α represents the rate of adjustment in treatment level. (As in our preceding application, it is possible to use a nonlinear control function in this context, and our model accommodates such generality.) Given the value of the unknown parameter $\theta\in\Theta$, the ideal control is $\psi(\theta)$, and on paths where $\Delta_t=\varepsilon$, the physician's absolute deviation from the ideal control is $|x_t-\psi(\theta)|=\alpha\varepsilon$. Hence, to minimize deviations from the ideal control, the physician should decrease Δ_t. As will be seen in §4, this strategy will result in incomplete learning in the case of a linear response curve $f(x,\theta)=\theta(x-x_0)$.

4 The Linear-Gaussian Model

In this section, we focus on a special case of the general response model (2.1) to illustrate the main salient features of the certainty-equivalence learning policy $\mathcal{C}$ and the incomplete learning

phenomenon. To that end, let the expected response curve be linear, $f(x, \theta) = \theta x$, and the noise terms be normally distributed, $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$. Then, we can re-express (2.1) as

$$y_t \;=\; \theta x_t + \epsilon_t \quad \text{for } t = 1, 2, \ldots \tag{4.1}$$

In this case, estimates are computed via ordinary least squares regression, with closed-form expression for $\hat{\theta}_{t+1}$:

$$\hat{\theta}_{t+1} \;=\; \frac{\sum_{s=1}^{t} x_s y_s}{\sum_{s=1}^{t} x_s^2}. \tag{4.2}$$

Because $\{\epsilon_t\}$ are normally distributed, the density of $\mathbb{P}_\theta$ in (2.4) is defined via the Gaussian density in this case. The response model (4.1) represents a static environment in the sense that the unknown parameter θ does not change over time. We will study this static case in the following subsection, and then consider changing environments where the unknown parameter can vary over time.

4.1 Incomplete learning in static environments

Our first task is to formalize the observations in Example 1, which suggests that certainty-equivalence can exhibit incomplete learning in a static environment. We deduce from (4.1) and (4.2) that

$$\hat{\theta}_{t+1} \;=\; \theta + \frac{M_t}{J_t} \quad \text{for } t = 1, 2, \ldots \tag{4.3}$$

where $M_t = \sum_{s=1}^{t} x_s \epsilon_s$ and $J_t = \sum_{s=1}^{t} x_s^2$. Based on the characterization of the estimator in (4.3), our following result shows that there are exactly two possible asymptotic outcomes for the policy $\mathcal{C}$ in a static environment.

Proposition 1 (convergence of estimator in static environments) *Let $\theta \in \mathbb{R}$, and assume that $\theta_t = \theta \neq \zeta$ for $t = 1, 2, \ldots$ Then, for any $\psi(\cdot) \in \Psi$,*

(i) *$\hat{\theta}_t \to \theta$ almost surely on $\{J_\infty = \infty\}$, and*

(ii) *$\hat{\theta}_t \to \zeta$ almost surely on $\{J_\infty < \infty\}$,*

where $\{\hat{\theta}_t\}$ is the sequence of certainty-equivalence estimates generated under $\mathcal{C}$, and $J_\infty = \lim_{t\to\infty} J_t$.

Proposition 1 categorizes the asymptotic learning outcomes based on whether $\{J_t\}$ diverges to ∞. Note that in this setting, J_t can be viewed as a measure of cumulative information, formally called the empirical *Fisher information* accumulated in the first t periods. Proposition 1 states that $\{\hat{\theta}_t\}$ identifies θ if and only if the cumulative information tends to ∞. Therefore, the asymptotic outcomes in a static environment are partitioned into two cases: (i) *consistency*, which occurs if $\{J_t\}$ diverges to ∞, and (ii) *incomplete learning*, which occurs if $\{J_t\}$ converges to a finite limit. By the continuity of $\psi(\cdot)$, we also deduce that the control sequence $\{x_t\}$ converges almost surely to $\psi(\theta)$ on $\{J_\infty = \infty\}$. However, on the event $\{J_\infty < \infty\}$, $\{x_t\}$ converges almost surely to the uninformative control $\psi(\zeta) = 0$, which is not necessarily equal to $\psi(\theta)$. The proof of Proposition 1 is based on showing that M_t is a square-integrable martingale and then applying the strong law of large numbers for martingales (see also Lai and Wei 1982, for a related application).

Our next result shows that in a static environment, $\{\hat{\theta}_t\}$ exhibits incomplete learning under the

certainty-equivalence learning policy $\mathcal{C}$.

Theorem 1 (incomplete learning in static environments) *Let $\psi(\cdot) \in \Psi$, $\theta \in \mathbb{R}$, and assume that $\theta_t = \theta \neq \zeta$ for $t = 1, 2, \ldots$ Then $\mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \to \zeta\} > 0$, where $\{\hat{\theta}_t\}$ is the sequence of certainty-equivalence estimates generated under $\mathcal{C}$.*

To see the intuition behind the incomplete learning result in Theorem 1, note that the decision rule in (2.3) creates a temporal dependency within the control sequence $\{x_t,\, t = 1, 2, \ldots\}$. If $\{x_t\}$ approaches the uninformative control, then the "signal quality" of the responses in (4.1) diminishes, and the learning slows down, thereby creating further tendency to choose a control in the vicinity of the uninformative control, leading to an estimate close to the uninformative estimate ζ. This vicious cycle leads the dynamical system to be attracted to the fixed point of incomplete learning.

An important consequence of Theorem 1 is the poor accuracy of the certainty-equivalence learning policy $\mathcal{C}$, which is expressed in the following result.

Corollary 1 (accuracy in static environments) *Let $\psi(\cdot) \in \Psi$, $\theta \in \mathbb{R}$, and assume that $\theta_t = \theta \neq \zeta$ for $t = 1, 2, \ldots$ Then there exists a positive constant δ such that the sequence of certainty-equivalence estimates $\{\hat{\theta}_t\}$ generated under $\mathcal{C}$ is not asymptotically ε-accurate for any $\varepsilon \in (0, \delta)$.*

The preceding result, in conjunction with Theorem 1, states that the eventual inaccuracy of $\{\hat{\theta}_t\}$ will stay above a certain positive value, namely $|1 - \zeta/\theta|$, with a positive probability $p_0 = \mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \to \zeta\}$. Letting $\delta = \min\{|1 - \zeta/\theta|, p_0\}$, we deduce that $\{\hat{\theta}_t\}$ is not asymptotically ε-accurate for any $\varepsilon < \delta$.

Further discussion of Example 1. The above analysis of incomplete learning helps us view Example 1 in a new light. In that example, the uninformative estimate is $\zeta = 1$. As shown in Figure 1, about one fifth of all sample paths of $\{\hat{\theta}_t\}$ converge to ζ in this setting, providing a numerical example of the incomplete learning result in Theorem 1. We can also measure the accuracy performance of $\mathcal{C}$ in Example 1. Figure 2 displays sample paths of the inaccuracy process $\{\Delta_t, t = 1, 2, \ldots\}$ under $\mathcal{C}$.

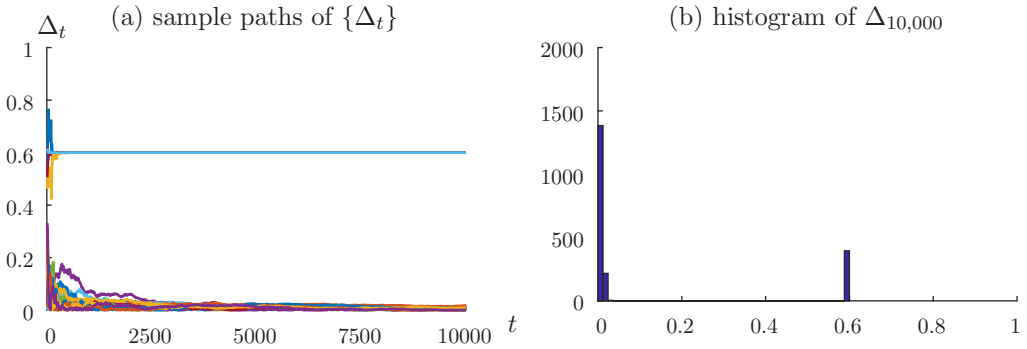

Figure 2: **Inaccuracy of certainty-equivalence learning in a static environment.** Panels (a) and (b) show sample paths of the inaccuracy process $\{\Delta_t\}$, and the histogram the inaccuracy in period 10,000, respectively, generated under the certainty-equivalence learning policy $\mathcal{C}$ in Example 1. In the first 10,000 periods, approximately 20% of the 2,000 sample paths converge to a positive value, namely 0.6.

Note that approximately 20% of the generated sample paths in Figure 2 end up with an inaccuracy of $|1 - \zeta/\theta| = |1 - 1/2.5| = 0.60$. Hence, we estimate that there is $p_0 = 0.20$ probability that the eventual inaccuracy of $\{\hat{\theta}_t\}$ will be more than 0.20 in Example 1. As a result, $\{\hat{\theta}_t\}$ is not asymptotically ε-accurate for any ε less than $\delta = 0.20$.

No uninformative estimate implies no incomplete learning. In a static environment with an uninformative estimate $\zeta \in \Theta$, we observe incomplete learning because ζ is reachable by the certainty-equivalence estimates $\{\hat{\theta}_t\}$ with some positive probability. But, if Θ is a strict subset of $\mathbb{R}$, there might be no uninformative estimate ζ in Θ. Note that, because $|\psi'(\theta)| \geq \ell > 0$ for all $\theta \in \Theta$, an uninformative estimate $\psi^{-1}(0)$ exists in $\mathbb{R}$, but may be outside Θ. The following result complements Theorem 1 by investigating settings where Θ is a strict subset of $\mathbb{R}$ containing no uninformative estimates.

Proposition 2 (no uninformative estimate implies no incomplete learning) *Let $\psi(\cdot) \in \Psi$, and $\theta \in \Theta \subseteq \mathbb{R}$. Assume that $\theta_t = \theta$ for $t = 1, 2, \ldots$, and there does not exist any $\zeta \in \Theta$ satisfying $\psi(\zeta) = 0$. Then, $\mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \to \theta\} = 1$, where $\{\hat{\theta}_t = \arg\min_{\theta \in \Theta} S_{t-1}(\theta),\ t = 2, 3, \ldots\}$ is the sequence of certainty-equivalence estimates generated under $\mathcal{C}$.*

Proposition 2 states that, in a static environment where there is no reachable uninformative estimate in Θ, the certainty-equivalence learning policy $\mathcal{C}$ is consistent. In what follows, we will further study the connection between incomplete learning and the reachability of the uninformative estimate to explore the broader extent of the incomplete learning phenomenon.

4.2 Incomplete learning in changing environments

4.2.1 A boundedly changing environment

We will now investigate a slight modification of Example 1 by letting $\{\theta_t\}$ vary within a *bounded interval* in a cyclical fashion, with the implicit question whether this type of change will prevent the system (4.1-4.2) into settling into the uninformative estimate and control.

Example 2: A boundedly changing environment. *Assume that $f(x, \theta) = \theta x$ for all $x \in \mathcal{X} = \mathbb{R}$ and $\theta \in \Theta = \mathbb{R}$, and that $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$ with $\sigma = 3$. Let $\mathcal{T}_+ = \bigcup_{k=0}^{\infty} \bigcup_{n=1}^{2000} \{4000k + n\}$, and $\{\theta_t,\ t = 1, 2, \ldots\}$ be such that $\theta_1 = 2.5$ and*

$$\theta_{t+1} - \theta_t \ = \ \begin{cases} +0.001 & \text{if } \ t \in \mathcal{T}_+ \\ -0.001 & \text{otherwise,} \end{cases}$$

for all $t \geq 1$. The decision maker sets the initial control as $x_1 = 1$, and subsequently uses the control function $\psi(\theta) = -1 + \theta$.

Compared to the original (static) example, Example 2 poses a slightly more difficult learning problem since the unknown parameter sequence $\{\theta_t\}$ keeps changing over time. As portrayed in Figure 3, allowing the unknown parameter to fluctuate within a bounded interval, we still observe that $\{\hat{\theta}_t\}$ converges to the spurious fixed point 1 on 20% of the sample paths as in Example 1.

Thus, the incomplete learning result we observed in Example 1 persists in Example 2.

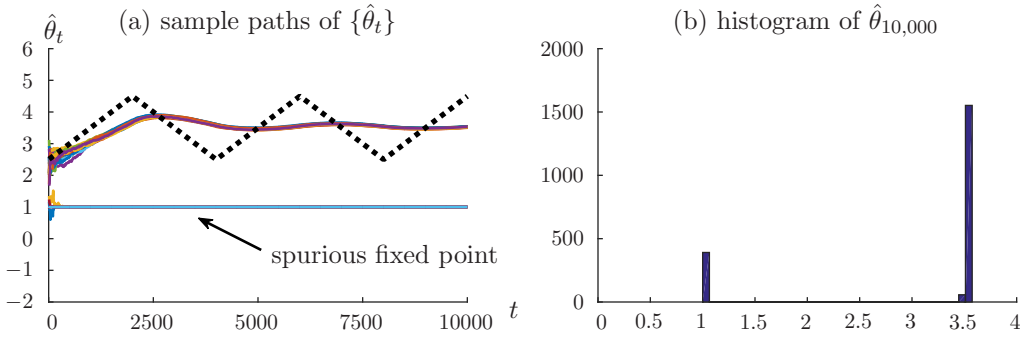

Figure 3: **Certainty-equivalence estimates in a boundedly changing environment.** Panels (a) and (b) depict sample paths of the estimate sequence $\{\hat{\theta}_t\}$ (*solid curves*), and the histogram of the estimate in period 10,000, respectively, generated under the certainty-equivalence learning policy $\mathcal{C}$ in Example 2. There are 2,000 sample paths in total. The values of $\{\theta_t\}$ are shown in the *dotted curve* in panel (a).

Our next task is to formalize the observation in Example 2, namely, that there is not sufficient temporal change in $\{\theta_t\}$ to avoid incomplete learning. For this purpose, in the next result, we extend Theorem 1 to environments that fluctuate in a bounded fashion.

Theorem 2 (incomplete learning in boundedly fluctuating environments) *Let $\psi(\cdot) \in \Psi$, and assume that*

$$\zeta - \kappa_1 \;\leq\; \theta_t \;\leq\; \zeta + \kappa_2 \qquad \text{for } t = 1, 2, \ldots,$$

and that $\theta_t \geq \zeta + \kappa_0$ eventually, where $\infty < -\kappa_1 < 0 < \kappa_0 < \kappa_2 < \infty$. Then $\mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \to \zeta\} > 0$, where $\{\hat{\theta}_t\}$ is the sequence of certainty-equivalence estimates generated under $\mathcal{C}$.

Remark 4 In the hypothesis of the preceding theorem, the condition that $\theta_t \geq \zeta + \kappa_0$ for sufficiently large t ensures that $\{\theta_t\}$ will eventually be confined to a bounded interval on one side of ζ. If this condition is violated and $\{\theta_t\}$ is allowed to visit both sides of ζ infinitely often, then it is possible to construct an example in which $\{\hat{\theta}_t\}$ fluctuates perpetually and moves arbitrarily close to ζ without converging to ζ (see Example 7 in Appendix C).

Discussion of the incomplete learning phenomenon. Theorem 2 shows that if the unknown parameter sequence $\{\theta_t\}$ is fluctuating within lower and upper bounds that are independent of time (as in Example 2), then there is a positive probability that the sequence of estimates $\{\hat{\theta}_t\}$ converges to the uninformative estimate ζ, and incomplete learning exists (with positive probability) under $\mathcal{C}$. As explained in the preceding subsection, incomplete learning depends on: (i) the reachability of the uninformative estimate ζ within the space of estimates; and (ii) the diminishing signal quality near ζ. Theorem 2 shows that, in boundedly fluctuating environments, ζ is still reachable by the certainty-equivalence estimates of $\mathcal{C}$ with positive probability, and the quality of the signals can diminish as in static environments.

Noting that the expected response $f(x_t, \theta_t)$ depends on two variables, namely the control x_t and the unknown parameter θ_t, we can employ the above analysis to compare how the changes in these

two variables affect incomplete learning under $\mathcal{C}$. To that end, let $V_{\boldsymbol{\theta}}(t) = \sum_{s=1}^{t}(\theta_s - \zeta)^2$ be the cumulative quadratic deviation of the parameter sequence from the uninformative estimate, and $V_{\boldsymbol{x}}(t) = \sum_{s=1}^{t}\big(x_s - \psi(\zeta)\big)^2$ be the cumulative quadratic deviation of the control sequence from the uninformative control. In the antecedent literature, it has been shown that if $V_{\boldsymbol{x}}(t)$ is linearly increasing in t then incomplete learning will not occur (see, e.g., Keskin and Zeevi 2014, §3.4). Based on this, one might expect that a similar result would hold for $V_{\boldsymbol{\theta}}(t)$. But, our preceding analysis shows that if $V_{\boldsymbol{\theta}}(t)$ increases linearly in t then incomplete learning persists. This identifies a significant difference in the manner in which the variations in $\{x_t\}$ and $\{\theta_t\}$ affect the incomplete learning phenomenon. While a linearly growing $V_{\boldsymbol{x}}(t)$ can eliminate incomplete learning, a linearly growing $V_{\boldsymbol{\theta}}(t)$ may not ensure a similar result. It is perhaps worth noting that this contrast is present also when the variations in $\{x_t\}$ and $\{\theta_t\}$ are measured as deviations from the historical average. Letting $\bar{V}_{\boldsymbol{\theta}}(t) = \sum_{s=1}^{t}(\theta_s - \bar{\theta}_t)^2$ and $\bar{V}_{\boldsymbol{x}}(t) = \sum_{s=1}^{t}(x_s - \bar{x}_t)^2$, where $\bar{\theta}_t = t^{-1}\sum_{s=1}^{t}\theta_s$ and $\bar{x}_t = t^{-1}\sum_{s=1}^{t}x_s$, we note that linear growth of $\bar{V}_{\boldsymbol{x}}(t)$ helps avoid incomplete learning (see Keskin and Zeevi 2014) whereas linear growth of $\bar{V}_{\boldsymbol{\theta}}(t)$ does not (as in Example 2 and Theorem 2). As a simple illustration of the above contrast, consider the piecewise-linear cyclical pattern of $\{\theta_t\}$ in Example 2. If $\{x_t\}$ follows a similar cyclical pattern, then (as explained above) there will be no incomplete learning. But, when $\{\theta_t\}$ exhibits a cyclical pattern as in Example 2, there is still incomplete learning.

4.2.2 A more volatile changing environment

Theorem 2 demonstrates that merely the existence of a changing environment is not sufficient for avoiding the incomplete learning phenomenon. Now, given that incomplete learning persists in *boundedly* changing environments, what happens in *unboundedly* fluctuating environments? To investigate this question, let us now consider an environment where the unknown parameter sequence $\{\theta_t\}$ changes in an unbounded and volatile fashion.

Example 3: A volatile environment. *Assume that* $f(x,\theta) = \theta x$ *for all* $x \in \mathcal{X} = \mathbb{R}$ *and* $\theta \in \Theta = \mathbb{R}$, *and that* $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0,\sigma^2)$ *with* $\sigma = 3$. *Let* $\{\theta_t,\, t = 1,2,\ldots\}$ *be a sequence such that* $\theta_t = \sum_{s=1}^{t}\xi_s$ *for all* t, *where* $\xi_t \overset{\text{iid}}{\sim} \text{Normal}(0,1)$. *The decision maker sets the initial control as* $x_1 = 1$, *and subsequently uses the control function* $\psi(\theta) = -1 + \theta$.

Figure 4 depicts the estimates under $\mathcal{C}$ in Example 3, where $\{\theta_t\}$ evolves as an unobservable random walk process. Because such a process would drift towards ζ infinitely often, the signal quality of the observations would decrease infinitely often, and ζ will be reachable by the estimates of $\mathcal{C}$. As a negative consequence of this fact, we observe that the probability of $\{\hat{\theta}_t\}$ converging to ζ increases dramatically in Example 3: compared to the 20% likelihood of incomplete learning in Example 1 (see Figure 1), we now estimate a 45% chance of incomplete learning (see Figure 4).

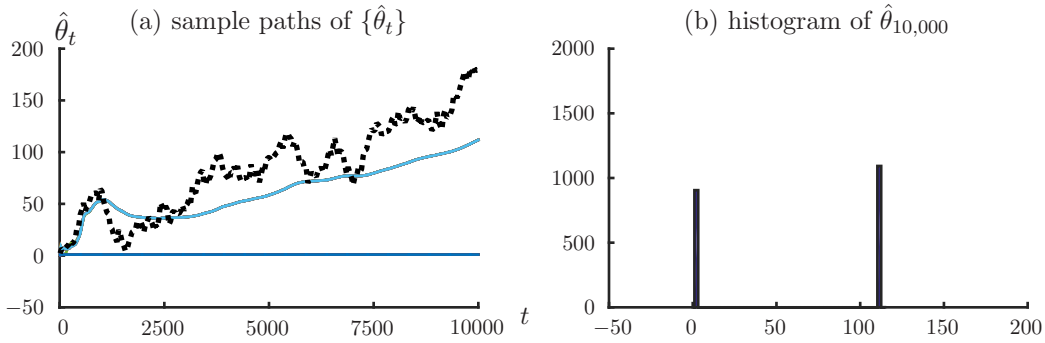

Figure 4: **Certainty-equivalence estimates in a volatile environment.** Panels (a) and (b) depict sample paths of the estimate sequence $\{\hat{\theta}_t\}$ (*solid curves*), and the histogram of the estimate in period 10,000, respectively, generated under $\mathcal{C}$ in Example 3. There are 2,000 sample paths in total. The values of $\{\theta_t\}$ are shown in the *dotted curve* in panel (a). Approximately 45% of the sample paths converge to ζ.

4.2.3 Environments drifting away from the uninformative estimate

Combining our observations in Examples 2 and 3, we note that: (i) bounded fluctuations in $\{\theta_t\}$ are not sufficient to render the uninformative estimate unreachable by the certainty-equivalence estimates; and (ii) making the fluctuations in $\{\theta_t\}$ unbounded and volatile does not necessarily render the uninformative estimate unreachable, as long as $\{\theta_t\}$ can drift towards ζ. Given these observations, we will now study environments where $\{\theta_t\}$ drifts away from ζ. To that end, consider the following example.

Example 4: A slowly and unboundedly changing environment. *Assume that* $f(x,\theta) = \theta x$ *for all* $x \in \mathcal{X} = \mathbb{R}$ *and* $\theta \in \Theta = \mathbb{R}$*, and that* $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0,\sigma^2)$ *with* $\sigma = 3$*. Let* $\{\theta_t,\, t = 1,2,\ldots\}$ *be an increasing sequence such that* $\theta_t = 1 + \sqrt{8\log(t+1)}$ *for all* t*. The decision maker sets the initial control as* $x_1 = 1$*, and subsequently uses the control function* $\psi(\theta) = -1 + \theta$*.*

In Example 4, $\{\theta_t\}$ keeps increasing without an upper bound. Somewhat surprisingly, the incomplete learning seems to be barely visible in this example. As seen in Figure 5, more than 96% of the sample paths keep track of the changing parameter sequence.

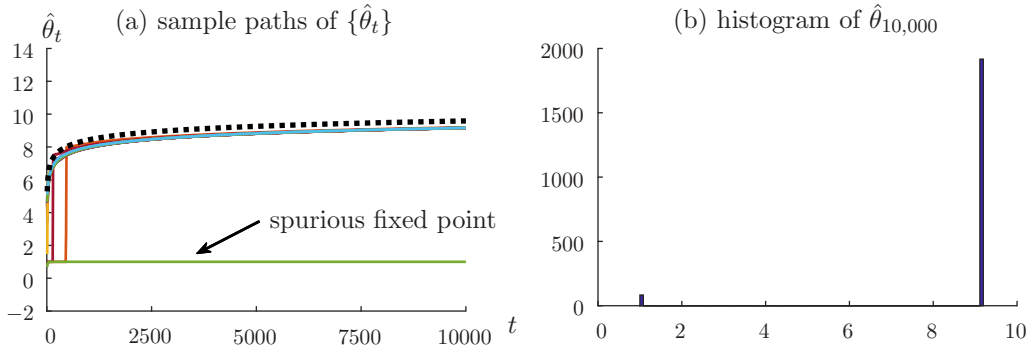

Figure 5: **Certainty-equivalence estimates in a slowly and unboundedly changing environment.** Panels (a) and (b) depict sample paths of the estimate sequence $\{\hat{\theta}_t\}$ (*solid curves*), and the histogram of the estimate in period 10,000, respectively, generated under the certainty-equivalence learning policy $\mathcal{C}$ in Example 4. There are 2,000 sample paths in total. The values of $\{\theta_t\}$ are shown in the *dotted curve* in panel (a).

To characterize settings as in Example 4, let $\{\theta_t^l,\, t = 1, 2, \ldots\}$ and $\{\theta_t^h,\, t = 1, 2, \ldots\}$ be two sequences that respectively designate lower and upper bounds for the unknown parameter sequence $\{\theta_t,\, t = 1, 2, \ldots\}$. We will assume that $\{\theta_t\}$ essentially takes values between the lower and upper bound processes $\{\theta_t^l\}$ and $\{\theta_t^h\}$, allowing for some violation of these bounds in the following sense.

Definition (evolution between lower and upper bound processes with tolerance) *Let* $\rho_{\boldsymbol{\theta}}^- = \sum_{t=1}^{\infty} \max\{\theta_t^l - \theta_t,\, 0\}$ *be the cumulative violation of the lower bound process* $\{\theta_t^l\}$*, and* $\rho_{\boldsymbol{\theta}} = \sum_{t=1}^{\infty} \max\{\theta_t - \theta_t^h,\, \theta_t^l - \theta_t,\, 0\}$ *be the cumulative violation of the lower and upper bound processes* $\{\theta_t^l\}$ *and* $\{\theta_t^h\}$*. Given* $R \geq 0$*, an unknown parameter sequence* $\{\theta_t\}$ *is said to evolve above* $\{\theta_t^l\}$ *with tolerance* R *if* $\rho_{\boldsymbol{\theta}}^- \leq R$*. In addition, if* $\rho_{\boldsymbol{\theta}} \leq R$*, then* $\{\theta_t\}$ *is said to evolve between* $\{\theta_t^l\}$ *and* $\{\theta_t^h\}$ *with tolerance* R*.*

Our next result covers a family of unboundedly changing environments in which the probability that $\{\hat{\theta}_t\}$ converges to ζ is suitably small.

Proposition 3 (learning in a slowly changing environment) *Let* $\psi(\cdot) \in \Psi$*,* $\varepsilon \in (0, \frac{1}{2})$*, and*

$$\theta_t^l \;=\; \zeta + \sqrt{\kappa_1 \log(t+1)}\,, \tag{4.4}$$

for $t = 1, 2, \ldots$*, where* $\kappa_1 \geq 32\sigma\sqrt{2\log(4/\varepsilon)}\big/(\ell \log 2)$*. Assume that* $\{\theta_t\}$ *evolves above the lower bound process* $\{\theta_t^l\}$ *with tolerance* $R \leq \varepsilon\sqrt{\kappa_1}\log\big(\frac{1-\varepsilon}{1-\varepsilon/2}\big)\big/\big(128\log(1-r)\big)$ *where* $r = 2^{-\ell^2\kappa_1^2\log 2/(512\sigma^2)}$*. Then there exists a positive constant* δ *such that the sequence of certainty-equivalence estimates* $\{\hat{\theta}_t\}$ *generated under* $\mathcal{C}$ *satisfies*

$$\mathbb{P}_{\boldsymbol{\theta}}\big\{\hat{\theta}_t \geq \zeta + \delta \text{ for } t = 1, 2, \ldots\big\} \;\geq\; 1 - \varepsilon. \tag{4.5}$$

Remark 5 The hypothesis of Proposition 3 describes a minimum rate at which $\{\theta_t\}$ moves away from ζ in the positive direction. By symmetry, we arrive at a similar conclusion if $\{\theta_t\}$ moves away from ζ at the same rate, but in the opposite direction: if $\theta_t^h = \zeta - \sqrt{\kappa_1 \log(t+1)}$ for all t, and $\sum_{t=1}^{\infty} \max\{\theta_t - \theta_t^h,\, 0\} \leq R$, then the conclusion of Proposition 3 becomes $\mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \leq \zeta - \delta \text{ for all } t\} \geq 1 - \varepsilon$ for the constants κ_1 and δ given above. We also note that, as $\varepsilon \to 0$ in Proposition 3, the upper bound on R converges to zero, while the constant δ approaches $\frac{1}{4}\sqrt{\kappa_1 \log 2}$.

The lower bound in (4.4) describes a sufficient condition for the existence of $\delta > 0$ such that the uninformative estimate ζ is not δ-reachable with probability at least $1 - \varepsilon$. (The particular sub-logarithmic growth rate is an artifact of our proof technique; generalized growth conditions for tracking and asymptotic accuracy are discussed in Theorem 7 in §6, as well as in §7). An important special case of the above result is $R = 0$, where $\{\theta_t\}$ moves away from ζ strictly above $\{\theta_t^l\}$. With $R > 0$, $\{\theta_t\}$ is allowed to move towards ζ with an eventually diminishing frequency.

Unlike the environments in Examples 2 and 3, certain changing environments (in which $\{\theta_t\}$ drifts away from ζ at a critical rate) can render the uninformative estimate essentially unreachable. Proposition 3 spells out a condition on the unknown parameter sequence $\{\theta_t\}$ that keeps the

estimate sequence $\{\hat{\theta}_t\}$ away from ζ with high probability. This makes the incomplete learning result, in which $\{\hat{\theta}_t\}$ converges to ζ, very unlikely under said condition. The main intuition behind this result is the following. If $\{\theta_t\}$ moves away from ζ, then the signal quality of observations will gradually increase because the relative magnitude of noise terms will decay. With higher signal quality, it is less likely that the sequence of estimates $\{\hat{\theta}_t\}$ induced by the certainty-equivalence learning policy will converge to the uninformative estimate. As a result, a changing environment can help avoid incomplete learning if it makes the uninformative estimate ζ gradually less reachable.

Our next goal is to study the implications of Proposition 3 on the accuracy of $\{\hat{\theta}_t\}$ in changing environments. To that end, we first decompose the estimation inaccuracy into two terms.

Proposition 4 (decomposition of estimation inaccuracy) *For any parameter sequence $\{\theta_t\}$, and $\psi(\cdot) \in \Psi$,*

$$1 - \frac{\hat{\theta}_{t+1}}{\theta_{t+1}} \;=\; \sum_{k=1}^{t} \frac{J_k}{J_t} \cdot \frac{\theta_{k+1} - \theta_k}{\theta_{t+1}} \;-\; \frac{M_t}{\theta_{t+1} J_t} \tag{4.6}$$

for $t = 1, 2, \ldots$, where $M_t = \sum_{s=1}^{t} x_s \epsilon_s$ and $J_t = \sum_{s=1}^{t} x_s^2$.

Remark 6 The preceding proposition extends the estimation equation (4.3) to changing environments; note that if $\theta_{k+1} = \theta_k$ for all k, then (4.6) reduces to (4.3).

The above decomposition provides a key insight into the accuracy of $\{\hat{\theta}_t\}$: the first term on the right hand side of (4.6) is influenced by the changes in the unknown parameter sequence $\{\theta_t\}$, while the second is driven by estimation noise. If $\{J_t\}$ grows at a sufficiently fast rate, the second term will vanish eventually. On the other hand, the magnitude of the first term (i.e., the effect of changing environment) is influenced by not only the growth rate of $\{J_t\}$ but also the changes in $\{\theta_t\}$. Roughly speaking, if $\{\theta_t\}$ drifts away from ζ at a critical rate, then the fraction $(\theta_{k+1} - \theta_k)/\theta_{t+1}$ in (4.6) will eventually offset the growth in $\{J_t\}$, making asymptotic accuracy possible (see the discussion following Theorem 3 for a more formal account).

Using the decomposition in Proposition 4, we show that the sub-logarithmic growth condition in Proposition 3 substantially improves the asymptotic accuracy of estimates, thereby avoiding a negative consequence of incomplete learning.

Theorem 3 (accuracy in a slowly changing environment) *Let $\psi(\cdot) \in \Psi$, $\varepsilon \in (0, \frac{1}{2})$, and*

$$\theta_t^l \;=\; \zeta + \sqrt{\kappa_1 \log(t+1)}\,, \tag{4.7a}$$

$$\theta_t^h \;=\; \zeta + \sqrt{\kappa_2 \log(t+1)}\,, \tag{4.7b}$$

for $t = 1, 2, \ldots$, where $\kappa_1 \geq 32\sigma\sqrt{2\log(4/\varepsilon)}\big/(\ell \log 2)$ and $\kappa_1 \leq \kappa_2 \leq \kappa_1/(1-\varepsilon/8)$. Assume that $\{\theta_t\}$ is eventually nondecreasing and evolves between the lower and upper bound processes $\{\theta_t^l\}$ and $\{\theta_t^h\}$ respectively, with tolerance $R \leq \varepsilon\kappa_1 \log\left(\frac{1-\varepsilon}{1-\varepsilon/2}\right)\big/\left(128\sqrt{\kappa_2}\log(1-r)\right)$ where $r = 2^{-\ell^2\kappa_1^2 \log 2/(512\sigma^2)}$. Then the sequence of certainty-equivalence estimates $\{\hat{\theta}_t\}$ generated under $\mathcal{C}$ is asymptotically ε-accurate.

Remark 7 The upper bound condition in Theorem 3 can be replaced by a total variation condition as follows. For all $s < t$, let $\mathcal{P}(s,t)$ be the set of all partitions of $\{s, s+1, \ldots, t\}$ and define $\mathcal{V}_{\boldsymbol{\theta}}(s,t) = \sup_{\{t_0,t_1,\ldots,t_K\}\in\mathcal{P}(s,t),K\geq 1} \left\{ \sum_{k=1}^{K} |\theta_{t_k} - \theta_{t_{k-1}}| \right\}$. If $\mathcal{V}_{\boldsymbol{\theta}}(s,t) \leq \sqrt{\kappa_2 \log(t+1)} - \sqrt{\kappa_2 \log(s+1)}$ for $s < t$, $\theta_1 \leq \zeta + \sqrt{\kappa_2 \log 2}$, and $\{\theta_t\}$ is eventually nondecreasing and evolves above $\{\theta_t^l\}$ with tolerance R, then θ_t would eventually be bounded above by θ_t^h.

Discussion and numerical illustrations. Theorem 3 states that the asymptotic inaccuracy of $\{\hat{\theta}_t\}$ becomes arbitrarily small in the family of slowly changing environments described in (4.7). This stands in stark contrast to Corollary 1 which proves that the asymptotic inaccuracy of $\{\hat{\theta}_t\}$ is always above a positive constant δ in static environments. The reason for this is the following: in the slowly changing environment given in Theorem 3, Proposition 3 implies that $\{\hat{\theta}_t\}$ remains bounded away from ζ by a positive margin with high probability. On this event, $\{J_t = \sum_{s=1}^{t} x_s^2,\ t = 1, 2, \ldots\}$ diverges to ∞, eliminating any possibility of incomplete learning by Proposition 1. Recalling the decomposition of inaccuracy in Proposition 4, this means that the effect of noise, which is given by the second term on the right hand side of (4.6), converges to zero. If the environment is changing slowly as in Theorem 3, then we can also characterize the maximum and minimum possible growth rates of $\{J_t\}$, and prove that the effect of said change, which is given by the first term on the right hand side of (4.6), becomes very small eventually.

Figure 6 demonstrates the accuracy of the certainty-equivalence learning policy in Example 4, which satisfies the hypotheses of Theorem 3. Observing that the inaccuracy Δ_t becomes less than 0.05 on more than 95% of the sample paths, we can deduce that the estimate sequence $\{\hat{\theta}_t\}$ is asymptotically ε-accurate for $\varepsilon = 0.05$ in this example. This is a significant improvement over the asymptotic inaccuracy of 0.20 observed in Example 1 (see Figure 2).

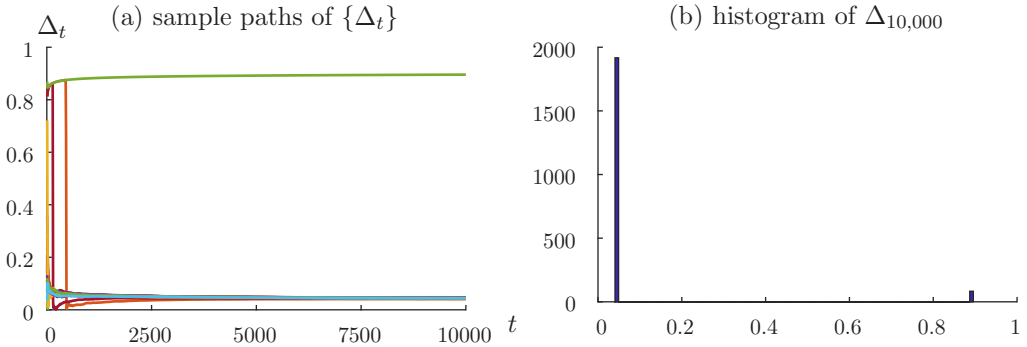

Figure 6: **Inaccuracy of certainty-equivalence learning in a slowly and unboundedly changing environment.** Panels (a) and (b) show sample paths of the inaccuracy process $\{\Delta_t\}$, and the histogram of the inaccuracy in period 10,000, respectively, generated under the certainty-equivalence learning policy $\mathcal{C}$ in Example 4. On approximately 96% of the 2,000 sample paths, the estimate $\hat{\theta}_{10,000}$ is ε-accurate for $\varepsilon = 0.05$.

The improved accuracy of $\{\hat{\theta}_t\}$ in the slowly changing environments described in Theorem 3 leads to another question: how does the certainty-equivalence learning policy behave in more quickly changing environments? Our next example addresses such settings.

Example 5: Another unboundedly changing environment. *Assume that* $f(x,\theta) = \theta x$ *for all* $x \in \mathcal{X} = \mathbb{R}$ *and* $\theta \in \Theta = \mathbb{R}$*, and that* $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0,\sigma^2)$ *with* $\sigma = 3$*. Let* $\{\theta_t,\, t = 1,2,\ldots\}$ *be an increasing sequence such that* $\theta_t = 1 + 2\sqrt{t}$ *for all* t*. The decision maker sets the initial control as* $x_1 = 1$*, and subsequently uses the control function* $\psi(\theta) = -1 + \theta$*.*

As shown in Figure 7, more than 95% of the sample paths of $\{\hat{\theta}_t\}$ avoid incomplete learning in Example 5, tracing the unknown parameter sequence $\{\theta_t\}$.

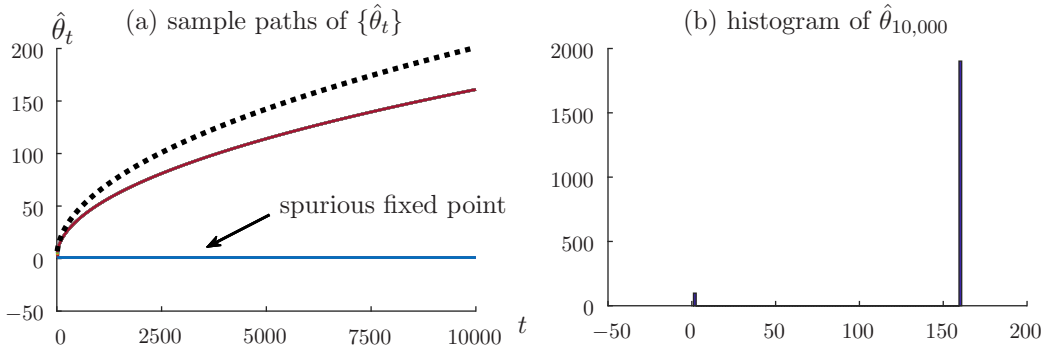

Figure 7: **Certainty-equivalence estimates in another unboundedly changing environment.** Panels (a) and (b) depict sample paths of the estimate sequence $\{\hat{\theta}_t\}$ (*solid curves*), and the histogram of the estimate in period 10,000, respectively, generated under $\mathcal{C}$ in Example 5. There are 2,000 sample paths in total. The values of $\{\theta_t\}$ are shown in the *dotted curve* in panel (a).

Compared to Example 4, $\{\theta_t\}$ moves away from ζ at a faster rate in Example 5, thereby increasing signal quality of observations and thus "helping" the certainty-equivalence learning policy avoid incomplete learning (see also §7 for further discussion of moderately changing environments and how they can facilitate learning).

5 Incomplete Learning in Nonlinear Models

In this section, we extend the analysis of incomplete learning in static environments to a family of nonlinear response models. For purposes of demonstration, let us consider the following example with nonlinear response.

Example 6: A static environment – nonlinear response. *Assume that* $f(x,\theta) = \frac{1}{1+e^{-\theta x}} + \frac{\theta x}{2}$ *for all* $x \in \mathcal{X} = \mathbb{R}$ *and* $\theta \in \Theta = \mathbb{R}$*, and that* $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0,\sigma^2)$ *with* $\sigma = 6$*. Let* $\{\theta_t,\, t = 1,2,\ldots\}$ *be a constant sequence with* $\theta_t = 2.5$ *for all* t*. The decision maker sets the initial control as* $x_1 = 1$*, and subsequently uses the control function* $\psi(\theta) = -1 + \theta$*.*

As shown in Figure 8, the above example exhibits another case of incomplete learning, where $\mathcal{C}$ can stop learning prematurely in static environments with nonlinear response structure.

The response model in Example 6 belongs to a family of nonlinear models called generalized linear models (GLMs). In these models, the response function is the composition of a known link function $g : \mathbb{R} \to \mathbb{R}$ and the linear function $x \mapsto \theta x$, whose parameter θ is unknown to the decision

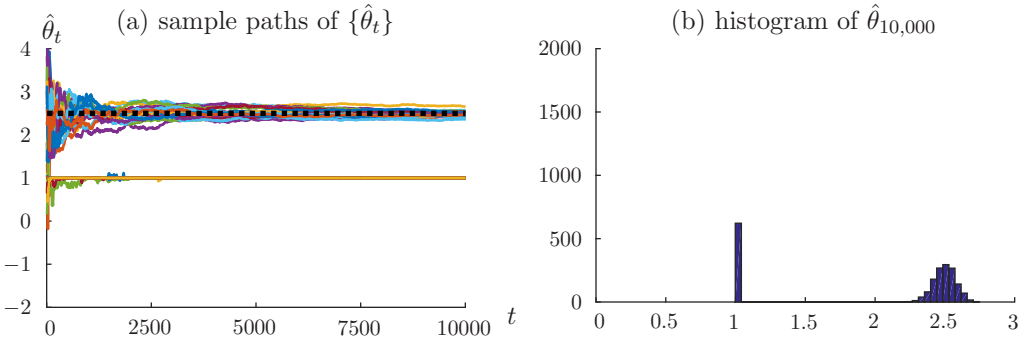

Figure 8: **Certainty-equivalence estimates in a static environment with nonlinear response.** Panels (a) and (b) depict sample paths of the estimate sequence $\{\hat{\theta}_t\}$, and the histogram of the estimate in period 10,000, respectively, generated under $\mathcal{C}$ in Example 6. There are 2,000 sample paths in total. Thirty percent of the sample paths of $\{\hat{\theta}_t\}$ converge to 1. The iterated nonlinear least squares estimates are computed via the Levenberg-Marquardt algorithm.

maker; that is, $f(x,\theta) = g(\theta x)$ for all $x \in \mathcal{X}$ and $\theta \in \Theta$. To generalize our analysis of incomplete learning, we assume that $g(\cdot)$ is a differentiable and increasing function such that $\tilde{\ell} \leq g'(\xi) \leq \tilde{L}$ for all $\xi \in \Xi = \{\theta x : (x,\theta) \in \mathcal{X} \times \Theta\}$, where $0 < \tilde{\ell} \leq \tilde{L} < \infty$. Note that, for the linear-Gaussian model studied in the preceding section, we have $g(\xi) = \xi$, which satisfies these properties with $\tilde{\ell} = \tilde{L} = 1$.

Our next result extends our analysis of incomplete learning to the GLMs described above.

Theorem 4 (learning in static environments with nonlinear response) *Let $\psi(\cdot) \in \Psi$, $\theta \in \Theta$, and assume that $\theta_t = \theta$ for $t = 1, 2, \ldots$ Denote by $\{\hat{\theta}_t\}$ the sequence of certainty-equivalence estimates generated under $\mathcal{C}$.*

(i) *If $\psi(\theta) \neq 0$ and there exists $\zeta \in \Theta$ satisfying $\psi(\zeta) = 0$, then $\mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \to \zeta\} > 0$.*

(ii) *If there does not exist any $\zeta \in \Theta$ satisfying $\psi(\zeta) = 0$, then $\mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \to \theta\} = 1$.*

The preceding theorem generalizes the analysis in §4.1: Theorem 4(i) states that, if there is an uninformative estimate, then there is a positive probability of incomplete learning in the context of GLMs; Theorem 4(ii) states that, if there is no such uninformative estimate, then the sequence of certainty-equivalence estimates will be consistent in our GLM setting. Thus, the intuition we derived via Theorem 1 and Proposition 2 in the context of the linear-Gaussian model remains valid in a broader context of nonlinear models.

6 A General Solution for Incomplete Learning

In this section, we extend the main ideas developed in the preceding section to the general response model (2.5) with $f : \mathcal{X} \times \Theta \to \mathbb{R}$ assumed to be a continuously differentiable function, and derive a unifying solution that has good accuracy performance in both static and slowly changing environments represented by Examples 1 and 4, respectively.

6.1 Formulation and intuition

To generally describe the incomplete learning phenomenon, we first need to extend our definitions of uninformative control and uninformative estimate to the general response model (2.5). Recall

that in the linear-Gaussian model the uninformative control is 0 and the uninformative estimate is $\zeta = \psi^{-1}(0)$. In general, the informativeness of controls depends on the shape of the response curve. If x_t is chosen such that $f_\theta(x_t, \hat{\theta}_t) = 0$, then

$$\frac{\partial S_t(\hat{\theta}_t)}{\partial \theta} \;=\; \frac{\partial S_{t-1}(\hat{\theta}_t)}{\partial \theta}, \tag{6.1}$$

where $\partial S_t(\theta)/\partial\theta = -2\sum_{s=1}^{t}\big(y_s - f(x_s,\theta)\big)\, f_\theta(x_s,\theta)$. By the estimation equation (2.2), we know that $\partial S_{t-1}(\hat{\theta}_t)/\partial\theta = 0$, which implies $\hat{\theta}_{t+1} = \hat{\theta}_t$ by invoking (2.2) once more. To identify such controls that fail to update the estimate $\hat{\theta}_t$, let

$$u(\theta) \;=\; \{x \in \mathcal{X} : f_\theta(x,\theta) = 0\} \quad \text{ for } \theta \in \Theta, \tag{6.2}$$

and assume that there exists a unique $\zeta \in \Theta$ satisfying $\psi(\zeta) \in u(\zeta)$. With slight abuse of notation, we will hereafter use $u(\zeta)$ to refer to the single element in that set. As in the linear-Gaussian model, if $\hat{\theta}_t = \zeta$ for some t then $x_s = \psi(\zeta)$ and $\hat{\theta}_s = \zeta$ for all $s > t$. Thus, extending our previous definitions, we refer to ζ as the *uninformative estimate* and $\psi(\zeta)$ as the *uninformative control*. We assume that all controls other than $\psi(\zeta)$ are informative in the following sense: given any $\delta > 0$ there exists a finite and positive constant c_δ such that for all x satisfying $|x - \psi(\zeta)| > \delta$ we have $\min_{\theta\in\Theta} |f_\theta(x,\theta)| > c_\delta$. Roughly speaking, this condition means that the controls that are different than the uninformative control make $f_\theta(x,\theta)$ distinct from zero, and provide information at a positive rate. (The particular rate of information accumulation will be identified explicitly below.)

To avoid incomplete learning in general, we also need to extend our intuition on how information accumulates. In the linear-Gaussian model, incomplete learning occurs if $x_t \to 0$, and the amount of information provided by choosing a control $x \in \mathcal{X}$ can be expressed as

$$I(x) \;=\; x^2 \quad \text{ for } x \in \mathcal{X}, \tag{6.3}$$

which is why we measured the cumulative information with $J_t = \sum_{s=1}^{t} x_s^2 = \sum_{s=1}^{t} I(x_s)$ in that case. In general, the rate of information accumulation depends on both the control and the estimate of the decision maker. With slight abuse of notation, let

$$I(x,\theta) \;=\; \big(f_\theta(x,\theta)\big)^2 \quad \text{ for } x \in \mathcal{X} \text{ and } \theta \in \Theta. \tag{6.4}$$

In our general response model, we measure the rate of information accumulation with (6.4), which is a generalization of (6.3). When this rate gets close zero, the estimate sequence $\{\hat{\theta}_t\}$ under the certainty-equivalence learning policy $\mathcal{C}$ runs the risk of "getting stuck" at ζ. We will now use the information rate in (6.4), and study the impact of limiting the number of observations used in estimation. For that purpose, define a least squares estimation function that uses the last w observations. Let $\varphi(w,t)$ be the minimizer of $S_{w,t}(\theta) = \sum_{s=t-w+1}^{t}\big(y_s - f(x_s,\theta)\big)^2$ where $1 \le w \le t$. As argued in (2.2), $\varphi(w,t)$ is given by

$$\frac{\partial S_{w,t}\big(\varphi(w,t)\big)}{\partial \theta} \;=\; 0\,, \tag{6.5}$$

where $\partial S_{w,t}(\theta)/\partial\theta = -2\sum_{s=t-w+1}^{t}\big(y_s - f(x_s,\theta)\big)\, f_\theta(x_s,\theta)$. For example, in the linear-Gaussian model, $\varphi(w,t)$ has the following closed-form expression:

$$\varphi(w,t) \;=\; \left(\sum_{s=t-w+1}^{t} y_s x_s\right) \Big/ \left(\sum_{s=t-w+1}^{t} x_s^2\right). \tag{6.6}$$

In general, the estimator $\varphi(w,t)$ has the same form as $\hat{\theta}_{t+1}$, but it only uses the observations from period $t-w+1$ to period t. This makes $\hat{\theta}_{t+1}$ a special case of $\varphi(w,t)$, simply because $\hat{\theta}_{t+1} = \varphi(t,t)$. In what follows, we will construct a sequence of estimation windows, $\{w_n,\, n = 1,2,\ldots\}$, that will be consecutively used in the estimation equation (6.5). Throughout the sequel, we will denote the cumulative sums of this sequence by $\tau_n = \sum_{i=1}^{n} w_i$ for all n.

Define $I^*(x) = \min_{\theta\in\Theta}\{I(x,\theta)\}$ for all $x \in \mathcal{X}$, and suppose that $K > 0$ is a sufficiently large constant satisfying $I(x,\theta) \le K\, I^*(x)$ for all $x \in \mathcal{X}$ and $\theta \in \Theta$. Let w_1 be a natural number, and $X_1 \in \mathcal{X}$ such that $I^*(X_1) > 0$. The decision maker chooses $x_t = X_1$ for $t = 1,2,\ldots,w_1$. After this initialization, the decision maker computes the following estimate at the end of period τ_n for all $n \ge 1$:

$$\widehat{\Theta}_{n+1} \;=\; \varphi(w_n,\tau_n). \tag{6.7}$$

Based on the most recent estimate in (6.7), compute

$$X_{n+1} \;=\; \psi\big(\widehat{\Theta}_{n+1}\big). \tag{6.8}$$

Because the noise terms $\{\epsilon_t\}$ are continuous random variables, we have $\mathbb{P}_{\boldsymbol{\theta}}\{I^*(X_{n+1}) = 0\} = 0$. Consequently, $I^*(X_{n+1}) > 0$ almost surely. Let w_{n+1} be the smallest integer satisfying

$$w_{n+1} \;\ge\; \nu \log(\tau_n + w_{n+1})/I^*(X_{n+1}), \tag{6.9}$$

where ν is a scale parameter. Having computed the next control X_{n+1} and the estimation window w_{n+1}, the policy chooses $x_t = X_{n+1}$ for $t = \tau_n+1,\ldots,\tau_n+w_{n+1}$. Based on this construction, we note that $\{\tau_n,\, n = 1,2,\ldots\}$ can be viewed as the subsequence of periods in which estimation windows are updated, and the repetitive use of the equations (6.7) and (6.8) provides a variant of the certainty-equivalence learning policy $\mathcal{C}$ defined in §2. While the control function $\psi(\cdot)$ is still employed in a certainty-equivalence manner, the estimate $\widehat{\Theta}_{n+1}$ no longer has unlimited memory. Thus, we will hereafter call this variant the *certainty-equivalence learning policy with limited memory*, and denote it by $\mathcal{C}^*$. Accordingly, we will denote by $\{\hat{\theta}_t^*,\, t = 1,2,\ldots\}$ the estimate sequence generated under $\mathcal{C}^*$, i.e., $\hat{\theta}_t^* = \widehat{\Theta}_{n+1}$ for $t = \tau_n+1,\ldots,\tau_n+w_{n+1}$ and $n = 1,2,\ldots$ We will also denote by $\{\Delta_t^*, t = 1,2,\ldots\}$ the inaccuracy process under $\mathcal{C}^*$.

6.2 Theory and illustrations

Avoiding incomplete learning. Our first result shows that limiting memory helps avoid incomplete learning under fairly general conditions. To express said conditions in a compact form, let us define a measure of how frequently the unknown parameter sequence $\{\theta_t\}$ occupies a neighborhood

of the uninformative control ζ. For $a, b > 0$, we define the *occupancy measure* μ as

$$\mu\big(\zeta - a, \zeta + b\big) \;:=\; \limsup_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} \mathbb{I}\big\{\zeta - a \le \theta_s \le \zeta + b\big\}, \tag{6.10}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function (i.e., given condition A, $\mathbb{I}\{A\} = 1$ if A holds, and 0 otherwise). We note that, despite the fact that the occupancy measure μ and the previously defined tolerance parameter R (which appeared in Proposition 3 and Theorem 3, and will re-appear in Theorem 7 below) are both related to the temporal evolution of $\{\theta_t\}$, they are fundamentally different concepts. The occupancy measure μ is related to our treatment of incomplete learning, which depends on the location of $\{\theta_t\}$ relative to ζ. On the other hand, R is related to our treatment of asymptotic accuracy in changing environments where $\{\theta_t\}$ slowly moves away from ζ. For any given $R > 0$, the slowly changing environments in Proposition 3 and Theorems 3 and 7 would imply $\mu\big(\zeta - a, \zeta + b\big) = 0$ for all $a, b > 0$.

Theorem 5 (no incomplete learning with limited memory) *Let $\psi(\cdot) \in \Psi$, and assume that there exist $a, b > 0$ such that $\theta_t \ge \zeta - a$ eventually, and $\mu\big(\zeta - a, \zeta + b\big) = 0$. Then $\mathbb{P}_{\boldsymbol{\theta}}\big\{\hat{\theta}^*_t \to \zeta\big\} = 0$, where $\{\hat{\theta}^*_t\}$ is the sequence of certainty-equivalence estimates generated under $\mathcal{C}^*$.*

We note that, by symmetry, the conclusion of Theorem 5 holds also if $\theta_t \ge \zeta - a$ is replaced by $\theta_t \le \zeta + b$ above.

A simple verbal paraphrase of Theorem 5 is that, curtailing the memory of the estimates, it is possible to entirely eliminate the incomplete learning problem in a broad class of environments. The hypothesis of Theorem 5 allows $\{\theta_t\}$ to become arbitrarily close to ζ infinitely often, but as long as $\{\theta_t\}$ does not frequently visit or jump over the neighborhood $(\zeta - a, \zeta + b)$, the policy $\mathcal{C}^*$ would not suffer from incomplete learning. This stands in stark contrast with Theorems 1 and 2, which demonstrate the incomplete learning of $\mathcal{C}$ in static and boundedly changing environments, even though $\{\theta_t\}$ is eventually bounded away from ζ in both of those results.

Remark 8 (the source of logarithmic window length) To avoid incomplete learning, the policy $\mathcal{C}^*$ employs a logarithmic scaling in the construction of the estimation windows $\{w_n\}$. This is ensured by the use of the logarithm function in (6.9). Our analysis indicates that, by the law of the iterated logarithm, Theorem 5 would hold under any scaling that dominates the iterated logarithm (see the proof of Theorem 5 for details). The logarithmic scaling in (6.9) satisfies this condition, and since it is possible to use another scaling that grows faster than the iterated logarithm, the function $\log(\cdot)$ in (6.9) can be replaced by, say, $(\log\log(\cdot))^2$.

Achieving asymptotic accuracy. Encouraged by Theorem 5, we now turn our attention to the accuracy of $\mathcal{C}^*$. In our next result, we show that $\mathcal{C}^*$ achieves asymptotic accuracy in static environments.

Theorem 6 (accuracy in static environments) *Let $\psi(\cdot) \in \Psi$, $\theta \in \mathbb{R}$, and assume that $\theta_t = \theta \neq \zeta$ for $t = 1, 2, \ldots$ Then, for any $\varepsilon \in (0, 1)$, the sequence of certainty-equivalence estimates $\{\hat{\theta}_t^*\}$ generated under $\mathcal{C}^*$ is asymptotically ε-accurate.*

The performance guarantee for $\mathcal{C}^*$ in Theorem 6 is a significant improvement over the performance of $\mathcal{C}$ in Theorem 1 and Corollary 1, which state that incomplete learning arises with positive probability. Figure 9 displays the evolution of the inaccuracy of $\mathcal{C}^*$. Comparing Figures 2 and 9, we note that limiting the estimation memory remarkably improves asymptotic accuracy.

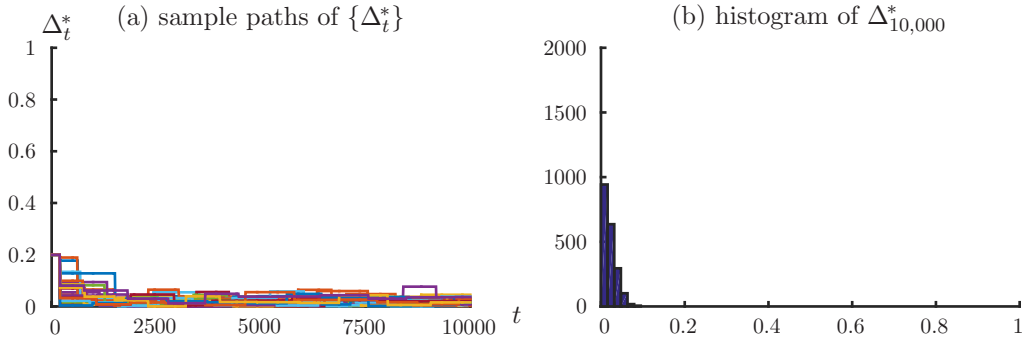

Figure 9: **Accuracy of certainty-equivalence learning with limited memory in a static environment.** Panels (a) and (b) show sample paths of the inaccuracy process $\{\Delta_t^*\}$, and the histogram of the inaccuracy in period 10,000, respectively, generated under $\mathcal{C}^*$ in Example 1. More than 95% of the 2,000 sample paths eventually achieve an inaccuracy Δ_t^* less than 0.05. For all sample paths, the initial control is $X_1 = 1$, and the initial estimation window is $w_1 = 200$. The scale parameter of $\mathcal{C}^*$ is $\nu = 250$.

The improved performance of $\mathcal{C}^*$ relies on avoiding the incomplete learning trap by gradually increasing the amount of information collected in an estimation window. In particular, if the control sequence gets close to the uninformative control, then $\mathcal{C}^*$ further adjusts its estimation window to maintain the "right rate" in information accumulation. As shown in Figure 10, adaptively adjusting memory resolves the incomplete learning problem associated with certainty-equivalence in static environments.

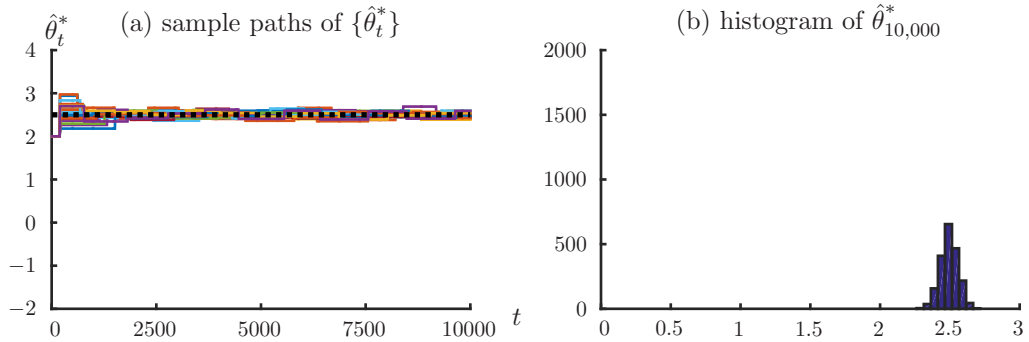

Figure 10: **Certainty-equivalence estimates with limited memory in a static environment.** Panels (a) and (b) depict sample paths of the estimate sequence $\{\hat{\theta}_t^*\}$, and the histogram of the estimate in period 10,000, respectively, generated under $\mathcal{C}^*$ in the setting of Example 1. (Problem parameters are the same as in Figure 9.)

Our final result shows that limiting memory also provides good accuracy performance in slowly

changing environments. To express this result compactly we define asymptotic tracking as follows.

Definition (asymptotic tracking) *The estimate sequence* $\{\hat{\theta}_t^*\}$ *is said to asymptotically track the unknown parameter sequence* $\{\theta_t\}$*, expressed as* $\hat{\theta}_t^* \asymp \theta_t$*, if there exist constants* η_1, η_2 *where* $0 < \eta_1 \leq 1 \leq \eta_2 < \infty$*, such that*

$$\liminf_{t\to\infty} \frac{\hat{\theta}_t^*}{\theta_t} \geq \eta_1 \quad \text{and} \quad \limsup_{t\to\infty} \frac{\hat{\theta}_t^*}{\theta_t} \leq \eta_2 \quad \text{almost surely.} \tag{6.11}$$

Theorem 7 (tracking accuracy in slowly changing environments) *Let* $\psi(\cdot) \in \Psi$*, and*

$$\theta_t^l = \zeta + \kappa_1 G(t)\,, \tag{6.12a}$$

$$\theta_t^h = \zeta + \kappa_2 G(t)\,, \tag{6.12b}$$

for $t = 1, 2, \ldots$*, where* $0 < \kappa_1 \leq \kappa_2$*. Assume that* $G(\cdot)$ *is concave and nondecreasing,* $G(t) \to \infty$*, and* $\{\theta_t\}$ *evolves between the lower and upper bound processes* $\{\theta_t^l\}$ *and* $\{\theta_t^h\}$ *respectively, with tolerance* $R > 0$*. Then* $\hat{\theta}_t^* \asymp \theta_t$*, where* $\{\hat{\theta}_t^*\}$ *is the sequence of certainty-equivalence estimates generated under* $\mathcal{C}^*$*.*

Remark 9 If κ_1 and κ_2 are sufficiently close to each other, then η_1 and η_2 are both close to 1, implying asymptotic accuracy. To be precise, if $\kappa_2 \leq (1 + \varepsilon/4)\kappa_1$ for $\varepsilon \in (0, \frac{1}{2})$, then $\{\hat{\theta}_t^*\}$ is asymptotically ε-accurate.

The preceding theorem states that $\mathcal{C}^*$ can "track" the unknown parameter in slowly changing environments. That is, as long as $\{\theta_t\}$ moves away from ζ at a slow growth rate described by the concave function $G(\cdot)$, $\{\hat{\theta}_t^*\}$ will attain the same order of magnitude as $\{\theta_t\}$. It is perhaps worth noting that the changing environment described in Theorem 7 is more general than the one described in Theorem 3, in the sense that the growth envelope in (6.12) is more general in its functional form, and the tolerance parameter R could be larger.

Figures 11 and 12 show that $\mathcal{C}^*$ can track the unknown parameter sequence $\{\theta_t\}$ in the slowly changing environment described in Example 4.

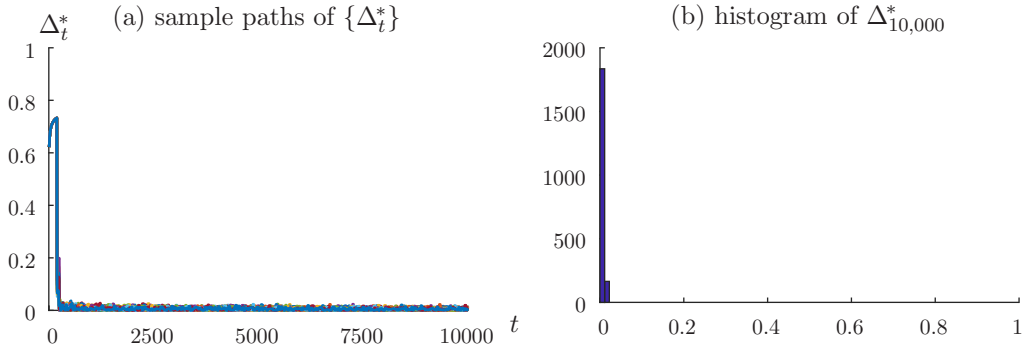

Figure 11: **Accuracy of certainty-equivalence learning with limited memory in a slowly changing environment.** Panels (a) and (b) show sample paths of the inaccuracy process $\{\Delta_t^*\}$, and the histogram of the inaccuracy in period 10,000, respectively, generated under $\mathcal{C}^*$ in Example 4. More than 99% of the 2,000 sample paths eventually achieve an inaccuracy Δ_t^* less than 0.02. For all sample paths, the initial control is $X_1 = 1$, and the initial estimation window is $w_1 = 200$. The scale parameter of $\mathcal{C}^*$ is $\nu = 250$.

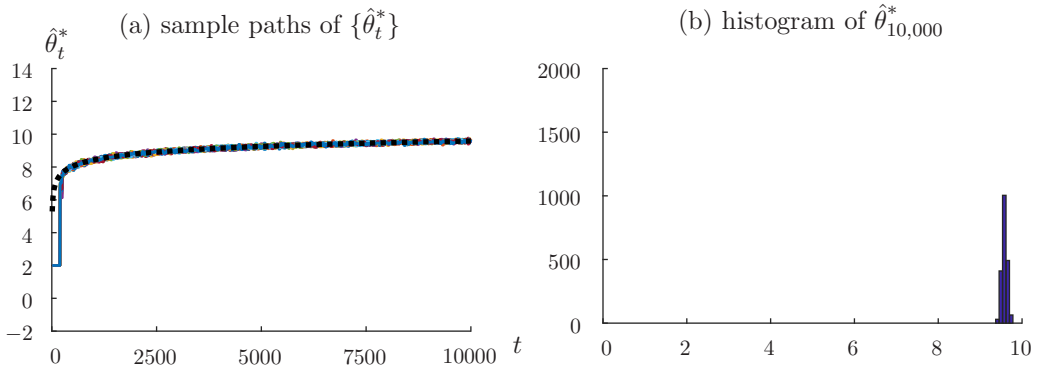

Figure 12: **Certainty-equivalence estimates with limited memory in a slowly changing environment.** Panels (a) and (b) depict sample paths of the estimate sequence $\{\hat{\theta}_t^*\}$, and the histogram of the estimate in period 10,000, respectively, generated under $\mathcal{C}^*$ in Example 4. (Problem parameters are the same as in Figure 11.)

Evolution of estimation windows in static and changing environments. The operating principle of $\mathcal{C}^*$ is to gradually increase the amount of information collected within an estimation window. In a static environment, this principle leads to increasing the size of the estimation window, w_n, because the signals received from the environment do not necessarily "grow stronger" over time. But, in the slowly changing environments described in Theorems 3 and 7, the strength of these signals increases as time progresses; hence it is possible for the amount of information collected in an estimation window to increase while keeping the window size w_n more or less stable. Figure 13 depicts this feature of $\mathcal{C}^*$ by comparing the evolution of its window size in Examples 1 and 4.

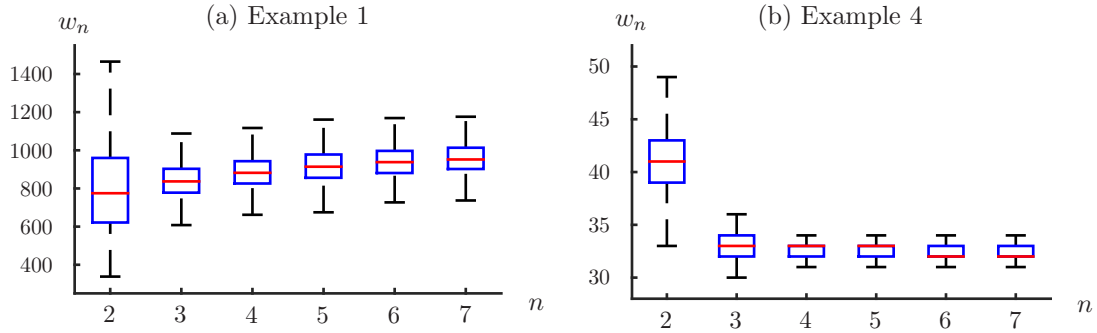

Figure 13: **Estimation window size.** The box-and-whiskers plots above show how the estimation windows w_n of $\mathcal{C}^*$ evolve in Examples 1 and 4, respectively. For every iteration $n \geq 2$, the box displays the lower quartile, median, and upper quartile values for the 2,000 sample paths generated, while each whisker extends either to the most extreme data point or else for a distance equal to 1.5 times the inter-quartile range, whichever is smaller. For all sample paths, the initial control is $X_1 = 1$, and the initial estimation window is $w_1 = 200$. The scale parameter of $\mathcal{C}^*$ is $\nu = 250$.

7 Concluding Remarks

Generalization from least squares to M-estimation. For concreteness and exposition purposes, we focused in this paper on (nonlinear) least squares estimators described by (2.2). We note that least squares estimation belongs to a broad family of estimation methods that are predicated

on optimizing an empirical objective function that serves as a proxy for an unobserved ideal objective. These are collectively referred to as *M-estimators*, and include least squares and maximum likelihood as well as several other well studied instances. Our general analysis in §6 can be extended to M-estimators that satisfy the following conditions. Let $\phi : \mathbb{R}^{2t} \to \Theta$ be an estimator function that maps the history of observed responses and controls to an estimate $\hat{\theta}_{t+1}$ as follows:

$$\hat{\theta}_{t+1} \;=\; \phi(x_1, y_1, \ldots, x_t, y_t) \;:=\; \arg\max_{\theta} \left\{ \sum_{s=1}^{t} \lambda\big(y_s - f(x_s, \theta)\big) \right\}, \tag{7.1}$$

where $\lambda(\cdot)$ is a twice differentiable and concave loss function such that: (i) $\mathbb{E}_{\boldsymbol{\theta}}\{\lambda'(\epsilon_t)\} = 0$ for all t; and (ii) there exists $c > 0$ satisfying $|\lambda''(z)| \geq c$ for all $z \in \mathbb{R}$. Let $\mathcal{L}_t(\theta) = \sum_{s=1}^{t} \lambda\big(y_s - f(x_s, \theta)\big)$. In this case, we replace (2.2) by

$$\frac{\partial \mathcal{L}_t(\hat{\theta}_{t+1})}{\partial \theta} \;=\; 0, \tag{7.2}$$

where $\partial \mathcal{L}_t(\theta)/\partial\theta = -\sum_{s=1}^{t} \lambda'\big(y_s - f(x_s, \theta)\big)\, f_\theta(x_s, \theta)$. A well-known example of the aforementioned M-estimation procedure is maximum likelihood estimation, for which $\lambda(\cdot)$ is the logarithm of the density of the noise terms $\{\epsilon_t\}$, and $\mathcal{L}_t(\cdot)$ is the log-likelihood function at the end of period t.

To extend the proofs of our general results in §6, one needs to study the optimality condition (7.2) and employ the mean value theorem as in the proof of Theorem 5. In the case of M-estimation, the generalized optimality conditions are expressed in terms of weighted sums, where the second derivatives of the loss function $\lambda(\cdot)$ serve as weights. As long as these second derivatives are *bounded away from zero*, our asymptotic analysis, proofs of Theorems 5-7, follows in the same manner. (For further details on this extension, please see the remark following the proof of Theorem 5 in Appendix E.)

Constrained estimation. In many applications of dynamic learning, the decision maker may know a priori that Θ, the feasible set of unknown model parameters, is a strict subset of $\mathbb{R}$. In this case, the unconstrained estimator $\hat{\theta}_t$ would be projected onto Θ to reduce inaccuracy, which leads to $\{\hat{\theta}_t\}$ taking values possibly on the boundary of Θ. We would like to emphasize that our proof of incomplete learning in Theorem 1 does not rely on the convergence of the estimate sequence $\{\hat{\theta}_t\}$ to the boundary of Θ, or the convergence of the control sequence $\{x_t\}$ to the boundary of $\mathcal{X}$. Instead, our analysis reveals that there is a positive probability that $\{\hat{\theta}_t\}$ and $\{x_t\}$ converge to a uninformative estimate-control pair ζ and $\psi(\zeta)$ that are in the *interior* of Θ and $\mathcal{X}$, respectively. This result holds even if Θ and $\mathcal{X}$ are both unbounded, and thus, it is distinct from most antecedent work on incomplete learning for certainty-equivalence, e.g., Lai and Robbins (1982), that rely on the uninformative control being on the boundary of $\mathcal{X}$.

Extension to martingale-difference noise. As shown in §5, our findings on the incomplete learning phenomenon hold for a family of nonlinear models called generalized linear models (GLMs); see Theorem 4(i-ii). We note that these findings can be further generalized to the case where the noise terms follow a martingale difference sequence. To be more precise, the statement of Theorem 4 remains valid if the noise terms $\{\epsilon_t\}$ form a square-integrable martingale difference sequence with

respect to the canonical filtration $\mathcal{F}_t = \sigma(\epsilon_1, \ldots, \epsilon_t)$. By using the strong law of large numbers for martingales instead of the classical strong law of large numbers (while applying the other proof arguments verbatim) in the proof of Theorem 4, one can cover this extended setting and show that, if $\{\epsilon_t\}$ is a square-integrable martingale difference sequence with conditional probability density function $h_{\epsilon_t}(\cdot|\mathcal{F}_{t-1})$ and support $\mathbb{R}$, then Theorem 4 still holds. Note that, in this extension, the density of $\mathbb{P}_{\boldsymbol{\theta}}$ is given by $h_\theta(y_1, \ldots, y_t) = h_{\epsilon_1}\big(y_1 - f(x_1, \theta)\big) \prod_{s=2}^{t} h_{\epsilon_s}\big(y_s - f(x_s, \theta)|\mathcal{F}_{s-1}\big)$ for $y_1, \ldots, y_t \in \mathbb{R}$.

Moderately changing environments: from sub-logarithmic to linear growth. In Theorem 3, we describe a specific family of slowly changing environments where the unknown parameter sequence $\{\theta_t\}$ drifts away from the uninformative estimate ζ at a slow, sub-logarithmic rate. Of course, if $\{\theta_t\}$ moves away from ζ at a faster sub-linear rate (e.g., at a rate of order t^α where $0 < \alpha < 1$), the decision maker would be receiving even stronger signals; hence it would be relatively easier to achieve asymptotic accuracy in such *moderately* changing environments. Theorem 7 provides the generalized growth conditions on $\{\theta_t\}$ for achieving asymptotic accuracy.

Comparison with antecedent work on certainty-equivalence policies. As discussed earlier in §1.2, incomplete learning of certainty-equivalence policies has been the subject of several studies. In particular, Lai and Robbins (1982) showed that, in a dynamic learning setting with a *compact* set of feasible controls, the control sequence of a certainty-equivalence learning policy can converge to the *boundary* of the feasible control set; see also den Boer and Zwart (2014) for a variant of this result in a dynamic pricing formulation. At first glance, one might think that incomplete learning is a boundary-related phenomenon that will disappear if the feasible control set is *unbounded*. However, unlike the Lai and Robbins (1982) result, we show that the incomplete learning phenomenon persists even in this setting. Our result is based on the observation that the controls of the certainty-equivalence learning policy can converge to an uninformative fixed point in the *interior* of the control set. The proof technique in Lai and Robbins (1982) relies on the control sequence taking a fixed boundary value after a finite number of periods, which is not possible in our setting. The added challenge in proving our result is to show that the certainty-equivalence controls can become "trapped" in a narrow range of interior values (rather than a fixed boundary value). To obtain our incomplete learning result, we employ the strong law of large numbers for martingales, and characterize the information growth patterns on sample paths with and without incomplete learning. In addition to the above, our study differs from the work by Lai and Robbins (1982) (and follow-up papers surveyed in §1.2) insofar as: (i) we consider several different changing environments to shed light on the extent of the incomplete learning phenomenon under certainty-equivalence policies; and (ii) we prove that limiting the estimation memory of certainty-equivalence policies can eliminate incomplete learning, improving accuracy and tracking performance.

We would also like to note that, in the literature on dynamic learning, there exist other ap-

proaches to modifying certainty-equivalence policies. In the vast majority of these approaches, the control function of the certainty-equivalence policy is modified to implement *forced exploration*. For example, in the widely studied multiarmed bandit problem, this idea dates back to Robbins (1952), later refined in the formulation of the upper confidence bound policies and randomization (ϵ-greedy) policies for bandit problems with arms indexed by an unknown parameter; see Lai and Robbins (1985) and Auer et al. (2002). We refer readers to Harrison et al. (2012), Broder and Rusmevichientong (2012), den Boer and Zwart (2014), Keskin and Zeevi (2014) for typical examples of such modifications in the context of the dynamic pricing problem described in §3; see also den Boer and Zwart (2015) for a simple modification of a certainty-equivalence policy in a dynamic pricing and learning problem with finite initial inventory. Unlike the studies above, our certainty-equivalence learning policy with limited memory (i.e., the policy $\mathcal{C}^*$) does not attempt to replace the control function to implement forced exploration; i.e., $\mathcal{C}^*$ employs the certainty-equivalence control function $\psi(\cdot)$ *all the time*, using limited estimation memory to avoid incomplete learning in lieu of forced exploration.

Comparison with antecedent work on learning in changing environments. In a recent study, Keskin and Zeevi (2016) analyzed the performance of various forced-exploration dynamic pricing policies in changing environments. In particular, they constructed and studied policies that employ moving windows, gradually decaying weights, and change-point detection tests. Our work differs from theirs in two major ways. First, the focus of Keskin and Zeevi (2016) is on forced exploration; all of their policies rely on a pre-determined schedule of explicit experiments with controls, rather than using certainty-equivalence. By contrast, our work studies the performance of certainty-equivalence policies *without* any forced exploration. Secondly, the aforementioned moving windows in Keskin and Zeevi (2016) are chosen deterministically in the beginning of the time horizon, whereas the estimation windows of the policy $\mathcal{C}^*$ are chosen adaptively. Consequently, the absence of forced exploration and the adaptive nature of estimation windows make our policies and their analysis distinct.

Multiple unknown parameters. In this paper, we have studied a family of response models characterized by a single unknown parameter, and shown that certainty-equivalence learning policies may exhibit incomplete learning. We should note that this phenomenon is observed in a broader context where the response model can depend on multiple unknown parameters. As an illustrative example, suppose that the response is given by $y_t = \alpha+\beta x_t+\epsilon_t$, where α and β are unknown scalars, and $x_t \in \mathcal{X} = \mathbb{R}$. In this case, the parameter vector is $\theta = (\alpha, \beta)$, which takes values in $\Theta \subseteq \mathbb{R}^2$. The definition of a certainty-equivalence learning policy in this setting is similar to the dynamical system in (2.2-2.3): at the end of each period t, a least squares estimate $\hat{\theta}_{t+1} = (\hat{\alpha}_{t+1}, \hat{\beta}_{t+1})$ is computed, and then a certainty-equivalence decision rule $\psi : \Theta \to \mathcal{X}$ maps the estimate $\hat{\theta}_{t+1}$ into the control x_{t+1} to be used in the following period, i.e., $x_{t+1} = \psi(\hat{\theta}_{t+1})$. (For example, a certainty-equivalence decision rule commonly used in dynamic pricing settings is $\psi(\alpha, \beta) = -\alpha/(2\beta)$; assuming that x_t

and y_t are the price and demand in period t, respectively, this decision rule would maximize the expected revenue function if the values of $\alpha > 0$ and $\beta < 0$ were known with certainty.)

Figure 14 shows the performance of certainty-equivalence learning in the aforementioned setting. As depicted in panel (a), the controls $\{x_t\}$ of the certainty-equivalence learning policy do not necessarily converge to $\psi(\theta)$, which equals 1.1 in this example. Instead, they converge to various values that are distinct from $\psi(\theta)$. The essential reason for this behavior is that the uninformative control varies over time. To be precise, the uninformative control that fails to update the least square estimate is the sample average of the first t controls, namely $\bar{x}_t = \sum_{s=1}^{t} x_s$. Because $\{\bar{x}_t\}$ evolves over time, the control sequence $\{x_t\}$ can "get stuck" at different values. The primary challenge in the analysis of the above multi-parameter setting is the time-varying nature of the uninformative estimate-control pair. This means that, in the case of incomplete learning, the location of the uninformative equilibrium is not fixed in advance, which makes the convergence results especially challenging. Changing uninformative estimate-control pairs present an interesting direction for future research on incomplete learning, and we believe that the analysis we develop in this paper would hopefully constitute a key step in facilitating that study.

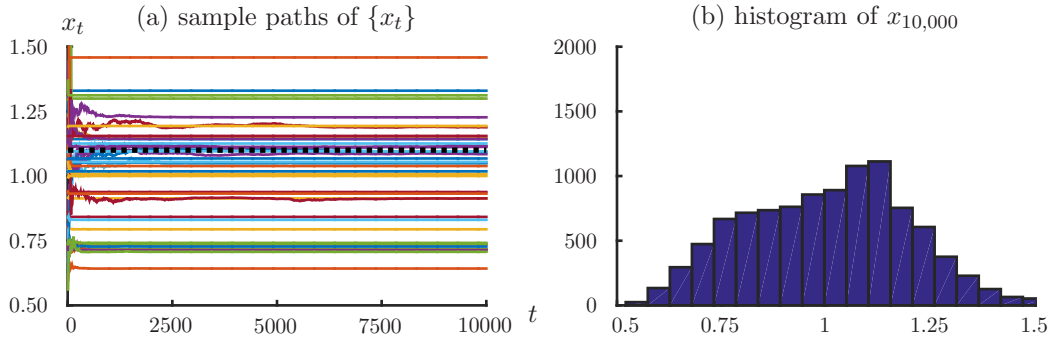

Figure 14: **Incomplete learning with multiple unknown parameters.** Panels (a) and (b) depict sample paths of the control sequence $\{x_t\}$, and the histogram of the control in period 10,000, respectively, under the certainty-equivalence learning policy in a static environment with two unknown parameters. The true values of the unknown parameters are $\alpha = 1.1$ and $\beta = -0.5$, and the standard deviation of noise terms is $\sigma = 0.5$. The first two controls are set to be $x_1 = 1$ and $x_2 = 2$, and subsequently are set using the control function $\psi(\alpha, \beta) = -\alpha/(2\beta)$.

Appendix A: Speed of Learning

On the paths in which the estimate sequence $\{\hat{\theta}_t\}$ exhibits incomplete learning, the rate of convergence to the uninformative estimate can differ from the rate of convergence on the set of paths where $\{\hat{\theta}_t\}$ converges to the true parameter. In Example 1, 20% of the sample paths exhibit incomplete learning whereas the remaining 80% exhibit consistency. For the latter sample paths, a log-log regression reveals that the mean squared error of estimates, namely $\mathbb{E}_{\boldsymbol{\theta}}(\hat{\theta}_t - \theta)^2$, decreases at a rate close to t^{-1} (with $R^2 = 0.99$, see panel (a) of Figure 15). This means that the cumulative

mean squared error $\sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\theta}}(\hat{\theta}_t - \theta)^2$, which can also be linked to the decision maker's "regret" under a quadratic loss function, grows proportional to $\log T$ on the sample paths that exhibit consistency. Because Proposition 1 implies that $x_t \to \psi(\theta)$ almost surely on these sample paths, the Fisher information $\{J_t = \sum_{s=1}^{t} x_s^2,\, t = 1, 2, \ldots\}$ increases linearly in the case of consistency, which explains why the mean squared error decays in inverse proportion to t. However, $\{\hat{\theta}_t\}$ cannot have the same convergence rate for incomplete learning because we know by Proposition 1 that $J_\infty < \infty$ almost surely in this case. For the sample paths with incomplete learning, $\mathbb{E}_{\boldsymbol{\theta}}(\hat{\theta}_t - \zeta)^2$ decays at a rate close to $e^{-0.01t}$ (with $R^2 = 0.97$, see panel (b) of Figure 15; the decay coefficient 0.01 is statistically significant with $p < 0.001$).

Consequently, as explained by Proposition 1, the convergence rate for incomplete learning is substantially faster than that for consistency. Note that, in the case of incomplete learning, the cumulative mean squared error $\sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\theta}}(\hat{\theta}_t - \theta)^2$ grows linearly in T, due to Markov's inequality and because $\hat{\theta}_t$ converges to ζ with positive probability.

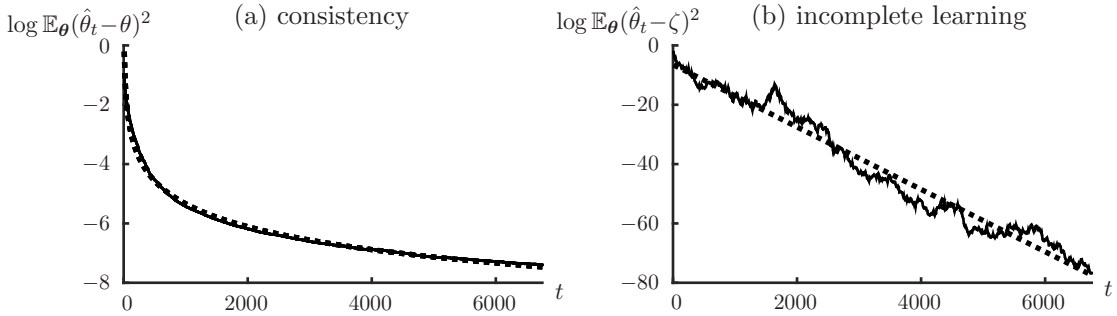

Figure 15: **Speed of convergence.** The solid curves in panels (a) and (b) depict the semi-log plots of $\mathbb{E}_{\boldsymbol{\theta}}(\hat{\theta}_t - \theta)^2$ in the case of consistency and $\mathbb{E}_{\boldsymbol{\theta}}(\hat{\theta}_t - \zeta)^2$ in the case of incomplete learning, respectively, in Example 1. The dotted curves in panels (a) and (b) display the best log-log and semi-log regression fits to the plots, respectively.

Appendix B: Proofs of the Results in §4

Proof of Proposition 1. Let $\mathcal{F}_t = \sigma(\epsilon_1, \ldots, \epsilon_t)$ for $t = 1, 2, \ldots$ We will first show by induction that x_t is square-integrable for $t = 1, 2, \ldots$ For the base step, note that x_1 is deterministic and hence square-integrable. Now, for the induction step, assume that $x_1, \ldots, x_t$ are square-integrable. By (2.3), $x_{t+1} = \psi(\hat{\theta}_{t+1})$. Using the mean value theorem and the fact that $|\psi'(\theta)| \le L < \infty$ for all $\theta \in \Theta$, we further deduce that

$$x_{t+1}^2 \;\le\; 2\big(\psi(\theta)\big)^2 + 2L^2(\hat{\theta}_{t+1} - \theta)^2 \;\overset{\text{(a)}}{=}\; 2\big(\psi(\theta)\big)^2 + \frac{2L^2 M_t^2}{J_t^2} \;\overset{\text{(b)}}{\le}\; 2\big(\psi(\theta)\big)^2 + \frac{2L^2 M_t^2}{x_1^4}, \qquad \text{(B.1)}$$

where: (a) follows from (4.3); and (b) follows because $x_1^2 \le \sum_{s=1}^{t} x_s^2 = J_t$. Recall that $x_1, \ldots, x_t$ are square-integrable by the induction hypothesis. Moreover, because $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$, $\epsilon_1, \ldots, \epsilon_t$ are also square-integrable, which implies that $M_t = \sum_{s=1}^{t} x_s \epsilon_s$ is square integrable, i.e., $\mathbb{E}_{\boldsymbol{\theta}} M_t^2 < \infty$. By (B.1), this implies that $\mathbb{E}_{\boldsymbol{\theta}}\, x_{t+1}^2 < \infty$, i.e., x_{t+1} is square-integrable. Consequently, since x_t

and ϵ_t are square-integrable for $t = 1, 2, \ldots$, the stochastic process $\{M_t,\, t = 1, 2, \ldots\}$ is a square-integrable and zero-mean martingale with respect to the filtration $\{\mathcal{F}_t,\, t = 1, 2, \ldots\}$. Let $V_t = \langle M \rangle_t$ denote the predictable compensator of $\{M_t,\, t = 1, 2, \ldots\}$; that is

$$V_t \;=\; \sum_{s=1}^{t} \mathbb{E}_{\boldsymbol{\theta}}[(M_s - M_{s-1})^2 | \mathcal{F}_{s-1}] \tag{B.2}$$

for all t. By the strong law of large numbers for martingales (see Williams 1991, pp. 122-124), we know that (i) $M_t/V_t \to 0$ almost surely on $\{V_\infty = \infty\}$, and (ii) M_t converges almost surely to a finite limit on $\{V_\infty < \infty\}$, where $V_\infty = \lim_{t\to\infty} V_t$. Now recall that, by (2.2), $\hat{\theta}_{t+1} \in m\mathcal{F}_t$ for all t. Since $x_{t+1} = \psi(\hat{\theta}_{t+1})$ by (2.3), and $\psi(\cdot)$ is a known deterministic function, we further deduce that $x_{t+1} \in m\mathcal{F}_t$ for all t. Consequently, $\mathbb{E}_\theta[x_{t+1}^2 | \mathcal{F}_t] = x_{t+1}^2$ for all t. Thus, we have $V_t = \sigma^2 \sum_{s=1}^{t} \mathbb{E}_{\boldsymbol{\theta}}[x_s^2 | \mathcal{F}_{s-1}] = \sigma^2 \sum_{s=1}^{t} x_s^2 = \sigma^2 J_t$, where $J_t = \sum_{s=1}^{t} x_s^2$. Therefore, V_∞ is finite if and only if J_∞ is finite. By (i) and equation (4.3), $\hat{\theta}_t \to \theta$ almost surely on the event $\{J_\infty = \infty\}$. However, on the event $\{J_\infty < \infty\}$, $x_t \to 0$ because $\sum_{s=1}^{\infty} x_s^2 < \infty$. Thus, $\hat{\theta}_t \to \psi^{-1}(0)$ almost surely on $\{J_\infty < \infty\}$. ■

Proof of Theorem 1. Because $\theta \neq \zeta$, we deduce that either $\theta > \zeta$ or $\theta < \zeta$.

Case 1. $\theta > \zeta$. In this case, the remainder of the proof is a special case of the proof of Theorem 2 with $Y_t = \theta J_t$, $\kappa_0 = (\theta - \zeta)/2$, and $\kappa_2 = 3(\theta - \zeta)/2$.

Case 2. $\theta < \zeta$. Let $\kappa_0 = (\zeta - \theta)/2$, $\kappa_1 = 3(\zeta - \theta)/2$, and $\kappa_2 > 0$. Consider the proof of Theorem 2, and redefine the indicator variable χ_s as $\chi_s = \mathbb{I}\{\zeta - \kappa_0 < \theta_s \leq \zeta + \kappa_2\}$. If $\zeta - \kappa_1 \leq \theta_t \leq \zeta + \kappa_2$ for $t = 1, 2, \ldots$, and $\theta_t \leq \zeta - \kappa_0$ eventually, then by the argument used to prove (B.8) we deduce that there exists a finite random variable N_1 such that $\hat{\theta}_t \leq \zeta - \kappa_0/2$ for all $t \geq N_1$. By symmetry, we repeat the arguments following (B.8) in a similar manner to show that the conclusion of Theorem 2 holds for the above choices of κ_0, κ_1, and κ_2. ■

Proof of Proposition 2. Because $\psi(\cdot)$ is monotone and there exists no $\zeta \in \Theta$ satisfying $\psi(\zeta) = 0$, we deduce that either $\psi(\theta) > 0$ for all $\theta \in \Theta$, or $\psi(\theta) < 0$ for all $\theta \in \Theta$. Assume without loss generality that $\psi(\theta) > 0$ for all $\theta \in \Theta$. Recall that $J_t = \sum_{s=1}^{t} x_s^2$, and that $J_\infty = \lim_{t\to\infty} J_t$. Suppose towards a contradiction that $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty < \infty\} > 0$. On the event $\{J_\infty < \infty\}$, we know that $x_t \to 0$ because $\sum_{s=1}^{\infty} x_s^2 < \infty$. In addition, as shown in the proof of Proposition 1, $M_t = \sum_{s=1}^{t} x_s \epsilon_s$ converges almost surely to a finite limit on $\{J_\infty < \infty\}$. By (4.3) and the fact that $J_\infty \geq x_1^2 > 0$, this implies that, on the event $\{J_\infty < \infty\}$, $\hat{\theta}_t \in \Theta$ converges to a finite limit in Θ. But, since $\psi(\theta) > 0$ for all $\theta \in \Theta$, this contradicts the fact that $x_t = \psi(\hat{\theta}_t)$ converges to 0 on $\{J_\infty < \infty\}$. Thus, $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty < \infty\} = 0$, and $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty = \infty\} = 1$. By Proposition 1(i), we conclude that $\hat{\theta}_t \to \theta$ almost surely. ■

Proof of Theorem 2. First note that, for every $t \geq 2$, by the mean value theorem there exists a random variable c_t on the line segment between $\hat{\theta}_t$ and ζ such that $x_t = \psi(\hat{\theta}_t) = \psi(\zeta) + \psi'(c_t)(\hat{\theta}_t - \zeta)$.

Because $\psi(\zeta) = 0$, we can express the control in period t as

$$x_t \;=\; \psi(\hat{\theta}_t) \;=\; L_t(\hat{\theta}_t - \zeta) \quad \text{ for all } t = 1, 2, \ldots, \tag{B.3}$$

where $L_t = \psi'(c_t)$. Secondly, the equations (4.1) and (4.2) imply that

$$\hat{\theta}_{t+1} \;=\; \frac{Y_t + M_t}{J_t} \quad \text{ for } t = 1, 2, \ldots \tag{B.4}$$

where $Y_t = \sum_{s=1}^{t} \theta_s x_s^2$, $J_t = \sum_{s=1}^{t} x_s^2$, and $M_t = \sum_{s=1}^{t} x_s \epsilon_s$. We can re-express the preceding identity as follows: $\hat{\theta}_{t+1} = \sum_{s=1}^{t} \lambda_{s,t}\, \theta_s + M_t/J_t$ for all t, where $\lambda_{s,t} = x_s^2/J_t$. Note that $\lambda_{s,t} \geq 0$ for all s and t, and that $\sum_{s=1}^{t} \lambda_{s,t} = 1$. Now, suppose towards a contradiction that $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty < \infty\} = 0$; hence $J_\infty = \infty$ almost surely. As argued in the proof of Proposition 1, this implies by the strong law of large numbers for martingales that M_t/J_t converges to zero almost surely. Thus, there exists a finite random variable N_0 such that we have the following with probability one:

$$\sum_{s=1}^{t} \lambda_{s,t}\, \theta_s - \kappa^* \;\leq\; \hat{\theta}_{t+1} \;\leq\; \sum_{s=1}^{t} \lambda_{s,t}\, \theta_s + \kappa^* \quad \text{ for all } t \geq N_0, \tag{B.5}$$

where $\kappa^* = \max\{\kappa_1, \kappa_2\}$. By (4.4), we know that $\zeta - \kappa_1 \leq \sum_{s=1}^{t} \lambda_{s,t}\, \theta_s \leq \zeta + \kappa_2$, implying $-2\kappa^* \leq \hat{\theta}_{t+1} - \zeta \leq 2\kappa^*$ for all $t \geq N_0$. Combining these inequalities with (B.3), we have $x_{t+1}^2 \leq (2L\kappa^*)^2$ for all $t \geq N_0$, where L is the upper bound on the derivative of $\psi(\cdot)$. Note further that

$$\begin{aligned}
\hat{\theta}_{t+1} \;&\overset{\text{(a)}}{=}\; \zeta + \kappa_0 + \frac{\sum_{s=1}^{t} (\theta_s - \zeta - \kappa_0)\, x_s^2\, (\chi_s + \bar{\chi}_s)}{\sum_{s=1}^{t} x_s^2} + \frac{M_t}{J_t} \\
&\geq\; \zeta + \kappa_0 - (\kappa_0 + \kappa_1) \cdot \frac{\sum_{s=1}^{t} x_s^2\, \chi_s}{\sum_{s=1}^{t} x_s^2} + (\kappa_0 - \kappa_0) \cdot \frac{\sum_{s=1}^{t} x_s^2\, \bar{\chi}_s}{\sum_{s=1}^{t} x_s^2} + \frac{M_t}{J_t} \\
&=\; \zeta + \kappa_0 - (\kappa_0 + \kappa_1) \cdot \frac{\sum_{s=1}^{t} x_s^2\, \chi_s}{\sum_{s=1}^{t} x_s^2} + \frac{M_t}{J_t}
\end{aligned} \tag{B.6}$$

for $t = 1, 2, \ldots$, where: $\chi_s = \mathbb{I}\{\zeta - \kappa_1 \leq \theta_s < \zeta + \kappa_0\}$, $\bar{\chi}_s = \mathbb{I}\{\zeta + \kappa_0 \leq \theta_s \leq \zeta + \kappa_2\} = 1 - \chi_s$, and (a) follows by (B.4) and (4.4). Because $x_{t+1}^2 \leq (2L\kappa^*)^2$ for all $t \geq N_0$, the preceding inequality implies that

$$\hat{\theta}_{t+1} \;\geq\; \zeta + \kappa_0 - \frac{C_0 \sum_{s=1}^{t} \chi_s}{J_t} + \frac{M_t}{J_t} \quad \text{ for all } t \geq N_0, \tag{B.7}$$

where $C_0 = (\kappa_0 + \kappa_1)(2L\kappa^*)^2$. Since $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty = \infty\} = \mathbb{P}_{\boldsymbol{\theta}}\{M_t/J_t \to 0\} = 1$, and $\theta_t \geq \zeta + \kappa_0$ eventually, we deduce that there exists a finite random variable $N_1 \geq N_0$ such that we have the following with probability one:

$$\hat{\theta}_t \;\geq\; \zeta + \frac{\kappa_0}{2} \quad \text{ for all } t \geq N_1. \tag{B.8}$$

Let $\delta > 0$. Define a stochastic process $\{\gamma_t,\, t = 1, 2, \ldots\}$ with

$$\gamma_t \;=\; \frac{\alpha_t L}{2t} + \frac{\beta_t J_{t-1}}{2Lt} \quad \text{ for all } t = 1, 2, \ldots, \tag{B.9}$$

where $\alpha_t = -\delta \max\{\theta_t - \zeta,\, \theta_{t+1} - \zeta\}$, and $\beta_t = \min\{1,\, |1 + \delta/(\hat{\theta}_t - \zeta)|,\, |1 + \delta/(\hat{\theta}_{t+1} - \zeta)|\}$. Note that (B.8) implies that β_t would almost surely converge to 1. Moreover, combining (B.3) and (B.8), we have $x_t \geq L_t \kappa_0/2 \geq \ell \kappa_0/2$ for all $t \geq N_1$, because $L_t \geq \ell$. Recalling the definition of γ_t in (B.9) and noting that $|\alpha_t| \leq \delta \max\{\kappa_1, \kappa_2\} = \delta \kappa^*$ for all t, we deduce that there is a finite random variable $N_2 \geq N_1$ such that

$$\gamma_t \;\geq\; \gamma \quad \text{for all } t \geq N_2, \tag{B.10}$$

where $\gamma = (\ell \kappa_0)^2/(8L) > 0$. Because the first control x_1 is chosen deterministically without using any observed data, we will complete the proof by finding a contradiction when x_1 is negative, and when it is not.

Case 1. $x_1 < 0$. Let $\overline{\epsilon}_t := t^{-1} \sum_{s=2}^{t} \epsilon_s$, and define an event A as follows:

$$A \;=\; \left\{ \begin{array}{l} (\zeta - \theta_1) x_1 \leq \epsilon_1 \leq (\zeta - \theta_1 - \delta) x_1 \\ |\overline{\epsilon}_t| \leq \gamma_t \text{ for all } t \geq 2 \end{array} \right\}.$$

First, we will show that $\mathbb{P}_{\boldsymbol{\theta}}\{A\} > 0$. It follows from the strong law of large numbers that $\mathbb{P}_{\boldsymbol{\theta}}\{|\overline{\epsilon}_t| > \gamma/2, \text{ i.o.}\} = 0$. By (B.10), we also have $\mathbb{P}_{\boldsymbol{\theta}}\{\gamma_t < \gamma/2, \text{ i.o.}\} = 0$. The preceding two facts lead to the conclusion that $\mathbb{P}_{\boldsymbol{\theta}}\{|\overline{\epsilon}_t| > \gamma_t, \text{ i.o.}\} = 0$. In other words, there exists a finite random variable τ such that one almost surely has $|\overline{\epsilon}_t| \leq \gamma_t$ for all $t \geq \tau$. Because the random variable τ is finite, it attains some finite value n with positive probability, implying that $\mathbb{P}_{\boldsymbol{\theta}}\{A\} \geq \mathbb{P}_{\boldsymbol{\theta}}\{A | \tau = n\}\, \mathbb{P}_{\boldsymbol{\theta}}\{\tau = n\} > 0$. Secondly, we will prove by induction that, on the event A, we have

$$\zeta - \delta \;\leq\; \hat{\theta}_t \;\leq\; \zeta \quad \text{for all } t \geq 2. \tag{B.11}$$

For the base step, note that the condition $(\zeta - \theta_1) x_1 \leq \epsilon_1 \leq (\zeta - \theta_1 - \delta) x_1$ for the event A holds if and only if $\zeta - \delta \leq \theta_1 + \epsilon_1/x_1 \leq \zeta$, because $x_1 < 0$. By (B.4), we know that $\hat{\theta}_2 = \theta_1 + \epsilon_1/x_1$. Thus, the condition $(\zeta - \theta_1) x_1 \leq \epsilon_1 \leq (\zeta - \theta_1 - \delta) x_1$ is equivalent to $\zeta - \delta \leq \hat{\theta}_2 \leq \zeta$. For the induction step, suppose that $\zeta - \delta \leq \hat{\theta}_s \leq \zeta$ for all $s \leq t$ in a given time period t. On the event A we have $\sum_{s=2}^{t} \epsilon_s \geq -t\gamma_t$ and $-\sum_{s=2}^{t-1} \epsilon_s \geq -(t-1)\gamma_{t-1}$, from which we deduce that $\epsilon_t \geq -t\gamma_t - (t-1)\gamma_{t-1}$. By (B.9), this implies that

$$\begin{aligned} \epsilon_t \;\geq\; -\frac{\beta_t J_{t-1} + \beta_{t-1} J_{t-2}}{2L} - \frac{(\alpha_t + \alpha_{t-1}) L}{2} &\overset{\text{(b)}}{\geq} -\frac{J_{t-1} + J_{t-2}}{2L} + (\theta_t - \zeta)\delta L \\ &\overset{\text{(c)}}{\geq} -\frac{J_{t-1}}{L} + (\theta_t - \zeta)\delta L, \end{aligned} \tag{B.12}$$

where: (b) follows because $\max\{\alpha_t, \alpha_{t-1}\} \leq -\delta(\theta_t - \zeta)$ and $\max\{\beta_t, \beta_{t-1}\} \leq 1$; and (c) follows because J_t is nondecreasing in t. By the induction hypothesis, we know that $\zeta - \delta \leq \hat{\theta}_t$; hence $-(\hat{\theta}_t - \zeta) \leq \delta$. This implies that $-(\theta_t - \zeta)(\hat{\theta}_t - \zeta) L \leq (\theta_t - \zeta)\delta L$. Consequently, we deduce from (B.12) that $\epsilon_t \geq -J_{t-1}/L - (\theta_t - \zeta)(\hat{\theta}_t - \zeta) L$. Because $L_t \leq L$ for all t, the preceding inequality implies $\epsilon_t \geq -J_{t-1}/L_t - (\theta_t - \zeta)(\hat{\theta}_t - \zeta) L_t$. By (B.3), we know that $L_t = x_t/(\hat{\theta}_t - \zeta)$. Therefore,

$\epsilon_t \geq -(\hat{\theta}_t - \zeta)J_{t-1}/x_t - (\theta_t - \zeta)x_t$. Recalling (B.4) we have $\hat{\theta}_t = (Y_{t-1} + M_{t-1})/J_{t-1}$, implying that

$$\epsilon_t \;\geq\; -\frac{Y_{t-1} - \zeta J_{t-1} + M_{t-1}}{x_t} - (\theta_t - \zeta)x_t \;\overset{\text{(d)}}{=}\; -\frac{Y_t - \zeta J_t + M_{t-1}}{x_t}, \tag{B.13}$$

where (d) follows because $Y_t = Y_{t-1} + \theta_t x_t^2$ and $J_t = J_{t-1} + x_t^2$. Since $x_t < 0$, (B.13) implies $-x_t\epsilon_t \geq Y_t - \zeta J_t + M_{t-1}$. Recalling that $M_t = M_{t-1} + x_t\epsilon_t$, this is equivalent to $(Y_t + M_t)/J_t \leq \zeta$, which implies that $\hat{\theta}_{t+1} \leq \zeta$ by (B.4). To complete the induction, we note that $\sum_{s=2}^{t} \epsilon_s \leq t\gamma_t$ and $-\sum_{s=2}^{t-1} \epsilon_s \leq (t-1)\gamma_{t-1}$ on the event A. Therefore, we have

$$\begin{aligned}\epsilon_t \;\leq\; t\gamma_t + (t-1)\gamma_{t-1} \;=\; \frac{\beta_t J_{t-1} + \beta_{t-1}J_{t-2}}{2L} + \frac{(\alpha_t + \alpha_{t-1})L}{2} &\overset{\text{(e)}}{\leq} \frac{|1 + \delta/(\hat{\theta}_t - \zeta)|\,(J_{t-1} + J_{t-2})}{2L} \\ &\overset{\text{(f)}}{\leq} \frac{|1 + \delta/(\hat{\theta}_t - \zeta)|\,J_{t-1}}{L},\end{aligned} \tag{B.14}$$

where: (e) follows because $\max\{\alpha_t, \alpha_{t-1}\} \leq 0$ and $\max\{\beta_t, \beta_{t-1}\} \leq |1 + \delta/(\hat{\theta}_t - \zeta)|$; and (f) follows because J_t is nondecreasing in t. Now, because $\zeta - \delta \leq \hat{\theta}_t \leq \zeta$ by the induction hypothesis, (B.14) implies that

$$\epsilon_t \;\leq\; -\frac{(\hat{\theta}_t - \zeta + \delta)J_{t-1}}{(\hat{\theta}_t - \zeta)L}. \tag{B.15}$$

Because $-(\theta_t - \zeta + \delta)x_t \geq 0$, we further obtain

$$\epsilon_t \;\leq\; -\frac{(\hat{\theta}_t - \zeta + \delta)J_{t-1}}{(\hat{\theta}_t - \zeta)L} - (\theta_t - \zeta + \delta)x_t.$$

Recalling that $L_t \leq L$ for all t, we deduce by (B.15) that

$$\epsilon_t \;\leq\; -\frac{(\hat{\theta}_t - \zeta + \delta)J_{t-1}}{(\hat{\theta}_t - \zeta)L_t} - (\theta_t - \zeta + \delta)x_t \;\overset{\text{(g)}}{=}\; -\frac{(\hat{\theta}_t - \zeta + \delta)J_{t-1}}{x_t} - (\theta_t - \zeta + \delta)x_t,$$

where (g) follows because $L_t = x_t/(\hat{\theta}_t - \zeta)$ by (B.3). The estimation equation (B.4) implies that $\hat{\theta}_t = (Y_{t-1} + M_{t-1})/J_{t-1}$, hence

$$\epsilon_t \;\leq\; -\frac{Y_{t-1} - (\zeta - \delta)J_{t-1} + M_{t-1}}{x_t} - (\theta_t - \zeta + \delta)x_t \;\overset{\text{(h)}}{=}\; -\frac{Y_t - (\zeta - \delta)J_t + M_{t-1}}{x_t}, \tag{B.16}$$

where (h) follows because $Y_t = Y_{t-1} + \theta_t x_t^2$ and $J_t = J_{t-1} + x_t^2$. Because $x_t < 0$, (B.16) implies that $-x_t\epsilon_t \leq Y_t - (\zeta - \delta)J_t + M_{t-1}$. Recalling $M_t = M_{t-1} + x_t\epsilon_t$, we deduce that $\zeta - \delta \leq (Y_t + M_t)/J_t$. By (B.4) this implies that $\zeta - \delta \leq \hat{\theta}_{t+1}$. As a result, $\mathbb{P}_{\boldsymbol{\theta}}\{A\} > 0$ and on the event A we have $\zeta - \delta \leq \hat{\theta}_t \leq \zeta$ for all $t \geq 2$. But then, this contradicts (B.8), which holds almost surely under the assumption that $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty < \infty\} = 0$.

Case 2. $x_1 \geq 0$. In the definition of the event A, replace the condition $(\zeta - \theta_1)x_1 \leq \epsilon_1 \leq (\zeta - \theta_1 - \delta)x_1$ with $(\zeta - \theta_1 - \delta)x_1 \leq \epsilon_1 \leq (\zeta - \theta_1)x_1$. This change ensures that $\zeta - \delta \leq \hat{\theta}_2 \leq \zeta$ on A. The rest of the proof follows by the same argument. ■

Proof of Proposition 3. We will complete the proof in three steps.

Step 1: Find a relaxed growth envelope for θ_t. Let $z = \varepsilon\sqrt{\kappa_1}/128 > 0$. Recalling that $\{\theta_t\}$ evolves above $\{\theta_t^l\}$ with tolerance R, we deduce that if $\theta_t^l - z \le \theta_t$ then

$$\zeta + \sqrt{\tilde{\kappa}_1 \log(t+1)} \;\le\; \theta_t \tag{B.17}$$

for all t, where $\tilde{\kappa}_1 = \big(\sqrt{\kappa_1} - z/\sqrt{\log 2}\big)^2$. That is, as long as θ_t does not violate the lower bound θ_t^l by more than z, θ_t will satisfy the slightly relaxed growth condition in (B.17). Because z is chosen sufficiently small, (B.17) implies by elementary algebra that $\kappa_1 - \varepsilon\kappa_1/32 \le \tilde{\kappa}_1 \le \kappa_1$. Within a tolerance of R, the unknown parameter θ_t can violate (B.17) in at most $N = \lceil R/z \rceil$ periods. Because $R \le z \log\big(\frac{1-\varepsilon}{1-\varepsilon/2}\big)/\log(1-r)$ where $r = 2^{-\ell^2\kappa_1^2 \log 2/(512\sigma^2)}$, we know that $N\log(1-r) \ge \log\big(\frac{1-\varepsilon}{1-\varepsilon/2}\big)$, i.e., $(1-r)^N \ge (1-\varepsilon)/(1-\varepsilon/2)$.

Step 2: Find a lower bound on the conditional probability that $\hat{\theta}_t$ remains bounded away from ζ. First, let $\delta = \frac{1}{4}\sqrt{\tilde{\kappa}_1 \log 2}$. To prove that $\{\hat{\theta}_t,\, t = 1, 2, \ldots\}$ stays above $\zeta + \delta$ with high probability, let us define $B_t = \{\hat{\theta}_t \ge \zeta + \delta\}$ for all t. We will derive an upper bound on $\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\}$ for $t \ge 2$. Note that the equations (4.1) and (4.2) imply

$$\hat{\theta}_{t+1} \;=\; \frac{Y_t + M_t}{J_t} \quad \text{for } t = 1, 2, \ldots,$$

where $Y_t = \sum_{s=1}^t \theta_s x_s^2$, $J_t = \sum_{s=1}^t x_s^2$, and $M_t = \sum_{s=1}^t x_s\epsilon_s$. Therefore, for $t = 2, 3, \ldots$, we have

$$\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\} \;=\; \mathbb{P}_{\boldsymbol{\theta}}\big\{\hat{\theta}_{t+1} < \zeta + \delta \,\big|\, B_t\big\} \;=\; \mathbb{P}_{\boldsymbol{\theta}}\big\{Y_t + M_t < (\zeta+\delta)J_t \,\big|\, B_t\big\}.$$

By the definition of the sums Y_t, M_t, and J_t, the preceding identity is equivalent to

$$\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\} \;=\; \mathbb{P}_{\boldsymbol{\theta}}\big\{Y_{t-1} + \theta_t x_t^2 + M_{t-1} + \epsilon_t x_t < (\zeta+\delta)(J_{t-1} + x_t^2) \,\big|\, B_t\big\}.$$

Recalling that $\hat{\theta}_t = (Y_{t-1} + M_{t-1})/J_{t-1}$, we can re-express the above identity as follows:

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\} \;&=\; \mathbb{P}_{\boldsymbol{\theta}}\big\{\hat{\theta}_t J_{t-1} + \theta_t x_t^2 + \epsilon_t x_t < (\zeta+\delta)(J_{t-1} + x_t^2) \,\big|\, B_t\big\} \\ &=\; \mathbb{P}_{\boldsymbol{\theta}}\big\{\epsilon_t x_t < (\zeta + \delta - \hat{\theta}_t)J_{t-1} + (\zeta + \delta - \theta_t)x_t^2 \,\big|\, B_t\big\}.\end{aligned}$$

Because $(\zeta + \delta - \hat{\theta}_t)J_{t-1} \le 0$ on the event $B_t = \{\hat{\theta}_t \ge \zeta + \delta\}$, we deduce that

$$\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\} \;\le\; \mathbb{P}_{\boldsymbol{\theta}}\big\{\epsilon_t x_t < (\zeta + \delta - \theta_t)x_t^2 \,\big|\, B_t\big\}. \tag{B.18}$$

We will now find an upper bound on the right hand side of (B.18) under the condition that θ_t satisfies (B.17). Noting that $x_t > 0$ on $B_t = \{\hat{\theta}_t \ge \zeta + \delta\}$, we deduce the following: for all $t \ge 2$ in which θ_t satisfies (B.17), we have

$$\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\} \;\le\; \mathbb{P}_{\boldsymbol{\theta}}\Big\{\epsilon_t < (\zeta + \delta - \theta_t)x_t \,\Big|\, B_t\Big\} \;\le\; \mathbb{P}_{\boldsymbol{\theta}}\Big\{\epsilon_t < \big(\delta - \sqrt{\tilde{\kappa}_1 \log(t+1)}\big)x_t \,\Big|\, B_t\Big\}. \tag{B.19}$$

Since $\delta = \frac{1}{4}\sqrt{\tilde{\kappa}_1 \log 2}$, we know that

$$\delta - \sqrt{\tilde{\kappa}_1 \log(t+1)} \;\le\; \delta - \tfrac{1}{2}\sqrt{\tilde{\kappa}_1 \log 2} - \tfrac{1}{2}\sqrt{\tilde{\kappa}_1 \log t} \;=\; -\tfrac{1}{4}\sqrt{\tilde{\kappa}_1 \log 2} - \tfrac{1}{2}\sqrt{\tilde{\kappa}_1 \log t}.$$

Moreover, on $B_t = \{\hat{\theta}_t \geq \zeta + \delta\}$, we also have $x_t \geq \psi(\zeta + \delta) \geq \ell\delta = \frac{1}{4}\ell\sqrt{\tilde{\kappa}_1 \log 2} > 0$ because $\psi'(\theta) \geq \ell > 0$. Thus, (B.19) implies that

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\} &\leq \mathbb{P}_{\boldsymbol{\theta}}\Big\{\epsilon_t < \big(-\tfrac{1}{4}\sqrt{\tilde{\kappa}_1 \log 2} - \tfrac{1}{2}\sqrt{\tilde{\kappa}_1 \log t}\big)\tfrac{1}{4}\ell\sqrt{\tilde{\kappa}_1 \log 2}\,\Big|\, B_t\Big\} \\
&\overset{\text{(a)}}{=} \mathbb{P}_{\boldsymbol{\theta}}\Big\{\epsilon_t < -\tfrac{1}{16}\ell\tilde{\kappa}_1 \log 2 - \tfrac{1}{8}\ell\tilde{\kappa}_1\sqrt{(\log 2)(\log t)}\Big\} \\
&\overset{\text{(b)}}{\leq} \exp\left(-\frac{\big(\frac{1}{16}\ell\tilde{\kappa}_1 \log 2 + \frac{1}{8}\ell\tilde{\kappa}_1\sqrt{(\log 2)(\log t)}\big)^2}{2\sigma^2}\right) \\
&\leq \exp\Big(-\tfrac{1}{512}\sigma^{-2}\ell^2\tilde{\kappa}_1^2(\log 2)^2 - \tfrac{1}{128}\sigma^{-2}\ell^2\tilde{\kappa}_1^2(\log 2)(\log t)\Big) \qquad \text{(B.20)}
\end{aligned}$$

for all $t \geq 2$ in which (B.17) holds, where: (a) follows by the independence of ϵ_t and $\hat{\theta}_t \in m\mathcal{F}_{t-1}$, and (b) follows by Markov's inequality and the fact that $\mathbb{E}_{\boldsymbol{\theta}}\{e^{k\epsilon_t}\} = e^{\frac{1}{2}k^2\sigma^2}$ for all $k \geq 0$. Note that (B.20) can be expressed in the following compact form:

$$\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\} \;\leq\; c\,t^{-q} \qquad \text{(B.21)}$$

for all $t \geq 2$ in which (B.17) holds, where $q = (\ell^2\tilde{\kappa}_1^2 \log 2)/(128\sigma^2)$ and $c = 2^{-q/4}$. Repeating the above arguments, we also deduce that $\mathbb{P}_{\boldsymbol{\theta}}\{B_2^c\} = \mathbb{P}_{\boldsymbol{\theta}}\big\{\hat{\theta}_2 < \zeta + \delta\big\} \leq c$. On the other hand, whenever the condition (B.17) is violated in period t, then we have a more relaxed lower bound on θ_t. In general, $\theta_t \geq \zeta + \sqrt{\kappa_1 \log(t+1)} - R \geq \zeta + \delta + \frac{1}{4}\sqrt{\kappa_1 \log 2}$, because $R \leq \frac{1}{2}\sqrt{\kappa_1 \log 2}$ and $\delta \leq \frac{1}{4}\sqrt{\kappa_1 \log 2}$. By (B.18) and the arguments used to derive (B.20), we deduce that $\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1}^c \,|\, B_t\} \leq \mathbb{P}_{\boldsymbol{\theta}}\big\{\epsilon_t < -\frac{1}{16}\ell\kappa_1 \log 2\big\} \leq \exp\big(-\frac{1}{512}\sigma^{-2}\ell^2\kappa_1^2(\log 2)^2\big) = 2^{-\ell^2\kappa_1^2 \log 2/(512\sigma^2)} = r$ for $t \geq 2$. As argued in Step 1, (B.24) can be violated in at most N periods, which implies that

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}\left\{\bigcap_{t=1}^{\infty}\big\{\hat{\theta}_{t+1} \geq \zeta + \delta\big\}\right\} &\geq \mathbb{P}_{\boldsymbol{\theta}}\big\{\hat{\theta}_2 \geq \zeta + \delta\big\}\prod_{t=2}^{\infty}\mathbb{P}_{\boldsymbol{\theta}}\Big\{\hat{\theta}_{t+1} \geq \zeta + \delta\,\Big|\,\hat{\theta}_t \geq \zeta + \delta\Big\} \\
&= \mathbb{P}_{\boldsymbol{\theta}}\{B_2\}\prod_{t=2}^{\infty}\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1} \,|\, B_t\} \\
&\overset{\text{(c)}}{\geq} (1-r)^N\prod_{t=1}^{\infty}\big(1 - c\,t^{-q}\big), \qquad \text{(B.22)}
\end{aligned}$$

where (c) follows because $\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1} \,|\, B_t\} \geq 1 - c\,t^{-q}$ in all but at most N periods, and $\mathbb{P}_{\boldsymbol{\theta}}\{B_{t+1} \,|\, B_t\} \geq 1 - r$ in all periods.

Step 3: Prove that $\hat{\theta}_t$ is very likely to be bounded away from ζ. Let $p_\infty = \prod_{t=1}^{\infty}\big(1 - c\,t^{-q}\big)$, and note that

$$\log p_\infty \;=\; \sum_{t=1}^{\infty}\log\big(1 - c\,t^{-q}\big) \;\overset{\text{(d)}}{\geq}\; \log(1-c)\sum_{t=1}^{\infty}t^{-q} \;\overset{\text{(e)}}{\geq}\; \log(1-c)\,\frac{\pi^2}{6} \;\geq\; 2\log(1-c),$$

where: (d) follows because $\log(x) \geq c^{-1}\log(1-c)\,(1-x)$ for $1 - c \leq x \leq 1$; and (e) follows because $\tilde{\kappa}_1 \geq \kappa_1/2 \geq 16\sigma/(\ell\sqrt{\log 2})$, which implies that $q = (\ell^2\tilde{\kappa}_1^2 \log 2)/(128\sigma^2) \geq 2$. The preceding

inequality implies that $p_\infty \geq (1-c)^2 \geq 1-2c$. Recalling that $\tilde{\kappa}_1 \geq \kappa_1/2 \geq 16\sigma\sqrt{2\log(4/\varepsilon)}/(\ell \log 2)$, we further get $c = 2^{-q/4} = 2^{-(\ell^2\tilde{\kappa}_1^2 \log 2)/(512\sigma^2)} \leq 2^{-\log(4/\varepsilon)/\log 2} = 2^{-\log_2(4/\varepsilon)} = \varepsilon/4$. Therefore,

$$\mathbb{P}_{\boldsymbol{\theta}}\left\{ \bigcap_{t=1}^{\infty} \{\hat{\theta}_{t+1} \geq \zeta + \delta\} \right\} \;\geq\; (1-r)^N(1-2c) \;\geq\; (1-r)^N(1-\varepsilon/2).$$

As shown in Step 1, $(1-r)^N \geq (1-\varepsilon)/(1-\varepsilon/2)$; hence $\mathbb{P}_{\boldsymbol{\theta}}\{ \bigcap_{t=1}^{\infty}\{\hat{\theta}_{t+1} \geq \zeta + \delta\}\} \geq 1-\varepsilon$. ■

Proof of Proposition 4. Let $\xi_s := \theta_s - \theta_{s-1}$ for $s = 1, 2, \ldots$, where we set $\theta_0 = 0$ without loss of generality. By (4.1) and (4.2), we know that

$$\begin{aligned}
\hat{\theta}_{t+1} \;=\; \frac{\sum_{s=1}^t \theta_s x_s^2}{\sum_{s=1}^t x_s^2} + \frac{\sum_{s=1}^t x_s\epsilon_s}{\sum_{s=1}^t x_s^2} \;&=\; \theta_{t+1} - \frac{\sum_{s=1}^t (\theta_{t+1}-\theta_s)x_s^2}{\sum_{s=1}^t x_s^2} + \frac{\sum_{s=1}^t x_s\epsilon_s}{\sum_{s=1}^t x_s^2} \\
&\overset{\text{(a)}}{=}\; \theta_{t+1} - \frac{\sum_{s=1}^t \sum_{k=s}^t \xi_{k+1} x_s^2}{\sum_{s=1}^t x_s^2} + \frac{\sum_{s=1}^t x_s\epsilon_s}{\sum_{s=1}^t x_s^2} \\
&=\; \theta_{t+1} - \frac{\sum_{s=1}^t \sum_{k=s}^t \xi_{k+1} x_s^2}{J_t} + \frac{M_t}{J_t} \qquad \text{(B.23)}
\end{aligned}$$

for $t = 1, 2, \ldots$, where $J_t = \sum_{s=1}^t x_s^2$, $M_t = \sum_{s=1}^t x_s\epsilon_s$, and (a) follows because $\theta_s = \sum_{k=1}^s \xi_k$ for all s. On the right hand side of (B.23), the numerator of the second term can be expressed as

$$\sum_{s=1}^t \sum_{k=s}^t \xi_{k+1} x_s^2 \;=\; \sum_{k=1}^t \sum_{s=1}^k \xi_{k+1} x_s^2 \;=\; \sum_{k=1}^t \xi_{k+1} \sum_{s=1}^k x_s^2 \;=\; \sum_{k=1}^t \xi_{k+1} J_k.$$

As a result, equation (B.23) becomes

$$\hat{\theta}_{t+1} \;=\; \theta_{t+1} \;-\; \sum_{k=1}^t \frac{J_k}{J_t}\,\xi_{k+1} \;+\; \frac{M_t}{J_t} \qquad \text{for } t = 1, 2, \ldots$$

This implies that

$$1 - \frac{\hat{\theta}_{t+1}}{\theta_{t+1}} \;=\; \sum_{k=1}^t \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \;-\; \frac{M_t}{\theta_{t+1} J_t} \qquad \text{for } t = 1, 2, \ldots \qquad ■$$

Proof of Theorem 3. Since $\kappa_1 \leq \kappa_2$, we have $R \leq \varepsilon\sqrt{\kappa_1}\log\big(\frac{1-\varepsilon}{1-\varepsilon/2}\big)/\big(128\log(1-r)\big)$ where $r = 2^{-\ell^2\kappa_1^2 \log 2/(512\sigma^2)}$. Thus, we deduce by Proposition 3 that $\mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\} \geq 1-\varepsilon$, where $\delta = \frac{1}{4}\sqrt{\tilde{\kappa}_1 \log 2}$, and $\tilde{\kappa}_1$ is the constant defined in Step 1 in the proof of Proposition 3. We will complete remainder of the proof in three steps.

Step 1: Find a relaxed growth envelope for θ_t. We extend Step 1 in the proof of Proposition 3 as follows. Let $z = \varepsilon\kappa_1/(128\sqrt{\kappa_2}) > 0$. Because $\{\theta_t\}$ evolves between the lower and upper bound processes $\{\theta_t^l\}$ and $\{\theta_t^h\}$ with tolerance R, we deduce that if $\theta_t^l - z \leq \theta_t \leq \theta_t^h + z$ then

$$\zeta + \sqrt{\tilde{\kappa}_1 \log(t+1)} \;\leq\; \theta_t \;\leq\; \zeta + \sqrt{\tilde{\kappa}_2 \log(t+1)} \qquad \text{(B.24)}$$

for all t, where $\tilde{\kappa}_1 = \big(\sqrt{\kappa_1} - z/\sqrt{\log 2}\big)^2$ and $\tilde{\kappa}_2 = \big(\sqrt{\kappa_2} + z/\sqrt{\log 2}\big)^2$; i.e., as long as θ_t does not violate the lower and upper bounds θ_t^l and θ_t^h by more than z, θ_t will satisfy (B.24). By elementary algebra, (B.24) implies that $\kappa_1 - \varepsilon\kappa_1/32 \leq \tilde{\kappa}_1 \leq \kappa_1$ and $\kappa_2 \leq \tilde{\kappa}_2 \leq \kappa_2 + \varepsilon\kappa_1/32$. Because $\kappa_1 \leq \kappa_2 \leq \kappa_1/(1-\varepsilon/8)$, we further deduce that $\tilde{\kappa}_1 \leq \tilde{\kappa}_2 \leq \tilde{\kappa}_1/(1-\varepsilon/4)$. Given tolerance R, θ_t can violate (B.24) in at most $N = \lceil R/z \rceil$ periods. As argued in Step 1 in the proof of Proposition 3, this implies that $(1-r)^N \geq (1-\varepsilon)/(1-\varepsilon/2)$.

Step 2: Derive lower and upper bounds J_t on $\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\}$. By definition, the following statements hold almost surely on $\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\}$: Because $\psi'(\theta) \geq \ell > 0$ for all $\theta \in \mathbb{R}$, we have $x_t \geq \psi(\zeta + \delta) \geq \ell\delta > 0$, which implies that

$$J_t \;\geq\; c_1 t \quad \text{for all } t, \tag{B.25}$$

where $c_1 = \ell^2\delta^2$. Thus, $J_\infty = \infty$, implying by the strong law of large numbers for martingales that $M_t/J_t \to 0$ as $t \to \infty$ (see (i) in the proof of Proposition 1). Recalling Proposition 4, we note that the second term on the right hand side of (4.6) converges to zero. On the other hand, the first term on the right hand side of (4.6) is between 0 and 1 for all t. Therefore, for any $\epsilon_0 > 0$ there is a finite random variable n_0 such that $-\epsilon_0 \leq 1 - \hat{\theta}_{t+1}/\theta_{t+1} \leq 1 + \epsilon_0$ for all $t \geq n_0$. Consequently, we have $-\epsilon_0\theta_{t+1} \leq \hat{\theta}_{t+1} \leq (1+\epsilon_0)\theta_{t+1}$, which implies by elementary algebra that $-\epsilon_0(\theta_{t+1} - \zeta) - (1+\epsilon_0)\zeta \leq \hat{\theta}_{t+1} - \zeta \leq (1+\epsilon_0)(\theta_{t+1} - \zeta) + \epsilon_0\zeta$ for $t \geq n_0$. Thus,

$$|\hat{\theta}_{t+1} - \zeta| \;\leq\; (1+2\epsilon_0)\max\{|\zeta|, |\theta_{t+1} - \zeta|\} \quad \text{for } t \geq n_0. \tag{B.26}$$

Now, recalling that $\{\theta_t\}$ evolves between the lower and upper bound processes $\{\theta_t^l\}$ and $\{\theta_t^h\}$ with tolerance R, we know that $\sum_{t=1}^{\infty} \max\{\theta_t - \theta_t^h,\, \theta_t^l - \theta_t,\, 0\} \leq R < \infty$, and hence there is a finite random variable n_1 such that (B.24) is satisfied for all $t \geq n_1$. Let $n_2 = \max\big\{n_0, n_1, \lceil \exp(4\zeta^2/\kappa_1) \rceil\big\}$, and note that $|\zeta| \leq \frac{1}{2}\sqrt{\tilde{\kappa}_1 \log(t+2)}$ for $t \geq n_2$. Because $\zeta + \sqrt{\tilde{\kappa}_1 \log(t+2)} \leq \theta_{t+1}$ for $t \geq n_1$, $\max\{|\zeta|, |\theta_{t+1} - \zeta|\} = |\theta_{t+1} - \zeta|$ for all $t \geq n_2$. We also know that $\theta_{t+1} \leq \zeta + \sqrt{\tilde{\kappa}_2 \log(t+2)}$ for $t \geq n_1$, which implies that $|\hat{\theta}_{t+1} - \zeta| \leq (1+2\epsilon_0)|\theta_{t+1} - \zeta| \leq (1+2\epsilon_0)\sqrt{\tilde{\kappa}_2 \log(t+2)}$ for all $t \geq n_2 \geq n_1$. Recalling that $x_{t+1} = \psi(\hat{\theta}_{t+1})$ and that $\psi'(\theta) \leq L$ for all $\theta \in \mathbb{R}$, we have $x_{t+1}^2 = \big(\psi(\hat{\theta}_{t+1})\big)^2 \leq K_0 \log(t+2)$ for $t \geq n_2$, where $K_0 = (1+2\epsilon_0)^2 L^2 \tilde{\kappa}_2$. This implies that

$$J_t \;=\; J_{n_2} + \sum_{s=n_2+1}^{t} x_s^2 \;\leq\; J_{n_2} + \sum_{s=n_2+1}^{t} K_0 \log(s+1) \;=\; J_{n_2} + K_0(t+2)\log(t+1)$$

for all $t > n_2$. Let $n_3 = \max\{n_2 + 1, J_{n_2}/K_0\}$. Then,

$$J_t \;\leq\; c_2(t+2)\log(t+1) \quad \text{for all } t \geq n_3, \tag{B.27}$$

where $c_2 = 2K_0$. Having characterized the growth rate of J_t in the lower and upper bounds in (B.25) and (B.27), respectively, we will now show that $\hat{\theta}_t$ is eventually ε-accurate on $\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\}$.

Step 3: Prove that inaccuracy is small on $\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\}$. To complete the proof, we will use the decomposition of inaccuracy in Proposition 4. Because $\{\theta_t\}$ is eventually nondecreasing and (B.24) holds for all $t \geq n_1$, we know that there exists a finite random variable n_4 such that $\theta_{t+1} \geq \theta_t \geq 0$ for all $t \geq n_4$. Let $\lambda = 1 - \varepsilon/4 \in (0,1)$, $m_t = \lceil t^\lambda \rceil$, and $n_5 = (\max\{n_3, n_4\})^{1/\lambda}$. For all $t \geq n_5$, m_t exceeds n_3 and n_4; hence the first term on the right hand side of (4.6) can be expressed as

$$\sum_{k=1}^{t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} = \sum_{k=1}^{n_4-1} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} + \sum_{k=n_4}^{m_t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} + \sum_{k=m_t+1}^{t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \quad \text{for all } t \geq n_5, \tag{B.28}$$

where $\xi_{k+1} = \theta_{k+1} - \theta_k$. For the first sum on the right hand side of (B.28), we note that

$$\sum_{k=1}^{n_4-1} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \leq \sum_{k=1}^{n_4-1} \frac{J_k}{J_t} \cdot \frac{|\xi_{k+1}|}{\theta_{t+1}} \overset{\text{(a)}}{\leq} \frac{J_{n_4}}{J_t} \sum_{k=1}^{n_4-1} \frac{|\xi_{k+1}|}{\theta_{t+1}} \overset{\text{(b)}}{\leq} \frac{K_1}{J_t} \overset{\text{(c)}}{\leq} \frac{K_1}{c_1 t} \quad \text{for all } t \geq n_5, \tag{B.29}$$

where: $K_1 = J_{n_4} \sum_{k=1}^{n_4-1} |\xi_{k+1}|/\theta_{n_4}$; (a) follows because J_t is nondecreasing in t; (b) follows because $\theta_{t+1} \geq \theta_{n_4}$; and (c) follows by (B.25). For the second sum on the right hand side of (B.28), we have

$$\sum_{k=n_4}^{m_t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \overset{\text{(d)}}{\leq} \frac{J_{m_t}}{J_t} \sum_{k=n_4}^{m_t} \frac{\xi_{k+1}}{\theta_{t+1}} \overset{\text{(e)}}{\leq} \frac{J_{m_t}}{J_t} \quad \text{for all } t \geq n_5, \tag{B.30}$$

where: (d) follows because J_t is nondecreasing in t and $\xi_{k+1} \geq 0$ for $k \geq n_4$; and (e) follows because $\sum_{k=n_4}^{m_t} \xi_{k+1} \leq \theta_{m_t} \leq \theta_{t+1}$. Using the bounds we found in Step 2, we deduce from (B.30) that

$$\sum_{k=n_4}^{m_t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \overset{\text{(f)}}{\leq} \frac{c_2(m_t+2)\log(m_t+1)}{c_1 t} \overset{\text{(g)}}{\leq} \frac{c_2(t^\lambda+3)\log(t^\lambda+2)}{c_1 t} \quad \text{for all } t \geq n_5, \tag{B.31}$$

where: (f) follows by (B.25) and (B.27); and (g) follows because $m_t = \lceil t^\lambda \rceil$. Now, note that as $t \to \infty$ the right hand side of (B.29) converges to zero. Because $\lambda < 1$, the right hand side of (B.31) similarly converges to zero. Hence there is a finite random variable $n_6 \geq n_5$ such that

$$\sum_{k=1}^{n_4-1} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} + \sum_{k=n_4}^{m_t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \leq \frac{\varepsilon}{4} \quad \text{for all } t \geq n_6. \tag{B.32}$$

For the third sum on the right hand side of (B.28), note that

$$\sum_{k=m_t+1}^{t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \overset{\text{(h)}}{\leq} \sum_{k=m_t+1}^{t} \frac{\xi_{k+1}}{\theta_{t+1}} \overset{\text{(i)}}{=} 1 - \frac{\sum_{k=0}^{m_t} \xi_{k+1}}{\sum_{k=0}^{t} \xi_{k+1}} = 1 - \frac{\theta_{m_t+1}}{\theta_{t+1}} \quad \text{for all } t \geq n_6, \tag{B.33}$$

where: (h) follows because $\xi_{k+1} \geq 0$ for $k \geq m_t \geq n_4$ and $J_k \leq J_t$ for $k \leq t$; and (i) follows because $\theta_{t+1} = \sum_{k=0}^{t} \xi_{k+1}$. Because (B.24) holds for all $t \geq m_t \geq n_1$, (B.33) implies

$$\sum_{k=m_t+1}^{t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \leq 1 - \frac{\zeta + \sqrt{\tilde{\kappa}_1 \log(m_t+2)}}{\zeta + \sqrt{\tilde{\kappa}_2 \log(t+2)}} = \frac{\sqrt{\tilde{\kappa}_2 \log(t+2)} - \sqrt{\tilde{\kappa}_1 \log(m_t+2)}}{\zeta + \sqrt{\tilde{\kappa}_2 \log(t+2)}} \quad \text{for all } t \geq n_6.$$

Noting that $m_t + 2 = \lceil t^\lambda \rceil + 2 \geq t^\lambda + 2 \geq (t+2)^\lambda$, we further obtain

$$\sum_{k=m_t+1}^{t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \;\leq\; \frac{\sqrt{\tilde{\kappa}_2 \log(t+2)} - \sqrt{\lambda \tilde{\kappa}_1 \log(t+2)}}{\zeta + \sqrt{\tilde{\kappa}_2 \log(t+2)}} \;=\; \frac{1 - \sqrt{\lambda \tilde{\kappa}_1 / \tilde{\kappa}_2}}{1 + \zeta / \sqrt{\tilde{\kappa}_2 \log(t+2)}} \quad \text{for all } t \geq n_6.$$

Recall that $|\zeta| \leq \frac{1}{2}\sqrt{\tilde{\kappa}_1 \log(t+2)} \leq \frac{1}{2}\sqrt{\tilde{\kappa}_2 \log(t+2)}$ for all $t \geq n_6 \geq n_2$. Thus,

$$\sum_{k=m_t+1}^{t} \frac{J_k}{J_t} \cdot \frac{\xi_{k+1}}{\theta_{t+1}} \;\leq\; 2\big(1 - \sqrt{\lambda \tilde{\kappa}_1 / \tilde{\kappa}_2}\big) \overset{\text{(j)}}{\leq} \frac{\varepsilon}{2} \quad \text{for all } t \geq n_6, \tag{B.34}$$

where (j) follows because $\lambda = 1 - \varepsilon/4$ and $\tilde{\kappa}_1 \geq (1 - \varepsilon/4)\tilde{\kappa}_2$. Combining (B.32) and (B.34), we deduce that the first term on the right hand side of (4.6) will eventually be between 0 and $3\varepsilon/4$ almost surely on $\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\}$. Recall that the second term on the right hand side of (4.6) converges to zero almost surely on $\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\}$, which implies that there is a finite random variable n_7 such that $|M_t/(\theta_{t+1} J_t)| \leq \varepsilon/4$ for all $t \geq n_7$. Therefore $|1 - \hat{\theta}_{t+1}/\theta_{t+1}| \leq \varepsilon$ for all $t \geq \max\{n_6, n_7\}$. Because the above statements hold almost surely on $\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\}$, we conclude that $\hat{\theta}_t$ is eventually ε-accurate on $\{\hat{\theta}_t \geq \zeta + \delta \text{ for all } t\}$, and hence $\{\hat{\theta}_t\}$ is asymptotically ε-accurate. ■

Appendix C: An Example with Cyclical Pattern of Estimates

The following example demonstrates that if the conditions of Theorem 2 are violated, then the certainty-equivalence estimates $\{\hat{\theta}_t\}$ can keep fluctuating without converging to ζ.

Example 7: Another boundedly changing environment. *Assume that $f(x, \theta) = \theta x$ for all $x \in \mathcal{X} = \mathbb{R}$ and $\theta \in \Theta = \mathbb{R}$. Let $t_n = 2^n$ for all $n = 1, 2, \ldots$, $\mathcal{T}_{\text{odd}} = \bigcup_{k=1}^{\infty} [t_{2k-1}, t_{2k})$, and $\mathcal{T}_{\text{even}} = \bigcup_{k=1}^{\infty} [t_{2k}, t_{2k+1})$. Construct a sequence $\{\theta_t,\, t = 1, 2, \ldots\}$ such that*

$$\theta_t \;=\; \begin{cases} 0.8 & \text{if } \; t \in \mathcal{T}_{\text{odd}}, \\ 1.2 & \text{if } \; t \in \mathcal{T}_{\text{even}}, \end{cases}$$

for all $t = 1, 2, \ldots$ The decision maker sets the initial control as $x_1 = 1$, and subsequently uses the control function $\psi(\theta) = -1 + \theta$. To observe the cyclical behavior of $\{\hat{\theta}_t\}$ crisply, suppose that $\epsilon_t \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2)$ with $\sigma = 0.01$.

Note that Example 7 does not satisfy the conditions of Theorem 2 because its unknown parameter sequence $\{\theta_t\}$ will visit both sides of ζ infinitely often. Figure 16 shows that the sample paths of $\{\hat{\theta}_t\}$ keep fluctuating in this example. Moreover, the sample paths get very close to ζ without converging to ζ; the running minima of all sample paths is approximately 1.0152.

Since we focus on incomplete learning in this paper, we rule out such cyclical patterns of $\{\hat{\theta}_t\}$ by assuming in Theorem 2 that $\{\theta_t\}$ will eventually be fluctuating in bounded interval on one side of ζ.

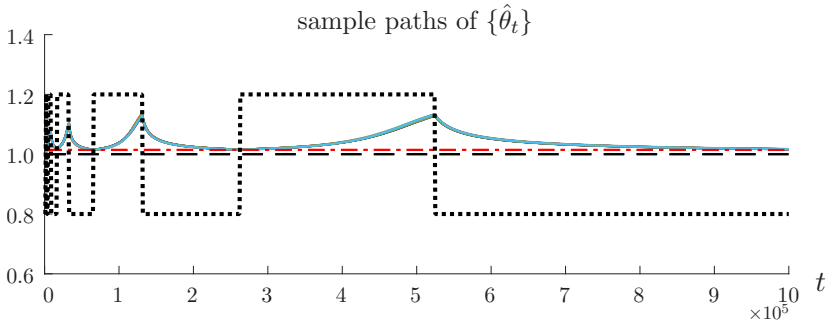

Figure 16: **Certainty-equivalence estimates in Example 7.** The sample paths of $\{\hat{\theta}_t\}$ are shown in the *solid curves*, and the values of $\{\theta_t\}$ are shown in the *dotted curve*. The *dashed line* shows the value of $\zeta = 1$, and the *dash-dotted line* shows the value of 1.0152.

Appendix D: Proof of Theorem 4

Proof of Theorem 4(i). Because $\theta \neq \zeta$, we have either $\theta > \zeta$ or $\theta < \zeta$. Assume without loss of generality that $\theta > \zeta$. We deduce from (2.2) that $\hat{\theta}_{t+1}$ satisfies

$$\sum_{s=1}^{t} \big(y_s - g(\hat{\theta}_{t+1} x_s)\big) g'(\hat{\theta}_{t+1} x_s)\, x_s \;=\; 0 \quad \text{for } t = 1, 2, \ldots \tag{D.1}$$

By (2.1) and the fact that $f(x, \theta) = g(\theta x)$ for $x \in \mathcal{X}$ and $\theta \in \Theta$, (D.1) implies that

$$\sum_{s=1}^{t} \big(g(\theta x_s) - g(\hat{\theta}_{t+1} x_s) + \epsilon_s\big) g'(\hat{\theta}_{t+1} x_s)\, x_s \;=\; 0 \quad \text{for } t = 1, 2, \ldots \tag{D.2}$$

Invoking the mean value theorem in (D.2) we deduce that, for all s and t, there exists a random variable $\bar{\varphi}(\hat{\theta}_{t+1}, x_s)$ between $\hat{\theta}_{t+1}$ and θ such that $g(\theta x_s) - g(\hat{\theta}_{t+1} x_s) = g'\big(\bar{\varphi}(\hat{\theta}_{t+1}, x_s) x_s\big)(\theta - \hat{\theta}_{t+1}) x_s$. Therefore, (D.2) can be expressed as follows:

$$(\theta - \hat{\theta}_{t+1}) \sum_{s=1}^{t} g'\big(\bar{\varphi}(\hat{\theta}_{t+1}, x_s) x_s\big) g'(\hat{\theta}_{t+1} x_s)\, x_s^2 + \sum_{s=1}^{t} g'(\hat{\theta}_{t+1} x_s)\, x_s \epsilon_s \;=\; 0 \quad \text{for } t = 1, 2, \ldots \tag{D.3}$$

For notational brevity, define $M_t(\vartheta) = \sum_{s=1}^{t} g'(\vartheta x_s) x_s \epsilon_s$, $J_t(\vartheta) = \sum_{s=1}^{t} \big(g'(\vartheta x_s)\, x_s\big)^2$, and $\tilde{J}_t(\vartheta) = \sum_{s=1}^{t} g'\big(\bar{\varphi}(\vartheta, x_s) x_s\big) g'(\vartheta x_s)\, x_s^2$ for $t = 1, 2, \ldots$ and $\vartheta \in \Theta$. Then, (D.3) implies that

$$\hat{\theta}_{t+1} \;=\; \theta + \frac{M_t(\hat{\theta}_{t+1})}{\tilde{J}_t(\hat{\theta}_{t+1})} \quad \text{for } t = 1, 2, \ldots \tag{D.4}$$

Suppose towards a contradiction that, for all $\vartheta \in \Theta$, $J_t(\vartheta)$ diverges to ∞ almost surely. Note that, for all $t = 1, 2, \ldots$, and $\vartheta \in \Theta$, we have the following by elementary algebra:

$$\begin{aligned} S_t(\vartheta) - S_t(\theta) \;&=\; \sum_{s=1}^{t} \big(y_s - g(\vartheta x_s)\big)^2 - \sum_{s=1}^{t} \big(y_s - g(\theta x_s)\big)^2 \\ &\overset{\text{(a)}}{=}\; \sum_{s=1}^{t} \big(g(\theta x_s) - g(\vartheta x_s) + \epsilon_s\big)^2 - \sum_{s=1}^{t} \epsilon_s^2 \\ &=\; \sum_{s=1}^{t} \big(g(\theta x_s) - g(\vartheta x_s)\big)^2 + 2 \sum_{s=1}^{t} \big(g(\theta x_s) - g(\vartheta x_s)\big) \epsilon_s, \end{aligned} \tag{D.5}$$

where: (a) follows by (2.1) and the fact that $f(x,\theta) = g(\theta x)$ for $x \in \mathcal{X}$ and $\theta \in \Theta$. For brevity, let $\mathcal{J}_t(\vartheta) = \sum_{s=1}^t \big(g(\theta x_s) - g(\vartheta x_s)\big)^2$, $\mathcal{M}_t(\vartheta) = \sum_{s=1}^t \big(g(\theta x_s) - g(\vartheta x_s)\big)\epsilon_s$, and note that

$$S_t(\vartheta) - S_t(\theta) \;=\; \mathcal{J}_t(\vartheta) + 2\mathcal{M}_t(\vartheta) \;=\; \mathcal{J}_t(\vartheta) \cdot \left(1 + \frac{2\mathcal{M}_t(\vartheta)}{\mathcal{J}_t(\vartheta)}\right) \tag{D.6}$$

for all $t = 1, 2, \ldots$, and $\vartheta \in \Theta$. Because $g'(\xi) \leq \tilde{L}$ for all $\xi \in \Xi = \{\theta x : (x,\theta) \in \mathcal{X} \times \Theta\}$, we deduce that, for all $\vartheta \in \Theta$, $\{\mathcal{M}_t(\vartheta),\, t = 1, 2, \ldots\}$ is a square-integrable and zero-mean martingale with respect to the filtration $\{\mathcal{F}_t,\, t = 1, 2, \ldots\}$, where $\mathcal{F}_t = \sigma(\epsilon_1, \ldots, \epsilon_t)$ for $t = 1, 2, \ldots$ Note that the predictable compensator of $\{\mathcal{M}_t(\vartheta),\, t = 1, 2, \ldots\}$ is $\mathcal{V}_t(\vartheta) = \sigma^2 \sum_{s=1}^t \big(g(\theta x_s) - g(\vartheta x_s)\big)^2 = \sigma^2 \mathcal{J}_t(\vartheta)$.

Let $\tilde{\delta} > 0$, and choose $\vartheta \in \Theta$ such that $\vartheta \leq \theta - \tilde{\delta}$. Since $J_t(\vartheta) \xrightarrow{\text{a.s.}} \infty$ as $t \to \infty$, and $\tilde{\ell} \leq g'(\xi) \leq \tilde{L}$ for all $\xi \in \Xi = \{\theta x : (x,\theta) \in \mathcal{X} \times \Theta\}$, we have $\mathcal{J}_t(\vartheta) \xrightarrow{\text{a.s.}} \infty$ as $t \to \infty$. By the strong law of large numbers for martingales (see Williams 1991, pp. 122-124), this further implies that $\mathcal{M}_t(\vartheta)/\mathcal{J}_t(\vartheta) \xrightarrow{\text{a.s.}} 0$ as $t \to \infty$. Combining this fact with (D.6), we deduce that, for all $\vartheta \in \Theta$ satisfying $\vartheta \leq \theta - \tilde{\delta}$, $S_t(\vartheta) - S_t(\theta) \xrightarrow{\text{a.s.}} \infty$ as $t \to \infty$. By symmetry, $S_t(\vartheta) - S_t(\theta) \xrightarrow{\text{a.s.}} \infty$ as $t \to \infty$ for all $\vartheta \in \Theta$ satisfying $\vartheta \geq \theta - \tilde{\delta}$. Thus, we have $S_t(\vartheta) - S_t(\theta) \xrightarrow{\text{a.s.}} \infty$ as $t \to \infty$ for all $\vartheta \in \Theta$ such that $|\vartheta - \theta| \geq \tilde{\delta}$. On the other hand, we also have $S_t(\vartheta) - S_t(\theta) = 0$ for all t, if $\vartheta = \theta$. Based on these facts, we conclude that, with probability one, any ϑ outside the $\tilde{\delta}$-neighborhood of θ cannot be the minimizer of $S_t(\cdot)$ as $t \to \infty$. Therefore, for all $\tilde{\delta} > 0$, the following holds with probability one: the estimator $\hat{\theta}_{t+1}$, which minimizes $S_t(\cdot)$, will eventually be located inside the $\tilde{\delta}$-neighborhood of θ. Since $\tilde{\delta} > 0$ was selected arbitrarily, we deduce that $\hat{\theta}_{t+1}$ converges to θ almost surely (this is a standard proof argument regarding the consistency of M-estimators; see, e.g., Wu 1981, Lemma 1). Now, recall that $x_t = \psi(\hat{\theta}_t) = L_t(\hat{\theta}_t - \zeta)$ for all t, where $L_t = \psi'(c_t)$ and c_t is between $\hat{\theta}_t$ and ζ. Let $\bar{L}_t = g'(\bar{\varphi}(\hat{\theta}_{t+1})x_t)$, $\tilde{L}_t = g'(\hat{\theta}_{t+1}x_t)$, $\tilde{a}_t = (\theta - \zeta)\min\{\bar{L}_t L_t(\hat{\theta}_t - \zeta)$, $\bar{L}_{t+1}L_{t+1}(\hat{\theta}_{t+1} - \zeta)\}$, $\tilde{b}_t = (\theta - \zeta)\min\{\tilde{J}_{t-1}(\hat{\theta}_{t+1})/(\tilde{L}_t L_t(\hat{\theta}_t - \zeta)), \tilde{J}_t(\hat{\theta}_{t+2})/(\tilde{L}_{t+1}L_{t+1}(\hat{\theta}_{t+1} - \zeta))\}$, $\tilde{c}_t = \min\{|M_{t-1}(\hat{\theta}_{t+1})|/(\tilde{L}_t L_t(\hat{\theta}_t - \zeta)), |M_t(\hat{\theta}_{t+2})|/(\tilde{L}_{t+1}L_{t+1}(\hat{\theta}_{t+1} - \zeta))\}$, and define a stochastic process $\{\tilde{\gamma}_t,\, t = 1, 2, \ldots\}$ such that

$$\tilde{\gamma}_t \;=\; \tfrac{1}{2t}(\tilde{a}_t + \tilde{b}_t + \tilde{c}_t) \quad \text{ for all } t = 1, 2, \ldots \tag{D.7}$$

Because $\hat{\theta}_{t+1} \to \theta$ almost surely, we deduce that $\tilde{a}_t/t \to 0$ and $\tilde{b}_t/t \to (\theta - \zeta)g'(\theta\psi(\theta))\psi(\theta)$ almost surely. Moreover, by the strong law of large numbers for martingales, $\tilde{c}_t/t \to 0$ almost surely. Letting $\tilde{\gamma} = (\theta - \zeta)g'(\theta\psi(\theta))\psi(\theta)/2 > 0$, we thus have $\tilde{\gamma}_t \to \tilde{\gamma}$ almost surely. As in the proof of Theorem 2, we consider the two possible cases for x_1.

Case 1. $x_1 < 0$. Let $\delta > 0$, $\overline{\epsilon}_t := t^{-1}\sum_{s=2}^t \epsilon_s$, and

$$\tilde{A} \;=\; \left\{ \begin{array}{l} (\zeta - \theta)\tilde{L}x_1 \leq \epsilon_1 \leq (\zeta - \theta - \delta)\tilde{\ell}x_1 \\ |\overline{\epsilon}_t| \leq \tilde{\gamma}_t \text{ for all } t \geq 2 \end{array} \right\}.$$

To see that $\mathbb{P}_{\boldsymbol{\theta}}\{\tilde{A}\} > 0$, we deduce by the strong law of large numbers that $\mathbb{P}_{\boldsymbol{\theta}}\{|\overline{\epsilon}_t| > \tilde{\gamma}/2, \text{ i.o.}\} = 0$. Because $\tilde{\gamma}_t \to \tilde{\gamma}$ almost surely, we also have $\mathbb{P}_{\boldsymbol{\theta}}\{\tilde{\gamma}_t < \tilde{\gamma}/2, \text{ i.o.}\} = 0$. Thus, $\mathbb{P}_{\boldsymbol{\theta}}\{|\overline{\epsilon}_t| > \tilde{\gamma}_t, \text{ i.o.}\} = 0$;

i.e., there is a finite random variable $\tilde{\tau}$ such that $|\overline{\epsilon}_t| \leq \tilde{\gamma}_t$ for all $t \geq \tilde{\tau}$, with probability one. Since $\tilde{\tau}$ attains some finite value n with positive probability, $\mathbb{P}_{\boldsymbol{\theta}}\{\tilde{A}\} \geq \mathbb{P}_{\boldsymbol{\theta}}\{\tilde{A}|\tilde{\tau}=n\}\,\mathbb{P}_{\boldsymbol{\theta}}\{\tilde{\tau}=n\} > 0$. In the remainder of our analysis in this case, we will prove by induction that, on $\tilde{A}$, we have $\zeta-\delta \leq \hat{\theta}_t \leq \zeta$ for all $t \geq 2$. For the base step, note that the condition $(\zeta-\theta)\tilde{L}x_1 \leq \epsilon_1 \leq (\zeta-\theta-\delta)\tilde{\ell}x_1$ implies that $\zeta-\delta \leq \hat{\theta}_2 \leq \zeta$. For the induction step, suppose that $\zeta-\delta \leq \hat{\theta}_s \leq \zeta$ for all $s \leq t$. Since $\sum_{s=2}^{t}\epsilon_s \geq -t\tilde{\gamma}_t$ and $-\sum_{s=2}^{t-1}\epsilon_s \geq -(t-1)\tilde{\gamma}_{t-1}$ on $\tilde{A}$, we have $\epsilon_t \geq -t\tilde{\gamma}_t-(t-1)\tilde{\gamma}_{t-1}$. By (D.7) and elementary algebra, this implies that

$$\begin{aligned}\epsilon_t \;&\geq\; -\bar{L}_t L_t(\theta-\zeta)(\hat{\theta}_t-\zeta) - \frac{(\theta-\zeta)\tilde{J}_{t-1}(\hat{\theta}_{t+1})}{\tilde{L}_t L_t(\hat{\theta}_t-\zeta)} - \frac{|M_{t-1}(\hat{\theta}_{t+1})|}{\tilde{L}_t L_t(\hat{\theta}_t-\zeta)} \\ &\overset{\text{(b)}}{\geq}\; -\bar{L}_t L(\theta-\zeta)(\hat{\theta}_t-\zeta) - \frac{(\theta-\zeta)\tilde{J}_{t-1}(\hat{\theta}_{t+1}) + M_{t-1}(\hat{\theta}_{t+1})}{\tilde{L}_t L_t(\hat{\theta}_t-\zeta)}, \end{aligned} \tag{D.8}$$

where (b) follows since $|y| \geq y$ for all $y \in \mathbb{R}$, and $\hat{\theta}_t \leq \zeta$. Because $x_t = L_t(\hat{\theta}_t-\zeta)$ for all t, we deduce from (D.8) that

$$\epsilon_t \;\geq\; -\bar{L}_t(\theta-\zeta)x_t - \frac{(\theta-\zeta)\tilde{J}_{t-1}(\hat{\theta}_{t+1}) + M_{t-1}(\hat{\theta}_{t+1})}{\tilde{L}_t x_t}. \tag{D.9}$$

Since $x_t < 0$, (D.9) implies that

$$(\theta-\zeta)\big(\tilde{J}_{t-1}(\hat{\theta}_{t+1}) + \bar{L}_t\tilde{L}_t x_t^2\big) + M_{t-1}(\hat{\theta}_{t+1}) + \tilde{L}_t x_t \epsilon_t \;\leq\; 0. \tag{D.10}$$

Because $\tilde{J}_t(\hat{\theta}_{t+1}) = \tilde{J}_{t-1}(\hat{\theta}_{t+1}) + \bar{L}_t\tilde{L}_t x_t^2$ and $M_t(\hat{\theta}_{t+1}) = M_{t-1}(\hat{\theta}_{t+1}) + \tilde{L}_t x_t \epsilon_t$, (D.10) is equivalent to $\theta + M_t(\hat{\theta}_{t+1})/\tilde{J}_t(\hat{\theta}_{t+1}) \leq \zeta$. By (D.4), this implies that $\hat{\theta}_{t+1} \leq \zeta$, which proves one of the inequalities in the induction step. For the other inequality, note that $\epsilon_t \leq t\tilde{\gamma}_t + (t-1)\tilde{\gamma}_{t-1}$ on $\tilde{A}$. Therefore, using (D.7) and elementary algebra, we obtain the following:

$$\begin{aligned}\epsilon_t \;&\leq\; \bar{L}_t L(\theta-\zeta)(\hat{\theta}_t-\zeta) + \frac{(\theta-\zeta)\tilde{J}_{t-1}(\hat{\theta}_{t+1})}{\tilde{L}_t L_t(\hat{\theta}_t-\zeta)} + \frac{|M_{t-1}(\hat{\theta}_{t+1})|}{\tilde{L}_t L_t(\hat{\theta}_t-\zeta)} \\ &\overset{\text{(c)}}{\leq}\; -\bar{L}_t L_t(\theta-\zeta+\delta)(\hat{\theta}_t-\zeta) - \frac{(\theta-\zeta+\delta)\tilde{J}_{t-1}(\hat{\theta}_{t+1}) + M_{t-1}(\hat{\theta}_{t+1})}{\tilde{L}_t L_t(\hat{\theta}_t-\zeta)}, \end{aligned} \tag{D.11}$$

where (c) follows because $-|y| \leq y$ for all $y \in \mathbb{R}$, $\hat{\theta}_t \leq \zeta$, and $\delta > 0$. Recalling that $x_t = L_t(\hat{\theta}_t-\zeta)$, we further deduce that

$$\epsilon_t \;\leq\; -\bar{L}_t(\theta-\zeta+\delta)x_t - \frac{(\theta-\zeta+\delta)\tilde{J}_{t-1}(\hat{\theta}_{t+1}) + M_{t-1}(\hat{\theta}_{t+1})}{\tilde{L}_t x_t}. \tag{D.12}$$

Combining (D.12) with the fact that $x_t < 0$, we have

$$(\theta-\zeta+\delta)\big(\tilde{J}_{t-1}(\hat{\theta}_{t+1}) + \bar{L}_t\tilde{L}_t x_t^2\big) + M_{t-1}(\hat{\theta}_{t+1}) + \tilde{L}_t x_t \epsilon_t \;\geq\; 0. \tag{D.13}$$

Thus, $\theta + M_t(\hat{\theta}_{t+1})/\tilde{J}_t(\hat{\theta}_{t+1}) \geq \zeta-\delta$, implying by (D.4) that $\hat{\theta}_{t+1} \geq \zeta-\delta$. This completes the induction. As a result, $\mathbb{P}_{\boldsymbol{\theta}}\{\tilde{A}\} > 0$, and on $\tilde{A}$, we have $\zeta-\delta \leq \hat{\theta}_t \leq \zeta$ for all $t \geq 2$. This

contradicts with the fact that $\hat{\theta}_{t+1} \to \theta$ almost surely. Therefore, there must exist $\vartheta \in \Theta$ such that $J_t(\vartheta)$ converges to a finite limit with positive probability. Since $g'(\xi) \geq \tilde{\ell}$ for all $\xi \in \Xi = \{\theta x : (x, \theta) \in \mathcal{X} \times \Theta\}$, we have $J_t(\vartheta) \geq \tilde{\ell}^2 \sum_{s=1}^{t} x_s^2$. Consequently, because $\lim_{t\to\infty} J_t(\vartheta) < \infty$ with positive probability, we conclude that x_t converges to zero with positive probability; i.e., $\mathbb{P}_{\boldsymbol{\theta}}\{x_t \to 0\} = \mathbb{P}_{\boldsymbol{\theta}}\{\hat{\theta}_t \to \zeta\} > 0$.

Case 2. $x_1 \geq 0$. In this case, we modify the definition of $\tilde{A}$ by replacing the condition $(\zeta - \theta)\tilde{L}x_1 \leq \epsilon_1 \leq (\zeta - \theta - \delta)\tilde{\ell}x_1$ with $(\zeta - \theta - \delta)\tilde{\ell}x_1 \leq \epsilon_1 \leq (\zeta - \theta)\tilde{L}x_1$ to ensure that $\zeta - \delta \leq \hat{\theta}_2 \leq \zeta$. The remainder of the proof follows by the same argument. ■

Proof of Theorem 4(ii). Since $\psi(\cdot)$ is monotone and there exists no $\zeta \in \Theta$ satisfying $\psi(\zeta) = 0$, we deduce that either $\psi(\vartheta) > 0$ for all $\vartheta \in \Theta$, or $\psi(\vartheta) < 0$ for all $\vartheta \in \Theta$. Assume without loss generality that $\psi(\vartheta) > 0$ for all $\vartheta \in \Theta$. For $t = 1, 2, \ldots$ and $\vartheta \in \Theta$, let $J_t(\vartheta) = \sum_{s=1}^{t} \big(g'(\vartheta x_s)\, x_s\big)^2$ and $J_\infty(\vartheta) = \lim_{t\to\infty} J_t(\vartheta)$. Suppose towards a contradiction that there exists $\tilde{\vartheta} \in \Theta$ such that $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty(\tilde{\vartheta}) < \infty\} > 0$. Because $g'(\xi) \geq \tilde{\ell}$ for all $\xi \in \Xi = \{\theta x : (x, \theta) \in \mathcal{X} \times \Theta\}$, we deduce that $J_\infty(\tilde{\vartheta}) \geq \tilde{\ell}^2 \sum_{s=1}^{t} x_s^2$ for all t. Therefore, on the event $\{J_\infty(\tilde{\vartheta}) < \infty\}$, we have $\sum_{s=1}^{\infty} x_s^2 < \infty$, which implies that $x_t \to 0$. Since $\psi(\cdot)$ is differentiable and monotone, this further implies that, on the event $\{J_\infty(\tilde{\vartheta}) < \infty\}$, $\hat{\theta}_t \in \Theta$ converges to a finite limit in Θ. But, since $\psi(\vartheta) > 0$ for all $\vartheta \in \Theta$, this contradicts the fact that $x_t = \psi(\hat{\theta}_t)$ converges to 0 on $\{J_\infty(\tilde{\vartheta}) < \infty\}$. Thus, $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty(\vartheta) < \infty\} = 0$ and $\mathbb{P}_{\boldsymbol{\theta}}\{J_\infty(\vartheta) = \infty\} = 1$ for all $\vartheta \in \Theta$. Using the argument following (D.6) in the proof of Theorem 4(i), we deduce that (a) the strong law of large numbers for martingales (see Williams 1991, pp. 122-124) and (b) the fact that $J_t(\vartheta) \to \infty$ almost surely for all $\vartheta \in \Theta$ jointly imply that $\inf_{\vartheta\in\Theta}\{S_t(\vartheta) - S_t(\theta) : |\vartheta - \theta| \geq \tilde{\delta}\} \to \infty$ for any given $\tilde{\delta} > 0$. Consequently, the estimator $\hat{\theta}_{t+1}$ that minimizes $S_t(\cdot)$ converges to θ almost surely. ■

Appendix E: Proofs of the Results in §6

For the sake of comparison with the results in §4, we would like to provide some high-level intuition for the results in §6. Roughly speaking, in the proofs of our incomplete learning results in Theorems 1 and 2, we considered how a few initial response realizations can make the certainty-equivalence estimate $\hat{\theta}_t$ "get stuck" in a small neighborhood of ζ. In the proof of Theorem 5, we will show that, by limiting the estimation memory, one can eliminate such negative impact of initial response realizations and hence avoid incomplete learning. In the proof of Theorem 6, we will show that, in static environments, this result can further be strengthened to establish the almost sure convergence of the certainty-equivalence estimates, resulting in asymptotic estimation accuracy. Finally, in the proof of Theorem 7, we study how the certainty-equivalence estimates evolve in changing environments and in the absence of incomplete learning. If the unknown parameter θ_t drifts away from ζ at a slow and concavely growing rate, the certainty-equivalence estimates can asymptotically track θ_t.

Proof of Theorem 5. Because $\widehat{\Theta}_{n+1} = \varphi(w_n, \tau_n)$ and $\tau_n = \tau_{n-1} + w_n$, we know by (6.5) that

$$-2 \sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \Big(y_s - f\big(x_s, \widehat{\Theta}_{n+1}\big) \Big) f_\theta\big(x_s, \widehat{\Theta}_{n+1}\big) \;=\; 0,$$

for $n = 1, 2, \ldots$ Recalling that $x_s = X_n$ for $s = \tau_{n-1} + 1, \ldots, \tau_{n-1} + w_n$, we have

$$\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \Big(y_s - f\big(X_n, \widehat{\Theta}_{n+1}\big) \Big) \;=\; 0, \tag{E.1}$$

for all n. Using the response model (2.5) we further obtain

$$\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \Big(f\big(X_n, \theta_s\big) - f\big(X_n, \widehat{\Theta}_{n+1}\big) + \epsilon_s \Big) \;=\; 0, \tag{E.2}$$

for all n. Now, by the mean value theorem there exists a random variable $C_{s,n}$ on the line segment between θ_s and $\widehat{\Theta}_{n+1}$ such that $f\big(X_n, \theta_s\big) - f\big(X_n, \widehat{\Theta}_{n+1}\big) = \big(\theta_s - \widehat{\Theta}_{n+1}\big)\, f_\theta\big(X_n, C_{s,n}\big)$. Consequently, (E.2) implies the following for all n:

$$\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \big(\theta_s - \widehat{\Theta}_{n+1}\big)\, f_\theta\big(X_n, C_{s,n}\big) + \mathcal{S}(w_n, \tau_{n-1}) \;=\; 0,$$

where $\mathcal{S}(w,t) = \sum_{s=t+1}^{t+w} \epsilon_s$. By elementary algebra, we can express the preceding identity as

$$\widehat{\Theta}_{n+1} \;=\; \sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \mu_{s,n}\, \theta_s \;+\; \left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} f_\theta\big(X_n, C_{s,n}\big) \right)^{-1} \mathcal{S}(w_n, \tau_{n-1}) \tag{E.3}$$

for all n, where

$$\mu_{s,n} \;=\; \left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} f_\theta\big(X_n, C_{s,n}\big) \right)^{-1} f_\theta\big(X_n, C_{s,n}\big).$$

Let us focus on the second term on the right hand side of (E.3). Note that

$$\begin{aligned}
\left| \left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} f_\theta\big(X_n, C_{s,n}\big) \right)^{-1} \mathcal{S}(w_n, \tau_{n-1}) \right| &\overset{\text{(a)}}{\leq} \left| \frac{\mathcal{S}(w_n, \tau_{n-1})}{w_n \sqrt{I^*(X_n)}} \right| \\
&= \left(\frac{2\sigma^2 \log\log w_n}{w_n\, I^*(X_n)} \right)^{1/2} \big| \mathcal{E}(w_n, \tau_{n-1}) \big| \\
&\overset{\text{(b)}}{\leq} \left(\frac{2\sigma^2 \log\log w_n}{\nu \log \tau_n} \right)^{1/2} \big| \mathcal{E}(w_n, \tau_{n-1}) \big| \\
&\overset{\text{(c)}}{\leq} \left(\frac{2\sigma^2 \log\log \tau_n}{\nu \log \tau_n} \right)^{1/2} \big| \mathcal{E}(w_n, \tau_{n-1}) \big|
\end{aligned}$$

for $n = 2, 3, \ldots$, where: $\mathcal{E}(w,t) = (2\sigma^2 w \log\log w)^{-1/2} \mathcal{S}(w,t)$; ν is the scale parameter of $\mathcal{C}^*$; (a) follows because $I(x,\theta) = \big(f_\theta(x,\theta)\big)^2$ and $I^*(X_n) = \min_{\theta\in\Theta}\{I(X_n,\theta)\}$; (b) follows because

$w_n \geq \nu \log(\tau_n)/I^*(X_n)$ for $n \geq 2$; and (c) follows because $w_n \leq \tau_n$. For all t, by the law of the iterated logarithm we have that $\limsup_{w\to\infty} |\mathcal{E}(w,t)| = \limsup_{w\to\infty} \big|(2\sigma^2 w \log\log w)^{-1/2}\mathcal{S}(w,t)\big| = 1$ almost surely. Noting that $\lim_{x\to\infty}\{\log\log(x)/\log(x)\} = 0$, and that $\lim_{n\to\infty}\{\tau_n\} = \infty$ almost surely, we consequently deduce that

$$\lim_{n\to\infty} \left| \left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} f_\theta\big(X_n, C_{s,n}\big) \right)^{-1} \mathcal{S}(w_n, \tau_{n-1}) \right| \;=\; 0, \qquad \text{almost surely.} \tag{E.4}$$

Let $\delta = b/3$. By (E.3) and (E.4), there is a finite random variable m_0 such that

$$\left| \widehat{\Theta}_{n+1} - \sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \mu_{s,n}\, \theta_s \right| \;\leq\; \delta, \tag{E.5}$$

for all $n \geq m_0$. Letting $\tilde{\delta} = \delta/\big((a+b)K\big)$, by the theorem's hypothesis there exists a subsequence of estimation windows $\{w_{n(k)},\, k = 1, 2, \ldots\}$ such that the following holds for all k:

$$\frac{1}{w_{n(k)}} \sum_{s=\tau_{n(k)-1}+1}^{\tau_{n(k)-1}+w_{n(k)}} \mathbb{I}\big\{\zeta - a \leq \theta_s \leq \zeta + b\big\} \;\leq\; \tilde{\delta},$$

and $\theta_s \geq \zeta - a$ for all $s \geq \tau_{n(k)-1} + 1$. Consequently, we have

$$\sum_{s=\tau_{n(k)-1}+1}^{\tau_{n(k)-1}+w_{n(k)}} \mu_{s,n(k)}\, \mathbb{I}\big\{\zeta - a \leq \theta_s \leq \zeta + b\big\} \;\leq\; \frac{\tilde{\delta}\, w_{n(k)} K \sqrt{I^*(X_n)}}{w_{n(k)}\sqrt{I^*(X_n)}} \;=\; \frac{b}{3(a+b)} \qquad \text{for all } k.$$

This implies that

$$\begin{aligned} \sum_{s=\tau_{n(k)-1}+1}^{\tau_{n(k)-1}+w_{n(k)}} \mu_{s,n(k)}\, \theta_s \;&\geq\; \frac{b}{3(a+b)}\,(\zeta - a) \;+\; \left(1 - \frac{b}{3(a+b)}\right)(\zeta + b) \\ &=\; \zeta + \frac{2b}{3} \;=\; \zeta + 2\delta \end{aligned} \tag{E.6}$$

for all k. By (E.5) and (E.6), we deduce that there is a finite random variable k_0 such that $\widehat{\Theta}_{n(k)+1} \geq \zeta + \delta$ for $k \geq k_0$. As a result, with probability one, $\widehat{\Theta}_{n+1}$ does not converge to ζ. ■

Remark (extension to M-estimators) As mentioned earlier, our general analysis in §6 can be extended to any M-estimator characterized by (7.1). To do this, the optimality condition in (E.1) is replaced by the following:

$$-\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \lambda'\Big(y_s - f\big(x_s, \widehat{\Theta}_{n+1}\big)\Big)\, f_\theta\big(x_s, \widehat{\Theta}_{n+1}\big) \;=\; 0,$$

for all n. As in the proof of Theorem 5, we recall that $x_s = X_n$ for $s = \tau_{n-1}+1, \ldots, \tau_{n-1}+w_n$, and then use the response model (2.5) and the mean value theorem on $f(\cdot)$ and $\lambda'(\cdot)$ to deduce that

$$\widehat{\Theta}_{n+1} \;=\; \sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \tilde{\mu}_{s,n}\, \theta_s \;+\; \left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \lambda''(\tilde{C}_{s,n}) f_\theta\big(X_n, C_{s,n}\big) \right)^{-1} \tilde{\mathcal{S}}(w_n, \tau_{n-1}), \tag{E.7}$$

for all n, where:

$$\tilde{\mu}_{s,n} \;=\; \left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \lambda''(\tilde{C}_{s,n}) f_\theta\big(X_n, C_{s,n}\big)\right)^{-1} \lambda''(\tilde{C}_{s,n}) f_\theta\big(X_n, C_{s,n}\big),$$

$$\tilde{\mathcal{S}}(w_n, \tau_{n-1}) \;=\; \sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \lambda'(\epsilon_s),$$

and $\tilde{C}_{s,n}$ is a random variable on the line segment between ϵ_s and $y_s - f\big(X_n, \widehat{\Theta}_{n+1}\big)$. To extend the proof of Theorem 5, note that

$$\begin{aligned}\left|\left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \lambda''(\tilde{C}_{s,n}) f_\theta\big(X_n, C_{s,n}\big)\right)^{-1} \tilde{\mathcal{S}}(w_n, \tau_{n-1})\right| &\overset{\text{(a)}}{\leq} \left|\frac{\tilde{\mathcal{S}}(w_n, \tau_{n-1})}{c\, w_n \sqrt{I^*(X_n)}}\right| \\ &\overset{\text{(b)}}{\leq} \left(\frac{2\sigma^2 \log\log \tau_n}{c^2\, \nu \log \tau_n}\right)^{1/2} \left|\tilde{\mathcal{E}}(w_n, \tau_{n-1})\right|\end{aligned}$$

for $n \geq 2$, where: $\tilde{\mathcal{E}}(w,t) = (2\sigma^2 w \log\log w)^{-1/2} \tilde{\mathcal{S}}(w,t)$; ν is the scale parameter of $\mathcal{C}^*$; (a) follows because $|\lambda''(z)| \geq c$ for all $z \in \mathbb{R}$, $I(x,\theta) = \big(f_\theta(x,\theta)\big)^2$ and $I^*(X_n) = \min_{\theta\in\Theta}\{I(X_n,\theta)\}$; and (b) follows because $\nu \log(\tau_n)/I^*(X_n) \leq w_n \leq \tau_n$ for $n \geq 2$. Using the law of the iterated logarithm and the fact that $\lim_{n\to\infty}\{\tau_n\} = \infty$ almost surely, as in the proof of Theorem 5, we deduce that

$$\lim_{n\to\infty} \left|\left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \lambda''(\tilde{C}_{s,n}) f_\theta\big(X_n, C_{s,n}\big)\right)^{-1} \tilde{\mathcal{S}}(w_n, \tau_{n-1})\right| \;=\; 0, \quad \text{almost surely.} \tag{E.8}$$

Replacing (E.4) with (E.8), and repeating the rest of the arguments in the proof of Theorem 5 in exactly the same way, we obtain the extension of Theorem 5 to the case of M-estimation. To generalize Theorems 6 and 7, we invoke (E.7) and (E.8) instead of (E.3) and (E.4), respectively, in the proofs of these theorems, and in accordance with that, we replace the weights $\{\mu_{s,n}\}$ with $\{\tilde{\mu}_{s,n}\}$ in the proof of Therorem 7. Noting that $\tilde{\mu}_{s,n} \in [0,1]$ and $\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \tilde{\mu}_{s,n} = 1$, we repeat the arguments in the proofs of Theorems 6 and 7 in exactly the same way to obtain the extended results.

Proof of Theorem 6. By the theorem's hypothesis, we have $\theta_t = \theta$ for all t. Using this fact and repeating the arguments used to derive (E.3), we deduce that

$$\widehat{\Theta}_{n+1} \;=\; \theta + \frac{\mathcal{S}(w_n, \tau_{n-1})}{w_n\, f_\theta\big(X_n, C_n\big)} \tag{E.9}$$

for all n, where C_n is a random variable on the line segment between θ and $\widehat{\Theta}_{n+1}$. By (E.4), we further obtain

$$\lim_{n\to\infty} \left|\frac{\mathcal{S}(w_n, \tau_{n-1})}{w_n\, f_\theta\big(X_n, C_n\big)}\right| \;=\; 0, \quad \text{almost surely.} \tag{E.10}$$

Thus, $\widehat{\Theta}_n \to \theta$ almost surely. As a result, with probability one, there exists a finite random variable N such that $|1 - \widehat{\Theta}_n/\theta| \leq \varepsilon$ for all $n \geq N$, which implies that $|1 - \hat{\theta}_t^*/\theta| \leq \varepsilon$ for all $t \geq \tau_N$. ■

Proof of Theorem 7. We will prove the theorem in the first three steps, and the remark that follows the theorem in the final step.

Step 1: Find a relaxed growth envelope for θ_t. Because $\{\theta_t\}$ evolves between the lower and upper bound processes $\{\theta_t^l\}$ and $\{\theta_t^h\}$ with tolerance R, there exists $z > 0$ such that $\theta_t^l - z \leq \theta_t \leq \theta_t^h + z$ for all but possibly finitely many t. Thus, there exists a natural number N_0 such that

$$\zeta + \kappa_1 G(t) - z \;\leq\; \theta_t \;\leq\; \zeta + \kappa_2 G(t) + z \tag{E.11}$$

for all $t \geq N_0$. Moreover, because $G(\cdot)$ is nondecreasing and $G(t) \to \infty$, there exists $N_1 \geq N_0$ such that $G(t) > \big(z + \max\{1 - \zeta, 0\}\big)/\kappa_1$ for all $t \geq N_1$.

Step 2: Derive an upper bound on the growth of w_n. Let $a(t)$ be the largest integer a satisfying $\tau_a \leq t$. Recalling that $\{\tau_n,\, n = 1, 2, \ldots\}$ is the subsequence of periods at which estimation windows are updated, the definition of $a(t)$ implies that $\tau_{a(t)}$ is the latest period before t, at which the estimation window is updated, and that $w_{a(t)} = \tau_{a(t)} - \tau_{a(t)-1}$ is the size of last estimation window before t. Letting $\delta = \frac{1}{2}\big(\kappa_1 G(N_1) - z\big) > 0$, we deduce from (E.3) and (E.4) that there is a finite random variable m_0 such that

$$\left| \widehat{\Theta}_{n+1} - \sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \mu_{s,n}\, \theta_s \right| \;\leq\; \delta,$$

for all $n \geq m_0$ almost surely. Now, let $m = \max\{m_0, a(N_1) + 1\}$. Because $\mu_{s,n} \in [0,1]$, $\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \mu_{s,n} = 1$, and $G(\cdot)$ is nondecreasing, we have that $\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \mu_{s,n}\, \theta_s \geq \zeta + \kappa_1 G(\tau_{n-1}) - z$ for all $n \geq m$. Therefore,

$$\widehat{\Theta}_{n+1} \;\geq\; \zeta + \kappa_1 G(\tau_{n-1}) - z - \delta \;\geq\; \zeta + \kappa_1 G(N_1) - z - \delta \;=\; \zeta + \delta,$$

for all $n \geq m$ almost surely. Because $\psi'(\theta) \geq \ell > 0$ for all $\theta \in \mathbb{R}$, we deduce that $X_n - \psi(\zeta) = \psi\big(\widehat{\Theta}_n\big) - \psi(\zeta) \geq \ell\delta > 0$ for all $n \geq m + 1$ almost surely. Thus, there exists a finite and positive constant c such that $I^*(X_n) > c$ all $n \geq m + 1$ almost surely. Recalling that w_n is the smallest integer satisfying $w_n \geq \nu \log(\tau_n)/I^*(X_n)$ for all $n \geq 2$, we further get $w_n - 1 < \nu \log(\tau_n)/I^*(X_n) \leq \nu \log(\tau_n)/c$ for all $n \geq m + 1$ almost surely. Thus,

$$w_n \;=\; \tau_n - \tau_{n-1} \;\leq\; 1 + \frac{\nu \log(\tau_n)}{c} \qquad \text{for all } n \geq m + 1 \text{ almost surely.} \tag{E.12}$$

Step 3: Prove that $\hat{\theta}^*_{t+1}$ has the same order of magnitude as θ_{t+1}. Note that (E.3) implies

$$\frac{\hat{\theta}^*_{t+1}}{\theta_{t+1}} \;=\; \sum_{s=\tau_{a(t)-1}+1}^{\tau_{a(t)-1}+w_{a(t)}} \frac{\mu_{s,a(t)}\, \theta_s}{\theta_{t+1}} \;-\; \left(\theta_{t+1} \sum_{s=\tau_{a(t)-1}+1}^{\tau_{a(t)-1}+w_{a(t)}} f_\theta\Big(X_{a(t)}, C_{s,a(t)}\Big) \right)^{-1} \mathcal{S}\Big(w_{a(t)}, \tau_{a(t)-1}\Big), \tag{E.13}$$

for all $t \geq \tau_{m+1}$. We will first prove that the second term on the right hand side of (E.13) converges to zero. By (E.4), there is a finite random variable N_2 such that

$$\left| \left(\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} f_\theta\big(X_n, C_{s,n}\big) \right)^{-1} \mathcal{S}(w_n, \tau_{n-1}) \right| \;\leq\; \frac{\varepsilon}{2},$$

for all $n \geq N_2$ almost surely. Because (E.11) holds for all $t \geq \tau_{m+1} \geq N_0$, we also know that $\theta_{t+1} \geq \zeta + \kappa_1 G(t) - z$ for all $t \geq \tau_{m+1}$, which implies that $\theta_{t+1} \geq 1$ for all $t \geq \tau_{m+1} \geq N_1$. Letting $N_3 = \max\{\tau_{m+1}, \tau_{N_2}\}$, we have

$$\left| \left(\theta_{t+1} \sum_{s=\tau_{a(t)-1}+1}^{\tau_{a(t)-1}+w_{a(t)}} f_\theta\Big(X_{a(t)}, C_{s,a(t)}\Big) \right)^{-1} \mathcal{S}\Big(w_{a(t)}, \tau_{a(t)-1}\Big) \right| \;\leq\; \frac{\varepsilon}{2}, \tag{E.14}$$

for all $t \geq N_3$ almost surely. Now, we will study the first term on the right hand side of (E.13). Recalling that $\mu_{s,n} \in [0,1]$, $\sum_{s=\tau_{n-1}+1}^{\tau_{n-1}+w_n} \mu_{s,n} = 1$, and $G(\cdot)$ is nondecreasing, we have

$$\sum_{s=\tau_{a(t)-1}+1}^{\tau_{a(t)-1}+w_{a(t)}} \frac{\mu_{s,a(t)}\,\theta_s}{\theta_{t+1}} \;\geq\; \frac{\zeta + \kappa_1 G\big(\tau_{a(t)-1}\big) - z}{\zeta + \kappa_2 G\big(\tau_{a(t)+1}\big) + z}, \tag{E.15}$$

for all $t \geq N_3$ almost surely. Let $\eta_1, \eta_2 > 0$ such that $\eta_1 < \kappa_1/\kappa_2 \leq 1 \leq \kappa_2/\kappa_1 < \eta_2$. Since $G(\cdot)$ is nondecreasing and $G(t) \to \infty$, there exists a finite random variable $N_4 \geq N_3$ such that $G(t) \geq \big((1-\eta_1)|\zeta| + (1+\eta_1)z\big)/(\kappa_1 - \eta_1\kappa_2)$ for all $t \geq N_4$. Moreover, invoking (E.12) for $n = a(t)$ and $n = a(t)+1$, we also get $\tau_{a(t)+1} - \tau_{a(t)-1} \leq 2 + 2\nu \log(\tau_{a(t)+1})/c$ for all $t \geq N_4$ almost surely; that is, $\tau_{a(t)-1} \geq \tau_{a(t)+1} - 2\nu \log\big(\tau_{a(t)+1}\big)/c - 2$ for all $t \geq N_4$ almost surely. Combining this with the fact that $\tau_{a(t)}$ diverges almost surely to ∞ as $t \to \infty$, we deduce that there exists a finite random variable $N_5 \geq N_4$ such that $\tau_{a(t)-1} \geq \lambda \tau_{a(t)+1} + (1-\lambda)N_4$ for all $t \geq N_5$ almost surely, where $\lambda = \eta_1\kappa_2/\kappa_1 < 1$. Because $G(\cdot)$ is concave and nondecreasing, this implies that

$$\begin{aligned} G\big(\tau_{a(t)-1}\big) \;&\geq\; \lambda G\big(\tau_{a(t)+1}\big) + (1-\lambda) G(N_4) \\ &\geq\; \lambda G\big(\tau_{a(t)+1}\big) + \big((1-\eta_1)|\zeta| + (1+\eta_1)z\big)/\kappa_1 \end{aligned} \tag{E.16}$$

for all $t \geq N_5$ almost surely. By elementary algebra, this further implies that the right hand side of (E.15) is greater than or equal to η_1 for all $t \geq N_5$. Thus, by (E.13) and (E.14), we conclude that $\liminf_{t\to\infty} \hat{\theta}^*_{t+1}/\theta_{t+1} \geq \eta_1$ almost surely. To complete the proof of (a), note that

$$\sum_{s=\tau_{a(t)-1}+1}^{\tau_{a(t)-1}+w_{a(t)}} \frac{\mu_{s,a(t)}\,\theta_s}{\theta_{t+1}} \;\leq\; \frac{\zeta + \kappa_2 G\big(\tau_{a(t)+1}\big) + z}{\zeta + \kappa_1 G\big(\tau_{a(t)-1}\big) - z}, \tag{E.17}$$

for all $t \geq N_3$ almost surely. Recall that $G(\cdot)$ is nondecreasing and $G(t) \to \infty$, implying that there exists $\tilde{N}_4 \geq N_3$ such that $G(t) \geq \big((\eta_2 - 1)|\zeta| + (1+\eta_2)z\big)/(\eta_2\kappa_1 - \kappa_2)$ for all $t \geq \tilde{N}_4$. Repeating the above arguments with $\lambda = \kappa_2/(\eta_2\kappa_1)$, we conclude that the right hand side of (E.17) is less than or equal to η_2, hence $\limsup_{t\to\infty} \hat{\theta}^*_{t+1}/\theta_{t+1} \leq \eta_2$ almost surely.

Step 4: Derive an upper bound on the eventual inaccuracy of $\hat{\theta}^*_{t+1}$. Note that $\kappa_2 \leq (1+\varepsilon/4)\kappa_1$ implies $\kappa_2 \leq \kappa_1/(1-\varepsilon/4)$. Thus, we can choose $\eta_1 = 1 - \varepsilon/2 < \kappa_1/\kappa_2$ and $\eta_2 = 1 + \varepsilon/2 > \kappa_2/\kappa_1$. Combining this with (E.14), we obtain the desired result. ■

Acknowledgment. The authors thank the area editor Bert Zwart, the associate editor, and three referees for their helpful comments that improved the presentation and structuring of the paper.

References

Agrawal, S. and Goyal, N. (2012), 'Analysis of Thompson Sampling for the Multi-armed Bandit Problem', *Proceedings of the 25th Annual Conference on Learning Theory* pp. 39.1–39.26.

Anderson, T. W. and Taylor, J. (1976), 'Some Experimental Results on the Statistical Properties of Least Squares Estimates in Control Problems', *Econometrica* **44**(6), 1289–1302.

Åström, K. and Wittenmark, B. (2013), *Adaptive Control*, Dover, Mineola, NY.

Auer, P., Cesa-Bianchi, N. and Fischer, P. (2002), 'Finite-time Analysis of the Multiarmed Bandit Problem', *Machine Learning* **47**(2), 235–256.

Auer, P., Ortner, R. and Szepesvári, C. (2007), 'Improved Rates for the Stochastic Continuum-armed Bandit Problem', *Proceedings of the International Conference on Computational Learning Theory* pp. 454–468.

Besbes, O. and Zeevi, A. (2015), 'On the (Surprising) Sufficiency of Linear Models for Dynamic Pricing with Demand Learning', *Management Science* **61**(4), 723–739.

Borkar, V. and Varaiya, P. (1979), 'Adaptive Control of Markov Chains, I: Finite Parameter Set', *IEEE Transactions on Automatic Control* **24**(6), 953–957.

Borkar, V. and Varaiya, P. (1982), 'Identification and Adaptive Control of Markov Chains', *SIAM Journal on Control and Optimization* **20**(4), 470–489.

Brezzi, M. and Lai, T. (2000), 'Incomplete Learning from Endogenous Data in Dynamic Allocation', *Econometrica* **68**(6), 1511–1516.

Brezzi, M. and Lai, T. (2002), 'Optimal Learning and Experimentation in Bandit Problems', *Journal of Economic Dynamics and Control* **27**(1), 87–108.

Broder, J. and Rusmevichientong, P. (2012), 'Dynamic Pricing under a General Parametric Choice Model', *Operations Research* **60**(4), 965–980.

Cheung, W., Simchi-Levi, D. and Wang, H. (2017), 'Dynamic Pricing and Demand Learning with Limited Price Experimentation'. Forthcoming in *Operations Research*.

den Boer, A. (2014), 'Dynamic Pricing with Multiple Products and Partially Specified Demand Distribution', *Mathematics of Operations Research* **39**(3), 863–888.

den Boer, A. and Zwart, B. (2014), 'Simultaneously Learning and Optimizing using Controlled Variance Pricing', *Management Science* **60**(3), 770–783.

den Boer, A. and Zwart, B. (2015), 'Dynamic Pricing and Learning with Finite Inventories', *Operations Research* **63**(4), 965–978.

Garcia, C., Prett, D. and Morari, M. (1989), 'Model Predictive Control: Theory and Practice–A Survey', *Automatica* **25**(3), 335–348.

Gosavi, A. (2009), 'Reinforcement Learning: A Tutorial Survey and Recent Advances', *INFORMS Journal on Computing* **21**(2), 178–192.

Harrison, J., Keskin, N. and Zeevi, A. (2012), 'Bayesian Dynamic Pricing Policies: Learning and Earning Under a Binary Prior Distribution', *Management Science* **58**(3), 570–586.

Jennrich, R. (1969), 'Asymptotic Properties of Non-linear Least Squares Estimators', *The Annals of Mathematical Statistics* **40**(2), 633–643.

Kaelbling, L., Littman, M. and Moore, A. (1996), 'Reinforcement Learning: A Survey', *Journal of Artificial Intelligence Research* **4**, 237–285.

Kalman, R. and Bucy, R. (1961), 'New Results in Linear Filtering and Prediction Theory', *Journal of Basic Engineering* **83**(1), 95–108.

Keskin, N. and Zeevi, A. (2014), 'Dynamic Pricing with an Unknown Demand Model: Asymptotically Optimal Semi-myopic Policies', *Operations Research* **62**(5), 1142–1167.

Keskin, N. and Zeevi, A. (2016), 'Chasing Demand: Learning and Earning in a Changing Environment', *Mathematics of Operations Research* **42**(2), 277–307.

Kiefer, J. and Wolfowitz, J. (1952), 'Stochastic Estimation of the Maximum of a Regression Function', *The Annals of Mathematical Statistics* **23**(3), 462–466.

Lai, T. (1994), 'Asymptotic Properties of Nonlinear Least Squares Estimates in Stochastic Regression Models', *The Annals of Statistics* **22**(4), 1917–1930.

Lai, T. (2003), 'Stochastic Approximation', *The Annals of Statistics* **31**(2), 391–406.

Lai, T. and Robbins, H. (1979), 'Adaptive Design and Stochastic Approximation', *Annals of Statistics* **7**(6), 1196–1221.

Lai, T. and Robbins, H. (1981), 'Consistency and Asymptotic Efficiency of Slope Estimates in Stochastic Approximation Schemes', *Z. Wahrscheinlichkeitstheorieverw. Gebiete* **56**(3), 329–360.

Lai, T. and Robbins, H. (1982), 'Iterated Least Squares in Multiperiod Control', *Advances in Applied Mathematics* **3**(1), 50–73.

Lai, T. and Robbins, H. (1985), 'Asymptotically Efficient Adaptive Allocation Rules', *Advances in Applied Mathematics* **6**(1), 4–22.

Lai, T. and Wei, C. (1982), 'Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems', *The Annals of Statistics* **10**(1), 154–166.

Lobo, M. and Boyd, S. (2003), 'Pricing and Learning with Uncertain Demand'. Working Paper. Stanford University, Stanford, CA.

Marquardt, D. (1963), 'An Algorithm for Least-squares Estimation of Nonlinear Parameters', *Journal of the Society for Industrial and Applied Mathematics* **11**(2), 431–441.

McLennan, A. (1984), 'Price Dispersion and Incomplete Learning in the Long Run', *Journal of Economic Dynamics and Control* **7**(3), 331–347.

Prescott, E. (1972), 'The Multi-Period Control Problem Under Uncertainty', *Econometrica* **40**(6), 1043–1058.

Robbins, H. (1952), 'Some Aspects of the Sequential Design of Experiments', *Bulletin of the American Mathematical Society* **58**(5), 527–535.

Rothschild, M. (1974), 'A Two-armed Bandit Theory of Market Pricing', *Journal of Economic Theory* **9**(2), 185–202.

Sutton, R. and Barto, A. (1998), *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA.

Thompson, W. (1933), 'On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples', *Biometrika* **25**(3), 285–294.

Williams, D. (1991), *Probability with Martingales*, Cambridge University Press, Cambridge, UK.

Wu, C.-F. (1981), 'Asymptotic Theory of Nonlinear Least Squares Estimation', *The Annals of Statistics* **9**(3), 501–513.

N. Bora Keskin is an Assistant Professor at the Fuqua School of Business at Duke University. His main research studies management problems that involve decision making under uncertainty. In particular he is interested in stochastic models, and their application to revenue management, dynamic pricing, demand learning, and product differentiation. Prior to his graduate education, he worked at McKinsey & Company as a consultant in banking and telecommunications industries.

Assaf Zeevi is Professor and holder of the Kravis chair at the Graduate School of Business, Columbia University. His research focuses on the formulation and analysis of mathematical models of complex systems, with particular research and teaching interests that lie in the intersection of operations research, statistics, computer science and economics. Recent application areas have been motivated by problems in healthcare analytics, dynamic pricing, recommendation engines and personalization, and the valuation and monetization of digital goods. He is the recipient of several research awards including a CAREER Award from the National Science Foundation, an IBM Faculty Award, Google Research Award, as well as several best paper recognitions. Assaf is a member of several editorial boards in his professional community, as well as several advisory boards for companies in the high technology sector.