On the (surprising) sufficiency of linear models for dynamic pricing with demand learning

Omar Besbes* Columbia University

Assaf Zeevi† Columbia University

first version: 01/2012, revised: 04/2013, 03/2014

Abstract

We consider a multi-period single product pricing problem with an unknown demand curve. The seller's objective is to adjust prices in each period so as to maximize cumulative expected revenues over a given finite time horizon; in doing so, the seller needs to resolve the tension between learning the unknown demand curve and maximizing earned revenues. The main question that we investigate is the following: how large of a revenue loss is incurred if the seller uses a simple parametric model which differs significantly (i.e., is *misspecified*) relative to the underlying demand curve. This "price of misspecification" is expected to be significant if the parametric model is overly restrictive. Somewhat surprisingly, we show (under reasonably general conditions) that this may not be the case.

Keywords: model mis-specification, inference, price optimization, revenue management, myopic pricing.

*Graduate School of Business, e-mail: ob2105@columbia.edu

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

1 Introduction

The famous industrial statistician George E. P. Box is widely credited for the saying:

> *"All models are wrong, but some are useful."*

A direct read of the first part of the statement is that all mathematical models are abstractions of reality, and as such can only capture some of its salient features; in other words, they are inherently *misspecified* relative to the true underlying system/phenomenon that is being studied. There exists a rather vast literature in statistics/econometrics that addresses this point, and develops extensions of classical estimation theory to misspecified contexts; cf. White (1996) and references therein. Yet, to the best of our knowledge, most OR/MS type studies, where decisions and control rules are usually a result of optimizing an objective function that explicitly builds on a system model, ignore the possibility of said model being incorrectly specified. Our focus in this paper is to illustrate the extent to which the latter part of Box's statement might apply in such settings. This will be done in the context of a prototypical dynamic decision making problem whose details are described next.

1.1 The problem and key questions

We consider a monopoly operating in a stationary demand environment that offers a product characterized by a set of attributes which are observable to customers. Over the time horizon of interest, the only attribute that the seller can modify is the price of the product, and this can only be done at pre-determined epochs; we index the periods between such epochs by $t \geq 1$. We let p_t denote the price offered during the t^{th} period, and D_t the corresponding realized demand. We assume that the mean value of D_t conditional on price is given by a deterministic function $\lambda(\cdot)$ (aka the *demand curve*). The seller's objective is to sequentially set prices with the intent of maximizing cumulative expected revenues.

We consider this dynamic optimization problem with the added complication that the true demand curve, $\lambda(\cdot)$, is not known to the seller. A common approach would then be to postulate a *demand model*, and over time jointly infer its structure from observed demand realizations, while concurrently optimizing revenues. This variant of dynamic pricing problems, often referred to as the problem of learning and earning, has a long and storied history, dating back to pioneering work of economists such as Rothschild (1974), and has been the focus of significant recent work in economics, computer science and operations research. Most of this work makes a significant simplifying assumption: the seller is assumed to know the structure of the demand curve, up to a finite number of unknown parameters. In other words, the demand model postulated by the seller is *well specified* with respect to the underlying demand curve. There are very few papers that avoid making this assumption, and they propose to address the potential for model misspecification using standard approaches in nonparametric statistics: by judiciously expanding the scope and complexity

of the model as further data becomes available (e.g., using higher degree polynomials, more complex smoothing splines etc), it is is possible to approximate a very broad class of functional relationships representing the underlying demand curve. Roughly speaking, misspecification is made to vanish, asymptotically. For further discussion the reader is referred to section 1.3 which contains a review of the relevant literature.

The focus of this paper is quite different. Rather than striving to eliminate misspecification in the manner described above, we make it even more pronounced by assuming the seller adopts an exceedingly simple parametric model for the demand curve, in particular, the widely used linear model. With that as a given, we would like to better understand whether, when, and to what extent does this simple and *incorrect* model support "good" pricing decisions; a more direct interpretation of Box's statement above. In so doing, we restrict attention to a simple class of pricing policies, which are abstracted away from practice, and in line with the typical policies designed for the well-specified cases. These policies operate in a semi-myopic manner: they loop through estimation and optimization steps, and price to optimize immediate revenues given current model estimates, while performing some minimal price experimentation. Despite their simple minded and incorrect predicate – a linear demand model – the aforementioned pricing policies are, somewhat surprisingly, quite "useful."

1.2 Main findings and qualitative insights

We start with the following thought experiment. Suppose that a "good" policy is constructed based on the linear model assumption in a well-specified setting (namely, when the underlying demand curve is linear as well). How will this policy perform in an environment in which the demand curve is no longer well specified, namely, when it differs from the linear modeling assumptions?

We first explore this question numerically and, quite surprisingly, find that the policy performs remarkably well over a reasonable range of scenarios, in spite of said misspecification. Motivated by these observations, the remainder of the paper explores the underlying theory that helps explain these numerical findings.

Mimicking this numerical experiment, the departure point for our theoretical investigation is a family of "good" policies designed in the well specified setting; when the demand curve is unknown but matches the modeling assumptions, such policies have been identified in the literature (see section 1.3) and have the semi-myopic structure discussed earlier. This broad family of *semi-myopic* pricing policies that are based on a linear demand model gives rise, under reasonably general conditions, to several interesting conclusions.

First, pricing decisions generated by this policy converge in probability, despite model misspecification, to the *optimal price* corresponding to the *true* underlying demand curve (Theorem 1).[1]

[1]One should emphasize here that the linear demand model that is adopted as a primitive in these policies is not

Second, going beyond the property of consistency outlined above, we prove a result that at first glance may seem rather remarkable: the above mentioned policies asymptotically accumulate revenues at a rate that is close to optimal (Theorem 2), despite their very simple structure and the fact that they are predicated on an incorrect demand model. To punctuate this point, note that even if the seller were to know a priori what is the parametric structure of the demand curve (assuming it even belongs to a parametric family), this knowledge, and the customization of the pricing policy to it, provides only limited (asymptotic) performance gains. [We emphasize here that the yardsticks we use to measure performance (consistency and growth rate of the regret) are those that are used in the well-specfied literature. In that sense, linear models are *sufficient* insofar as they allow to achieve performance, as measured by those yardsticks, that is similar to the one that would be achieved with knowledge of the true model class.]

The above theory allows us to tease out the key ingredient in mitigating the impact of misspecification: roughly speaking, the interleaving of estimation and optimization cycles in the "proximity" of the perceived optimal price point results in price steps that are in the direction of the gradient of the true underlying revenue function. To bring this point to full focus, we further discuss the relationship with other gradient methods in §3.3. In the present paper, we focus on fitting a linear model for concreteness and given its widespread use. The phenomenon above -that the impact of misspecification is mitigated by the interplay of estimation and optimization loops- would also arise under different models, and we comment on this in §5.3.

To complement the above findings, we highlight some potential pitfalls associated with model mis-specification and their implications. Roughly speaking, we demonstrate that if the demand model is not sufficiently flexible (e.g., has only a single degree of freedom), the positive behavior reported above does not continue to hold; the sequence of resulting prices might converge to a strictly sub-optimal value or even oscillate over time (see Proposition 1 in §5.1 and the discussion that follows). The oscillatory behavior illustrates that even in a completely stable (stationary) environment, a monopolist that is regularly re-calibrating its model might (falsely) conclude that the demand environment is changing temporally, while the price changes are in fact driven by a mismatch between the adopted model and the ambient demand curve.

Summarizing, the high level contribution of this paper is two fold. From a theory standpoint, the paper identifies how incorrect models may lead to correct pricing decisions under fairly general assumptions. From a more practical viewpoint, it provides some justification for the prevalent use of simple parametric models, as it establishes that the "price of misspecification" may not be as high as one might expect. In particular, it highlights the role of the estimation/optimization cycles, that are typically core elements of any pricing algorithm, in mitigating the impact of misspecification. It is worth noting that the conclusions of the paper may extend beyond the pricing application;

crucial. Similar results hold for many parametric classes of demand models, including many commonly used families such as exponential and logit. We further comment on this point in Section 5.3.

this point will be discussed in further detail in Section 2 and the proofs.

The remainder of the paper. We finish this section with a review of related work. The next section formulates the problem and presents a motivating experiment. Section 3 establishes the main result on consistency of pricing decisions derived from misspecified models, while Section 4 analyzes the more refined revenue-optimality properties of the class of policies under consideration. A discussion of the main findings, modeling assumptions and future directions is presented in Section 5.3. All the proofs of the results are collected in Appendices A, B and C.

1.3 Review of related work

As alluded to earlier, one of the first papers to formulate and study the dynamic pricing problem with an unknown demand curve was that of Rothschild (1974), which used a bandit-type formulation to study optimal pricing strategies. There have been extensive follow ups, extensions and generalizations of this work, primarily in the economics literature. The problem has received significant recent attention in the OR/MS community, primarily focusing on the setting where the seller knows the structure of the unknown demand curve up to some finite (and small) number of parameters. The focus of these papers has mostly been on the design of policies that suitably balance the exploration-exploitation trade off inherent to the problem; see, e.g., Broder and Rusmevichientong (2012), Harrison et al. (2012, 2011), and den Boer and Zwart (2013). In the terminology of the present paper, all of the above studies consider the well-specified case. There are far fewer studies that consider the situation where the demand curve can not be represented as a function that is parametrized with a finite number of parameters. Besbes and Zeevi (2009) considers this problem, and proposes to address it using standard approaches from nonparametric statistics, namely, building a sequence of models that are in essence "finitely parametrized," and judiciously grow the complexity of these models as more demand observations become available; see also Wang et al. (2011) for further improvements on those results. We refer the reader to Araman and Caldentey (2010) for a recent review paper on the topic. The issue of model misspecification, and its potential negative implications has surfaced in several recent papers: see, for example, Cachon and Kök (2007) in a newsvendor context, and Mersereau and Zhang (2012).

Our work differs markedly from the streams of work outlined above. Taking as our departure point that most models tend to be misspecified, the present paper attempts to provide some explanation for the reasons simple models might perform reasonably well in a broad set of scenarios. Philosophically, this is somewhat related, at least in spirit, to a study by Dawes (1979) that emphasizes the usefulness of improper linear models in the context of clinical prediction. One of the main points that the current paper attempts to elucidate is the fundamental distinction between capturing the "correct" model and arriving at the "correct" decision; a point that was the focus in Besbes et al. (2010), Chehrazi and Weber (2010) and Kao et al. (2009).

Most closely related to our work are probably Cooper et al. (2006) and Cooper et al. (2009). These papers also focus on the interplay between misspecification and decisions in the context of estimation-optimization cycles. The emphasis in Cooper et al. (2006) is on potential negative aspects of misspecification (the spiral down effect) in the context of capacity booking problems; but the authors also identify some special cases in which decisions end up being optimal despite the presence of misspecification. From a somewhat different angle, Cooper et al. (2009) establish that in an oligopoly setting, players that ignore the impact of their competitors decisions (leading to some form of misspecification), may end up in a *better* equilibrium in comparison to the one that would arise had they predicated their actions on the true model of competition. In the present paper, we establish that in pricing problems, the impact of misspecification is mitigated (if not completely eliminated) due to judiciously designed estimation/optimization cycles, and shed some theoretical light on the main elements contributing to this phenomenon.

2 Problem Formulation and Motivating Experiment

2.1 The model

We consider a multi-period single product pricing problem where a seller (acting as a monopolist) needs to set prices, p_t, in each period $t = 1, 2, \ldots$, chosen from a set of feasible prices given by the interval $[p^{(l)}, p^{(h)}]$. As described in section 1, the aggregate market response at time t to the posted price p_t is given by

$$D_t = \lambda(p_t) \; + \; \varepsilon_t, \qquad t \geq 1, \tag{1}$$

where $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$ is a deterministic function representing the mean demand conditional on the prevailing price, and ε_t is a zero-mean random variable with finite variance representing demand shocks. We assume that the random variables ε_t, $t \geq 1$ are independent and identically distributed. We denote by Π the class of all *admissible* pricing policies available for use by the seller. Each policy is represented by a sequence $\pi = (p_1, p_2, \ldots)$, where each entry within the sequence is restricted to depend only on past demand observations and past decisions; namely, p_t is adapted to the filtration generated by $(p_1, \ldots, p_{t-1}, D_1, \ldots, D_{t-1})$, for $t \geq 1$.

In our setting the demand curve, $\lambda(\cdot)$, is *not known* and the seller can only learn about it indirectly by observing market response to offered prices. In other words, the seller is faced with the joint problem of learning demand while concurrently trying to maximize revenues; the so-called learning and earning problem. We assume throughout that $\lambda(\cdot)$ is positive, strictly decreasing and twice continuously differentiable on the price domain $[p^{(l)}, p^{(h)}]$ and denote by $\lambda^{-1}(\cdot)$ its inverse on $[\lambda(p^{(h)}), \lambda(p^{(l)})]$. In addition, we assume that the revenue function $r(p) = p\lambda(p)$ admits a unique maximizer $p^* \in (p^{(l)}, p^{(h)})$. Instances of families of demand functions satisfying such assumptions

are given in Example 1 in Section 3. Clearly if the seller were to know $\lambda(\cdot)$ prior to the start of the selling season, he would simply set $p_t \equiv p^*$ for all times $t \geq 1$, hence maximizing the per period expected revenues.

For mathematical purposes and for the results to come, it is not necessary to assume that $\lambda(p_t) + \varepsilon_t \geq 0$ almost surely for all possible prices. However, this is an assumption that would hold in the pricing application we consider.

2.2 Performance metric

The efficacy of any admissible policy will be measured in two ways. The first, and more rudimentary measure, examines the long term behavior of the prices generated by the policy and assesses whether the sequence of prices converge to the true optimal price $p^* = \arg\max\{r(p)\}$. More specifically, a pricing policy is said to be *consistent* if

$$p_t \ \to \ p^* \qquad \text{as } t \to \infty, \tag{2}$$

in probability.

It is fairly clear that absence of consistency renders essentially no hope of maximizing cumulative expected revenues. At the same time, it is important to note that consistency focuses on the asymptotic behavior of the decision variable, and thus has little to say on how said decisions impact generated revenues over any finite time horizon. To address that, we will also evaluate the expected cumulative revenues generated by a policy $\pi = (p_1, p_2, \ldots)$ over a given time horizon T

$$\mathbb{E}^\pi \left[\sum_{t=1}^{T} p_t \ D_t \right], \tag{3}$$

where $\mathbb{E}^\pi[\cdot]$ denotes the expectation operator with respect to the true underlying statistical model (1), under π. In particular, we will compare those to the revenues generated by an oracle that knows the ambient demand curve. More specifically, we define the *regret* of any admissible policy $\pi \in \Pi$ as follows:

$$R(\pi, T) \ \ = \ \ p^* \lambda(p^*) T \ - \ \mathbb{E}^\pi \Big[\sum_{t=1}^{T} p_t D_t \Big]. \tag{4}$$

Clearly the smaller the regret, the better the performance of a given policy, as the oracle revenues (first term on the RHS above) are a strict upper bound on the performance of any pricing policy. The magnitude of the regret, and in particular the way it scales as the time horizon increases, provides a more refined lens to view the performance of a given policy.

2.3 The class of pricing policies

We will focus on pricing policies whose salient features are: i.) modeling the demand curve with a *linear function* whose two parameters need to be inferred from demand observations; and ii.)

determining prices at judicious time instants, called *recalibration points*, by essentially maximizing a proxy of the revenue function $r(\cdot)$ which is constructed from the estimated linear demand function. More specifically, the proposed policies operate in stages, the terminal point of each stage corresponding to a recalibration point. At the commencement of each stage, which we index by i for $i = 1, 2, ...$, the seller has an estimate of (what he considers to be) the optimal price $\widehat{p}_i$. The seller then sets 2 prices to be used at stage i, the values of which are suitable perturbations of $\widehat{p}_i$. Each such price will be used for I_i periods. At the end of stage i, estimates of the model parameters (α, β) are updated using least squares regression based on a subset of past observations, whose indices will be denoted by $\mathcal{T}_i$. The seller then computes the next estimate of the revenue maximizing price $\widehat{p}_{i+1}$, and the process repeats indefinitely, or until the end of the time horizon. Let $\mathcal{P} : \mathbb{R} \to \mathbb{R}$ be the projection operator on $[p^{(l)}, p^{(h)}]$, defined for all $x \in \mathbb{R}$ as $\mathcal{P}(x) = \min\{\max\{x, p^{(l)}\}, p^{(h)}\}$. The algorithm below provides a detailed description.

Semi-myopic pricing scheme: $\widehat{\pi}(\widehat{p}_1, \{I_i, \delta_i, \mathcal{T}_i : i \geq 1\})$

Set $t_1 = 0$

For $i \geq 1$

Step 1: Pricing and information collection

Set prices

$$\begin{aligned} p_t &= \widehat{p}_i, \qquad t = t+1, ..., t+I_i \\ p_t &= \widehat{p}_i + \delta_i, \qquad t = t+I_i+1, ..., t+2I_i \end{aligned}$$

Set $t_{i+1} = t_i + 2I_i$

Step 2: Recalibration

$$(\widehat{\alpha}_{i+1}, \widehat{\beta}_{i+1}) = \arg\min_{\alpha,\beta} \Big\{ \sum_{t \in \mathcal{T}_i} \big[D_t - (\alpha - \beta p_t) \big]^2 \Big\} \tag{5}$$

Step 3: Reoptimization

$$\widehat{p}_{i+1} = \mathcal{P}\left(\frac{\widehat{\alpha}_{i+1}}{2\widehat{\beta}_{i+1}} \right) \tag{6}$$

The above family of policies is predicated on a simple (*linear*) demand model, which is most likely misspecified relative to the ambient demand curve. In addition, it combines estimation and optimization in a manner that is effectively identical to how these elements would be executed in the well specified setting. To that end, the structure proposed above is abstracted away from common practice in applied revenue management. The common working assumption there is to adopt a simple parametric family as a demand model, linear models being a prototypical example,

and "solving" the dynamic optimization problem (see (3) above) by separating and cycling between estimation and optimization, essentially invoking a type of certainty equivalence principle, while concurrently conducting proper price experimentation.

In recent papers on learning and pricing, Broder and Rusmevichientong (2012) and den Boer and Zwart (2013) propose policies that are shown to have desirable theoretical properties, in the sense that they achieve (or almost achieve) the minimum rate of growth of the regret (4) with respect to the time horizon. More broadly, Harrison et al. (2011) establish simple sufficient conditions for a policy to achieve the best possible growth rate of regret. In essence, that theory stipulates that myopic-type pricing policies are "optimal" provided that statistical information on the parameter estimates (in the sense of Fisher) accumulates at a suitable rate, but the deviations that this forces from myopic decisions do not happen too frequently. In particular, when $I_i = 1$, $\mathcal{T}_i = \{1, ..., t_{i+1}\}$ and $\delta_i = t^{-1/4}$, the policy above is among the simplest instances that essentially satisfies these two sufficient conditions in Harrison et al. (2011, Theorem 2) [2], and can be thought of as a simplified version of the controlled variance policy proposed in den Boer and Zwart (2013).

2.4 An illustrative numerical experiment

The main question that we will start pursuing, numerically in this section and later theoretically, is: what is the impact of basing a pricing policy on an incorrect demand specification, specifically a linear model, when the true underlying demand curve (according to which observations are generated) is different.

Consider the policy $\widehat{\pi}$, which is designed based on the premise that the linear model is *well-specified*. We examine the performance of this policy in three demand curve environments (for the sake of simplicity, these are normalized to be between 0 and 1):

- linear $\mathcal{L}_1 = \{(\alpha - \beta p)^+\} : \alpha \in [\underline{\alpha}, \overline{\alpha}], \beta \in [\underline{\beta}, \overline{\beta}]\}$ where $[\underline{\alpha}, \overline{\alpha}] = [0.8, 1]$ and $[\underline{\beta}, \overline{\beta}] = [0.2, 1]$; [a well specified setting]
- exponential $\mathcal{L}_2 = \{\exp\{\alpha - \beta p\} : \alpha \in [\underline{\alpha}, \overline{\alpha}], \beta \in [\underline{\beta}, \overline{\beta}]\}\}$ where $[\underline{\alpha}, \overline{\alpha}] = [-0.2, 0]$ and $[\underline{\beta}, \overline{\beta}] = [0.3, 1]$; [a misspecified setting]
- logit $\mathcal{L}_3 = \{\exp\{\alpha - \beta p\}/(1 + \exp\{\alpha - \beta p\})^{-1} : \alpha \in [\underline{\alpha}, \overline{\alpha}], \beta \in [\underline{\beta}, \overline{\beta}]\}\}$ where $[\underline{\alpha}, \overline{\alpha}] = [0, 1]$ and $[\underline{\beta}, \overline{\beta}] = [0.5, 1]$. [a misspecified setting]

For each of the above specifications, $\mathcal{L}_i$ $i = 1, 2, 3$, we take 500 draws from the parameters α and β according to a uniform distribution on $[\underline{\alpha}, \overline{\alpha}]$ and $[\underline{\beta}, \overline{\beta}]$, respectively. Each draw determines the parameters for the demand curve in that particular instance. We then pit that against our proposed

[2] A formal verification is provided in Remark C1 in Appendix C for the case in which ε_t's are uniformly bounded almost surely.

policy, which is oblivious to the correct specification and is predicated on the two-parameter linear model. We simulate a sample path under the ambient demand curve and policy, and compute the fraction of oracle revenues that are achieved:

$$\frac{\sum_{t=1}^{T} p_t D_t}{p^* \lambda(p^*) T} \quad ,$$

the higher this ratio is, the better the performance of the policy. Note that there are two sources of loss: *statistical error* that stems from real time inference of model parameters; and a *misspecification error* that impacts performance in the latter two demand curve instances, when the linear model used for the policy is incorrectly specified.

In Table 1, we report the average ratio over the 500 instances for each case. The random variables ε_t are assumed to be normally distributed with standard deviation σ and we assume that the policy uses block size $I_i = 1$, $\mathcal{T}_i = \{1, ..., t_{i+1}\}$, with initial price $\widehat{p}_1 = 1$ and with $\delta_t = \rho t^{-1/4}$. We test different values of noise variance σ^2 and tuning parameter ρ. Throughout, we fix the price domain to be $[p^{(l)}, p^{(h)}] = [0, 5]$.

		well-specified			**mis-specified**					
demand functions		$\mathcal{L}_1$ (linear)			$\mathcal{L}_2$ (exponential)			$\mathcal{L}_3$ (logit)		
		time periods (T)			time periods (T)			time periods (T)		
	ρ	100	500	10^3	100	500	10^3	100	500	10^3
$\sigma = 0.25$	0.25	0.90	0.94	0.95	0.91	0.94	0.95	0.84	0.90	0.92
	0.5	0.87	0.93	0.95	0.93	0.96	0.96	0.87	0.93	0.95
	0.75	0.79	0.88	0.91	0.94	0.96	0.97	0.91	0.95	0.96
$\sigma = 0.5$	0.25	0.83	0.89	0.91	0.82	0.87	0.89	0.69	0.77	0.80
	0.5	0.80	0.88	0.91	0.87	0.92	0.93	0.76	0.84	0.87
	0.75	0.74	0.84	0.87	0.94	0.96	0.97	0.81	0.95	0.96

Table 1: **The impact of misspecification.** Fraction of optimal (oracle) revenues achieved by the linear-based pricing policy, averaged over a set of 500 random test instances. The standard error of the mean was always below 0.0125.

Focusing on the columns of Table 1 corresponding to the well-specified case (draws from the linear class $\mathcal{L}_1$), one observes that the policy performs very well. This is not surprising and simply confirms the theory and numerical experiments developed in recent literature; see Harrison et al. (2011), Broder and Rusmevichientong (2012) and den Boer and Zwart (2013).

Turning attention to columns corresponding to the mis-specified cases, corresponding to $\mathcal{L}_2$ or $\mathcal{L}_3$, the performance of the policy is still surprisingly very good, and on similar order to the fraction of optimal revenues achieved in the well-specified setting. In other words, in terms of the two sources of revenue losses highlighted above, the extent of losses that stem from misspecification appear surprisingly small. The rest of the paper focuses on identifying the drivers for this phenomenon

and developing pertinent theory.

Remark 1 (A more general problem formulation) The dynamic pricing problem presented above is a special instance of a more general class of problems in which the decision-maker seeks a decision x in some feasible compact set $\mathcal{X} \subset \mathbb{R}$, to optimize an objective function that directly depends on some unknown response function $\lambda(x)$. That is,

$$\max_{x \in \mathcal{X}} G(x, \lambda(x)) \quad . \tag{7}$$

Here $G(\cdot, \cdot)$ is the objective function and conditional on selecting x_t in period t, the decision-maker has only access to noisy observations of $\lambda(x_t)$ given by $\lambda(x_t) + \varepsilon_t$, with ε_t iid random variables with zero mean and finite variance. In the pricing problem, $G(x, y) = xy$. While the paper focuses on the pricing application, we keep the proofs at an abstract level and show how these apply to (7) above under appropriate conditions imposed on the mapping $G(\cdot, \cdot)$.

3 Consistency of the price process

Motivated by the illustrative numerical example detailed in the previous section, we now seek to develop some theory to buttress the observations gleaned from that experiment. For the purposes of the analysis, we will focus on policies $\widehat{\pi}$ that "forget" about past data. In particular, from here on we assume that $\mathcal{T}_i = \{t_i + 1, ..., t_{i+1}\}$, i.e., the parameter recalibration step (5) uses only the most recent data. This restriction is made for tractability purposes and enables us to highlight the main effects at play in a transparent fashion.

3.1 Theory

For the purpose of our main result, we impose the following conditions.

Assumption 1 *i.) For some* $\rho > 0$, $\mathbb{E}[\exp\{s\varepsilon_1\}] < \infty$ *for all* $s \in (-\rho, \rho)$.

ii.) $(1/2)\, \lambda(p)\, |\lambda''(p)| \,/\, (\lambda'(p))^2 \,<\, 1$ *for all* $p \in [p^{(l)}, p^{(h)}]$.

The first condition ensures that the demand shock distribution is suitably "light tailed," which greatly facilitates analysis (examples of standard distributions satisfying this property include Bernoulli, Normal, Exponential, and Poisson). The latter condition imposes some shape restrictions on the true underlying demand function. While there is no direct economic interpretation of this condition, it is satisfied for a large class of widely used demand functions, as seen below.

Example 1 (Models satisfying Assumption 1 *ii.)***)** .

– Linear models. If $\lambda(p) = a - bp$, then note that $(1/2)\lambda(p)|\lambda''(p)|/(\lambda'(p))^2 = 0$ and the assumption is always satisfied.

– Exponential models. If $\lambda(p) = \exp\{a - bp\}$, then $(1/2)\lambda(p)|\lambda''(p)|/(\lambda'(p))^2 \leq (1/2) < 1$ and the assumption is always satisfied.

– Logit models. If $\lambda(p) = \exp\{a - bp\}/(1 + \exp\{a - bp\})$, then $\lambda'(p) = -b\lambda(p)[1 - \lambda(p)]$ and

$$\frac{1}{2}\frac{\lambda(p)|\lambda''(p)|}{(\lambda'(p))^2} \quad = \quad \frac{1}{2}\left|\frac{1 - 2\lambda(p)}{1 - \lambda(p)}\right|,$$

hence the assumption is satisfied as long as $\lambda(p) < 3/4$ for all $p \in [p^{(l)}, p^{(h)}]$.

We next analyze in detail the sequence of prices generated by the class of policies. In the context of the semi-myopic pricing schemes, the least squares estimates of (α, β) are given by:

$$\widehat{\beta}_{i+1} \quad = \quad -\frac{\sum_{t=t_i+1}^{t_{i+1}}(p_t - \overline{p}_i)(D_t - \overline{D}_i)}{\sum_{t=t_i+1}^{t_{i+1}}(p_t - \overline{p}_i)^2} \tag{8}$$

$$\widehat{\alpha}_{i+1} \quad = \quad \overline{D}_i + \widehat{\beta}_{i+1}\overline{p}_i, \tag{9}$$

where

$$\overline{D}_i \; = \; \frac{1}{2I_i}\sum_{t=t_i+1}^{t_{i+1}} D_t, \qquad \text{and} \qquad \overline{p}_i \; = \; \frac{1}{2I_i}\sum_{t=t_i+1}^{t_{i+1}} p_t.$$

The next result shows that if the sequence of batch sizes corresponding to recalibration points is suitably chosen, the resulting sequence of prices $\{\widehat{p}_t : t \geq 1\}$ will be consistent.

Theorem 1 (consistency) *Let Assumption 1 hold. Suppose that in the linear-model semi-myopic policy $\widehat{\pi}$, $\mathcal{T}_i = \{t_i + 1, ..., t_{i+1}\}$, $\delta_i \to 0$ and $\delta_i I_i^{1/2}/\log(I_i) \to \infty$ as $i \to \infty$. Then, for any initial price $\widehat{p}_1$, the sequence of prices $\{p_t : t \geq 1\}$ generated by $\widehat{\pi}$ converges in probability to the true revenue maximizing price p^*.*

In other words, under the above conditions, a two-parameter linear model, in conjunction with the rather simple structure of the semi-myopic pricing policy, guarantees that the resulting sequence of prices recovers the *optimal* price corresponding to the true (and unknown) underlying demand curve, regardless of the functional form of the latter.

3.2 Basic intuition underlying Theorem 1

The proof of Theorem 1 relies on establishing that the mapping from the estimate of the price decision in stage i, $\widehat{p}_i$ to the estimate of the price decision in stage $i+1$, $\widehat{p}_{i+1}$, is some perturbation of a contraction; and that the contraction admits p^* as a unique fixed point. This enables us to establish the convergence of $\widehat{p}_i$ to p^*. To flesh out some of the key ideas and intuition that underlie the result, it will be conducive to consider a setting in which $\varepsilon_t = 0$ for all $t \geq 1$ and assume that $I_i = 1$.

Suppose that the demand curve is given by a logit function $\lambda(p) = \exp\{4.1-p\}/(1+\exp\{4.1-p\})$. Put $\delta_i = i^{-1}$. With an initial price estimate of $\widehat{p}_1 = 8$, we depict in Figure 1(a) the true demand curve as well as the estimated demand models at times 2, 3 and 30. The corresponding revenue functions are shown in Figure 1(b).

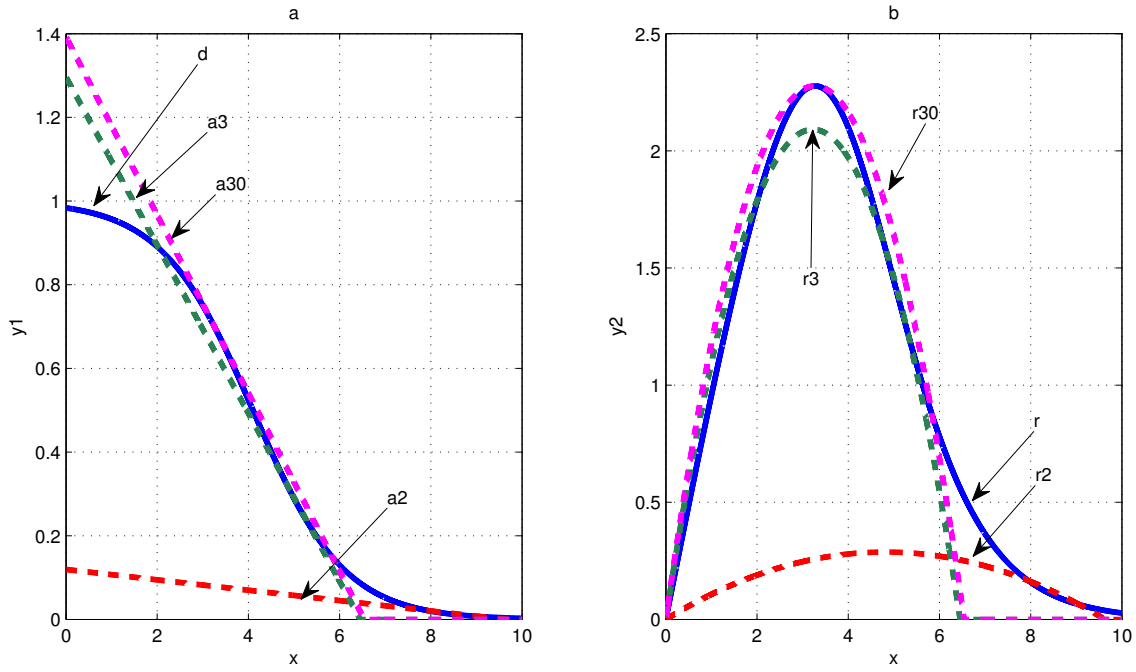

Figure 1: **Convergence of the price process when fitting two parameters.** The demand curve is given by a logit $\exp\{4.1-p\}/(1+\exp\{4.1-p\})$. The fitted model is linear given by $\alpha-\beta p$. The price process is started at various initial prices.

While the fitted demand model and corresponding revenue function do not coincide with their true underlying counterparts, we do observe that as time goes by, the fitted revenue function and true revenue function grow closer in the region around p^*, the price that maximizes revenues of the true revenue function. This "local convergence" is exactly what allows the seller to tease out a near-optimal price decision from a wrong model.

From a technical standpoint, Assumption $1ii.$) is the key ingredient in establishing a contraction-type property of the mapping that generates the pricing decisions. Among other things, this ensures systematic convergence of the price sequence regardless of the initial condition. Having said that, the question of why the limit is in fact $p^* = \arg\max_{p\in[p^{(l)},p^{(h)}]}\{p\lambda(p)\}$ remains open.

Why is the limit point p^*, the maximizer of the true revenue function? We outline

the intuition behind this below. Absent noise, it is straightforward to establish that

$$\widehat{\beta}_{i+1} = -\frac{\lambda(\widehat{p}_i + \delta_i) - \lambda(\widehat{p}_i)}{\delta_i}, \tag{10}$$

$$\widehat{\alpha}_{i+1} = \widehat{\beta}_{i+1}\widehat{p}_i + \lambda(\widehat{p}_i), \tag{11}$$

$$\widehat{p}_{i+1} = \frac{\widehat{\alpha}_{i+1}}{2\widehat{\beta}_{i+1}}. \tag{12}$$

Now suppose that $\widehat{p}_i$ converges to some limit $\widetilde{p}$. By (10), it must be that the sequence $\widehat{\beta}_i$ converges to $\widetilde{\beta}$, where $\widetilde{\beta} = -\lambda'(\widetilde{p})$. Similarly, the above, in conjunction with (11) implies that $\widehat{\alpha}_i$ converges to $\widetilde{\alpha}$, where $\widetilde{\alpha} = \widetilde{\beta}\widetilde{p} + \lambda(\widetilde{p})$. Equation (12) now implies that $\widetilde{p}$ must satisfy the following

$$\widetilde{p} = \frac{\widetilde{\alpha}}{2\widetilde{\beta}} = \frac{\widetilde{p}}{2} - \lambda(\widetilde{p})\frac{1}{2\lambda'(\widetilde{p})},$$

i.e., $\widetilde{p}$ satisfies $\widetilde{p}+\lambda(\widetilde{p})/\lambda'(\widetilde{p}) = 0$, which is exactly the first order condition for revenue maximization when the demand function is $\lambda(p)$. This equation admits p^* as a unique solution, by assumption. Hence, it must be that $\widetilde{p} = p^*$, and the limit can only be the price that maximizes the true revenue function.

The above arguments ignore the noise associated with the observations. Note that in general,

$$\widehat{\beta}_{i+1} = -\lambda'(\widehat{p}_i) + O(\delta_i) + \frac{1}{\delta_i}\frac{1}{2I_i}\sum_{t=t_i+1}^{t_{i+1}} \epsilon_t \ ,$$

where we use $O(\delta)$ to represent a quantity that is of order δ. Now, $(2I_i)^{-1}\sum_{t=t_i+1}^{t_{i+1}} \epsilon_t$ will be "close" to zero, and using an exponential bound we can show that, with high probability it will be bounded above by a factor of $\log(I_i)/I_i^{1/2}$. As a result, as long as δ_i converges to zero and the batch sizes are such that $\delta_i^{-1}\log(I_i)/I_i^{1/2}$ converges to zero as i grows large (the condition required in the theorem), then $\widehat{\beta}_{i+1} \approx -\lambda'(\widehat{p}_i)$ as i grows large. Roughly speaking, this condition ensures that the deterministic skeleton argument we presented first continues to hold when one accounts for the noise associated with the observations.

A different lens through which to view the convergence to p^*. Let us examine closely the price sequence produced by the linear-based semi-myopic pricing policies, focusing for transparency on the case where there is no noise ($\epsilon_t \equiv 0$). Approximating terms up to δ_i factors, the estimation step leads to

$$\begin{aligned}\widehat{\beta}_{i+1} &\approx -\lambda'(\widehat{p}_i)\\ \widehat{\alpha}_{i+1} &\approx \lambda(\widehat{p}_i) + \widehat{\beta}_{i+1}\widehat{p}_i.\end{aligned}$$

The optimization step yields the price in stage $i+1$, which is given by (assuming it is interior)

$$\begin{aligned}\widehat{p}_{i+1} &\approx \frac{\widehat{p}_i}{2} - \frac{\lambda(\widehat{p}_i)}{\lambda'(\widehat{p}_i)}\\ &= \widehat{p}_i + \frac{1}{-2\lambda'(\widehat{p}_i)}[\lambda(\widehat{p}_i) + \widehat{p}_i\lambda'(\widehat{p}_i)].\end{aligned}$$

In other words, the price sequence roughly satisfies the recursion

$$\widehat{p}_{i+1} \approx \widehat{p}_i + \frac{1}{-2\lambda'(\widehat{p}_i)} r'(\widehat{p}_i), \qquad i \geq 1. \tag{13}$$

Hence, in each iteration, the estimation/optimization cycle produces a price point that follows the gradient of the true revenue function, and the step size is approximately given by $1/|2\lambda'(p_i)|$. Consequently, in spite of model mis-specification, the estimation-optimization cycles naturally yield a direction of improvement in the underlying objective function. We further explore connections to gradient methods in the next subsection.[3]

3.3 Relation to gradient methods

Given the discussion above, a natural question is whether (13) is a variant of classical stochastic approximation. The method of stochastic approximation, specifically the Kiefer-Wolfowitz (KW) algorithm of Kiefer and Wolfowitz (1952), is perhaps the most general "model-free" approach for solving the dynamic pricing problem under demand model uncertainty. In particular, fixing an initial price p_1, that method produces the following sequence of price updates

$$p_{i+1} = p_i + a_i \frac{(p_i + c_i)D_i(p_i + c_i) - (p_i - c_i)D_i'(p_i - c_i)}{2c_i}, \qquad i \geq 1$$

where: $\{a_i : i = 1, 2, \ldots\}$ is the *step size* sequence; $\{c_i : i = 1, 2, \ldots\}$ is the gradient differencing sequence; and D_i, D_i' are two successive (independent) demand observations evaluated at the input prices $p_i + c_i$ and $p_i - c_i$, respectively. (Recall, conditionally, realized demand is given by $D_t = \lambda(p_t) + \varepsilon_t$, where $\lambda(\cdot)$ is the true underlying demand function.) The recursion above is effectively a steepest ascent algorithm which seeks to optimize the objective function $r(p) := p\lambda(p)$, using noisy *estimates* of the gradient of the revenue function in lieu of direct gradient observations. Under suitable conditions on the primitive sequences, and assuming strong concavity of the underlying objective function, the price process can be shown to converge to $p_* = \arg\min\{r(p)\}$.

Restricting attention again to the noise free setting, let us contrast (13) with the (noise free) stochastic approximation price recursion

$$p_{i+1} \approx p_i + a_i r'(p_i), \qquad i \geq 1,$$

where in the above the notation '$\approx$' is due to the gradient approximation (also used in deriving (13)). Comparing the above and (13), we see that, like its stochastic approximation counterpart, the linear-based semi-myopic policy we analyze prescribes prices that ultimately follow the direction

[3] In the absence of noise and derivative approximation, determining conditions for convergence of the price process is akin to determining the stability of the dynamic price process. Let $H(p) = p - \frac{1}{2\lambda'(p)} r'(p)$. The price process is locally stable if $H'(p^*) < 1$ and globally stable if $|H'(p)| < 1$ for all p in the price domain. The latter is exactly what Assumption 1*ii*.) ensures.

of the gradient, however, the method is *not local* and *does not rely* on a prescribed sequence of step sizes. In particular, while the tuning sequence $\{a_i\}$ in the context of stochastic approximation is typically specified to be proportional to $\{1/i\}$, and hence is monotonically decreasing, the step size in (13) may be "large" and does not necessarily shrink with the number of iterations; for further connections with stochastic approximation variants with non-vanishing step size see, e.g., Nemirovski et al. (2009).

In settings where second order (Hessian) information is available, a prevalent iterative (deterministic) optimization scheme is given by Newton's method and its variants (see, e.g., Bertsekas (1999)). Adapting this to our context, the sequence of price iterates generated by the method would be given by

$$p_{i+1} = p_i - r'(p_i)/r''(p_i), \quad i \geq 1.$$

In essence, the recursion is predicated on approximating the objective function by a quadratic function $r(p) \approx r(p_i) + r'(p_i)(p - p_i) + (1/2)r''(p_i)(p - p_i)^2$, using the local curvature parameters evaluated at p_i, and selecting the next price p_{i+1} so as to maximize this approximation. Note that here too the step size is variable and not pre-determined as in the KW stochastic approximation scheme. In contrast to the (zero-noise) recursion corresponding to the linear-based semi-myopic policy, here the step size is inversely proportional to the (negative) second derivative of the objective function, while in the former the step size is inversely proportional to twice the gradient value. To better understand where this is derived from, recall that the approximation underlying Newton's method is given by $r(p^*) + (1/2)r''(p^*)(p - p^*)^2$, and relies on the *correctly specified* first and second derivatives of the objective function evaluated at p^*. In contrast, the linear-based semi-myopic policy is a "first order" method that uses the misspecified demand model as a primitive. To that end, the approximation of the revenue function at p^* discussed in (10)-(12), is given by $\tilde{r}_{p^*}(p) \approx \big([\lambda(p^*) - p^*\lambda'(p^*)] + \lambda'(p^*)p\big)p$. Note that $\tilde{r}''_{p^*}(p^*) = 2\lambda'(p^*)$, which differs in general from the second-order approximation $r''(p^*) = 2\lambda'(p^*) + p^*\lambda''(p^*)$.

In summary, the interplay of estimation and optimization loops leads the firm to follow the (noisy) gradient of the true underlying revenue function. This quite notable in that here, we do not attempt to construct a policy that is robust to mis-specification in the first place. Rather, the starting point is to understand whether policies that are typically used for learning and earning (and designed for well-specified scenarios) suffer from mis-specification.

4 Revenue optimality

Having established consistency of price estimates under fairly general conditions, we next investigate the efficacy of these pricing decisions as measured by cumulative revenue performance. Recall, from (4), the regret $R(\pi, T)$ measures the gap between the performance of an oracle that has access to

the true underlying demand curve $\lambda(\cdot)$, and the performance of any given (admissible) policy.

Let I_0 denote some positive integer and $\nu > 1$ some positive number, and define the following sequence of block sizes.

$$I_i \;=\; \left\lfloor \nu^i I_0 \right\rfloor, \qquad i = 1, 2, ... \tag{14}$$

Theorem 2 (revenue optimality) *Let Assumption 1 hold. Suppose that in the linear-model semi-myopic policy $\widehat{\pi}$, $\mathcal{T}_i = \{t_i + 1, ..., t_{i+1}\}$ and the block sequence is selected as in* (14) *with $\delta_i = I_i^{-1/4}$, $i \geq 1$. Then, for any initial price $\widehat{p}_1$, the sequence of prices $\{p_t : t \geq 1\}$ generated by $\widehat{\pi}$ satisfies*

$$R(\widehat{\pi}, T) \;\leq\; C\,(\log T)^2\,\sqrt{T}$$

for some positive constant C and all $T \geq 2$.

The result above establishes that the average revenue loss per period, $R(\widehat{\pi}, T)/T$, converges to zero. In other words, the revenue gap between the oracle and the proposed policy vanishes. The obvious question is whether the *size* of this gap can be improved upon; the smaller the size of the gap, the better the performance of the policy. It stands to reason that if one has prior knowledge on the structure of the demand curve, for example if it belongs to a parametric family that is known a priori to the designer of the policy, then the size of this gap could be reduced considerably. Surprisingly, this intuition is essentially false. In a recent paper, Harrison et al. (2011) consider a well specified setting where the demand curve is a linear function of two unknown parameters, and the seller *knows* this structure, up to the values of the parameters. They prove that *no policy* can have a revenue gap (regret) which is smaller than order-$\sqrt{T}$ (uniformly over all parameter values of the demand curve).[4] In light of this result, and that the growth rate of $(\log T)^2$ is negligible compared to $\sqrt{T}$, the somewhat surprising conclusion is that the revenue performance of the semi-myopic linear-based pricing policies achieves a growth rate in terms of regret that is close to optimal, *whether or not* the linear model is misspecified relative to the underlying demand curve.

The result above provides further theoretical evidence of the limited impact of model misspecification within the context of our dynamic pricing problem. The specific structure of the policy that is used in making this point is rather crude, insofar as it discards "most" of past observations to re-estimate parameters, and concurrently re-optimizes prices relatively rarely. We should clarify that this structure is imposed for mathematical tractability; recalling the numerical experiments presented in §2.4, it appears that aggregating all the data, (while re-solving often) results in finite time performance that is on par with, or close to, that achieved in the well-specified setting. With regard to the tuning of the policy in the above result, and the intuition underlying this, recall

[4] We conjecture that the $(\log T)^2$ term in the upper bound in Theorem 2 is an artifact of proof technique, and that it is possible to improve the bound by eliminating this term.

from Theorem 1 that to have consistency we need that $\delta_i^{-1}\log(I_i)/I_i^{1/2}$ converges to zero, which is satisfied with the above policy tuning specification. The proof of Theorem 2 establishes that the regret in the i^{th} batch is bounded by $O(\log(I_{i-1})I_{i-1}^{-1/2}I_i + \delta_i^2 I_i)$. The first error term stems from the error still present from the previous inference batch and the second source of error stems from the losses due to price experimentation in the current batch. The selection of δ_i taken in Theorem 2 minimizes the growth rate of this loss, up to logarithmic terms. This selection of δ_i could be seen as ensuring that the cumulative squared price variation used in any batch of size I_i is of order $\sqrt{I_i}$, which is a similar requirement to the one made in the well-specified case (see Harrison et al. (2011)).

Numerical Performance. We have conducted experiments over a large set of scenarios with regard to the constant associated with Theorem 2. The constant appeared to lie somewhere between 1 and 12 for the policy that does not aggregate all observations (the one analyzed theoretically) and decreases for the policy that aggregates all observations. In the theoretical policy proposed, there are three sources of losses: one due to the lack of aggregation of data; one due to misspecification; and another due to the structure of the policy that updates decisions relatively rarely (only order $\log T$ times, as opposed to, say, order T for the CVP policy in den Boer and Zwart (2013)). Testing the exact performance of such a policy does not allow to tease out the different sources of losses, and hence provides limited information. Already in the well-specified case, Broder and Rusmevichientong (2012) separate the theoretical analysis of MLE-cycle (which only uses exploration samples) and the practical testing of the policies which then aggregate all samples: performance is significantly improved when aggregating all the data. We observe the same here. If one were to tune of the parameters of the policy, a possible approach would be to generate a large number of demand curve scenarios that one could face; simulate the performance under those for the time horizon of interest using different parameter selections; and pick the parameters that lead to the best performance among those scenarios.

5 Discussion

5.1 The need for two parameters

In earlier sections we have witnessed that a simple family of policies which is predicated on a two-parameter linear model, can effectively achieve excellent performance (both theoretically as well as numerically) despite the fact that the underlying demand curve is not linear. It turns out that having two degrees of freedom in the model used by the policy is critical.

Consider the case where the model used by the policy is restricted to a single parameter. For concreteness, suppose the value of the intercept α is fixed a priori, i.e., $\widehat{\alpha}_t = \widetilde{\alpha}$ for all $t \geq 1$. To isolate the effect of misspecification, and disentangle it from the fact that observations of the

demand curve are confounded by statistical noise, we again consider the zero noise setting, $\varepsilon_t \equiv 0$ for all $t \geq 0$.

Consider the simple semi-myopic pricing policy described in the previous section, with stages of length $I_i = 1, i \geq 1$, which uses only one price at each stage given by $\widehat{p}_i$. (Note that in the well specified case, if a single parameter is unknown then there is no need for two prices to be used for model calibration.) It is then straightforward to establish that the sequence of estimates and prices satisfies for $t \geq 1$

$$\widehat{\beta}_{t+1} = \frac{\widetilde{\alpha} - \lambda(\widehat{p}_t)}{\widehat{p}_t}, \tag{15}$$

$$\widehat{p}_{t+1} = \mathcal{P}\left(\frac{(\widetilde{\alpha}/2)\widehat{p}_t}{\widetilde{\alpha} - \lambda(\widehat{p}_t)}\right). \tag{16}$$

Behavior of the price process. Based on (16), it is evident that behavior of the sequence of prices generated by the policy will be determined by properties of the mapping $x \mapsto \mathcal{P}\left(\frac{(\widetilde{\alpha}/2)x}{\widetilde{\alpha}-\lambda(x)}\right)$, and if prices converge, it can only be to a fixed point of this mapping. This allows to characterize the unique possible limit point associated with the sequence $\{\widehat{p}_t : t \geq 1\}$.

Proposition 1 (limit points) *Suppose that $\lambda(p^{(l)}) < \widetilde{\alpha}$. Consider the sequence of prices $\{\widehat{p}_t : t \geq 1\}$ generated by the semi-myopic pricing policy with $I_i = 1$. If $\{\widehat{p}_t : t \geq 1\}$ converges, then the only possible limit point is $\check{p}$ given by*

$$\check{p} = \mathcal{P}\Big(\lambda^{-1}\big(\widetilde{\alpha}/2\big)\Big). \tag{17}$$

The condition $\lambda(p^{(l)}) < \widetilde{\alpha}$ precludes the situation where the true demand curve and estimated demand model may never cross on the price domain for any parameter value (if $\lambda(\cdot)$ decreases "slowly"). It is straightforward to check that if $\lambda(p) = \widetilde{\alpha} - \beta p$ for some $\beta > 0$, and hence the model is well-specified, then $\check{p} = p^*$ and thus the limit point is the price that maximizes the true revenue function. If however the model is mis-specified, then $\check{p}$ and p^* will in general *differ* and hence even if prices converge, the generated revenues will be strictly *sub-optimal* in almost all time periods.

The result above leaves open the question of establishing convergence. In Example 2 in Appendix B, we illustrate that convergence may not take place and the price path may oscillate indefinitely over time. The reader may question whether such an example, via the choice of demand models and specific parameter values, is pathological in some way and potentially not representative. It is in fact possible to further analyze the properties of the price process, and establish a relationship between the *local stability* of that process and the *elasticity* of the underlying demand curve at the limiting price, namely, $\mathcal{E}_\lambda(\check{p}) = -\check{p}\lambda'(\check{p})/\lambda(\check{p})$. In particular, it can be shown that whenever $\mathcal{E}_\lambda(\check{p}) < 2$, the price process will be locally stable and whenever $\mathcal{E}_\lambda(\check{p}) > 2$, it will be unstable. We do not document the proof of these results here as the purpose of this section is mainly illustrative.

We note that the convergence of prices to a sub-optimal price can be interpreted as a form of "spiral-down effect," related in spirit to the study of Cooper et al. (2006) that analyze a booking limit capacity allocation problem. There the failure to properly model the distributions of arrival classes is the main driver behind this behavior.

It is worthwhile to contrast the observations above to existing results regarding the lack of convergence of prices to an optimal price in the presence of demand learning. Convergence of the price process to a sub-optimal limit point (i.e., not the optimal price) may take place even in the more favorable setting where the demand curve belongs to a parametric family, and the seller therefore knows the structure of this demand curve up to the value of some finite number of unknown parameter. The driver behind this *incomplete learning* phenomenon is essentially the presence of an indeterminate equilibrium; roughly speaking, this is a point in parameter space that serves as an attractor for the dynamical system generating price updates, and at that point no further information can be learned about the unknown parameter values. The reader is referred to McLennan (1984) and Harrison et al. (2012) for further discussion. In the context of the present section, the phenomenon is very different: the driver of suboptimal pricing as outlined in Proposition 1 is primarily the mismatch between the functional form of the (true) demand curve, and the demand model used by the seller. In particular, if the demand model was well-specified, convergence to p^* would always take place (in a single iteration when there is no noise). The oscillatory behavior discussed above appears to be a novel phenomenon, at least in the context of monopoly pricing. Somehow relatedly, in the context of competition, rich structures of price best response dynamics have been reported in Puu (1991), and more recently in Cooper et al. (2009) where firms ignore the presence of competition when selecting their prices.

5.2 Aggregation of past data for inference

For illustrative purposes, we focused on one particular semi-myopic scheme in which estimation of model parameters is based on data only from the most recent "batch." From a practical standpoint, this can be viewed as an extreme case of exponential smoothing (a scheme that weighs down past data, and is used heavily in revenue management applications). From an analysis perspective, it enables us to decouple batch periods and facilitates dealing with dependencies in the observation process. Having said that, an interesting question is what would happen when all past data is used to fit the demand model (say, again, a linear model). In other words, instead of recalibrating according to (5), one would now recalibrate using all past observations

$$(\widehat{\alpha}_{i+1}, \widehat{\beta}_{i+1}) \quad = \quad \arg\min_{\alpha,\beta} \Big\{ \sum_{t=1}^{t_{i+1}} \big[D_t - (\alpha - \beta p_t) \big]^2 \Big\}. \tag{18}$$

As an illustration of some of the possible consequences, we consider in Figure 2 the histograms of the estimate of the optimal price $\widehat{p}_t$ after 10^4 periods, and for different starting prices. Here the

(true) underlying demand curve is taken to be exponential, given by $\exp\{-0.5p\}$, for which the optimal price, maximizing the corresponding revenue function, is $p^* = 2$.

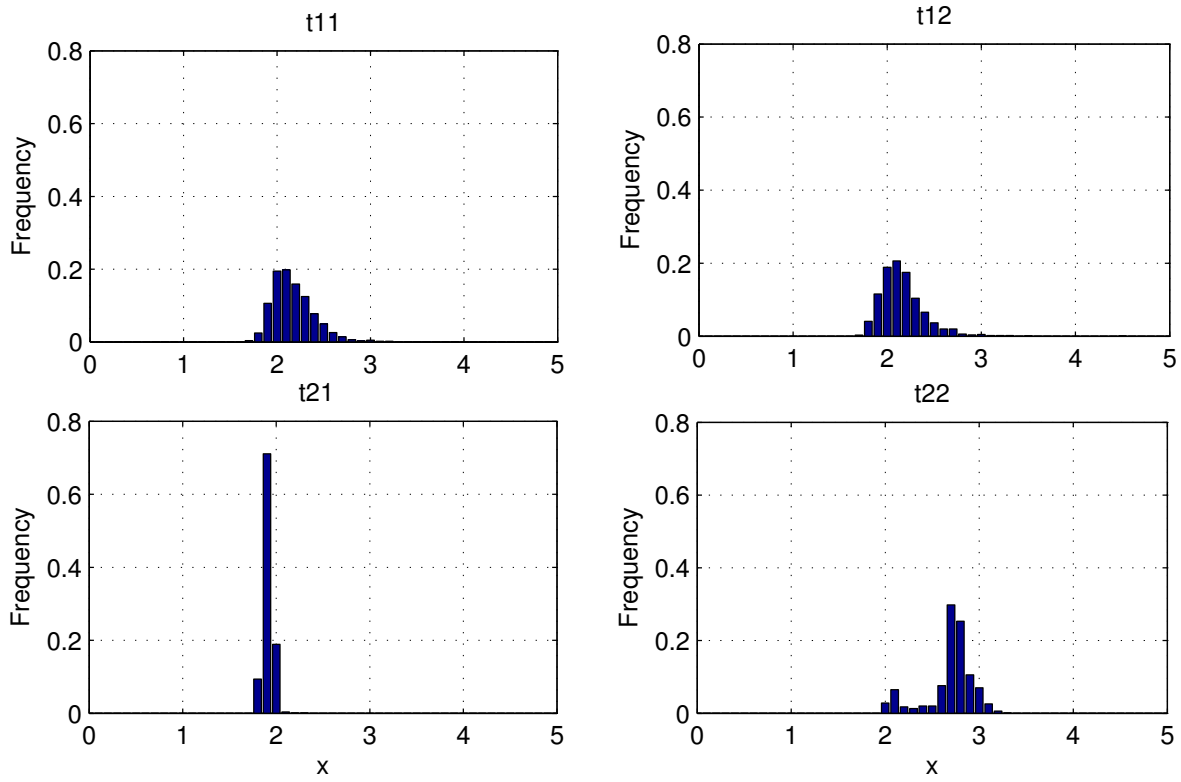

Figure 2: **The impact of aggregation.** The figure depicts the histogram of $\widehat{p}_t$ after 10^4 periods based on 10^3 simulations for the pricing scheme that uses only most recent observations versus a scheme that aggregates all observations. The true demand curve is exponential given by $\exp\{-0.5p\}$, and the fitted model is linear given by $\alpha - \beta p$; $\sigma = 0.1$.

Focusing on the two top panels, corresponding to the cases in which no aggregation takes place, we observe that the impact of the initial price $\widehat{p}_1$ dissipates and the empirical distribution concentrates around the optimal price p^* (as highlighted earlier in the paper). When all observations are used (the two bottom panels), the variance of the estimate $\widehat{p}_t$ appears to be much lower than the case with no aggregation. However, the impact of the initial price does not dissipate anymore: after 10^4 periods the distribution of $\hat{p}_t$ has a mode of 2.75, which is strictly suboptimal, when $\hat{p}_1 = 4.5$. This suggests that schemes that use all existing data might be sensitive to initial conditions (the initial price) in the presence of misspecification. Exploring properties of schemes that use exponential smoothing for past data seems like an important practical avenue of future research.

5.3 Discussion of modeling assumptions and directions for future research

On the linear modeling assumption. The main results derived in this paper, both consistency and revenue-optimality, are predicated on the choice of a linear function to model the unknown demand curve. Our focus on linear models stems from their ubiquitous presence, both in academic studies as well as in the practice of revenue management. A close inspection of the proofs reveals that the main results can be extended straightforwardly, exactly along the same lines, if the policy uses a *generalized linear model*, e.g., exponential $\ell(p;\alpha,\beta) = \exp\{\alpha - \beta p\}$, logit $\ell(p;\alpha,\beta) = \exp\{\alpha - \beta p\}(1+\exp\{\alpha-\beta p\})^{-1}$ and the like. The only difference would be the conditions under which global convergence takes place; for example, it is possible to show that if the inference class is exponential, then the main results continue to hold, provided that the added condition $\lambda(p)\lambda''(p) \geq (\lambda'(p))^2$ is satisfied. On the positive front, this highlights that the virtues of the estimation/optimization cycles do not rely on the linear structure. However, this also begs the following question: acknowledging the possibility of model misspecification, which is the "best" misspecified model? This topic lies in the general domain of model selection and offers an interesting avenue for future research.

Which policies have potential to work under misspecification? More specifically, will every policy which is designed to work optimally or near-optimally in the well specified case, perform well also in the misspecified setting. This a very general question that is beyond the scope of the present paper. As we indicated earlier, antecedent literature mostly provides anecdotal findings, for example Cooper et al. (2006) highlight specific conditions under which it is possible that decisions predicated on a misspecified model converge to an optimal value for the capacity booking problem. Our present work identifies two key ingredients that appear necessary in order for a policy to mitigate the impact of misspecification: i.) the inference model should be sufficiently rich (in the present case, it was critical to have two degrees of freedom); and *ii.*) in blending estimation and optimization, inference should be based on observations "around" the optimal decision point (given the postulated model). This allows us, among other things, to examine recent policies developed in the well specified setting, to determine whether there is hope in transferring them to misspecified scenarios. For example, the CVP policy of den Boer and Zwart (2013) has the above mentioned property, but the basic version of MLE-cycle analyzed in Broder and Rusmevichientong (2012) would not satisfy said condition, as the inference is based on only two fixed experimentation prices. The MLE-cycle policy that bases inference on all past data (also analyzed in Broder and Rusmevichientong (2012)), might again have the potential to mitigate the impact of misspecification, as it does possess the "localization" property. The broader question of porting "good" policies from the well specified to the misspecified setting for general joint learning and optimization problems is worthy of further investigation.

References

Araman, V. F. and Caldentey, R. A. (2010), 'Revenue management with incomplete demand information', forthcoming in Encyclopedia of Operations Research, Wiley .

Bertsekas, D. P. (1999), Nonlinear Programming, Athena Scientific, Belmont, Massachusetts.

Besbes, O., Phillips, R. and Zeevi, A. (2010), 'Testing the validity of a demand model: an operations perspective', Manufacturing & Service Operations Management **12**, 162–183.

Besbes, O. and Zeevi, A. (2009), 'Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms', Operations Research **57**(6), 1407–1420.

Broder, J. and Rusmevichientong, P. (2012), 'Dynamic pricing under a general parametric choice model', Operations Research **60**(4), 965–980.

Cachon, G. and Kök, A. (2007), 'How to (and how not to) estimate the salvage value in the newsvendor model', Manufacturing & Service Operations Management **9**, 276–290.

Chehrazi, N. and Weber, T. A. (2010), 'Monotone approximation of decision problems', Operations Research **58**, 1158–1177.

Cooper, W. L., Homem-de-Mello, T. and Kleywegt, A. J. (2006), 'Models of the spiral-down effect in revenue management', Operations Research **54**, 968–987.

Cooper, W. L., Homem-de-Mello, T. and Kleywegt, A. J. (2009), 'Learning and pricing using models that do not explicitly account for competition', working paper, Univ. of Minnesota .

Dawes, R. M. (1979), 'The robust beauty of improper linear models in decision making', American Psychologist **34**, 571–582.

den Boer, A. and Zwart, B. (2013), 'Simultaneously learning and optimizing using controlled variance pricing', forthcoming in Management Science .

Harrison, J. M., Keskin, B. N. and Zeevi, A. (2011), 'Dynamic pricing with an unknown linear demand model: Asymptotically optimal semi-myopic policies', working paper, Stanford Univ .

Harrison, J. M., Keskin, B. N. and Zeevi, A. (2012), 'Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution', Management Science **58**(3), 570–586.

Kao, Y., Van Roy, B. and Yan, X. (2009), 'Directed regression', Advances in Neural Information Processing Systems, MIT Press **22**, 889–897.

Kiefer, J. and Wolfowitz, J. (1952), 'Stochastic estimation of the maximum of a regression function', Annals of Mathematical Statistics **23**, 462–466.

McLennan, A. (1984), 'Price dispersion and incomplete learning in the long run', Journal of Economic Dynamics and Control **7**, 331–347.

Mersereau, A. and Zhang, D. (2012), 'Markdown pricing with unknown fraction of strategic customers', Manufacturing & Service Operations Management **14**(3), 355–370.

Nemirovski, A., Juditsky, A., Lan, A. G. and Shapiro, A. (2009), 'Robust stochastic approximation approach to stochastic programming', SIAM J. Optimization **19**, 1574–1609.

Puu, T. (1991), 'Chaos in duopoly pricing', Chaos, Solitons & Fractals **1**(6), 573–581.

Rothschild, M. (1974), 'A two-armed bandit theory of market pricing', Journal of Economic Theory **9**, 185–202.

Wang, Z., Deng, S. and Ye, Y. (2011), 'Close the gaps: A learning-while-doing algorithm for a class of single-product revenue management problems', working paper, Stanford Univ .

White, H. (1996), Estimation, Inference and Specification Analysis, Cambridge University Press.

A Proofs for Sections 3 and 4

Preliminaries. For later use, we define $\sigma = \sqrt{\mathbb{E}\varepsilon_1^2}$,

$$\begin{aligned} m_0 &= \min_{p\in[p^{(l)},p^{(h)}]} |\lambda(p)|, \qquad m_1 = \min_{p\in[p^{(l)},p^{(h)}]} |\lambda'(p)|, \\ M_1 &= \max_{p\in[p^{(l)},p^{(h)}]} |\lambda'(p)|, \qquad M_2 = \max_{p\in[p^{(l)},p^{(h)}]} |\lambda''(p)|. \end{aligned}$$

Note that under the assumptions on $\lambda(\cdot)$, $m_0, m_1 > 0$ and $M_1, M_2 < \infty$.

We will prove the conclusions of Theorems 1 and 2 under more general conditions, corresponding to formulation (7), where $\mathcal{X} = [p^{(l)}, p^{(h)}]$ and the same assumptions on $\lambda(\cdot)$ hold. As mentioned in §2, the pricing example is a special case of (7) with $G(x, y) = xy$. In particular, we will further assume that $G(\cdot, \cdot)$ satisfies the following conditions:

1. $G(x, y)$ is continuously differentiable on $[p^{(l)}, p^{(h)}] \times [0, +\infty)$.

2. Let $\bar{G}(p) := G(p, \lambda(p))$. $\bar{G}(p)$ is twice differentiable with bounded second derivative, unimodal, and admits a unique interior maximizer denoted by p^* in $[p^{(l)}, p^{(h)}]$.

3. For any $\alpha > 0$, $\beta > 0$, $G(x, \alpha - \beta x)$ is unimodal and admits a unique maximizer $h(\alpha, \beta)$ in $[0, +\infty)$. Furthermore h is continuously differentiable with bounded partial derivatives on $[m_0/2, +\infty) \times [m_1/2, +\infty)$.

Note under the assumptions of the main text, all these assumptions hold in the special case in which $G(x, y) = xy$. Let

$$\begin{aligned} \check{\alpha}(p) &:= \lambda(p) - \lambda'(p)p \\ \check{\beta}(p) &:= -\lambda'(p). \end{aligned}$$

Let h_α denote the derivative of h with respect to its first argument and h_β denote the derivative of h with respect to its second argument. We also make the following assumption.

Assumption A1

$$\max_{p \in [p^{(l)}, p^{(h)}]} \left\{ \left| \lambda''(p) \left[p h_\alpha(\check{\alpha}(p), \check{\beta}(p)) - h_\beta(\check{\alpha}(p), \check{\beta}(p)) \right] \right| \right\} < 1.$$

Note that in the pricing problem (when $G(x, y) = xy$) studied in the main text, this assumption reduces to Assumption 1 *ii.*). Of course, we want to emphasize that for each general instance of problem (7), Assumption A1 may take a different form. For the pricing problem, it holds under many models and parameter combinations for usual demand curves as highlighted in Example 1.

We denote by $\widehat{\pi}_G(\widehat{p}_1, \{I_i, \delta_i : i \geq 1\})$ the policy that mimics $\widehat{\pi}$, with the exception that when reoptimization is conducted, one attempts to maximize $G(p, \widehat{\alpha}_{i+1} - \widehat{\beta}_{i+1} p)$. In other words, Step 3 is replaced by:

Step 3: Reoptimization

$$\widehat{p}_{i+1} = \mathcal{P}\left(h(\widehat{\alpha}_{i+1}, \widehat{\beta}_{i+1}) \right) \tag{A1}$$

Theorem A1 *Let Assumption A1 hold. Suppose that in the linear-model semi-myopic policy $\widehat{\pi}_G$, $\mathcal{T}_i = \{t_i + 1, ..., t_{i+1}\}$ $\delta_i \to 0$ and $\delta_i I_i^{1/2} / \log(I_i) \to \infty$ as $i \to \infty$. Then, for any initial decision $\widehat{p}_1$, the sequence $\{p_t : t \geq 1\}$ generated by $\widehat{\pi}_G$ converges in probability to the decision p^* that maximizes $G(p, \lambda(p))$.*

Proof of Theorem 1. Since Assumption A1 reduces to Assumption 1 *ii.*) in this setting and since $h(\alpha, \beta) = \alpha/(2\beta)$, this result follows from Theorem A1. ■

Proof of Theorem A1. We will establish L^2 convergence of the sequence $\widehat{p}_i$ to p^*. The proof analyzes the expected deviations from p^*, $\mathbb{E}(\widehat{p}_{i+1} - p^*)^2$, and how they relate to those in the previous

iteration $\mathbb{E}(\widehat{p}_i - p^*)^2$ when i is sufficiently large. In particular, we will establish that the expected deviations shrink geometrically fast, up to a correcting factor due to the noise in the system. To show the latter, we show that the probability that the actual deviations do not behave as such is appropriately small.

Define

$$W_i^1 \;=\; \frac{1}{I_i}\sum_{j=t_i+1}^{t_i+I_i}\varepsilon_j, \qquad W_i^2 \;=\; \frac{1}{I_i}\sum_{j=t_i+I_i+1}^{t_{i+1}}\varepsilon_j.$$

Let $a_i = \sigma\big(2\log(I_i)\big)^{1/2}$ and

$$\mathcal{A}_i = \Big\{\omega : |W_i^j| \le a_i I_i^{-1/2},\; j = 1,2\Big\}.$$

$$\begin{aligned}\mathbb{E}(\widehat{p}_{i+1} - p^*)^2 &\le \mathbb{E}\big[(\widehat{p}_{i+1} - p^*)^2\big|\mathcal{A}_i\big]\mathbb{P}\{\mathcal{A}_i\} + \mathbb{E}\big[(\widehat{p}_{i+1} - p^*)^2\big|\mathcal{A}_i^c\big]\mathbb{P}\{\mathcal{A}_i^c\}\\ &\le \mathbb{E}\big[(\widehat{p}_{i+1} - p^*)^2\big|\mathcal{A}_i\big] + |p^{(h)} - p^{(l)}|^2\mathbb{P}\{\mathcal{A}_i^c\}. \qquad \text{(A2)}\end{aligned}$$

Next, we analyze $\mathbb{E}\big[(\widehat{p}_{i+1} - p^*)^2\big|\mathcal{A}_i\big]$ and $\mathbb{P}\{\mathcal{A}_i^c\}$ separately.

Analysis of $\mathbb{E}\big[(\widehat{p}_{i+1} - p^*)^2\big|\mathcal{A}_i\big]$: We will establish that for all $\omega \in \mathcal{A}_i$, $|\widehat{p}_{i+1} - p^*| \le \gamma|\widehat{p}_i - p^*| + v_i$ where $\gamma < 1$ will be specified later, and v_i is an appropriately shrinking sequence to be specified in the analysis below.

A simple derivation yields that

$$\begin{aligned}\widehat{\beta}_{i+1} &= -\frac{\lambda(\widehat{p}_i + \delta_i) - \lambda(\widehat{p}_i)}{\delta_i} - \frac{1}{\delta_i}\Big[-W_i^1 + W_i^2\Big] \qquad \text{(A3)}\\ \widehat{\alpha}_{i+1} &= \overline{D}_i + \widehat{\beta}_{i+1}\overline{p}_i. \qquad \text{(A4)}\end{aligned}$$

Given this, the recursion for decisions may be written as $\widehat{p}_{i+1} = \mathcal{P}\Big(h(\overline{D}_i + \widehat{\beta}_{i+1}\overline{p}_i, \widehat{\beta}_{i+1})\Big)$, or alternatively

$$\widehat{p}_{i+1} \;=\; \mathcal{P}\Big(h(\breve{\alpha}(\widehat{p}_i), \breve{\beta}(\widehat{p}_i)) + Z_i\Big), \qquad \text{(A5)}$$

with

$$Z_i \;=\; h(\overline{D}_i + \widehat{\beta}_{i+1}\overline{p}_i, \widehat{\beta}_{i+1}) - h(\breve{\alpha}(\widehat{p}_i), \breve{\beta}(\widehat{p}_i)).$$

Next, we analyze Z_i. First note that

$$\begin{aligned}\widehat{\beta}_{i+1} &= -\frac{\lambda(\widehat{p}_i + \delta_i) - \lambda(\widehat{p}_i)}{\delta_i} - \frac{1}{\delta_i}\Big[-W_i^1 + W_i^2\Big]\\ &= -[\lambda'(\widehat{p}_i) + \frac{1}{2}\lambda''(q_i)\delta_i] - \frac{1}{\delta_i}\Big[-W_i^1 + W_i^2\Big]\\ &= \breve{\beta}(\widehat{p}_i) + Z_i^1, \qquad \text{(A6)}\end{aligned}$$

where the second equality follows from Taylor's theorem applied to $\lambda(\cdot)$ with $q_i \in [\widehat{p}_i, \widehat{p}_i + \delta_i]$ and where

$$Z_i^1 = -\frac{1}{2}\lambda''(q_i)\delta_i - \frac{1}{\delta_i}\Big[-W_i^1 + W_i^2\Big].$$

Similarly, one has that $\widehat{\alpha}_{i+1}$ is a "perturbation" of $\check{\alpha}(\widehat{p}_i)$ in the follwing sense

$$\begin{aligned}
\widehat{\alpha}_{i+1} &= \overline{D}_i + \widehat{\beta}_{i+1}\overline{p}_i \\
&= \frac{1}{2}\lambda(\widehat{p}_i) + \frac{1}{2}\lambda(\widehat{p}_i + \delta_i) + W_i^1 + W_i^2 \\
&= \lambda(\widehat{p}_i) + \frac{1}{2}\lambda'(q_i')\delta_i + W_i^1 + W_i^2 - \lambda'(\widehat{p}_i)\widehat{p}_i - \lambda'(\widehat{p}_i)\frac{\delta_i}{2} + Z_i^1(\widehat{p}_i + \delta_i/2) \\
&= \check{\alpha}(\widehat{p}_i) + Z_i^2, \qquad \text{(A7)}
\end{aligned}$$

where the third equality follows from Taylor's theorem applied to $\lambda(\cdot)$ with $q_i' \in [\widehat{p}_i, \widehat{p}_i + \delta_i]$, and where

$$Z_i^2 = \frac{1}{2}\lambda'(q_i')\delta_i + W_i^1 + W_i^2 - \lambda'(\widehat{p}_i)\frac{\delta_i}{2} + Z_i^1(\widehat{p}_i + \delta_i/2).$$

Let $Y_i^1 = \frac{1}{2}M_2\delta_i + 2a_iI_i^{-1/2}\delta_i - 1$ and $Y_i^2 = M_1\delta_i + 2a_iI_i^{-1/2} + Y_i^1(p^{(h)} + \delta_i/2)$. Note that Y_i^1 and Y_i^2 converge to zero as i grows to infinity since, by assumption, δ_i and $\delta_i^{-1}a_iI_i^{-1/2}$ converge to zero. Let $i_0 = \min\big\{i \geq 1 \,:\, Y_i^1 < m_1/2 \text{ and } Y_i^2 < m_0/2\big\}$, For $i \geq i_0$ and $\omega \in \mathcal{A}_i$, $|Z_i^1| < m_1/2$ and $|Z_i^2| < m_0/2$. By assumption, h is continuously differentiable around $(\check{\alpha}(\widehat{p}_i), \check{\beta}(\widehat{p}_i))$, with bounded partial derivatives. Putting together (A6) and (A7), one obtains

$$Z_i = h(\check{\alpha}(\widehat{p}_i) + Z_i^1, \check{\beta}(\widehat{p}_i) + Z_i^2) - h(\check{\alpha}(\widehat{p}_i), \check{\beta}(\widehat{p}_i)) = h_\alpha(a_i, b_i)Z_i^1 + h_\beta(a_i, b_i)Z_i^2,$$

for some (a_i, b_i) on the line segment joining $(\check{\alpha}(\widehat{p}_i), \check{\beta}(\widehat{p}_i))$ to $(\check{\alpha}(\widehat{p}_i) + Z_i^1, \check{\beta}(\widehat{p}_i) + Z_i^2)$. Note that $|Z_i^1| \leq K_1[\delta_i + \delta_i^{-1}a_iI_i^{-1/2}]$ and $|Z_i^2| \leq K_2[\delta_i + \delta_i^{-1}a_iI_i^{-1/2}]$ for some positive constants K_1 and K_2 and hence for some positive K_3,

$$|Z_i| \leq K_3[\delta_i + \delta_i^{-1}a_iI_i^{-1/2}].$$

On another hand, we have $h(\check{\alpha}(\widehat{p}_i), \check{\beta}(\widehat{p}_i)) = h(\check{\alpha}(p^*), \check{\beta}(p^*)) + \rho'(q_i'')(\widehat{p}_i - p^*)$ for some $q_i'' \in [\widehat{p}_i, \widehat{p}_i + \delta_i]$ where $\rho(p) = h(\check{\alpha}(p), \check{\beta}(p))$.

The next lemma, whose proof is deferred to Appendix C, establishes that p^* is a fixed point of $h(\check{\alpha}(p), \check{\beta}(p))$.

Lemma A1 $h(\check{\alpha}(p^*), \check{\beta}(p^*)) = p^*$.

We deduce that

$$h(\check{\alpha}(\widehat{p}_i), \check{\beta}(\widehat{p}_i)) = p^* + \rho'(q_i'')(\widehat{p}_i - p^*).$$

Let

$$\gamma \;=\; \max_{p\in[p^{(l)},p^{(h)}]}\rho'(p) \;=\; \max_{p\in[p^{(l)},p^{(h)}]}\Big\{\Big|\lambda''(p)\Big[ph_\alpha(\check{\alpha}(p),\check{\beta}(p))-h_\beta(\check{\alpha}(p),\check{\beta}(p))\Big]\Big|\Big\}. \tag{A8}$$

Note that since $h(\cdot,\cdot)$ is continuously differentiable and $\lambda(\cdot)$ is twice continuously differentiable, the maximum above is achieved and Assumption A1 implies that $\gamma<1$.

We obtain that for all $\omega\in\mathcal{A}_i$,

$$|\widehat{p}_{i+1}-p^*| \;=\; |\mathcal{P}(h(\overline{D}_i+\widehat{\beta}_{i+1}\overline{p}_i,\widehat{\beta}_{i+1}))-p^*| \;\le\; |h(\overline{D}_i+\widehat{\beta}_{i+1}\overline{p}_i,\widehat{\beta}_{i+1}))-p^*| \;\le\; \gamma|\widehat{p}_i-p^*|+v_i, \tag{A9}$$

where $v_i=K_3[\delta_i+\delta_i^{-1}a_iI_i^{-1/2}]$.

Analysis of $\mathbb{P}\{\mathcal{A}_i^c\}$: We use the following lemma, whose proof, deferred to Appendix C, relies on a large deviations argument.

Lemma A2 *For some suitably large constant $K_4>0$, for $j=1,2$,*

$$\mathbb{P}\{W_i^j>a_iI_i^{-1/2}\} \;\le\; \frac{K_4}{I_i}, \quad \textit{for all } i\ge 1.$$

From the above, we deduce that

$$\mathbb{P}\{\mathcal{A}_i^c\} \;=\; \mathbb{P}\{\max_{j=1,2}|W_i^j|>a_iI_i^{-1/2}\} \overset{(a)}{\le} 2\mathbb{P}\{|W_i^1|>a_iI_i^{-1/2}\} \overset{(b)}{\le} \frac{4K_4}{I_i}, \tag{A10}$$

where (a) follows from a union bound and (b) follows from Lemma A2.

Bounding $\mathbb{E}(\widehat{p}_{i+1}-p^*)^2$: Using (A2), (A9) as well as (A10), one obtains

$$\mathbb{E}(\widehat{p}_{i+1}-p^*)^2 \;\le\; \gamma^2\mathbb{E}(\widehat{p}_i-p^*)^2+v_i^2+2\gamma\mathbb{E}|\widehat{p}_{i+1}-p^*|v_i+|p^{(h)}-p^{(l)}|^2\frac{4K_4}{I_i}.$$

Noting that

$$2\mathbb{E}|\widehat{p}_i-p^*|v_i\le 2[\gamma/(1-\gamma^2)]v_i^2+[(1-\gamma^2)/(2\gamma)]\mathbb{E}(\widehat{p}_i-p^*)^2,$$

one has that

$$\mathbb{E}(\widehat{p}_{i+1}-p^*)^2 \;\le\; \frac{1+\gamma^2}{2}\mathbb{E}(\widehat{p}_i-p^*)^2+w_i,$$

where

$$w_i=(1+2\gamma/(1-\gamma^2))v_i^2+4|p^{(h)}-p^{(l)}|^2K_4/I_i. \tag{A11}$$

Take i sufficiently large such that w_j is decreasing for all $j\ge i-\lceil i/2\rceil$. Let $\eta=(1+\gamma^2)/2$ and

$j(i) = \lceil i/2 \rceil$.

$$\begin{aligned}
\mathbb{E}(\widehat{p}_{i+1} - p^*)^2 &\leq \eta^i (p_1 - p^*)^2 + \sum_{j=0}^{i-1} \eta^j w_{i-j} \\
&\leq \eta^i (p_1 - p^*)^2 + \sum_{j=0}^{j(i)} \eta^j w_{i-j} + \sum_{j=j(i)+1}^{i-1} \eta^j w_{i-j}, \\
&\leq \eta^i (p_1 - p^*)^2 + \frac{w_{i-j(i)}}{1-\eta} + \eta^{j(i)+1} \sum_{j=1}^{i} w_j. \qquad \text{(A12)}
\end{aligned}$$

Since $\eta < 1$ and $w_i \to 0$ as $i \to \infty$, one obtains that $w_{i-j(i)} \to 0$ and $\eta^{j(i)+1} \sum_{j=1}^{i} w_j \to 0$ as $i \to \infty$. Hence

$$\mathbb{E}(\widehat{p}_{i+1} - p^*)^2 \to 0.$$

Since L^2 convergence implies convergence in probability, the result follows and the proof is complete. ■

Theorem A2 (revenue optimality) *Let Assumption A1 hold. Suppose that in the linear-model semi-myopic policy $\widehat{\pi}_G$, $\mathcal{T}_i = \{t_i + 1, ..., t_{i+1}\}$ and the block sequence is selected as in* (14) *with $\delta_i = I_i^{-1/4}$, $i \geq 1$. Then, for any initial decision $\widehat{p}_1$, the sequence of decisions $\{p_t : t \geq 1\}$ generated by $\widehat{\pi}_G$ satisfies*

$$\mathbb{E}\Big[\sum_{t=1}^{T} [\bar{G}(p^*) - \bar{G}(p_t)]\Big] \leq C \max\{1, \sigma^2\} (\log T)^2 \sqrt{T}$$

for some positive constant C independent of σ, and all $T \geq 2$.

Proof of Theorem 2. Let r denote the mapping $p \mapsto p\lambda(p)$ and note that $r(\cdot)$ is twice continuously differentiable with second derivative bounded by $(2M_1 + p^{(h)} M_2)$, where M_1 and M_2 where defined at the start of the Appendix. The result follows from applying Theorem A2 to the special case of interest. ■

Proof of Theorem A2. Note that throughout the proofs, all constants introduced $C_1, C_2, ...$ are constants that do not depend on σ. Fix a time horizon T and let $k = \inf\{j \geq 1 : 2\sum_{i=1}^{j} I_i \geq T\}$. The regret after T periods is given by

$$R(\widehat{\pi}_G, T) = \mathbb{E}\Big[\sum_{t=1}^{T} [\bar{G}(p^*) - \bar{G}(p_t)]\Big]$$

By assumption, $\bar{G}$ is twice differentiable with bounded second derivative. We deduce, through a Taylor expansion that

$$|\bar{G}(p^*) - \bar{G}(p)| \leq K(p - p^*)^2.$$

Hence,

$$\begin{aligned} R(\widehat{\pi}_G, T) &\leq \mathbb{E}\Big[\sum_{i=1}^{k}\Big(\big[\bar{G}(p^*) - \bar{G}(\widehat{p}_i)\big] + \big[\bar{G}(p^*) - \bar{G}(\widehat{p}_i + \delta_i)\big]\Big) I_i\Big] \\ &\leq K\sum_{i=1}^{k}\Big(\mathbb{E}(\widehat{p}_i - p^*)^2 + \mathbb{E}(\widehat{p}_i + \delta_i - p^*)^2\Big) I_i \\ &= K\sum_{i=1}^{k}\Big(2\mathbb{E}(\widehat{p}_i - p^*)^2 + \delta_i^2 + 2\mathbb{E}|\widehat{p}_i - p^*|\delta_i\Big) I_i \\ &\leq K\sum_{i=1}^{k}\Big(2\mathbb{E}(\widehat{p}_i - p^*)^2 + \delta_i^2 + 2[\mathbb{E}(\widehat{p}_i - p^*)^2]^{1/2}\delta_i\Big) I_i. \end{aligned}$$

Using equation (A12) from the proof of Theorem 1, one has that

$$\mathbb{E}(\widehat{p}_i - p^*)^2 \leq \eta^i(\widehat{p}_1 - p^*)^2 + \frac{w_{i-j(i)}}{1-\eta} + \eta^{j(i)+1}\sum_{j=1}^{i} w_j,$$

with the sequence $\{w_i : i \geq 1\}$ defined in (A11). In particular, noting that with the selection of parameters assumed in the theorem, $w_i \leq C_1 \max\{1, \sigma^2\}(\log I_i) I_i^{-1/2}$ for some suitably large positive constant C_1, we deduce that for some positive constant $C_2 > 0$, one has that

$$\mathbb{E}(\widehat{p}_i - p^*)^2 \leq C_2 \max\{1, \sigma^2\}(\log I_i) I_i^{-1/2}.$$

Hence, for some suitable constant $C_3 > 0$,

$$R(\widehat{\pi}_G, T) \leq C_3 K \max\{1, \sigma^2\}\sum_{i=1}^{k}[(\log I_i) I_i^{-1/2} + \delta_i^2] I_i \leq C_4 \max\{1, \sigma^2\}\sum_{i=1}^{k}(1 + \log I_i) I_i^{1/2}.$$

Bounding each term in the sum by the last one, one obtains that the regret is bounded by $C_4 k \max\{1, \sigma^2\}(1 + \log I_k) I_k^{1/2}$. Noting that $k \leq C_5 \log T$ for some $C_5 > 0$, one obtains

$$R(\widehat{\pi}_G, T) \leq C_6 \max\{1, \sigma^2\}(\log T)^2 T^{1/2}.$$

This completes the proof. ■

B Supplement to Section 5.1

Proof of Proposition 1. Let

$$W(p) = \frac{(\widetilde{\alpha}/2)p}{\widetilde{\alpha} - \lambda(p)}, \qquad \text{and} \qquad \widetilde{W}(p) = \mathcal{P}\left(W(p)\right).$$

Then, the recursion for the decisions may be rewritten as

$$\widehat{p}_{t+1} = \widetilde{W}(\widehat{p}_t), \qquad t \geq 1.$$

Since $\widetilde{W}(\cdot)$ is continuous on $[p^{(l)}, p^{(h)}]$, the only possible limit points of the sequence $\{\widehat{p}_t : t \geq 1\}$ are fixed points of $\widetilde{W}(\cdot)$. We next establish that $\widetilde{W}(\cdot)$ has exactly one fixed point in $[p^{(l)}, p^{(h)}]$ and this fixed point is given by $\check{p}$ defined in (17).

A fixed point of $W(\cdot)$ in $(0, +\infty)$ needs to satisfy

$$W(p)/p = 1.$$

Since $\lambda(\cdot)$ is assumed to be decreasing on $[p^{(l)}, p^{(h)}]$, $W(p)/p = (\widetilde{\alpha}/2)/(\widetilde{\alpha} - \lambda(p))$ is decreasing, which implies that $W(p)/p = 1$ has at most one solution in $(0, +\infty)$.

If $\widetilde{\alpha}/2 \in (\lambda(p^{(h)}), \lambda(p^{(l)}))$, then $W(\cdot)$ admits a fixed point in $(p^{(l)}, p^{(h)})$, and the latter is given by $\lambda^{-1}\big(\widetilde{\alpha}/2\big) = \check{p}$. Noting that $W(p^{(h)})/p^{(h)} < 1$ and $W(p^{(l)})/p^{(l)} > 1$, $\check{p}$ is also the unique fixed point of $\widetilde{W}(\cdot)$ in $[p^{(l)}, p^{(h)}]$.

If $\widetilde{\alpha}/2 \geq \lambda(p^{(l)})$, then $W(p^{(l)})/p^{(l)} \leq 1$. In such a case, $W(\cdot)$ does not admit any fixed point in $(p^{(l)}, +\infty)$ and the only fixed point of $\widetilde{W}(\cdot)$ on $[p^{(l)}, p^{(h)}]$ is $p^{(l)}$. If $\widetilde{\alpha}/2 \leq \lambda(p^{(h)})$, then $W(p^{(h)})/p^{(h)} \geq 1$. In such a case, $W(\cdot)$ does not admit any fixed point in $(0, p^{(h)})$ and the only fixed point of $\widetilde{W}(\cdot)$ on $[p^{(l)}, p^{(h)}]$ is $p^{(h)}$. This completes the proof. ■

Example 2 (price behavior) To illustrate the possible issues that may arise, consider the following illustrative example. The price domain is taken to be $[p^{(l)}, p^{(h)}] = [0, 10]$. We assume the true demand model is of logit-form, $\lambda(p) = \exp\{a - bp\}/(1 + \exp\{a - bp\})$ with $a > 0$ and $b > 0$, and that the seller fits the linear model, $1 - \beta p$; i.e., the only parameter that is inferred is β. A simple calculation yields that the only possible limit point is $\check{p} = a/b$, and that the revenue maximizing price is given by $p^* = b^{-1}W(\exp\{-1 + a\})$, where $W(\cdot)$ is the Lambert W function (the inverse of $x \mapsto x \exp\{x\}$). Clearly $\check{p}$ and p^* need not coincide.

In Figure 3, we depict two simulation runs, each including 30 price iterates. In both runs, the initial price, $\widehat{p}_1$, is taken to be p^*, namely, the optimal price. In the first case, appearing on panel (a), the underlying demand function is a logit with parameters $a = 3$ and $b = 1$, with $p^* \approx 2.55$ and $\check{p} = 3$. We observe that the price iterates converge to $\check{p}$. In the second case, appearing on panel (b), the underlying demand function is a logit with parameters $a = 4.1$ and $b = 1$, with $p^* \approx 3.27$ and $\check{p} = 4.1$. We observe that the price iterates *do not converge*, rather they oscillate around $\check{p}$. Note that the oscillations occur despite the fact that the underlying demand environment is stationary.

C Proofs of Auxiliary Results

Proof of lemma A1. Let

$$\begin{aligned} \alpha^* &= \check{\alpha}(p^*) = \lambda(p^*) - \lambda'(p^*)p^* \\ \beta^* &= \check{\beta}(p^*) = -\lambda'(p^*), \end{aligned}$$

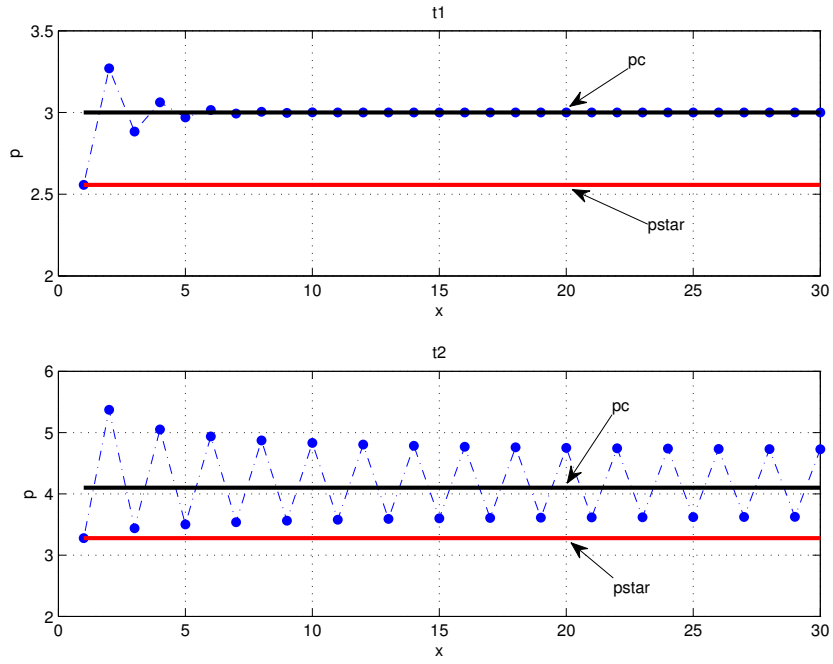

Figure 3: **Behavior of the price process.** Prices converge to $\check{p} \neq p^*$ in (a) and never settle in (b). The true model is a logit model given by $\exp\{3-p\}/(1+\exp\{3-p\})$ in (a) and by $\exp\{4.1-p\}/(1+\exp\{4.1-p\})$ in (b). The fitted model is linear given by $1-\beta p$. The price process is started at the price that maximizes the profit rate associated with the true model, p^*.

p^* is an interior maximum to $G(p,\lambda(p))$ and uniquely satisfies

$$G_x(p^*,\lambda(p^*)) + \lambda'(p^*)G_y(p^*,\lambda(p^*)) = 0. \tag{C13}$$

On the other hand, the unique maximizer of $G(p,\alpha^*-\beta^* p)$ satisfies

$$G_x(p,\alpha^*-\beta^* p) - \beta^* G_y(p^*,\alpha^*-\beta^* p) = 0. \tag{C14}$$

Note that $\alpha^*-\beta^* p^* = \lambda(p^*)$ and hence, by (C13), $p=p^*$ solves (C14), i.e., $h(\alpha^*,\beta^*)=p^*$. ■

Proof of Lemma A2. Fix $j\in\{1,2\}$. For $s\in(-\rho,\rho)$, define

$$\psi(s) = \log \mathbb{E}[\exp\{s\varepsilon_1\}].$$

Note that for any $s\in(-\rho,\rho)$ and any $x>0$, Markov's inequality yields that

$$\mathbb{P}\{W_i^j > x\} \quad \leq \quad \exp\{I_i(\psi(s)-sx)\}$$

Select i sufficiently large so that $a_i I_i^{-1/2} < \sigma\rho$. Fix $x = a_i I_i^{-1/2} = 2\sigma(\log I_i)^{1/2} I_i^{-1/2}$ and let $s^* = x/\sigma^2$. A third order Taylor expansion around 0 yields that for some $\widetilde{s}\in[0,s^*]$

$$\psi(s^*) \quad = \quad \frac{1}{2}\sigma^2(s^*)^2 + \frac{1}{6}\psi'''(\widetilde{s})(s^*)^3.$$

This implies that

$$\psi(s^*) - s^* x \geq -\frac{1}{2}\frac{x^2}{\sigma^2} - C_4 \frac{x^3}{\sigma^6},$$

where $C_4 = \max_{s\in[-\rho,\rho]}\{|\psi'''(s)|\}$, which in turn yields

$$\mathbb{P}\{W_i^1 > x\} \;\leq\; \exp\Big\{-I_i(-\frac{1}{2}\frac{x^2}{\sigma^2} - C_4 x^3/\sigma^6)\Big\}.$$

Substituting the value of x, one obtains for some suitably large constant $C_5 > 0$,

$$\mathbb{P}\{W_i^1 > a_i I_i^{-1/2}\} \;\leq\; \exp\Big\{-\log I_i + (C_4/\sigma^6) I_i^{-1/2} (\log I_i)^{3/2}\Big\} \;\leq\; \frac{C_5}{I_i}.$$

This completes the proof. ■

Remark C1 (Verification) We verify that the policy $\widehat{\pi}$ (with $I_i = 1$, $\mathcal{T}_i = \{1, ..., t_{i+1}\}$ and $\delta_i = t^{-1/4}$) satisfies sufficient conditions for minimax optimality when ε_t's are uniformly bounded almost surely, namely the two sufficient conditions in Harrison et al. (2011, Theorem 2) are satisfied. Suppose the support of ε_t is in $[-U, U]$ for some $U > 0$.

i.) We first check the information accumulation condition, and upper bound $\sum_{s=1}^{2t}(p_s - \bar{p}_{2t})^2$.

$$\begin{aligned}
\sum_{s=1}^{2t}(p_s - \bar{p}_{2t})^2 &= \sum_{s=2}^{2t}(1 - s^{-1})(p_s - \bar{p}_{s-1})^2 \\
&\geq \sum_{s=2}^{t}(p_{2s-1} - \bar{p}_{2s-2})^2 + \sum_{s=2}^{t}(p_{2s} - \bar{p}_{2s-1})^2 \\
&= \sum_{s=2}^{t} A_s^2 + \sum_{s=2}^{t} B_s^2,
\end{aligned}$$

where $A_s = p_{2s-1} - \bar{p}_{2s-2}$ and $B_s = p_{2s} - \bar{p}_{2s-1}$.

Suppose first $A_s \leq -\delta_s/2$. Then $A_s^2 \geq \delta_s^2/4$.

Suppose now $A_s > -\delta_s/2$. Note that $B_s = p_{2s} - \bar{p}_{2s-1} = p_{2s-1} + \delta_s - \bar{p}_{2s-2} + (1/(2s-2))(p_{2s-1} - \bar{p}_{2s-2}) \geq \delta_s - (1 + 1/(2s-2))\delta_s/2$. Hence, for $s \geq 2$, $B_s \geq \delta_s/4$. We deduce that $B_s^2 \geq \delta_s^2/16$.

Hence, we have that $\sum_{s=2}^{t} A_s^2 + B_s^2 \geq \kappa_0\sqrt{t}$ for some $\kappa_0 > 0$.

ii.) We now bound the deviations from the greedy solution: $\sum_{s=1}^{2t}(\varphi(\alpha_s, \beta_s) - p_{s+1})^2$, where $\varphi(\alpha, \beta) = \mathcal{P}(\alpha/(2\beta))$, and (α_s, β_s) are the least squares estimates based on all observations up to and including time s. We define $(\alpha_0, \beta_0) := (1, 1)$.

$$\begin{aligned}
\sum_{s=1}^{2t}(\varphi(\alpha_s, \beta_s) - p_{s+1})^2 &= \sum_{s=1}^{t}(\varphi(\alpha_{2s-2}, \beta_{2s-2}) - p_{2s-1})^2 + \sum_{s=1}^{t}(\varphi(\alpha_{2s-1}, \beta_{2s-1}) - p_{2s})^2 \\
&= \sum_{s=1}^{t}(\varphi(\alpha_{2s-1}, \beta_{2s-1}) - \varphi(\alpha_{2s-2}, \beta_{2s-2}) - \delta_s)^2.
\end{aligned}$$

Next, we evaluate $\varphi(\alpha_{2s-1},\beta_{2s-1})-\varphi(\alpha_{2s-2},\beta_{2s-2})$.

Let $u_s=\sum_{i=1}^s(p_i-\bar{p}_s)\varepsilon_i$ and $A_s=\sum_{i=1}^s(p_i-\bar{p}_s)^2$. Then, standard derivations lead to

$$\begin{aligned}\beta_{s+1}-\beta &= \frac{u_{s+1}}{A_{s+1}}\\ &= \frac{u_s}{A_s}\frac{1}{1+(A_{s+1}-A_s)/A_s}+\frac{u_{s+1}-u_s}{A_{s+1}}\end{aligned}$$

We know from i) that $A_s\geq\kappa\sqrt{s}$ almost surely. We deduce that $(A_{s+1}-A_s)/A_s\leq|p^{(h)}-p^{(l)}+\delta_1|^2/(\kappa_0\sqrt{s+1})$ almost surely. In addition, since ε_t is assumed to have finite support, $|u_{s+1}-u_s|/A_{s+1}=\leq|p^{(h)}-p^{(l)}|C_1/(\kappa_0\sqrt{s+1})$ for some $C_1>0$. We deduce that

$$|\beta_{s+1}-\beta_s| \leq \frac{C_2}{\sqrt{s}},$$

for some constant $C_2>0$. It follows that

$$\begin{aligned}|\alpha_{s+1}-\alpha_s| &= |\bar{D}_{s+1}-\bar{D}_s+\beta_{s+1}\bar{p}_{s+1}-\beta_s\bar{p}_s|\\ &\leq \frac{2U}{s+1}+|\beta_{s+1}-\beta_s|+\frac{1}{s+1}\beta_{s+1}p^{(h)}.\end{aligned}$$

We deduce that

$$|\alpha_{s+1}-\alpha_s| \leq \frac{C_3}{\sqrt{s}}.$$

Concluding, one has for some $C_4>0$

$$|\varphi(\alpha_{2s-1},\beta_{2s-1})-\varphi(\alpha_{2s-2},\beta_{2s-2})| \leq |\frac{\alpha_{s+1}}{2\beta_{s+1}}-\frac{\alpha_s}{2\beta_s}| \leq \frac{C_4}{\sqrt{s}}.$$

This implies that form some $\kappa_1>0$,

$$\sum_{s=1}^{2t}(\varphi(\alpha_s,\beta_s)-p_{s+1})^2 \leq \sum_{s=1}^{2t}(C_4s^{-1/2}+\delta_s)^2 \leq \kappa_1\sqrt{s}.$$