Uncertainty Quantification for Demand Prediction in Contextual Dynamic Pricing

Yining Wang
Warrington College of Business, University of Florida, 1384 Union Road Bryan, Gainesville, Florida, 32611, USA
yining.wang@warrington.ufl.edu

Xi Chen
Leonard N. Stern School of Business, New York University, 44 West 4th Street, New York, New York 10012, USA
xc13@stern.nyu.edu

Xiangyu Chang*
Department of Information Management and E-business, School of Management, Xi’an Jiaotong University, No. 28, Xianning West Rd, Shaanxi, 710049, China
xiangyuchang@xjtu.edu.cn

Dongdong Ge*
School of Information Management and Engineering, Shanghai University of Finance and Economics, No. 100, Wudong Road, Yangpu District, Shanghai, 200433, China
ge.dongdong@mail.shufe.edu.cn

Data-driven sequential decision has found a wide range of applications in modern operations management, such as dynamic pricing, inventory control, and assortment optimization. Most existing research on data-driven sequential decision focuses on designing an online policy to maximize revenue. However, the research on uncertainty quantification on the underlying true model function (e.g., demand function), a critical problem for practitioners, has not been well explored. In this study, using the problem of demand function prediction in dynamic pricing as the motivating example, we study the problem of constructing accurate confidence intervals for the demand function. The main challenge is that sequentially collected data lead to significant distributional bias in the maximum likelihood estimator or the empirical risk minimization estimate, making classical statistical approaches such as the Wald’s test no longer valid. We address this challenge by developing a debiased approach and provide the asymptotic normality guarantee of the debiased estimator. Based this the debiased estimator, we provide both point-wise and uniform confidence intervals of the demand function.

Key words: adaptive data; asymptotic normality; confidence interval; dynamic pricing; data-driven sequential decision

History: Received: August 2020; Accepted: December 2020 by Annabelle Qi Feng, after 1 round of revision
*Corresponding author.

1. Introduction

In recent years, data-driven sequential decision-making has received a lot of attentions and finds a wide range of applications in operations management, such as dynamic inventory control (see, e.g., Chen and Plambeck 2008, Chen et al. 2019a, b; Huh et al. 2011, Lei et al. 2019), dynamic pricing (see, e.g., Besbes and Zeevi 2009, 2015, Broder and Rusmevichientong 2012, Chen et al. 2019c, Wang et al. 2014), dynamic assortment optimization (see, e.g., Agrawal et al. 2019, Chen et al. 2020b, c, Rusmevichientong and Topaloglu 2012, Saure and Zeevi 2013, Wang et al. 2018). Take the personalized/contextual dynamic pricing as an example; it is usually assumed that the underlying demand, which is a function of the price and customer’s contextual information, follows a certain probabilistic model with unknown parameters. Over a finite time selling horizon of length T, at each time period, one customer arrives. The seller observes the characteristic of the customer and makes the price decision. Then the arriving customer makes the purchase decision based on the posted price. The seller will observe the purchase decision, update her knowledge about the demand model, and might change the price policy accordingly for future customers. The key challenge in dynamic pricing is to accurately estimate the underlying model parameter in demand function, which will then be used to determine prices later on. Existing literature on dynamic pricing only constructs a point estimator of the underlying model parameter, that is, estimating the parameter by a single number or a vector, without quantifying the uncertainty in the estimator. Uncertainty quantification is very useful for practitioners. It is highly desirable for the seller to obtain confidence intervals of the
underlying demand function, which is guaranteed to cover the true demand function with $1 - \alpha$ probability (also known as the confidence level, e.g., $\alpha = 0.05$).

Although construction of confidence interval has been a classical topic in statistics (Stigler 2002), the existing results in statistical literature mainly deal with independent and non-adaptive data. The behavior of sequentially collected data is quite different from independent data. In particular, in the (contextual) dynamic pricing problem both the decision (e.g., the price) and the collected customers’ contextual information at each time period are adaptive, which heavily correlate with information obtained in previous periods. Due to the sequential dependence, estimators computed from adaptively collected data might have severe distributional bias even when the sample size goes to infinity (Deshpande et al. 2018, 2019). Such a bias makes the classical approach of constructing confidence intervals (e.g., Wald’s test, see Chapter 17 of Keener 2010) no longer valid.

The main goal of our paper is to construct a debiased estimator that is asymptotically normal centered at the true model parameter with a simple covariance matrix structure. Based on the proposed debiased estimator, we construct both point-wise confidence intervals (i.e., confidence intervals valid for any given decision variable (price) and contextual information) and uniform confidence intervals (i.e., confidence intervals uniformly valid for all decision variables and contextual information). To highlight our main idea, we will consider the problem of constructing confidence intervals for demand function in dynamic pricing, which is one of the most important data-driven sequential decision problems in revenue management. More specifically, accurate uncertainty quantification of demand rates could benefit data-driven operations management in the following ways:

1. Confidence intervals of demand rates (uniformly for all prices) automatically lead to accurate quantification of variability of the revenue, as the revenue is simply a product of the posted price (known to the retailer) and the unknown demand rate. The accurate quantization of uncertainty in the revenue could provide essential information to the retailer, for example when the retailer would like to take a more risk-averse approach to minimize the probability of declined revenue instead of solely maximizing his/her average revenue.

2. The accurate quantification of variability in the demand rates could also provide valuable information for the retailer to make inventory decisions, as the inventory could decide on the stock order-up-to levels in new selling periods such that in order to keep the probability of out-of-stock at a manageable level. Such measurement/estimation of out-of-stock probabilities can only be made accurate by the unbiased uncertainty quantification approach proposed in this study.

In particular, we study a stylized personalized dynamic pricing model in which there are $T$ selling periods. At each selling period $t \in \{1, \ldots, T\}$, a potential customer comes with an observable personal context vector $x_t$. Instead of assuming $x_t$ are independent across time periods as in existing literature (e.g., Chen et al. 2020a, Miao et al. 2019), we allow $x_t$ to depend on information from previous selling periods. This is a more practical scenario since a customer’s contextual information might be heavily correlated with previous prices and realized demands. For example, a consecutive time periods of posted lower price or higher demands will attract new customers from a different population, whose contextual information will be different from the previous customers. By observing the contextual information $x_t$ of the arriving customer, the seller decides the price $p_t$ and the customer decides on a realized demand. We assume the demand of the arriving customer follows a general probabilistic model,

$$d_t = f(p_t, x_t; \theta_0) + \xi_t,$$  \hfill (1)

where $f$ is a parametric function parameterized by $\theta_0$ with a known form (e.g., linear or logistic), $\theta_0 \in \mathbb{R}^d$ is an unknown parameter vector that models the demand behaviors, and $\xi_t$ are zero mean, conditionally independent (conditioning on $p_t$ and $x_t$) noise variables. A typical objective of the retailer is to maximize his/her expected revenue, or more specifically

$$\max_{p_t \in [p_{\min}, p_{\max}]} \mathbb{E}[f(p_t, x_t; \theta_0)],$$

without knowing the model $\theta_0$ a priori. In this study, our goal is to construct confidence intervals for both the true model parameter $\theta_0$ and the underlying demand function $f$ (see the definition in section 1.1).

The demand model in Equation (1) is very general and covers two widely used demand models: the linear model and the logistic model. In the linear model, $d_t$ is modeled as

$$d_t = \langle \phi(p_t, x_t), \theta_0 \rangle + \xi_t,$$  \hfill (2)

where $\phi : (p_t, x_t) \mapsto \phi_t \in \mathbb{R}^d$ is a known feature map for the price and contextual information, and $\xi_t \sim \mathcal{N}(0, \nu^2)$ are noise variables. In the logistic regression model, $d_t \in \{0, 1\}$ is a binary demand realized according to the logistic model.

Wang, Chen, Chang, and Ge: Confidence Intervals for Demand Prediction
Production and Operations Management 30(6), pp. 1703–1717, © 2020 Production and Operations Management Society
The main objective of this study is to quantify the uncertainty for the learned demand function from purchase data on dynamically, adaptively chosen prices and contexts. Namely, we will construct two types of confidence intervals of the underlying demand function \(f\), \(\phi\)-point-wise confidence intervals and uniform confidence intervals, which are introduced as follows.

For a pre-specified confidence level \(1 - \alpha\) at the end of \(T\) time periods, where \(\alpha \in (0, 1)\) is usually a small constant such as 0.1 or 0.05, our goal is to construct upper and lower confidence interval edges \(\ell_a(p, x), u_a(p, x)\), such that for any given price \(p\), context \(x\), and \(\theta_0\),

\[
\lim_{T \to \infty} \Pr[\ell_a(p, x) \leq f(p, x; \theta_0) \leq u_a(p, x)] = 1 - \alpha. \tag{4}
\]

Note that, in Equation (4), both \(\ell_a(p, x), u_a(p, x)\) depend on \(T\) and are likely to become more accurate as \(T \to \infty\). The confidence interval in Equation (4) is known as the point-wise confidence interval since it holds for a fixed price \(p\) and context vector \(x\).

In many applications, we are also interested in confidence intervals \(L_a(\cdot, \cdot), U_a(\cdot, \cdot)\) with uniform coverage. More specifically, for a pre-determined confidence level \(1 - \alpha, \alpha \in (0, 1)\), \(L_a, U_a\) satisfy for all \(\theta_0\) that

\[
\lim_{T \to \infty} \Pr[\forall p \in [p_{\text{min}}, p_{\text{max}}], \forall x \in X, L_a(p, x) \leq f(p, x; \theta_0) \leq U_a(p, x)] = 1 - \alpha. \tag{5}
\]

where \(X\) is a certain compact subset of \(\mathbb{R}^d\) as the domain of all context vectors.

To construct these confidence intervals, we also provide the confidence interval of the model true parameter \(\theta_0\), which might have its own independent interest in practice.

As we mentioned, the main difficulty in constructing these confidence intervals lies in the two dependency structures of the price and contexts. Therefore, in contrast to the non-adaptive case where the maximum likelihood estimator (MLE) is unbiased, the MLE based on the adaptive data will have a significant distributional bias. In the next subsection, we briefly discuss two popular contextual dynamic pricing algorithms in the literature to better illustrate the adaptive data collection process. We also explain in section 3 why the classical construction of confidence intervals fails in our problem.

1.2. Online Policies for Contextual Dynamic Pricing

We mention two popular online policies for the contextual dynamic pricing problem. The \(\epsilon\)-greedy policy. An \(\epsilon\)-greedy policy (Watkins 1989) has a parameter \(\epsilon \in (0, 1)\) to balance the tradeoff between exploration and exploitation. At each selling period \(t \in [T]\), with probability \(\epsilon\), a price \(p_t \in [p_{\text{min}}, p_{\text{max}}]\) is selected uniformly at random for exploration. With probability \(1 - \epsilon\), the exploitation price \(p_t = \arg \max_{p \in [p_{\text{min}}, p_{\text{max}}]} \exp\{\phi(p_t, x_t, \theta_{t-1})\} \) is set based on the current estimate \(\theta_{t-1}\).
\[ \hat{\theta}_{t-1} = \arg \min_{\theta \in \mathbb{R}^d} \sum_{i=1}^{n} \rho(d_i, p_t, x_i; \theta) + \lambda \|\theta\|_2^2, \quad (6) \]

which is the regularized empirical-risk minimization (ERM) using sales data from prior selling episodes. Here \( \rho \) is a certain risk function depending on the particular class of the underlying demand model \( f \). For example, for the linear demand model, the least-squares function is commonly used:

\[ \rho(d_i, p_t, x_i; \theta) = (d_i - \langle p_t(x_i), \theta \rangle)^2. \]

For the logistic demand model, the negative log-likelihood function is often adopted,

\[ \rho(d_i, p_t, x_i; \theta) = -d_i \log(f(p_t(x_i); \theta)) - (1 - d_i) \log(1 - f(p_t(x_i); \theta)). \]

A common choice of \( \rho \) would be the negative log-likelihood function. In principle, the risk function \( \rho \) should be selected such that the underlying true model \( \theta_0 \) minimizes the \( \rho \) function in expectation. Detailed assumptions on \( \rho \) will be given in section 2.

The Upper-Confidence Bound (UCB) policy. In the UCB policy (or more specifically the LinUCB policy for linear or generalized linear contextual bandits (Abbasi-Yadkori et al. 2012, Filippi et al. 2010, Rusmevichientong and Tsitsiklis 2010), a regularized MLE \( \hat{\theta}_{t-1} \) is calculated for every selling period in (6). Afterwards, an offered price \( p_t \) is selected to maximize an upper bound of the demand function \( f \), or more specifically

\[ p_t = \max_{p \in [p_{\min}, p_{\max}]} \left\{ p \times \min \left\{ 1, f(p, x; \hat{\theta}_{t-1}) + \text{CI}_t(p, x) \right\} \right\}, \quad (7) \]

where \( \text{CI}_t(\cdot, \cdot) \) is a certain form of confidence bound such that with high probability \( f(p, x; \hat{\theta}_{t-1}) + \text{CI}_t(p, x) \geq f(p, x; \theta_0) \) for all \( p \) and \( x \), where \( \theta_0 \) is the underlying true model parameter. We refer the readers to the works of Abbasi-Yadkori et al. (2012), Rusmevichientong and Tsitsiklis (2010), Filippi et al. (2010) for the different variants of \( \text{CI}_t(\cdot, \cdot) \) forms in linear and generalized linear contextual bandits.

While the UCB policy naturally constructs “upper confidence bounds,” such constructed confidence bounds are inadequate for the use of predicting reasonable demand ranges because the upper confidence bound gives too wide intervals to be useful. In fact, confidence bounds in UCB are constructed using concentration inequalities, in which the constants are far from tight. Given the pre-specified confidence level \( 1 - \alpha \), our goal is to construct demand confidence intervals that have statistically accurate coverage as defined in Equations (4) and (5), allowing potential users to understand exactly the range of expected demands at certain confidence levels.

### 1.3. Related Works

Data-driven sequential decision-making has been extensively studied for revenue and inventory management problems with unknown or changing environments. In most existing literature, effective online policies are developed to maximize revenues. However, how to provide accurate confidence intervals for the key underlying probabilistic model parameters (e.g., demand function or utility parameters) have not been well explored in the literature. Recently, the work of Ban (2020) considered the construction of confidence intervals (for the demand functions) in an inventory control model. Compared to approaches proposed in this study, the work of Ban (2020) derives asymptotic normality of certain SAA strategies, while our approach de-biases general empirical-risk minimizers so that the constructed confidence intervals are applicable to a wide range of online policies, such as \( \varepsilon \)-greedy, upper confidence bounds or Thompson sampling. Technically, the limiting distributions in Ban (2020) were established using Stein’s methods, while our proposed approach is inspired by the one-step estimators in asymptotic statistics (Van der Vaart 2000).

Recently, the de-biased estimator has been extensively investigated in high-dimensional penalized estimators (Javanmard and Montanari 2014, Van de Geer et al. 2014, Wang et al. 2019, Zhang and Zhang 2014) since the regularization (e.g., \( \ell_1 \)-penalty in Lasso (Tibshirani 1996)) leads to the bias in the estimator. However, these works only deal with non-adaptively collected data and thus cannot be applied to our setting. The recent works of Deshpande et al. (2018, 2019) applied the de-biasing approach to confidence intervals of adaptively collected data, including multi-armed and linear contextual bandit problems. While the works of Deshpande et al. (2018, 2019) mainly focus on linear models, this study provides confidence intervals for general parametric models \( f(p, x; \theta) \). The extension to general parametric model classes poses some unique technical challenges, such as the sequential estimation of Fisher’s information matrix. In fact, the sequential estimation of Fisher’s information matrix is vital to the de-biasing procedure for generalized linear demand models, because the Fisher’s information matrix characterizes the way the estimated model \( \theta \) deviates from the true model parameter \( \theta_0 \). Unlike the pure linear model as studied in Deshpande et al. (2018, 2019), in most generalized linear demand models (including the most popular Logistic regression model), the Fisher’s information matrix depends on the true model parameter \( \theta_0 \). As a result, new methods need to be designed to accurately
estimate the information matrix in order to arrive at accurate confidence intervals. Further details are given in our section 4.

1.4. Notations and Paper Organization
Throughout this study we adopt the following asymptotic notations. For sequences \( \{a_n\} \) and \( \{b_n\} \), we write \( a_n = O(b_n) \) or \( b_n = \Omega(a_n) \) if \( \limsup_{n \to \infty} |a_n| / |b_n| < \infty \); we write \( a_n = o(b_n) \) or \( b_n = \omega(a_n) \) if \( \lim_{n \to \infty} |a_n| / |b_n| = 0 \).

The rest of the paper is organized as follows: in section 2 we list the assumptions made in this study, including discussion on why the imposed assumptions are useful and relevant; in section 3 we review the classical approach of Wald’s intervals for constructing confidence intervals, and explain why such a classical approach fails in contextual dynamic pricing problems; in section 4 we propose the de-biased approach and demonstrate, through both theoretical and empirical analysis, that our proposed confidence intervals are accurate in dynamic pricing. Finally, in section 5 we conclude the paper by mentioning several future directions for research. Proofs of some technical lemmas are deferred to the supplementary material.

2. Models and Assumptions
In this section we state assumptions that will be imposed throughout of this study. Most of the assumptions are standard in the literature of dynamic pricing or contextual bandits. There are, however, a few additional assumptions for the specific purposes of building accurate confidence intervals, which are often made in statistical literature.

2.1. Assumptions on the Demand Model \( f \)
We first list assumptions on the underlying demand function \( f \) (i.e., the mean of the demand), as well as assumptions on the underlying true parameter \( \theta_0 \).

(A1) For \( t = 1, \ldots, T \), \( p_t \in [p_{\min}, p_{\max}] \) and \( x_t \in \mathbb{X} \subseteq \mathbb{R}^d \) for some compact \( \mathbb{X} \), and \( \theta_0 \in \mathbb{\Theta} \subseteq \mathbb{R}^d \) for some known compact parameter class \( \mathbb{\Theta} \).

(A2) The demand function \( f \) is twice continuously differentiable with respect to \( x \), and furthermore \( f(p, x; \theta), \norm{\nabla_x f(p, x; \theta)}_2, \norm{\nabla_{\theta \theta} f(p, x; \theta)}_{\infty} \) are bounded and cannot be arbitrarily large. The two examples \( f(p, x; \theta) = \langle \phi(p, x), \theta \rangle \) (linear regression model) and \( f(p, x; \theta) = \exp \{ \langle \phi(p, x), \theta \rangle \} / (1 + \exp \{ \langle \phi(p, x), \theta \rangle \} \) (logistic regression model) satisfy both conditions, provided that the feature map \( \phi(p, x) \) is bounded.

2.2. Assumptions on the Noise Variables \( \{\xi_t\} \)
Recall that the noise variable \( \xi_t \) is defined as
\[
\xi_t := d_t - \mathbb{E}[d_t | x_t, p_t; \theta_0] = d_t - f(p_t, x_t; \theta_0),
\]
which is the difference between the realized demand and its (conditional) expectation. We list assumptions on the noise variables \( \{\xi_t\}_{t=1}^T \) across the \( T \) selling periods.

(B1) \( \{\xi_t\}_{t=1}^T \) are independent, centered, and bounded sub-Gaussian random variables;

(B2) There exists a known variance function \( \nu(\cdot; \theta) \) such that
\[
\mathbb{E}[\xi_t^2 | p_t, x_t] = \nu(p_t, x_t; \theta_0)^2,
\]
\( \nu(p, x, \theta) < \infty \) for all \( p, x, \theta \in \mathbb{\Theta} \) and Lipschitz continuous with respect to \( \theta \); \( 0 < \inf_{p, x} \nu(p, x; \theta_0) \leq \sup_{p, x} \nu(p, x; \theta_0) < \infty \).

In the above assumptions, (B1) is a standard assumption that the noise variables are all centered and sub-Gaussian with light tails, conditioned on the offered price \( p_t \) and the context vector \( x_t \). (B2) imposes further assumptions on the variance of the noise variables. In particular, it assumes that the conditional variance of \( \xi_t \) (conditioned on \( p_t \) and \( x_t \)) is bounded, never zero, and smooth. Such an assumption is useful in demand models \( f \) which are inherently heteroscedastic. For example, in the logistic demand model where \( d_t \in \{0, 1\} \) is a Bernoulli variable with \( \Pr[d_t = 1 | p_t, x_t, \theta] = f(p_t, x_t; \theta) = \exp \{ \langle \phi(p_t, x_t), \theta \rangle \} / (1 + \exp \{ \langle \phi(p_t, x_t), \theta \rangle \} \}, \) it is easy to verify that \( \nu^2(p_t, x_t; \theta) = \exp \{ \langle \phi(p_t, x_t), \theta \rangle \} / (1 + \exp \{ \langle \phi(p_t, x_t), \theta \rangle \} )^2 \), and all conditions in Assumption (B2) hold true. It is worthwhile noting that our Assumption (B2) also allows for the demand model to have stochastic demand noises that interact with the posted prices. For example, consider a demand model with an explicit interaction term \( d = \phi \theta_0 + px \) for \( \mathbb{E}[x] = 0 \) and \( \mathbb{E}[x^2] = 0.5 \). In this example, with the definition \( \xi = px \), it holds that \( \mathbb{E}[\xi^2 x] = 0 \) and \( \mathbb{E}[\xi^2] = \nu(p, x; \theta_0)^2 \), with \( \nu(p, x; \theta_0)^2 = 0.5p^2 \) depending on the posted price \( p \).

2.3. Assumptions on the Risk Function \( \rho \)
The empirical risk minimization problem in Equation (6) is the workhorse of our model estimates \( \hat{\theta} \). As discussed, popular risk functions \( \rho \) include the least-squares loss function \( \rho(d_t, p_t, x_t; \theta) = (d_t - f(p_t, x_t; \theta))^2 \) and the negative log-likelihood function \( \rho(d_t, p_t, x_t; \theta) = - \log(\mathbb{P}(d_t | p_t, x_t)) \). Below we give a list of assumptions imposed on the risk function \( \rho \) so that the ERM estimates satisfy desired properties.
(C1) The risk function $\rho$ is three times continuously differentiable with respect to $\theta$, and furthermore $|\rho(d,p,x;\theta)|, ||\nabla_0 \rho(d,p,x;\theta)||_{L_1}, \|\nabla_2 \rho(d,p,x;\theta)\|_{L_2}, \|\nabla_3 \rho(d,p,x;\theta)\|_{L_3} < \infty$ for all $d, p, x$ and $\theta$.

(C2) For all $p, x$, $E_{\theta} = \rho(p,x,\theta)|\nabla_0 \rho(p,x;\theta)| = 0$.

Here in Assumption (C1), $\nabla^3 \rho \theta$ is a symmetric $dx \times dx$ tensor, and its operator norm $\|\nabla^3 \rho \theta\|_{L_3}$ is defined as $\|\nabla^3 \rho \theta\|_{L_3} = \sup_{d:|z| \leq 1} (\nabla^3 \rho \theta(z,z,z))$. For the linear demand model and least-squares losses $\rho(d,p,x;\theta) = (d - \langle \phi(p,x), \theta \rangle)^2$, Assumption (C1) is implied by the boundedness of $\phi(p,x)$; for other parametric models (e.g., the logistic regression model) and the negative log-likelihood loss $\rho(d,p,x;\theta) = -\log P(d,p,x;\theta)$, Assumption (C1) are standard conditions used in the analysis of maximum likelihood estimator. Finally, Assumption (C2) means that the true model parameter $\theta_0$ is a stationary point of the loss function $\rho$, which is satisfied by both the least-squares loss function and the negative log-likelihood loss function. In statistical literature, $\nabla_0 \rho = -\nabla_0 \log P$ is known as (the negative of) the score function, whose expectation is zero under $\theta_0$.

2.4. Assumptions on the Contextual Pricing Model

At last, we state an assumption on the behavior of the contexts $\{x_t\}_{t=1}^T$ under the contextual pricing model $\Omega$.

(D1) There exists a positive constant $\kappa_0 > 0$ such that, for any selling period $t$ and filtration $\mathcal{F}_{t-1} = \{x_{t-1}, p_{t-1}, d_{t-1}\}$, it holds that $\lambda_{\min}(\mathbb{E}_{x_t \sim \xi_t(\theta)} [\nabla_0 \rho(d_t,p_t,x_t;\theta) \nabla_0 \rho(d_t,p_t,x_t;\theta)^T]) \geq \kappa_0$ and $\lambda_{\min}(\mathbb{E}_{x_t \sim \xi_t(\theta)} [\nabla_2 \rho \theta(d_t,p_t,x_t;\theta)]) \geq \kappa_0$ for all $d, p$ and $\theta$, which could potentially depend on $x_t$.

Assumption (D1) concerns two quantities: the (expected) outer product of demand gradients $\nabla_0 \nabla_0^T$, which by definition is always positive semi-definite, and the (expected) Hessian of the loss function $\nabla^2 \rho \theta$, which can theoretically be any symmetric matrix but is in general positive semi-definite for common loss functions like the least squares or negative log-likelihoods. Assumption (D1) then assumes, essentially, that both quantities $\mathbb{E}[\nabla_0 \nabla_0^T]$ and $\mathbb{E}[\nabla^2 \rho \theta]$ are positive definite in a “strict” sense, by lower bounding the least eigenvalues of both $\mathbb{E}[\nabla_0 \nabla_0^T]$ and $\mathbb{E}[\nabla^2 \rho \theta]$ by a positive constant $\kappa_0$. Since both expectations are conditioned upon the adaptively chosen prices $\{p_t\}$ and context vectors $\{x_t\}$, in Assumption (D1) we assume that the lower bound on the smallest eigenvalues holds for any such chosen prices-contexts in prior selling periods. Finally, we remark that the exact value of $\kappa_0$ does not need to be known, as it is only used in the theoretical analysis of the validity of confidence intervals constructed by our proposed algorithm.

We also remark that Assumption (D1) is essentially the only assumption we impose on the filtration and pricing process in this study, making the uncertainty quantification protocol proposed in this study applicable to a wide range of dynamic personalized pricing algorithms, such as the upper-confidence bound (UCB) and the Thompson sampling algorithms.

3. Limitation of Classical Wald’s Intervals

In classical parametric statistics with i.i.d. data points, the Wald’s interval is a standard approach toward building asymptotic estimation or confidence intervals on maximum likelihood estimates. In this section, we review the approach of Wald’s interval in the context of contextual dynamic pricing, and discuss why such a classical method cannot be directly applied because of the feedback structures presented in our problem.

Suppose after $T$ selling periods the offered prices, purchase activities and customers’ context vectors are $\{(p_t,d_t,x_t)\}_{t=1}^T$. Let $\hat{\theta}$ be the maximum likelihood estimate

$$\hat{\theta} = \arg \min_\theta - \sum_{t=1}^T \log P(d_t|x_t,p_t;\theta),$$

(10)

which is equivalent to Equation (6) with $\lambda = 0$ and $\rho(d_t,x_t,p_t,\theta) = -\log P(d_t|x_t,p_t;\theta)$. Using classical statistics theory (see, e.g., Van der Vaart 2000), if $(d_t,x_t,p_t)$ are statistically independent, then under mild regularity conditions it holds that

$$\left[\hat{I}_T(\hat{\theta})\right]^{1/2} (\hat{\theta} - \theta_0) \overset{d}{\rightarrow} \mathcal{N}(0, I_{d \times d}), \quad T \rightarrow \infty,$$

(11)

where $\hat{I}_T(\theta) = -\sum_{t=1}^T \nabla_0 \rho \theta(d_t|x_t,p_t;\theta)$ is the sample Fisher’s information matrix. With Equation (11), using the Delta’s method we have for fixed $p, x$

$$f(p,x;\theta) = f(p,x;\theta_0) \overset{d}{\rightarrow} \mathcal{N}(0, \sigma_\theta^2)$$

(12)

where $\sigma_\theta^2 = \nabla_0 f(p,x;\hat{\theta}) [\hat{I}_T(\hat{\theta})]^{-1} \nabla_0 f(p,x;\hat{\theta})$.

A confidence interval on $f(p,x;\theta_0)$ can then be constructed as

$$\ell_{\alpha}^{\text{classical}}(p,x) = f(p,x;\hat{\theta}) - \frac{z_{\alpha/2} \sigma_{px}}{\sqrt{T}}$$

(13)
where $z_{a/2} = \Phi^{-1}(1-a/2)$ is the $(1-a/2)$-quantile of a standard normal random variable $Z \sim N(0,1)$ and $\Phi(\cdot)$ denotes the cumulative distribution function of $Z$, that is, $\Pr(Z > z_{a/2}) = a/2$.

While the Wald’s interval is a general purpose and the most classical approach of constructing confidence intervals, one of the key assumptions made in the construction of the Wald’s interval is the statistical independence among the collected data $\{(p_t, x_t, d_t)\}_{t=1}^T$ across selling periods $t = 1, \ldots, T$. It is known that, without such independence assumptions, the Wald’s interval could be significantly biased, as in the case of multi-armed bandit predictions (Deshpande et al., 2018) and least-squares estimation in non-mixing time series (Lai and Wei, 1982).

To better illustrate the failure of Wald’s test for adaptively collected data, in Figure 1, we plot the empirical distributions of the normalized estimation and prediction errors of predicted demands $|f(p, x; \theta) - f(p, x; \theta_0)|/\sigma_{px}$ for the Wald’s interval approach. In particular, we consider the simple logistic demand model $f(p, x; \theta_0) = \phi(p, x)\theta_0/(1 + \phi(p, x)\theta_0)$, with $d = 2$, $\phi(p, x) = (0.9 + 0.1p, x_1)$ and $\theta_0 = (-1, 1)$. The price range is $p \in [0,1]$. The context generating process $\mathcal{C}_t$ is designed as $x_{t+1} = z_{t+1}/\max(1, |z_{t+1}|)$, where $z_{t+1} = z_t + d_t - f(p_t, x_t; \theta_0)$ and $z_1 = 0$. The empirical distributions are obtained with 5000 independent trials, each with $T = 10000$ selling periods and prices determined by the LinUCB algorithm as described in section 1.2. The top panels in Figure 1 depict the distributions of two coordinates of $\xi$ and the bottom panels are normalized demand prediction errors $|f(p, x; \theta) - f(p, x; \theta_0)|/\sigma_{px}$ for the cases of $p = 0.5, x = 0$; $p = 0.5, x = 1$; $p = 1, x = 1$, respectively. One can easily see that, in contextual dynamic pricing the confidence intervals constructed for both the estimation errors $\theta - \theta_0$ and the prediction error (of demands) $f(p, x; \theta) - f(p, x; \theta_0)$ deviate significantly from the desired limiting distributions $\mathcal{N}(0, I)$ (see (11)) and $\mathcal{N}(0, 1)$ (see (12)), calling for more sophisticated methods to construct accurate confidence intervals.

**4. Main Algorithm and Analysis**

The pseudo-code of our proposed algorithm for constructing confidence intervals of the demand function $f$ is given in Algorithm 1.
Algorithm 1: The main algorithm for constructing demand confidence intervals

1. **Input:** prices, purchases and contexts over \( T \) selling periods \((p_0, n, x_i)\) from \( \mathbb{E}^n \);
2. Compute a “pilot” estimate \( \hat{\theta} \) using the ERM in Eq. (9) with \( \lambda = 0 \);
3. Compute the “whitening” matrix \( W \in \mathbb{R}^{d 	imes T} \) by invoking the WHITEN procedure in Algorithm 2 in Sec. 4.3.
4. Compute the de-biased estimate \( \hat{\theta} = \hat{\theta} + W(d - \hat{f}) \), where \( d = (d_1, \ldots, d_T) \) and \( \hat{f} = (f(p_1, x_1), \ldots, f(p_T, x_T)) \)
5. Construction of point-wise confidence intervals for fixed \( p, x \) and confidence level \( 1 - \alpha \), construct the point-wise confidence interval

\[
\left[ \left( \hat{\theta}^{\text{lower}}(p,x), \hat{\theta}^{\text{upper}}(p,x) \right) = \left[ f(p,x) \pm \epsilon_x \right] \right]
\]

where \( \hat{\theta}^{\text{lower}}(p,x) = f(p,x) - \epsilon_x \) and \( \hat{\theta}^{\text{upper}}(p,x) = f(p,x) + \epsilon_x \) (\( \epsilon_x \) is defined in (3)).

6. Construction of uniform confidence intervals first obtain \( M \) independent Monte-Carlo samples of \((g_1, \ldots, g_M) \sim \mathbb{N}(0, W(W^T))^T\); for every \( m \) \((1 \leq m \leq M)\) compute

\[
\left( \hat{\theta}^{\text{lower}}(p,x), \hat{\theta}^{\text{upper}}(p,x) \right) = f(p,x) \pm \epsilon_x
\]

where \( \hat{\theta}^{\text{lower}}(p,x) = f(p,x) - \epsilon_x \) and \( \hat{\theta}^{\text{upper}}(p,x) = f(p,x) + \epsilon_x \) (\( \epsilon_x \) is defined in (3)).

At a high level, the objective of Algorithm 1 is to construct accurate confidence intervals in both the “point-wise” sense (i.e., confidence intervals for the expected demand \( f(p,x) \theta_0 \)) (4) for a single price \( p \) and context \( x \) and the “uniform” sense (i.e., confidence intervals in (5) for \( f(p,x) \theta_0 \)) that hold uniformly (i.e., over all possible prices and contexts). The input to Algorithm 1 is the historical price, context, and demand data over \( T \) selling periods, during which an adaptive dynamic pricing strategy is used. The adaptivity of the pricing strategy means that the demands and prices are highly correlated, and therefore the basic Wald’s intervals cannot be directly applied, as discussed in the previous section.

The key idea behind our proposed approach is the idea of “de-biasing” the empirical risk estimate \( \hat{\theta} \) (also termed as the “pilot” estimate in Algorithm 1). More specifically, built upon the biased pilot estimate \( \hat{\theta} \), we construct a \( d \times T \) “whitening” matrix \( W \) satisfying certain correlation and norm conditions (the procedure of constructing such a whitening matrix is presented in Algorithm 2 and section 4.3), a de-biased estimate \( \hat{\theta} \) is computed by adding the bias-correction term \( W(d - \hat{f}) \) to the ERM estimate \( \hat{\theta} \), or more specifically

\[
\hat{\theta} = \hat{\theta} + W(d - \hat{f}), \tag{14}
\]

where \( d = (d_1, \ldots, d_T) \), \( \hat{f} = (f(p_1, x_1), \ldots, f(p_T, x_T)) \).

For example, in the linear demand case, \( \hat{f} = (\hat{\phi}(p_1, x_1), \ldots, \hat{\phi}(p_T, x_T)) \), while in the logistic case,

\[
\hat{f} = \left( \frac{\exp(\phi(p_1, x_1), \hat{\theta})}{1 - \exp(\phi(p_1, x_1), \hat{\theta})}, \ldots, \frac{\exp(\phi(p_T, x_T), \hat{\theta})}{1 - \exp(\phi(p_T, x_T), \hat{\theta})} \right).
\]

With the bias correction, it can be proved that the bias contained in \( \hat{\theta} \) can be dominated by the main error terms that are asymptotically normal, as shown in Theorem 1 later. With the asymptotic normality of \( \hat{\theta} - \theta_0 \), both point-wise and uniform confidence intervals can be constructed using either the Delta’s method in Equation (12) or Monte-Carlo methods, as shown in Steps 5 and 6 of Algorithm 1.

In the rest of this section we provide a rigorous analysis of the proposed confidence intervals in Algorithm 1. In section 4.1, we perform a bias-variance decomposition analysis of the debiased estimate \( \hat{\theta} \) and prove in Theorem 1 that, under certain conditions, \( \hat{\theta} - \theta_0 \) is asymptotically normally distributed; in section 4.2 we upper bound the estimation error of the pilot estimate \( \hat{\theta} \), and in section 4.3 we propose a procedure of constructing the whitening matrix \( W \in \mathbb{R}^{d \times T} \) such that the conditions in Theorem 1 are satisfied. Finally, in section 4.4 we prove that both point-wise and uniform confidence intervals are asymptotically level-(1 - \( \alpha \)) theoretically establishing the accuracy of constructed intervals.

4.1. Analysis of the De-Biased Estimator

In Step 4 of Algorithm 1, a de-biased estimate \( \hat{\theta} \) is constructed based on the biased ERM estimate \( \hat{\theta} \) and a certain “whitening matrix” \( W \in \mathbb{R}^{d \times T} \). In this section we analyze the asymptotic distributional properties of \( \hat{\theta} \) based on certain conditions on \( W \). The question of how to obtain a whitening matrix \( W \) satisfying the desired conditions will be discussed in the next section.

For notational simplicity, we denote the gradient at time \( t \) by \( \nabla \theta_t := \nabla \theta_t(p_t, \hat{x}_t; \hat{\theta}) \in \mathbb{R}^{d} \) and \( G_t := (g_1, \ldots, g_T) \in \mathbb{R}^{d}. \) Also recall the definition of \( \xi_t \) for \( t = 1, \ldots, T \) in Equation (8). The following lemma shows a bias-variance decomposition \( \hat{\theta} - \theta_0 \).

**Lemma 1.** The estimation error \( \hat{\theta} - \theta_0 \) can be decomposed to \( \hat{\theta} - \theta_0 = b + v \), where the bias term \( b \) satisfies \( ||b||_2 \leq ||I_d - WG||_{op} ||\hat{\theta} - \theta_0||_2 + O(||\hat{\theta} - \theta_0||^2) \) almost surely and the variance \( v = W \xi, \xi = (\xi_1, \ldots, \xi_T) \in \mathbb{R}^T \).

**Proof of Lemma 1.** Recall the definition that \( \hat{\theta} = \hat{\theta} + W(d - \hat{f}) \), where \( d = (d_1, \ldots, d_T) \in \mathbb{R}^T \) and \( \hat{f} = (f(p_1, x_1; \hat{\theta}), \ldots, f(p_T, x_T; \hat{\theta})) \in \mathbb{R}^T \). Define also \( f = (f(p_1, x_1; \theta_0), \ldots, f(p_T, x_T; \theta_0)) \). By definition, \( d = f + \xi. \) Subsequently,

\[
\hat{\theta} - \theta_0 = \hat{\theta} - \theta_0 + W(f - \hat{f}) + W(d - f) = \hat{\theta} - \theta_0 + W(f - \hat{f}) + W \xi.
\]

Next, by Taylor expansion and the smoothness of \( f \) (see Assumption (2)), we have for every \( t \) that \( f(p_t, x_t; \hat{\theta}) - f(p_t, x_t; \theta_0) = \langle \nabla \theta f(p_t, x_t; \theta), \hat{\theta} - \theta_0 \rangle + O(||\hat{\theta} - \theta_0||^2) \). Hence, \( f - f = G(\hat{\theta} - \theta_0) + O(||\hat{\theta} - \theta_0||^2) \). We then have
\[ \hat{\theta}^d - \theta_0 = (I - W G)(\hat{\theta}^d - \theta_0) + O(||\hat{\theta}^d - \theta_0||_2^2) + W \xi, \]

which completes the proof.

Our next lemma shows that, when the bias term \( b \) is sufficiently small, the error \( \hat{\theta}^d - \theta_0 \) converges in distribution to a multivariate Gaussian distribution.

**Theorem 1.** Suppose the following conditions hold:

1. The non-anticipativity condition: the \( t \)-th column of \( W_t \), \( \omega_t \), is measurable conditioned on \( \{\xi_r, p_r, d_r, x_r, \omega_r\}_{r < t} \cup \{x_t, p_t\} \);
2. \( \mathbb{E}[\sum_{t=1}^T ||\omega_t||_2^2] \rightarrow 0 \) as \( T \rightarrow \infty \);
3. Let \( D = \text{diag}(\nu^2) \in \mathbb{R}^{T \times T} \) be a diagonal matrix with \( \nu = (\nu(p_1, x_1; \theta_0), \ldots, \nu(p_T, x_T; \theta_0)) \in \mathbb{R}^T \);

\[
\max \{||I - W G||_\text{op}, ||\hat{\theta}^d - \theta_0||_2, ||\hat{\theta}^d - \theta_0||_2^2\} \nu, \text{as } T \rightarrow \infty.
\]

Then it holds that

\[
(W D W^T)^{-1/2} (\hat{\theta}^d - \theta_0) \overset{d}{\rightarrow} \mathcal{N}(0, I_d) \text{ as } T \rightarrow \infty. \tag{16}
\]

The key idea behind the proof of Theorem 1 mainly involves two steps. The first step is to show that, under the non-anticipativity conditions imposed on \( W \), the variance term \( W \xi \) converges in distribution to a normal distribution using martingale CLT type arguments. The second step shows that the bias term \( b \) is asymptotically dominated by \( W \xi \), and therefore the entire estimation error \( \hat{\theta}^d - \theta_0 \) converges in distribution to a normal distribution. The complete proof is given below.

**Proof of Theorem 1.** Adopt the decomposition of \( \hat{\theta}^d - \theta_0 \) in Lemma 1. By definition, \( \nu = W \xi = \sum_{t=1}^T \xi_t \omega_t \).

For every \( t \leq T \) define \( S_t := \sum_{r=1}^t \xi_r \omega_r \) and \( S_0 := 0 \).

Because \( \mathbb{E}[\xi_r | \omega_r, \xi_{r-1}, \omega_{r-1}, \ldots, \xi_1, \omega_1] = \mathbb{E}[\xi_r | p_r, \omega_t] = 0 \) by the non-anticipativity condition, we know that \( \{S_t - S_{t-1}\}_t \) is a martingale. The following lemma shows how the characteristic functions of \( \{S_t\} \) converge to the characteristic function of \( \mathcal{N}(0, W D W^T) \).

**Lemma 2.** Let \( z \sim \mathcal{N}(0, I_d) \) be a fresh sample from the standard \( d \)-dimensional Gaussian distribution. Define also \( \tilde{v} := (W D W^T)^{-1/2} v \). Then for any \( a \in \mathbb{R}^d \), \( ||a||_2 \leq 1 \), it holds that

\[
\mathbb{E}[\exp\{ia^T \tilde{v}\}] = \exp\{-||a||_2^2/2\} \leq \mathbb{E} \left[ \sum_{t=1}^T O(||\omega_t||_2^2) \right].
\]

The proof of Lemma 2 is based on standard Fourier-analytic approaches (Billingsley 2008, Brown 1971, Lai and Wei 1982), and is deferred to the supplementary material. Lemma 2 shows that the characteristic function of \( \tilde{v} := (W D W^T)^{-1/2} v \) converges point-wise to the characteristic function of \( z \sim \mathcal{N}(0, I_d) \), provided that \( \mathbb{E}[\sum_{t=1}^T ||\omega_t||_2^2] \rightarrow 0 \) as \( T \rightarrow \infty \).

By Levy’s continuity theorem, this implies \( \tilde{v} \overset{d}{\rightarrow} \mathcal{N}(0, I_d) \), or more specifically

\[
(W D W^T)^{-1/2} v \overset{d}{\rightarrow} \mathcal{N}(0, I_d). \tag{18}
\]

Because \( \text{tr}(W D W^T)/d \geq a_{\min}(W D W^T) \) and \( d \) is treated as a constant in this study, the third condition in Theorem 1 would imply that \( ||\hat{b}||^2/\text{tr}(W D W^T) \overset{p}{\rightarrow} 0 \) as \( T \rightarrow \infty \). This implies that

\[
(W D W^T)^{-1/2} [/2(||\hat{\theta}^d - \theta_0 - v||^2_2) \overset{p}{\rightarrow} 0 \text{ as } T \rightarrow \infty. \]

Consequently, \( (W D W^T)^{-1/2} (\hat{\theta}^d - \theta_0) \overset{d}{\rightarrow} \mathcal{N}(0, I_d) \) by Slutsky’s theorem.

With Theorem 1 demonstrating the asymptotic normality of \( \hat{\theta}^d - \theta_0 \), it is easy to derive the asymptotic normality of the demand prediction error \( f(p, x; \hat{\theta}^d) - f(p, x; \theta_0) \) as well. More specifically, we have the following result:
COROLLARY 1. Let \( p, x \) be fixed and all conditions in Theorem 1 are satisfied. Suppose also that all assumptions listed in section 2 hold. Then we have that

\[
\left| f(p, x; \hat{\theta}^d) - f(p, x; \theta_0) \right| / \hat{\sigma}_{px} \xrightarrow{d} \mathcal{N}(0, 1),
\]

where \( \hat{\sigma}_{px} \) is defined in Step 5 of Algorithm 1.

The proof of Corollary 1 is quite standard, by using local Taylor expansions at \( f(p, x; \theta_0) \) and invoking Slutsky’s theorem. For completeness, we give the proof of Corollary 1 in the supplementary materials.

To illustrate the normality of the de-biased estimator \( \hat{\theta}^d \) and the corresponding predicted demand function \( f(p, x; \hat{\theta}^d) \), Figure 2 plots the empirical distributions of the (normalized) estimation errors and demand prediction errors based on \( \hat{\theta}^d \). Apart from the difference in model and variance estimates, the plots in Figure 2 and those in Figure 1 are produced using exactly the same experimental and model parameter settings. Comparing Figure 2 against Figure 1, we can see that the empirical distributions of the errors of de-biased estimates align much more closely with the desired limiting distributions \( \mathcal{N}(0, I_t) \) and \( \mathcal{N}(0, 1) \), and there is no significant deviates in either high-density or tail regions. This justifies the validity of confidence intervals constructed using \( \hat{\theta}^d \), as we shall discuss in details in section 4.4 later.

4.2. Analysis of the Pilot Estimate \( \hat{\theta}^P \)

From the conditions listed in Theorem 1 in the previous section (see item 3), it is essential to upper bound the deviation of \( \hat{\theta}^P \) from the true underlying model \( \theta_0 \). In this section we analyze how close the pilot estimate \( \hat{\theta}^P \) is from the underlying true model \( \theta_0 \) in terms of \( \| \hat{\theta}^P - \theta_0 \|_2 \). We will prove a more general result applicable to the empirical risk minimizer (ERM) at any time period \( t \). More specifically, for every \( t \) we define

\[
\hat{\theta}^P_t := \arg \min_{\theta \in \Theta} \sum_{t' < t} \rho(d_{t'}, p_{t'}, x_{t'}; \theta),
\]

as the ERM on the data collected during time periods prior to \( t \). Clearly, our target \( \hat{\theta}^P = \hat{\theta}^P_{T+1} \).

LEMMA 3. Suppose all assumptions in section 2 hold. Then for any \( t \geq d \log d \), it holds that \( \| \hat{\theta}^P_t - \theta_0 \|_2 = O_P(\sqrt{d \log t / t}) \).

Figure 2  Empirical Distributions of the Normalized Estimation and Prediction Errors from the De-Biased Approach in Algorithm 1. The Experimental Setting is Identical to the One in Figure 1 [Color figure can be viewed at wileyonlinelibrary.com]

(a) Empirical distributions of \( \varepsilon_1, \varepsilon_2 \), where \( \varepsilon = (\varepsilon_1, \varepsilon_2) = (WDW^T)^{-1/2}(\hat{\theta}^d - \theta_0) \).

(b) Empirical distributions of \( |f(p, x; \hat{\theta}^d) - f(p, x; \theta_0)| / \hat{\sigma}_{px} \), where (from left to right) the price and contexts are \( (p, x) = (0.5, 0), (0.5, 1), (1, 1) \), respectively.
At a higher level, Lemma 3 establishes the error upper bounds for the pilot estimator $\hat{\theta}_t^p$. The error bound is subsequently used in analyzing the de-biased estimator $\hat{\theta}_t^d$. The proof of Lemma 3 is based on the standard argument of self-normalized martingale empirical processes and its applications in online contextual bandits, see, for example, the works of Rusmevichientong and Tsitsiklis (2010), Abbasi-Yadkori et al. (2012), Filippi et al. (2010). Concentration inequalities for matrix martingales are also involved (Tropp 2012). We defer the complete technical proof to the supplementary material.

4.3. The Whitening Procedure

The de-biased estimate $\hat{\theta}_t^d$ is constructed using a "whitening" matrix $W \in \mathbb{R}^{d \times T}$ to counteract the bias inherent in the pilot ERM estimator $\hat{\theta}_t$. The conditions in Theorem 1 suggest that $W$ needs to satisfy three properties:

1. $W = (w_1, \ldots, w_T)$ should be constructed such that $w_t \mid p_t, x_t, \tilde{z}_{t-1}$ is measurable, where $\tilde{z}_{t-1} = \{(d'_t, p_t, x_t)\}_{t < j}$; in other words, the computation of $w_t$ should only involve $\tilde{z}_{t-1}$ and $p_t, x_t$;

2. The norms of each column of $W$, or $\|w_t\|_2$, should be relatively evenly distributed, so that $\mathbb{E}[^{\sum_{t=1}^T\|w_t\|_2^2}_T] \rightarrow 0$ holds; (Because $W$ should be close to $I$, the more evenly distributed $\{w_t\}$ are the more likely the sum of their cubic norms is small, as in the AM-GM inequality.).

3. $W$ should be as close to $I_{d \times d}$ as possible, in order to fix the bias in $\hat{\theta}_t^d$.

Algorithm 2: The whitening procedure.

```
1: Input: historical data $\{(x_t, d_t, w_t)\}_{t=1}^T$, incremental parameter $\eta = T^{-1}$, $v \in (1/2, 1)$;
2: Initialize: $Z = I_{d \times d}$;
3: for $t = 1, 2, \ldots, T$ do:
4:     Compute $\hat{\theta}_t = \arg\min_{\theta} \sum_{t'<t} p(t', p_t, x_t, \theta);
5:     Compute $w_t = (E\nabla f(p_t, x_t, \hat{\theta}))/(E\nabla^2 f(p_t, x_t, \hat{\theta}))$;
6:     If $\|w_t\|_2 > \eta$ then normalize $w_t \leftarrow \eta \|w_t\|_2$;
7:     Update $Z \leftarrow Z - w_t (E\nabla f(p_t, x_t, \hat{\theta}))^T$;
8: end
9: Output: the whitening matrix $W = (w_1, \ldots, w_T) \in \mathbb{R}^{d \times T}$.
```

Our procedure of constructing the whitening matrix $W$ is outlined in Algorithm 2. Now we provide the intuition behind Algorithm 2. For the ease of discussion, let us pretend for now that $\hat{\theta}_t^d \equiv \theta_0$, which implies that $\nabla f(p_t, x_t, \theta_0) \approx \nabla f(p_t, x_t, \theta_0) = g_t$ (i.e., the $t$-th row of the matrix $G \in \mathbb{R}^{T \times d}$). Intuitively, to find a whitening matrix $W \in \mathbb{R}^{d \times T}$ such that $\|I_d - WG\|_{op}$ is as small as possible (see (16)), one simply sets $W = G^\dagger$, the Moore–Penrose pseudo-inverse of $G$. Since $T \gg d$, we know that $WG$ is precisely $I_d$ if $G$ has full column ranks.

Such an approach, however, violates the first two conditions in Theorem 1. First, because $W = G^\dagger$ depends on the entire matrix $G$, the $t$-th column of $W$, $w_t$, may not be measurable under the filtration of prior history (i.e., utilizing the information from later time periods). Furthermore, the columns of $W = G^\dagger$ might be particularly large if $G$ is ill-conditioned, jeopardizing the $\mathbb{E}[^{\sum_{t=1}^T\|w_t\|_2^2}] \rightarrow 0$ condition in Theorem 1.

To address the above-mentioned challenges, one cannot simply set $W = G^\dagger$ but must construct or optimize such a $W$ in a sequential way to avoid certain statistical correlation between $W$ and $G$. Starting from $Z_0 = I_{d \times d}$, at each time period $t$ the column $w_t$ of $W$ is constructed sequentially so as to satisfy both non-anticipativity and small-norm conditions. More specifically, let $Z_t := I_{d \times d} - \sum_{t' < t} w_t' G_{t'}^\dagger$ be the "remainder" of the identity matrix after the first $(t-1)$ time periods. Our objective is to reduce the norm of $Z_t$ as much as possible at each time period, so that $\|Z_{t+1}\|_{op}$ is close to zero. At time $t$, however, the constructed column $w_t$ cannot be computed using the previous time periods, and should not use any information from $\{G_{t'}\}_{t' < t}$ in order to satisfy the non-anticipativity condition in Theorem 1. Furthermore, the norm of $w_t$ should not be too large. Taking both constraints into consideration, the column $w_t$ could be computed as the optimal solution to the following constrained optimization problem:

$$w_t = \arg\min_{w \in \mathbb{R}^d} \|Z_t - wG_t^\dagger\|_{op} \text{ s.t. } \|w\|_2 \leq \eta,$$

where $\eta > 0$ is a small constant upper bounding the magnitude of $w_t$ (we will discuss the choice of $\eta$ in the next paragraph). It is easy to verify that, the solution to Equation (19) is precisely the $w_t$ computed in Algorithm 2. In particular, if the projection of $Z_t$ onto the direction of $g_t$, $Z_t g_t / \|g_t\|_2$, is small, then $w_t$ is simply the projection $Z_t g_t / \|g_t\|_2$ so that $\|Z_t - w_t g_t^\dagger\|_{op}$ is minimized. On the other hand, if $Z_t g_t / \|g_t\|_2$ is too large then the projection is again projected to the $\ell_2$ ball of radius $\eta$, so that $\|w_t\|_2 \leq \eta$ is always satisfied.

From the above discussion, the role of $\eta$ is important. If $\eta$ is too large, then the condition $\mathbb{E}[^{\sum_{t=1}^T\|w_t\|_2^2}] \rightarrow 0$ could be violated, invalidating the limiting distribution analysis in Theorem 1. More detailed calculations show that $\eta$ needs to satisfy $\eta = o(T^{-1/3})$ for $\mathbb{E}[^{\sum_{t=1}^T\|w_t\|_2^2}] \rightarrow 0$ to hold. On the other hand, if $\eta$ is too small then at the end $\|Z_{t+1}\|_{op}$ might be too large, violating the third condition in Theorem 1.
(by having a very large discrepancy \( \|I_d - WG\|_{op} \)).

More involved calculations (see, e.g., Corollary 2 below) show that \( \eta \) needs to satisfy \( \eta = \Omega(1/T) \) and \( \eta = o(T^{-1/2-\delta}) \) for \( \|I_d - WG\|_{op} \) to be sufficiently small. To summarize, we recommend the scaling of \( \eta = T^{-1/2-\delta} \) with \( \delta \in (1/2, 1) \). Our theoretical analysis shows that with \( \delta \in (1/2, 1) \) the main limiting distribution results will hold.

Our next lemma shows that, under our assumptions in section 2, the discrepancy \( \|I - WG\|_{op} \) can be effectively upper bounded when \( \eta \) is set appropriately.

**Lemma 4.** Suppose all assumptions made in section 2 hold, and \( \eta \) satisfies \( \eta \to \infty \). Then

\[
\|I_d - WG\|_{op} = O_p(\eta \sqrt{T}).
\]

The proof of Lemma 4 can be roughly divided into two steps: the first step is to prove that \( \|Z_{t+1}\|_{op} \) is sufficiently small under the assumed \( \eta \) scaling in Lemma 4, and the second step is to upper bound the discrepancy between \( G = (\nabla \sigma(p_t, x_t; \theta_0)) \), and its estimate \( \tilde{G} = (\nabla \sigma(p_t, x_t; \tilde{\theta}_t)) \). The complete proof is given below.

Proof of Lemma 4. For clarity we use the symbol \( \bar{w}_t \) for the vector computed at Step 5 of Algorithm 2, and \( \bar{v}_t \) for the normalized vector after Step 6 of Algorithm 2. For every \( t \leq T \) define \( u_t := \nabla \sigma(p_t, x_t; \tilde{\theta}_t) \) and \( Z_t := I_d - \sum_{t' < t} \bar{w}_{t'} u_{t'}^\top \), which coincides with the \( Z \) matrix at the beginning of iteration \( t \). According to Algorithm 2, \( w_t = (Z_t u_t) / \|u_t\|_2^2 \) is the projection of \( Z_t \) onto the direction of \( u_t \). Moreover, \( Z_{t+1} \) can be written as \( Z_{t+1} = Z_t - \bar{w}_t u_t^\top \), where \( \bar{w}_t = w_t \) if \( \|u_t\|_2 \leq \eta \) and \( \bar{w}_t = \eta_0 u_t / \|u_t\|_2 \) if \( \|u_t\|_2 > \eta \). Using the Pythagorean theorem we have that

\[
\|Z_t\|_2^2 - \|Z_{t+1}\|_2^2 = \|\bar{w}_t u_t^\top\|_2^2 = \|\bar{w}_t\|_2^2 \|u_t\|_2^2. \tag{20}
\]

Define

\[
R(Z_t, u_t) := \|Z_t u_t\|_2 / \|u_t\|_2 = \sqrt{u_t^\top (Z_t Z_t^\top) u_t / \|u_t\|_2^2},
\]

which is always between \( \sigma_{\min}(Z_t) \) and \( \sigma_{\max}(Z_t) \) (the smallest and largest singular values of \( Z_t \)). The case of \( \|u_t\|_2 > \eta \) corresponds to \( R(Z_t, u_t) / \|u_t\|_2 > \eta \). In this case, because \( \|\bar{w}_t\|_2 = \eta \) we have that

\[
\|Z_t\|_2^2 - \|Z_{t+1}\|_2^2 = \eta \|u_t\|_2^4 \times 1\{R(Z_t, u_t) > \eta \|u_t\|_2\}. \tag{21}
\]

Now let \( T_0 \leq T \) be the smallest integer such that \( R(Z_{T_0}, u_{T_0}) \leq \eta \|u_{T_0}\|_2 \). If such a \( T_0 \) exists, then

\[
\|Z_{T+1}\|_{op} \leq \|Z_{T_0}\|_{op} \leq R(Z_{T_0}, u_{T_0}) \leq \eta \|u_{T_0}\|_2 \leq O(\eta), \tag{22}
\]

where the first equality holds because the right-hand side of Equation (20) is always non-negative, and the last inequality holds thanks to Assumption (A2) that \( \|u_{T_0}\|_2 \) are bounded. We next show that such a \( T_0 \) always exists for sufficiently large \( T \). Assume the contrary. Then by telescoping both sides of Equation (21) from \( t = 1 \) to \( t = T \) we have

\[
\mathbb{E}[\|Z_{T+1}\|_2^2] \leq d - \eta \mathbb{E} \left[ \sum_{t=1}^T \mathbb{E}[\|u_t\|_2^2] \delta_{t-1, p_t, x_t, 1} \right] \leq d - \Omega(\eta T), \tag{23}
\]

where \( \delta_{t-1} = \{(d_{t'}, p_{t'}, x_{t'})\}_{t' < t} \) and the last inequality holds thanks to Assumption (D1). Since \( \eta T \to \infty \), Equation (23) suggests that \( \mathbb{E}[\|Z_{T+1}\|_2^2] < 0 \) for sufficiently large \( T \), which is the desired contradiction.

With Equation (22), it remains to upper bound the discrepancy between \( Z_{T+1} \) and \( I - WG \). By definition,

\[
Z_{T+1} = I - \sum_{t=1}^T \bar{w}_t \otimes \nabla \sigma(p_t, x_t, \tilde{\theta}_t), \quad \text{and} \quad I - WG = I - \sum_{t=1}^T \bar{w}_t \otimes \nabla \sigma(p_t, x_t, \theta_0'),\]

where \( a \otimes b \) means the outer product \( ab^\top \). Hence,

\[
\|Z_{T+1} - (I - WG)\| \leq \sum_{t=1}^T \|\bar{w}_t\|_2 \times O(\|\theta_t - \theta_0\|_2) \leq O(\eta \sqrt{T}). \tag{24}
\]

Combining Equations (22)-(24) we have

\[
\|I - WG\|_{op} \leq O(\eta + \eta \sqrt{T}) = O(\eta \sqrt{T}),
\]

which is to be demonstrated.

With Lemma 4 (and Lemma 3 for pilot estimator), it is easy to establish the following corollary showing that all conditions of Theorem 1 are satisfied with appropriate scaling of \( \eta \). The proof will be deferred to the supplementary material.

**Corollary 2.** Suppose all assumptions in section 2 hold true and \( \eta T \to \infty \), \( \eta^{1/2+\delta/2} \to 0 \) for some \( \delta > 0 \). Then all conditions of Theorem 1 is satisfied.

Finally we remark on the time complexity of Algorithm 2. There are \( T \) iterations in the main loop. At each iteration, a maximum likelihood estimation is calculated and then simple matrix updates are carried out. The matrix updates at each iteration has time complexity \( O(d^2) \). The time complexity at each iteration for maximum likelihood estimation varies depending on the precise algorithm to be used, but in general the time complexity will be linear in \( T \).
In this section we justify the construction of point-val\textsuperscript{e} theorem shows that the constructed confidence intervals for the expected demand on a fixed pair of price \(p\) and context vector \(x\). The following theorem shows that the constructed confidence interval \([\eta_{a,\text{debiased}}(p,x),\eta_{a,\text{debiased}}(p,x)]\) is asymptotically accurate as \(T \to \infty\).

**THEOREM 2.** For any given \(\alpha \in (0,1)\), let \([\eta_{a,\text{debiased}}(p,x),\eta_{a,\text{debiased}}(p,x)]\) be constructed as in Step 5 of Algorithm 1. Suppose also that all assumptions listed in section 2 hold, and the parameter \(\eta\) satisfies \(\eta T \to \infty\) and \(\eta T^{1/2+\delta} \to 0\) for some \(\delta > 0\). Then

\[
\lim_{T \to \infty} \Pr_{\eta} \left[ \eta_{a,\text{debiased}}(p,x) \leq f(p,x;\theta_0) \leq \eta_{a,\text{debiased}}(p,x) \right] = 1 - \alpha.
\]

Theorem 2 directly follows from Corollary 1 in section 4.1, which establishes that \(f(p,x;\theta_0) - f(p,x;\theta_0)/\hat{\sigma}_{p,x}^2\) converges in distribution to \(\mathcal{N}(0,1)\). To verify the validity of the constructed confidence intervals \([\eta_{a,\text{debiased}},\eta_{a,\text{debiased}}]\) numerically, we plot the calibration results for both the Wald’s interval and the de-biased approach with confidence levels \(1 - \alpha \in [0.7,0.95]\) using the experimental setting in Figure 1. The results are shown in Figure 3. For each confidence level \(1 - \alpha \in [0.7,0.95]\), we report the coverage rates (i.e., the relative frequency of \(f(p,x;\theta_0)\) falling into the constructed confidence intervals) for both approaches. The closer the coverage is to the target confidence level \(1 - \alpha\), the more accurate the constructed confidence intervals are.

As we can see in Figure 3, the baseline method (built on Wald’s intervals) suffers from significant under-coverage, with the coverage at level \(1 - \alpha = 0.7\) sometimes even below 0.55. On the other hand, the under-coverage effect of our proposed de-biased approach is minimal and most of the time upper bounded by 5%, making it significantly more accurate compared to the baseline method.

Apart from point-wise confidence intervals, in practical applications it is also important to construct confidence intervals for the entire demand function \(f(\cdot, \cdot; \theta_0)\), so that the expected demand of any incoming customer and any offered price can be effectively quantified. Our next theorem validates the accuracy of the uniform confidence intervals \([\eta_{a,\text{debiased}}(\cdot, \cdot),\eta_{a,\text{debiased}}(\cdot, \cdot)]\) constructed in Step 6 of our proposed Algorithm 1.

**THEOREM 3.** For any given \(\alpha \in (0,1)\), let \([\eta_{a,\text{debiased}}(\cdot, \cdot),\eta_{a,\text{debiased}}(\cdot, \cdot)]\) be constructed as in Step 2 of Algorithm 1. Suppose also that all assumptions listed in section 2 hold, and the parameter \(\eta\) satisfies \(\eta T \to \infty\) and \(\eta T^{1/2+\delta} \to 0\) for some \(\delta > 0\). Then

\[
\lim_{T \to \infty} \lim_{M \to \infty} \Pr_{\eta} \left[ \forall \theta \in [\theta_{\min}, \theta_{\max}], \forall x \in \mathcal{X}, \eta_{a,\text{debiased}}(p,x) \leq f(p,x;\theta_0) \leq \eta_{a,\text{debiased}}(p,x) \right] = 1 - \alpha.
\]
Figure 4 Coverage Results for the Wald’s Interval Approach (the red curves) and the De-biased Approach (the blue curves) for Confidence Intervals uniformly Over all \( p \in [0,1] \) and \( x \in [-1,1] \), with the Confidence level \( 1 - \alpha \) Ranging from 0.7 to 0.95. The x-axis is the Targeting \((1 - \alpha)\) Confidence Level and y-axis is the Empirical Coverage Rate over 5000 Independent Trials. The Dashed Black Curve Indicates the Perfect Coverage, that is, \( y=x \). The red Curve is the Classical Confidence Intervals Using the Wald’s Approach and the Blue Curve is the Intervals Using the De-Biased Approach (Color figure can be viewed at wileyonlinelibrary.com)

In this study we proposed a de-biased approach to construct accurate confidence intervals for the unknown demand curve based on dynamically adjusted prices and potentially sequentially/temporally correlated customer contexts. We also illustrate that the traditional method for independent data leads to a significant bias, which is invalid for the construction of confidence intervals. The developed confidence intervals are asymptotically level-\((1-\alpha)\) (i.e., cover the true demand curve with probability \( 1-\alpha \)), which is verified both theoretically and numerically.

One potential future direction is to develop location-sensitive uniform confidence intervals for the demand curve \( f(\cdot, \cdot; \theta_0) \). In particular, if we compare the point-wise confidence intervals \([\hat{\alpha}_a, U_{\alpha}^{\text{debiased}}]\) with the uniform ones \([\hat{I}_{\alpha}^{\text{debiased}}, U_{\alpha}^{\text{debiased}}]\) constructed in Algorithm 1, we can see that the confidence interval lengths \(|\hat{\alpha}_a^{\text{debiased}}(p,x) - U_{\alpha}^{\text{debiased}}(p,x)|\) differ for different price and context vector pairs (since \( \hat{\sigma}_{px}^{\text{debiased}} \) depends on \( p \) and \( x \)), while the lengths \(|\hat{I}_{\alpha}^{\text{debiased}}(p,x) - U_{\alpha}^{\text{debiased}}(p,x)|\) remain the same for all \( p \) and \( x \). It is thus an interesting question whether location-dependent confidence intervals (whose lengths depend on the particular values of \( p, x \)) can be constructed uniformly for the demand curve, satisfying \( \Pr[\forall p, x, L_a(p,x) \leq f(p,x;\theta_0) \leq U_a(p,x)] \to 1-\alpha \).

**5. Conclusion and Future Directions**

In this study we proposed a de-biased approach to construct accurate confidence intervals for the unknown demand curve based on dynamically adjusted prices and potentially sequentially/temporally correlated customer contexts. We also illustrate that the traditional method for independent data leads to a significant bias, which is invalid for the construction of confidence intervals. The developed confidence intervals are asymptotically level-\((1-\alpha)\) (i.e., cover the true demand curve with probability \( 1-\alpha \)), which is verified both theoretically and numerically.

One potential future direction is to develop location-sensitive uniform confidence intervals for the demand curve \( f(\cdot, \cdot; \theta_0) \). In particular, if we compare the point-wise confidence intervals \([\hat{\alpha}_a, U_{\alpha}^{\text{debiased}}]\) with the uniform ones \([\hat{I}_{\alpha}^{\text{debiased}}, U_{\alpha}^{\text{debiased}}]\) constructed in Algorithm 1, we can see that the confidence interval lengths \(|\hat{\alpha}_a^{\text{debiased}}(p,x) - U_{\alpha}^{\text{debiased}}(p,x)|\) differ for different price and context vector pairs (since \( \hat{\sigma}_{px}^{\text{debiased}} \) depends on \( p \) and \( x \)), while the lengths \(|\hat{I}_{\alpha}^{\text{debiased}}(p,x) - U_{\alpha}^{\text{debiased}}(p,x)|\) remain the same for all \( p \) and \( x \). It is thus an interesting question whether location-dependent confidence intervals (whose lengths depend on the particular values of \( p, x \)) can be constructed uniformly for the demand curve, satisfying \( \Pr[\forall p, x, L_a(p,x) \leq f(p,x;\theta_0) \leq U_a(p,x)] \to 1-\alpha \).

**Acknowledgment**

We thank the Department Editor, Senior Editor, and anonymous referees for their detailed and constructive comments that considerably improve the quality of this study. Xiangyu Chang thank the support from the NSFC under Grants 11771012 and 61502342. Dongdong Ge thank the support from the NSFC under Grants 11471205 and 11831002. Xiangyu Chang and Dongdong Ge are the co-corresponding authors of the paper.

**Note**

\(^1\)The delta’s method asserts that if \( \sqrt{n}(X_n - \beta) \xrightarrow{d} \mathcal{N}(0,\Sigma) \) then \( \sqrt{n}(g(X_n) - g(\beta)) \xrightarrow{d} \mathcal{N}(\Sigma g(\beta)^{\top} \Sigma g(\beta)) \). See for example the reference of Van der Vaart (2000).

**References**

Abbasi-Yadkori, Y., D. Pal, C. Szepesvari. 2012. Online-to-confidence-set conversions and application to sparse stochastic bandits. Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS) 1–9.

Agrawal, S., V. Avandhanula, V. Goyal, A. Zeevi. 2019. MNL-bandit: A dynamic learning approach to assortment selection. *Oper. Res.* 67(5): 1453–1485.

Ban, G. Y. 2020. Confidence intervals for data-driven inventory policies with demand censoring. *Oper. Res.* 68(2): 309–326.

Besbes, O., A. Zeevi. 2009. Dynamic pricing without knowing the demand function: Risk bounds and nearoptimal algorithms. *Oper. Res.* 57(6): 1407–1420.

Besbes, O., A. Zeevi. 2015. On the surprising sufficiency of linear models for dynamic pricing with demand learning. *Management Sci.* 61(4): 723–739.

Billingsley, P. 2008. *Probability and Measure*. John Wiley & Sons, Hoboken.
Broder, J., P. Rusmevichientong. 2012. Dynamic pricing under a general parametric choice model. Oper. Res. 60(4): 965–980.

Brown, B. M. 1971. Martingale central limit theorems. Ann. Math. Stat. 42(1): 59–66.

Chen, L., E. L. Plambeck. 2008. Dynamic inventory management with learning about the demand distribution and substitution probability. Manuf. Serv. Oper. Manag. 10(2): 236–256.

Chen, B., X. Chao, H. S. Ahn. 2019a. Coordinating pricing and inventory replenishment with nonparametric demand learning. Oper. Res. 67(4): 1035–1052.

Chen, B., X. Chao, C. Shi. 2019b. Nonparametric Learning Algorithms for Joint Pricing and Inventory Control with Lost-Sales and Censored Demand. SSRN: Available at https://ssrn.com/abstract=2836057 (accessed date January 8, 2021).

Chen, Q., S. Jasin, I. Duenyas. 2019c. A nonparametric self-adjusting control for joint learning and optimization of multi-product pricing with finite resource capacity. Math. Oper. Res. 44(2): 601–631.

Chen, X., Z. Owen, C. Pixton, D. Simchi-Levi. 2020a. A statistical learning approach to personalization in revenue management. Management Sci. (in press).

Chen, X., Y. Wang, Y. Zhou. 2020b. Dynamic assortment optimization with changing contextual information. J. Mach. Learn. Res. (in press).

Chen, X., Y. Wang, Y. Zhou. 2020c. Dynamic assortment selection under nested logit models. Prod. Oper. Manag. (in press).

Deshpande, Y., L. Mackey, V. Syrgkanis, M. Taddy. 2018 Accurate inference for adaptive linear models. Proceedings of the International Conference on Machine Learning (ICML).

Deshpande, Y., A. Javanmard, M. Mehrabi. 2019. Online debiasing for adaptively collected high-dimensional data. arXiv preprint arXiv:1911.01040.

Filippi, S., O. Cappe, A. Garivier, C. Szepesvári. 2010. Parametric bandits: The generalized linear case. Proceedings of the Advances in Neural Information Processing Systems (NeurIPS).

Huh, W. T., R. Levi, P. Rusmevichientong, J. B. Orlin. 2011. Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. Oper. Res. 59(4): 929–941.

Javanmard, A., A. Montanari. 2014. Confidence intervals and hypothesis testing for high-dimensional regression. J. Mach. Learn. Res. 15(1): 2869–2909.

Keener, R. W. 2010. Theoretical Statistics. Springer, Berlin.

Lai, T. L., C. Z. Wei. 1982. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Ann. Statist. 10(1): 154–166.

Lei, Y., Jasin, S., A. Sinha. 2019. Near-Optimal Bisection Search for Nonparametric Dynamic Pricing with Inventory Constraint, Ross School of Business Paper No. 1252, Available at SSRN: https://ssrn.com/abstract=2509425 (accessed date January 8, 2021).

Miao, S., X. Chen, X. Chao, J. Liu, Y. Zhang. 2019. Context-based dynamic pricing with online clustering. arXiv preprint arXiv:1902.06199.

Qiang, S., M. Bayati. 2016. Dynamic pricing with demand covariates. Technical report, Stanford University.

Rusmevichientong, P., H. Topaloglu. 2012. Robust assortment optimization in revenue management under the multinomial logit choice model. Oper. Res. 60(4): 865–882.

Rusmevichientong, P., J. N. Tsitsiklis. 2010. Linearly parameterized bandits. Math. Oper. Res. 35(2): 395–411.

Saure, D., A. Zeevi. 2013. Optimal dynamic assortment planning with demand learning. Manuf. Serv. Oper. Manag. 15(3): 387–404.

Stigler, S. M. 2002. Statistics on the Table: The History of Statistical Concepts and Methods. Harvard University Press, Cambridge.

Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. J. Roy. Stat. Soc. B, 58(1): 267–288.

Tropp, J. A. 2012. User-friendly tail bounds for sums of random matrices. Found. Comput. Math. 12(4): 389–434.

Van de Geer, S., P. Buhlmann, Y. Ritov, R. Dezeure. 2014. On asymptotically optimal confidence regions and tests for high-dimensional models. Ann. Statist. 42(3): 1166–1202.

Van der Vaart, A. W. 2000. Asymptotic Statistics, vol. 3. Cambridge university press, Cambridge.

Wang, Z., S. Deng, Y. Ye. 2014. Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. Oper. Res. 62(2): 318–331.

Wang, Y., X. Chen, Y. Zhou. 2018. Near-optimal policies for dynamic multinomial logit assortment selection models. Proceedings of the Advances in Neural Information Processing Systems (NeurIPS).

Wang, Y., J. Wang, S. Balakrishnan, A. Singh. 2019. Rate optimal estimation and confidence intervals for high-dimensional regression with missing covariates. J. Mult. Anal. 174: 104526.

Watkins, C. J. C. H. 1989. Learning from delayed rewards. Ph.D. thesis, King’s College, Cambridge.

Zhang, C. H., S. S. Zhang. 2014. Confidence intervals for low dimensional parameters in high dimensional linear models. J. Roy. Stat. Soc. B 76(1): 217–242.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Proofs of Statements