8-2015

Empirical Bayes Prediction for the Multivariate Newsvendor Loss Function

Gourab Mukherjee
University of Southern California

Lawrence D. Brown
University of Pennsylvania

Paat Rusmevichientong
University of Pennsylvania

Empirical Bayes Prediction for the Multivariate Newsvendor Loss Function

Abstract

We develop a novel Empirical Bayes methodology for prediction under check loss in high-dimensional Gaussian models. The check loss is a piecewise linear loss function having differential weights for measuring the amount of underestimation or overestimation. Prediction under it differs in fundamental aspects from estimation or prediction under weighted-quadratic losses. Because of the nature of this loss, our inferential target is a pre-chosen quantile of the predictive distribution rather than the mean of the predictive distribution. We develop a new method for constructing uniformly efficient asymptotic risk estimates which are then minimized to produce effective linear shrinkage predictive rules. In calculating the magnitude and direction of shrinkage, our proposed predictive rules incorporate the asymmetric nature of the loss function and are shown to be asymptotically optimal. Using numerical experiments we compare the performance of our method with traditional Empirical Bayes procedures and obtain encouraging results.

Disciplines

Physical Sciences and Mathematics

Submitted to the Annals of Statistics

EMPIRICAL BAYES PREDICTION FOR THE MULTIVARIATE NEWSVENDOR LOSS FUNCTION

By Gourab Mukherjee*, Lawrence D. Brown† and Paat Rusmevichientong*

*University of Southern California** *and University of Pennsylvania*†

Motivated by an application in inventory management, we consider the multi-product newsvendor problem of finding the optimal stocking levels that minimize the total backorder and lost sales costs. We focus on a setting where we have a large number of products and observe only noisy estimates of the underlying demand. We develop an Empirical Bayes methodology for predicting stocking levels, using data-adaptive linear shrinkage strategies which are constructed by minimizing uniformly efficient asymptotic risk estimates. In calculating the magnitude and direction of shrinkage, our proposed predictive rules incorporate the asymmetric nature of the piecewise linear newsvendor loss function and are shown to be asymptotically optimal. Using simulated data, we study the non-asymptotic performance of our method and obtain encouraging results. The asymptotic risk estimation method and the proof techniques developed here can also be used to formulate optimal empirical Bayes predictive strategies for general piecewise linear and related asymmetric loss functions.

1. Introduction. The *newsvendor problem* of determining optimal stocking levels is one of the classical problems in the literature on inventory management (Arrow, Harris and Marschak, 1951, Choi, 2012). We study the traditional newsvendor problem (Scarf, 1959) in the modern big data regime and consider the inventory management problem of a vendor who sells a large number of products. The demand distribution of each product is unknown and must be estimated from data. Our problem is motivated by modern-day online retailers who carry hundreds to millions of products. The historical sales and projected demand varies greatly across different products, and the company must decide the stocking quantity of each product. We consider a one-period setting, where based on the observed demand in the previous period, the firm must determine the stocking quantity of each product in the next period. The observed demand in the previous period provides a basis for estimating the demand in the next period. The firm has to balance the tradeoffs between stocking too much and incurring high inventory cost versus stocking too little and suffering lost sales. Our objective is to develop a data-driven policy that minimizes the total inventory and lost sales costs across all products.

Keywords and phrases: Shrinkage estimators, Empirical Bayes prediction, Multivariate newsvendor problem, Asymptotic optimality, Uniformly efficient risk estimates, Oracle inequality, Pin-ball loss, Piecewise linear loss, Hermite polynomials

When the demand distribution of each product is known in advance, this reduces to the classical newsvendor problem, which has been thoroughly studied in the literature; see, for example, Karlin and Scarf (1958). It is well-known that the optimal stocking quantity of each product is the newsvendor fractile – a ratio involving per-unit inventory and lost-sales costs – of the corresponding demand distribution, and we can treat each product independently. The challenges in our setting are the fact that demand distributions are unknown and must be estimated from data and the firm faces a large number of products.

From a statistical perspective the problem reduces to prediction of a future demand under a loss, sometimes referred to as the "check-loss" (Koenker and Bassett Jr, 1978) function. This loss is linear in the amount of underestimation (lost sales) or overestimation (over-stocking). The weights for these two linear segments differ. Since there are many products this is a multivariate statistical prediction problem with independent coordinate problems. In common with many other multivariate problems we find that empirical Bayes (shrinkage) can provide better performance than simple coordinate-wise rules; see James and Stein (1961), Zhang (2003), and Greenshtein and Ritov (2009) for some background. However, prediction under the loss function here differs in fundamental respects from estimation or prediction under the weighted quadratic losses considered in most of the previous literature. This necessitates different strategies for creation of effective empirical Bayes predictors.

Our algorithm begins with formulation of the problem via a Gaussian hierarchical Bayes structure, with unknown hyperparameters. We then develop an estimate of the hyperparameters that is adapted to the shape of the predictive loss. This estimate of the hyperparameters is converted to a prediction of demand via substitution in the Bayes formula for predictive risk. This yields a demand prediction for each product that we prove is overall asymptotically optimal as the number of products grows increasingly large. The hyperparameter estimator involves an appropriate use of Hermite polynomial expansions for the relevant stochastic functions. Cai et al. (2011) used such an expansion for a different, though somewhat related, problem involving estimation of the L_1 norm of an unknown mean vector. In other respects our derivation logically resembles that of Xie, Kou and Brown (2012, 2015) who constructed empirical Bayes estimators built from an unbiased estimate of risk. However their problem involved estimation under quadratic loss, and the mathematical formulas they used provide exactly unbiased estimates of risk, and are quite different from those we develop.

The remainder of Section 1 describes our basic setup and gives formal statements of our main asymptotic results. Section 2 provides further details. It explains the general mathematical structure of our asymptotic risk estimation methodology and sketches the proof techniques used to prove the main theorems about it. Sections 4 and 5 contain further narrative to explain the proofs of the main results, but technical details are deferred to the Appendices. Section 3 reports on some simulations. These clarify the nature of our estimator and provide some confidence that it performs well even when the number of products is not extremely large.

1.1. *Basic Setup.* We adopt the statistical prediction analysis framework of Aitchison and Dunsmore (1976) and Geisser (1993) and consider a one-step Gaussian predictive model. We have n products indexed by $i = 1, \ldots, n$, and for each i, the observed historical demand X_i and the unobserved future demand Y_i are distributed according to a normal distribution with an unknown mean θ_i; that is,

$$X_i = \theta_i + \sqrt{\nu_{p,i}} \cdot \epsilon_{1,i} \quad \text{for } i = 1, 2, \ldots, n \tag{1.1}$$

$$Y_i = \theta_i + \sqrt{\nu_{f,i}} \cdot \epsilon_{2,i} \quad \text{for } i = 1, 2, \ldots, n \ , \tag{1.2}$$

where the noise terms $\{\epsilon_{j,i} : j = 1, 2; i = 1, \ldots, n\}$ are i.i.d. from a standard normal distribution, and the past and future variances $\nu_{p,i}$, $\nu_{f,i}$ are known for all i. Note that, in multivariate notation $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_p)$ and $\boldsymbol{Y} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_f)$ where $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Sigma}_f$ are n dimensional diagonal matrices whose i^{th} entries are $\nu_{p,i}$ and $\nu_{f,i}$, respectively. If the mean θ_i were known, then the observed past demand X_i and future demand Y_i would be independent of each other.

Our objective is to compute an estimate $\hat{\boldsymbol{q}} = \{\hat{q}_i(\boldsymbol{X}) : 1 \leq i \leq n\}$ based on the past data $\boldsymbol{X}$ such that $\hat{\boldsymbol{q}}$ optimally predicts $\boldsymbol{Y}$ under a piecewise-linear predictive loss function, which reflects the newsvendor's cost. As a convention, we use bold font to denote vectors and matrices, while regular font denotes scalars. For ease of exposition, we will use $\widehat{\ }$ to denote data-dependent estimates, and we will sometimes write $\hat{\boldsymbol{q}}$ or its univariate version $\hat{q}_i$ without an explicit reference to $\boldsymbol{X}$.

Predictive Loss Function: For each product i, we assume that each unit of inventory incurs a holding cost $h_i > 0$, and each unit of lost sale incurs a cost of $b_i > 0$. When we estimate the future demand Y_i by $\hat{q}_i$, the loss corresponding to the i^{th} product is $b_i \cdot (Y_i - \hat{q}_i)^+ + h_i \cdot (\hat{q}_i - Y_i)^+$. It is a piecewise linear loss function (see Chapter 11.2.3 of Press, 2009). The loss is related to the pin-ball loss function (Steinwart and Christmann, 2011), which is widely used in statistics and machine learning for estimating conditional quantiles. For each $\boldsymbol{X} = \boldsymbol{x}$, the associated predictive loss is given by

$$l_i(\theta_i, \hat{q}_i(\boldsymbol{x})) \ = \ \mathbb{E}_{Y_i \sim N(\theta_i \, , \, \nu_{f,i})} \left[b_i (Y_i - \hat{q}_i(\boldsymbol{x}))^+ + h_i (\hat{q}_i(\boldsymbol{x}) - Y_i)^+ \right] \ ,$$

where the expectation is taken over the distribution of the future demand Y_i only. We use the notation $N(\mu, \nu)$ to denote a normal random variable with mean μ and variance ν. Note that $l_i(\theta_i, \hat{q}_i)$ measures the total expected inventory and lost sales costs associated with product i. Since Y_i is normally distributed with mean θ_i, it follows from Lemma 2.2 that

$$\begin{aligned} &l_i(\theta_i, \hat{q}_i) = \sqrt{\nu_{f,i}} \, (b_i + h_i) \, G\big(\, (\hat{q}_i - \theta_i)/\sqrt{\nu_{f,i}} \, , \, b_i/(b_i + h_i) \, \big), \text{ where} \\ &G(w, \beta) = \phi(w) + w\Phi(w) - \beta w \quad \text{for } w \in \mathbb{R}, \beta \in [0, 1] \, , \end{aligned} \tag{1.3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal PDF and CDF, respectively. Thus, given the data $\boldsymbol{X}$, the total cost associated with n stocking quantities $\hat{\boldsymbol{q}}$ is

$$L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}) = \frac{1}{n} \sum_{i=1}^{n} l_i(\theta_i, \hat{q}_i) \ .$$

To determine the stocking quantities, we want to minimize the expected loss $\mathbb{E}_{\boldsymbol{X}}\left[L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}})\right]$ over the class of estimators $\hat{\boldsymbol{q}}$ for **all** values of $\boldsymbol{\theta}$. If $\boldsymbol{\theta}$ were known, then by Lemma 2.2, the optimal stocking quantity for each product i is given by $\theta_i + \sqrt{\nu_{f,i}}\,\Phi^{-1}(b_i/(b_i+h_i))$. In absence of such knowledge, we consider hierarchical modeling and the related Empirical Bayes (EB) approach (Robbins, 1964, Zhang, 2003). This is a popular statistical method for combining information and conducting simultaneous inference on multiple parameters that are connected by the structure of the problem (Efron and Morris, 1973a,b, Good, 1980).

Hierarchical Modeling and Predictive Risk. We consider the conjugate hierarchical model and put a prior distribution $\pi_{\eta,\tau}$ on each θ_i, under which $\theta_1, \theta_2, \ldots, \theta_n$ are i.i.d. from $N(\eta, \tau)$ distribution. Here, η and τ are the *unknown* location and scale hyperparameters, respectively. The *predictive risk* associated with our estimator $\hat{\boldsymbol{q}}$ is defined by

$$R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}) = \mathbb{E}_{\boldsymbol{X}\sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_p)}\left[L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}})\right] \ ,$$

where the expectation is taken over $\boldsymbol{X}$. Note that the expectation over $\boldsymbol{Y}$ is already included in L via the definition of the loss ℓ_i. By Lemma 2.3, the Bayes estimate – the unique minimizer of the integrated Bayes risk $B_n(\eta, \tau) = \int R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}})\pi_{\eta,\tau}(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ – is given for $i = 1, \ldots, n$ by

$$\hat{q}_i^{\mathsf{Bayes}}(\eta, \tau) = \alpha_i(\tau)X_i + (1-\alpha_i(\tau))\eta + \sqrt{\nu_{f,i} + \alpha_i(\tau)\nu_{p,i}}\,\Phi^{-1}(b_i/(b_i+h_i)), \tag{1.4}$$

where, for all i, $\alpha_i(\tau) = \tau/(\tau + \nu_{p,i})$ denotes the shrinkage factor of product i.

Standard parametric Empirical Bayes methods (Efron and Morris, 1973b, Lindley, 1962, Morris, 1983, Stein, 1962) suggest using the marginal distribution of $\boldsymbol{X}$ to estimate the unknown hyperparameters. In this paper, inspired by Stein's Unbiased Risk Estimation (SURE) approach of constructing shrinkage estimators (Stein, 1981), we consider an alternative estimation method. Afterwards, in Section 1.2, we show that our method outperforms standard parametric EB methods which are based on the popular maximum likelihood and method of moments approaches.

Class of Shrinkage Estimators: The Bayes estimates defined in (1.4) are based on the conjugate Gaussian prior and constitute a class of linear estimators (Johnstone, 2013). When the hyperparameters are estimated from data, they form a class of adaptive linear estimators. Note that these estimates themselves are not linear but are derived from linear estimators by the estimation of tuning parameters, which, in this case, correspond to the shrinkage factor $\alpha_i(\tau)$ and the direction of shrinkage η. Motivated by the form of the Bayes estimate in (1.4), we study the estimation problem in the following three specific classes of shrinkage estimators:

- **Shrinkage governed by Origin-centric priors:** Here, $\eta = 0$ and τ is estimated based on the past data $\boldsymbol{X}$. Shrinkage here is governed by mean-zero priors. This class of estimators is denoted by $\mathcal{S}^0 = \{\hat{\boldsymbol{q}}(\tau) \mid \tau \in [0, \infty]\}$, where for each τ, $\hat{\boldsymbol{q}}(\tau) = \{\hat{q}_i(\tau) : i = 1, \ldots, n\}$, and for all i,

$$\hat{q}_i(\tau) = \alpha_i(\tau)X_i + \sqrt{\nu_{f,i} + \alpha_i(\tau)\nu_{p,i}}\,\Phi^{-1}\left(b_i/(b_i+h_i)\right) \ .$$

We can generalize $\mathcal{S}^0$ by considering shrinkage based on priors with an a priori chosen location η_0. The corresponding class of shrinkage estimators $\mathcal{S}^A(\eta_0) = \{\hat{\boldsymbol{q}}(\eta_0, \tau) | \tau \in [0, \infty]\}$, where η_0 is a prefixed location, consists of

$$\hat{q}_i(\eta_0, \tau) = \alpha_i(\tau) X_i + (1 - \alpha_i(\tau))\eta_0 + \sqrt{\nu_{f,i} + \alpha_i(\tau)\nu_{p,i}}\, \Phi^{-1}\left(b_i/(b_i + h_i)\right) .$$

As these estimators are location equivariant (Lehmann and Casella, 1998) the estimation problem in $\mathcal{S}^A(\eta_0)$ for any fixed η_0 reduces to an estimation problem in $\mathcal{S}^0$. So, we do not discuss shrinkage classes based on a priori centric priors as separate cases.

- **Shrinkage governed by Grand Mean centric priors:** In this case, $\eta = \bar{X}_n := n^{-1}\sum_{i=1}^n X_i$, and τ is estimated based on the past data. Shrinkage here is governed by priors centering near the grand mean of the past $\boldsymbol{X}$. This class of estimators is denoted by $\mathcal{S}^G = \left\{\hat{\boldsymbol{q}}^G(\tau) \mid \tau \in [0, \infty]\right\}$, where for all $\tau \in [0, \infty]$ and $i = 1, \ldots, n$,

$$\hat{q}_i^G(\tau) = \alpha_i(\tau) X_i + (1 - \alpha_i(\tau))\bar{X}_n + \sqrt{\nu_{f,i} + \alpha_i(\tau)\nu_{p,i}}\, \Phi^{-1}\left(b_i/(b_i + h_i)\right) .$$

- **General Data-Driven Shrinkage:** In the final case, we consider the general class of shrinkage estimators where both η and τ are simultaneously estimated. We shrink towards a data-driven location while simultaneously optimizing the shrinkage factor; this class is denoted by $\mathcal{S} = \{\hat{\boldsymbol{q}}(\eta, \tau) \mid \eta \in \mathbb{R}, \tau \in [0, \infty]\}$, where

$$\hat{q}_i(\eta, \tau) = \alpha_i(\tau) X_i + (1 - \alpha_i(\tau))\eta + \sqrt{\nu_{f,i} + \alpha_i(\tau)\nu_{p,i}}\, \Phi^{-1}\left(b_i/(b_i + h_i)\right) .$$

1.2. *Main Results.* For ease of understanding, we first describe the results for the class $\mathcal{S}^0$ where the direction of shrinkage is governed by mean-zero priors so that $\eta = 0$. The results for the other cases are stated afterwards; see Section 1.5. By definition, estimators in $\mathcal{S}^0$ are of the form: for $i = 1, \ldots, n$,

$$\hat{q}_i(\tau) = \alpha_i(\tau) X_i + \sqrt{\nu_{f,i} + \alpha_i(\tau)\nu_{p,i}}\, \Phi^{-1}\left(b_i/(b_i + h_i)\right) , \tag{1.5}$$

where $\alpha_i(\tau) = \tau/(\tau + \nu_{p,i})$ is the shrinkage factor, and the tuning parameter τ varies from $[0, \infty]$. We next describe the reasonable and mild conditions that we impose on the problem structure. These assumptions mainly facilitate the rigorousness of the theoretical proofs and can be further relaxed for practical use.

Assumptions

A1. Reasonable holding and lost-sales costs. To avoid degeneracies in our loss function, which can be handled easily but require separate case by case inspections, we impose the following condition on the lost sales and holding costs:

$$0 < \inf_i b_i/(b_i + h_i) \le \sup_i b_i/(b_i + h_i) < 1 \quad \text{and} \quad \sup_i (b_i + h_i) < \infty .$$

A2. Sufficient historical data. We assume the following upper bound on the ratio of the past to future variances for all the products:

$$\sup_i \ \nu_{p,i}/\nu_{f,i} < 1/(4e) \ . \tag{1.6}$$

To understand the implication of this assumption, consider an example where we want to predict the monthly demand of the products. We assume that the parameters in the predictive model are invariant over time. If we have independent historical sales data for each of the m previous months, using sufficiency in the Gaussian setup, we can reduce this multi-sample past data problem to a single-sample problem by averaging. The past variance of the averaged model is proportional to m^{-1}, and in this case, we will have $\nu_{p,i}/\nu_{f,i} = m^{-1}$ for each i. So, if we have a lot of historical data, the above condition will be satisfied. In this example, we will need at least 11 months of historical data to predict 1 month of demand.

Conditions of this form are not new in the predictive literature, as the ratio of the past to future variability controls the role of estimation accuracy in predictive models (George, Liang and Xu, 2006, Mukherjee and Johnstone, 2015). We emphasize, however, that this condition is not necessary; it simply facilitates a simple and clean theoretical proof of our results. As shown in our simulation experiments in Section 3, the condition can be greatly relaxed in applications. Afterwards, we discuss in further detail the nature and possible relaxation of this constraint. Also, to avoid degeneracies of the loss function, we impose a very mild but natural assumption on the future variances: $\sup_i \nu_{f,i} < \infty$.

A3. Bounded mean demand. We assume that the mean of the true demands of all the products is bounded; that is,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |\theta_i| < \infty. \tag{1.7}$$

Our Proposed Shrinkage Estimate: The predictive risk of estimators $\hat{\boldsymbol{q}}(\tau)$ of the form (1.5) is given by $R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) = \mathbb{E}_{\boldsymbol{\theta}}\left[L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))\right]$, where the expectation is taken over $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_p)$. Following Stein (1981), the idea of minimizing unbiased estimates of risk to obtain efficient estimates of tuning parameters has a considerable history in statistics (George and Strawderman, 2012, Hoffmann, 2000, Mallows, 1973, Stigler, 1990). However, as shown in Equation (1.3), our loss function $l(\cdot,\cdot)$ is *not* quadratic, so a direct construction of unbiased risk estimates is difficult. Instead, we approximate the risk function $\tau \mapsto R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))$ by an *Asymptotic Risk Estimator* (ARE) function $\tau \mapsto \widehat{\mathrm{ARE}}_n(\tau)$, which may be *biased*, but it approximates the true risk function *uniformly well for all* τ, particularly when the number of products is large. Note that $\widehat{\mathrm{ARE}}_n(\tau)$ depends *only* on the observed historical demand $\boldsymbol{X}$ and τ and is *independent* of $\boldsymbol{\theta}$. The estimation procedure is fairly complicated and is built on a Hermite polynomial expansion of the risk. It is described in the next subsection. Afterward, we show that our risk estimation method not only adapts to the data but also does a better job in adapting to the

shape of the loss function when compared with the widely used Empirical Bayes MLE (EBML) or method of moments (EBMM) estimates. The main result of this paper is stated in the following theorem.

THEOREM 1.1 (Uniform Approximation). *Under Assumptions A1-A2, for all $\boldsymbol{\theta}$ satisfying Assumption A3 and for all estimates $\hat{\boldsymbol{q}}(\tau) \in \mathcal{S}^0$,*

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \mathbb{E}|\widehat{\mathrm{ARE}}_n(\tau) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))| \;=\; 0 \;,$$

where the expectation is taken over the random variable $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \Sigma_p)$.

We propose an estimate of the tuning parameter τ for the class of shrinkage estimates $\mathcal{S}^0$ as follows:

(ARE Estimate) $$\hat{\tau}_n^{\mathrm{ARE}} = \operatorname*{arg\,min}_{\tau\in[0,\infty]} \widehat{\mathrm{ARE}}_n(\tau) \;.$$

Theorem 1.1 shows that the average distance between $\widehat{\mathrm{ARE}}$ and the actual loss is asymptotically uniformly negligible; we expect that minimizing $\widehat{\mathrm{ARE}}$ would lead to an estimate with competitive performance. To facilitate our discussion of the risk properties of our ARE Estimate, we next introduce the oracle loss (OR) hyperparameter

$$\tau_n^{\mathsf{OR}} = \operatorname*{arg\,min}_{\tau\in[0,\infty]} L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) \;.$$

Note that τ_n^{OR} depends not only on $\boldsymbol{x}$ but also on the unknown $\boldsymbol{\theta}$. So, it is not an estimator. Rather, it serves as the the theoretical benchmark of estimation accuracy because no estimator in $\mathcal{S}^0$ can have smaller risk than $\hat{\boldsymbol{q}}\left(\tau_n^{\mathsf{OR}}\right)$. Note that $\hat{\boldsymbol{q}}^{\mathsf{Bayes}} \in \mathcal{S}_0$, and thus, even if the correct hyperparameter τ were known, the estimator $\hat{\boldsymbol{q}}\left(\tau_n^{\mathsf{OR}}\right)$ is as good as the Bayes estimator. The following theorem shows that our proposed estimator is asymptotically nearly as good as the oracle loss estimator.

THEOREM 1.2 (Oracle Optimality in Predictive Loss). *Under Assumptions A1-A2, for all $\boldsymbol{\theta}$ satisfying Assumption A3 and for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left\{ L_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\mathrm{ARE}})\right) \;\geq\; L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau_n^{\mathsf{OR}})) \,+\, \epsilon \right\} = 0 \;.$$

The above theorem shows that the loss of our proposed estimator converges in probability to the optimum oracle value $L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau_n^{\mathsf{OR}}))$. We also show that, under the same conditions, it is asymptotically as good as τ_n^{OR} in terms of the risk (expected loss).

THEOREM 1.3 (Oracle Optimality in Predictive Risk). *Under Assumptions A1-A2 and for all $\boldsymbol{\theta}$ satisfying Assumption A3,*

$$\lim_{n\to\infty} R_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\mathrm{ARE}})\right) - \mathbb{E}\left[L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau_n^{\mathsf{OR}})) \right] = 0 \;.$$

We extend the implications of the preceding theorems to show that our proposed estimator is as good as any other estimator in $\mathcal{S}^0$ in terms of both the loss and risk.

COROLLARY 1.1. *Under Assumptions A1-A2, for all $\boldsymbol{\theta}$ satisfying Assumption A3, for any $\epsilon > 0$, and any estimator $\hat{\tau}_n \geq 0$,*

I. $\lim_{n\to\infty} \mathbb{P}\left\{L_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\mathrm{ARE}})\right) \;\geq\; L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n)) + \epsilon\right\} = 0$

II. $\lim_{n\to\infty} R_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\mathrm{ARE}})\right) - R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n)) \;\leq\; 0.$

Next, we present two very popular, standard EB approaches for choosing estimators in $\mathcal{S}^0$. The Empirical Bayes ML (EBML) estimator $\hat{\boldsymbol{q}}(\hat{\tau}^{\mathrm{ML}})$ is built by maximizing the marginal likelihood of $\boldsymbol{X}$ while the method of moments (EBMM) estimator $\hat{\boldsymbol{q}}(\hat{\tau}^{\mathrm{MM}})$ is based on the moments of the marginal distribution of $\boldsymbol{X}$. Following Xie, Kou and Brown (2012, Section 2) the hyperparameter estimates are given by

$$\begin{aligned} \hat{\tau}_n^{\mathrm{ML}} &= \operatorname*{arg\,min}_{\tau\in[0,\infty]} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i^2}{\tau+\nu_{p,i}} + \log(\tau+\nu_{p,i})\right) \\ \hat{\tau}_n^{\mathrm{MM}} &= \max\left\{\frac{1}{n}\sum_{i=1}^{p}\left(X_i^2 - \nu_{p,i}\right)\;,\; 0\right\} \end{aligned} \quad . \tag{1.8}$$

For standard EB estimates $\hat{\boldsymbol{q}}(\hat{\tau}_n^{\mathrm{EB}})$, such as those in (1.8) the hyperparameter estimate $\hat{\tau}_n^{\mathrm{EB}}$ does not depend on the shape of the individual loss functions $\{(b_i, h_i) : 1 \leq i \leq n\}$. We provide a complete definition of $\widehat{\mathrm{ARE}}_n$ and $\hat{\tau}_n^{\mathrm{ARE}}$ in the next section from where it will be evident that our asymptotically optimal estimator $\hat{\tau}_n^{\mathrm{ARE}}$ depends on the ratios $\{b_i/(b_i+h_i) : 1 \leq i \leq n\}$ in an essential way that remains important as $n \to \infty$. So, even asymptotically, the ML and MM estimates do not always agree with $\hat{\tau}^{\mathrm{ARE}}$, particularly in cases when the ratios are not all the same. By Theorems 1.1 and 1.2, it follows that any estimator as efficient as the OR estimator must asymptotically agree with $\hat{\tau}^{\mathrm{ARE}}$. Hence, unlike our proposed ARE based estimator, EBML and EBMM are not generally asymptotically optimal in the class of estimators $\mathcal{S}^0$. In Section 3.1, we provide an explicit numerical example to demonstrate the sub-optimal behavior of the EBML and EBMM estimators.

1.3. *Construction of Asymptotic Risk Estimates.* In this section, we describe the details for the construction of the Asymptotic Risk Estimation (ARE) function $\tau \mapsto \widehat{\mathrm{ARE}}_n(\tau)$, which is the core of our estimation methodology. The estimators in class $\mathcal{S}^0$ are coordinatewise rules, and the risk of such an estimate $\hat{q}(\tau)$ is

$$R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) = \frac{1}{n}\sum_{i=1}^{n} r_i(\theta_i, \hat{q}_i(\tau)) \;,$$

where $r_i(\theta_i, \hat{q}_i(\tau))$ is the risk associated with the i^{th} product. By Lemma 2.3, we have that

$$r_i(\theta_i, \hat{q}_i(\tau)) = (b_i+h_i)\sqrt{\nu_{f,i} + \nu_{p,i}\alpha_i^2(\tau)}\, G\left(c_i(\tau) + d_i(\tau)\theta_i, \tilde{b}_i\right) \;, \tag{1.9}$$

where for all i, $\alpha_i(\tau) = \tau/(\tau + \nu_{p,i})$, $\tilde{b}_i = b_i/(b_i + h_i)$, and

$$c_i(\tau) = \sqrt{\frac{1+\alpha_i(\tau)\nu_{p,i}}{1+\alpha_i(\tau)^2\nu_{p,i}}}\,\Phi^{-1}(\tilde{b}_i) \quad \text{and} \quad d_i(\tau) = -\frac{1-\alpha_i(\tau)}{\sqrt{\nu_{f,i}+\nu_{p,i}\alpha_i(\tau)^2}}\,.$$

The function $G(\cdot)$ is the same function as that associated with the predictive loss and was defined in (1.3). The dependence of $c_i(\tau)$ and $d_i(\tau)$ on τ is only through α_i. We propose an estimate $\widehat{\mathrm{ARE}}_n(\tau)$ of the multivariate risk $R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))$ by using the data-driven coordinate-wise estimate $\hat{T}_i(X_i,\tau)$ of $G(c_i(\tau)+d_i(\tau)\theta_i;\tilde{b}_i)$; that is,

$$\widehat{\mathrm{ARE}}_n(\tau) = \frac{1}{n}\sum_{i=1}^{n}(b_i+h_i)\sqrt{\nu_{f,i}+\nu_{p,i}\alpha_i(\tau)^2}\,\hat{T}_i(X_i,\tau)\,. \tag{1.10}$$

Defining the Coordinatewise Estimate $\hat{T}_i(X_i,\tau)$ – Heuristic Idea. Temporarily keeping the dependence on τ and i implicit, we now describe how we develop a data-driven estimate of the non-linear functional $G(c+d\,\theta,\tilde{b})$ of the unknown parameter θ.

If $|c+d\theta|$ is not too large we approximate the functional by $G_K(c+d\,\theta,\tilde{b})$ – its K order Taylor series expansion around 0:

$$G_K(c+d\,\theta,\tilde{b}) = G(0,\tilde{b}) + G'(0,\tilde{b})(c+d\,\theta) + \phi(0)\sum_{k=0}^{K-2}\frac{(-1)^k H_k(0)}{(k+2)!}\,(c+d\,\theta)^{k+2}\,,$$

where H_k is the k^{th} order probabilists' Hermite polynomial (Thangavelu, 1993). If $W \sim N(\mu,\nu)$ denotes a normal random variable with mean μ and variance ν, then the truncated functional G_K can be estimated unbiasedly using the following property of Hermite polynomials (for proof see Chihara, 2011):

LEMMA 1.1. *If $W \sim N(\mu,\nu)$, then $\nu^{k/2}\,\mathbb{E}_\mu\{H_k(W/\sqrt{\nu})\} = \mu^k$ for $k \geq 1$.*

Now, if $|c+d\theta|$ is large, then the truncated Taylor's expansion $G_K(\cdot)$ would not be a good approximation of $G(c+d\,\theta,\tilde{b})$. However, in that case, as shown in Lemma 2.4, we can use linear approximations with

$$G(c+d\,\theta,\tilde{b}) \;\approx\; (1-\tilde{b})(c+d\,\theta)^+ + \tilde{b}(c+d\,\theta)^-\,,$$

and their corresponding unbiased estimates can be used. Note that for all $x \in \mathbb{R}$, $x^+ = \max\{x,0\}$ and $x^- = \max\{-x,0\}$.

The Details. We need to combine the aforementioned estimates together in a data-driven framework. For this purpose, we use threshold estimates. We use the idea of *sample splitting*. We use the observed data to create two independent samples by adding white noise $\boldsymbol{Z} = \{Z_i : i = 1,\ldots,n\}$ and define

$$U_i = X_i + \sqrt{\nu_{p,i}}Z_i, \quad V_i = X_i - \sqrt{\nu_{p,i}}Z_i \text{ for } i = 1,\ldots,n.$$

Noting that U_i and V_i are independent, we will use V_i to determine whether or not $c_i(\tau)+d_i(\tau)\,\theta$ is large, and then use U_i to estimate $G(c_i(\tau)+d_i(\tau)\,\theta, \tilde{b})$ appropriately. For any fixed $\tau \in [0,\infty]$ and $i = 1,\ldots,n$, we transform

$$U_i(\tau) = c_i(\tau) + d_i(\tau)U_i, \quad V_i(\tau) = c_i(\tau) + d_i(\tau)V_i.$$

Note that $U_i(\tau) \sim N(c_i(\tau) + d_i(\tau)\theta_i\,,\, 2\nu_{p,i}\, d_i^2(\tau))$. By Lemma 1.1, we construct an unbiased estimate of $G_K(c_i(\tau) + d_i(\tau)\theta_i, \tilde{b}_i)$ as

$$\begin{aligned} S_i(U_i(\tau)) =& G(0,\tilde{b}_i) + G'(0,\tilde{b}_i)U_i(\tau) \\ &+ \phi(0) \sum_{k=0}^{K_n(i)-2} \frac{(-1)^k H_k(0)}{(k+2)!} \big(2\nu_{p,i} d_i^2(\tau)\big)^{(k+2)/2} H_{k+2}\bigg(\frac{U_i(\tau)}{(2\nu_{p,i} d_i^2(\tau))^{1/2}}\bigg). \end{aligned}$$

We use a further truncation on this unbiased estimate by restricting its absolute value to n. The truncated version

$$\begin{aligned} \tilde{S}_i(U_i(\tau)) &= S_i(U_i(\tau))\, I\{|S_i(U_i(\tau))| \le n\} \,+\, n\, I\{S_i(U_i(\tau)) > n\} \,-\, n\, I\{S_i(U_i(\tau)) < -n\} \\ &= \text{sign}\,(S_i(U_i(\tau)))\, \min\{|S_i(U_i(\tau))|\,,\, n\} \end{aligned}$$

is biased. But, because of its restricted growth, it is easier to control its variance, which greatly facilitates our analysis.

Threshold Estimates. For each product i, we then construct the following coordinatewise threshold estimates:

$$\hat{T}_i(X_i, Z_i, \tau) = \begin{cases} -\tilde{b}_i\, U_i(\tau) & \text{if} \quad V_i(\tau) < -\lambda_n(i) \\ \tilde{S}_i(U_i(\tau)) & \text{if} \quad -\lambda_n(i) \le V_i(\tau) \le \lambda_n(i) \\ (1-\tilde{b}_i)\, U_i(\tau) & \text{if} \quad V_i(\tau) > \lambda_n(i) \end{cases} \quad \text{for } i = 1,\ldots,n$$

with the threshold parameter

$$\lambda_n(i) = \gamma(i)\sqrt{2\log n}\,, \tag{1.11}$$

where $\gamma(i)$ is any positive number less than $\big(\,1/\sqrt{4e} - \sqrt{\nu_{p,i}/\nu_{f,i}}\,\big)$. Assumption A2 ensures the existence of $\gamma(i)$ because $\nu_{p,i}/\nu_{f,i} < 1/(4e)$ for all i.

The other tuning parameter that we have used in our construction process is the truncation parameter $K_n(i)$, which is involved in the approximation of G and is used in the estimate $\tilde{S}$. We select a choice of $K_n(i)$ that is independent of $\tau \in [0,\infty]$, and is given by

$$K_n(i) = 1 + \left\lceil e^2 \Big(\,\gamma(i) + \sqrt{2\nu_{p,i}/\nu_{f,i}}\,\Big)^2 (2\log n) \right\rceil. \tag{1.12}$$

Rao-Blackwellization. $\hat{T}_i(X_i, Z_i, \tau)$ are randomized estimators as they depend on the user-added noise $\boldsymbol{Z}$. And so, in the final step of the risk estimation procedure we apply Rao-Blackwell adjustment (Lehmann and Casella, 1998) to get $\hat{T}_i(X_i,\tau) = \mathbb{E}\left[\hat{T}_i(X_i, Z_i, \tau)|\boldsymbol{X}\right]$. Here, the expectation is over the distribution of $\boldsymbol{Z}$, which is independent of $\boldsymbol{X}$ and follows $N(0, I_n)$.

1.3.1. *Bias and Variance of the coordinatewise Risk Estimates.* The key result that allows us to establish Theorem 1.1 is the following proposition (for proof see Section 2.3) on estimation of the univariate risk components $G(c_i(\tau) + d_i(\tau)\theta_i, \tilde{b}_i)$ defined in (1.9). It shows that the bias of $\hat{T}_i(X_i, Z_i, \tau)$ as an estimate of $G(c_i(\tau) + d_i(\tau)\theta_i, \tilde{b}_i)$ converges to zero as $n \to \infty$. The scaled variance of each of the univariate threshold estimates $\hat{T}_i(X_i, Z_i, \tau)$ also converges to zero.

PROPOSITION 1.1. *Under Assumptions A1-A2, we have for all* $i = 1, \ldots, n$

I. $\lim_{n\to\infty} \sup_{\tau\in[0,\infty],\, \theta_i\in\mathbb{R}} \mathrm{Bias}_{\theta_i}(\hat{T}_i(X_i, Z_i, \tau)) = 0$,

II. $\lim_{n\to\infty} \sup_{\tau\in[0,\infty],\, \theta_i\in\mathbb{R}} n^{-1}\mathrm{Var}_{\theta_i}(\hat{T}_i(X_i, Z_i, \tau)) = 0$,

where the random vectors $\boldsymbol{X}$ *and* $\boldsymbol{Z}$ *are independent, with* $\boldsymbol{X}$ *following* (1.1) *and* $\boldsymbol{Z}$ *has* $N(0, I)$ *distribution.*

1.4. *Background and Previous Work.* Our work is connected to two streams of literature: inventory management in operations research and empirical Bayes estimation in statistics. We will briefly discuss the connection to each stream of literature. The newsvendor problem appeared in Edgeworth (1888) in connection with optimizing cash reserves in a bank. Based on a subsequent formulation in Arrow, Harris and Marschak (1951), the classical inventory theory has been developed assuming the demand distribution is known in advance, and the optimal solution is the newsvendor quantile of the underlying demand distribution (Karlin and Scarf, 1958). In contrast, here we work in a predictive setup where the demand distribution is unknown and must be estimated from past data.

Within the inventory literature, when the information on the demand distribution is not available, the most common approach is the use of Bayesian updates. Under this approach, the inventory manager has limited access to demand information; in particular, she knows the family of distributions to which the underlying demand belongs, but she is uncertain about its parameters. She has an initial prior belief regarding the uncertainty of the parameter values, and this belief is continually updated based on historical realized demands by computing posterior distributions. Early papers such as Scarf (1959), Scarf (1960), Karlin (1960) and Iglehart (1964) consider cases where the demand distribution belongs to the exponential and range families. Other papers that incorporate the Bayesian approach into stochastic inventory models include Murray and Silver (1966), Chang and Fyffe (1971) and Azoury (1985). It turns out that a simple myopic inventory policy based on a critical fractile is optimal or near-optimal (Lovejoy, 1990). All of the above references to Bayesian updates assume that the distribution of the prior belief is known and given in advance so that Bayesian updates can be computed explicitly. In contrast, in our paper, we assume that the prior distribution of θ_i is $N(\eta, \tau)$ for all i, but the parameters η and τ are *unknown* and must be estimated from data. Note that when η and τ are unknown, traditional Bayesian updates cannot be computed, and thus, the existing inventory literature is not applicable.

Here, we follow the compound decision theory framework introduced in Robbins (1985). In the statistics literature, there has been substantial research on the construction of linear EB estimates in such frameworks (Morris, 1983, Zhang, 2003). Since the seminal work by James and Stein (1961), shrinkage estimators are widely used in real-world applications (Efron and Morris, 1975). Stein's shrinkage is related to hierarchical empirical Bayes methods (Stein, 1962), and several related parametric empirical Bayes estimators have been developed (Efron and Morris, 1973b). As such, Stein's Unbiased Risk Estimate (SURE) is one of the most popular methods for obtaining the estimate of tuning parameters. Donoho and Johnstone (1995) used SURE to choose the threshold parameter in their SureShrink method. However, most of these developments have been under quadratic loss or other associated loss functions (Berger, 1976, Brown, 1975, Dey and Srinivasan, 1985), which admit unbiased risk estimates. DasGupta and Sinha (1999) discussed the role of Steinien shrinkage under the L_1 loss, which is related to our predictive loss only when $b = h$. Usually in inventory management problems $b \gg h$, very unlike the quadratic loss regime, we need to incorporate the asymmetric nature of the loss function. As we need to construct risk estimates that are adapted to the shape of the newsvendor loss function, we need to develop new methods for efficiently estimating the risk functionals associated with our class of shrinkage estimators. In our construction, we concentrate on obtaining uniform convergence of the estimation error over the range of the associated hyperparameters. This enables us to efficiently fine-tune the shrinkage parameters through minimization over the class of risk estimates. Finally, in contrast to quadratic loss results (Xie, Kou and Brown, 2012, Section 3), we develop a more flexible moment-based concentration approach that translates our risk estimation efficiency into the decision theoretic optimality of the proposed shrinkage estimator.

1.5. *Further Results.* We now describe our results for efficient estimation in class $\mathcal{S}$, where we shrink towards a data-driven direction η, and the hyperparameters η and τ are simultaneously estimated. The predictive risk of estimators $\hat{\boldsymbol{q}}(\eta, \tau)$ of the form (1.4) is given by $R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta, \tau)) = \mathbb{E}_{\boldsymbol{\theta}}\left[L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta, \tau))\right]$. We estimate the risk function by $(\eta, \tau) \mapsto \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta, \tau)$. The estimation procedure and the detailed proof for the results in this section are presented in Section 4. We estimate the tuning parameters τ and η for the class of shrinkage estimates $\mathcal{S}$ as

$$\left(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}}\right) = \underset{\tau\in[0,\infty],\ \eta\in\mathbb{R}:\ |\eta|\le M_n}{\arg\min} \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta, \tau) \ ,$$

where $M_n = \max\{|X_i| : 1 \le i \le n\}$. We restrict the shrinkage location η parameter to within the range $[-M_n, M_n]$ because no sensible shrinkage estimator would attempt to shrink toward a location that lies outside the range of the data. The oracle loss estimator here is given by

$$(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}}) = \underset{\tau\in[0,\infty],\ \eta\in\mathbb{R}:\ |\eta|\le M_n}{\arg\min} L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta, \tau)) \ .$$

Here, we need a stronger assumption than Assumption A3 in (1.7). Instead of ℓ_1 bounded demands, we now assume that there exists $\delta > 0$ such that the $(2+\delta)^{\text{th}}$

moment of true demands of all the products is bounded:

$$\textit{Assumption A3}'\textit{(Bounded Demand):} \quad \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |\theta_i|^{\delta+2} < \infty \text{ for some } \delta > 0.$$

The following theorem shows that our risk estimates estimate the true loss uniformly well.

THEOREM 1.4. *Under Assumptions A1-A2, for all $\boldsymbol{\theta}$ satisfying Assumption A3′ and for all estimates $\hat{\boldsymbol{q}}(\eta,\tau) \in \mathcal{S}$,*

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty],\, \eta\in\mathbb{R}:\, |\eta|\le M_n} \mathbb{E}\left| \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta,\tau) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta,\tau)) \right| = 0 \ .$$

Based on the above theorem, we derive the decision theoretic optimality of our proposed estimator. The following two theorems show that our estimator is asymptotically nearly as good as the oracle loss estimator, whereas the corollary shows that it is as good as any other estimator in $\mathcal{S}$.

THEOREM 1.5. *Under Assumptions A1-A2, and for all $\boldsymbol{\theta}$ satisfying Assumption A3′, we have, for any fixed $\epsilon > 0$,*

$$\lim_{n\to\infty} P\left\{ L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}})\big) \ge L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}})\big) + \epsilon \right\} = 0 \ .$$

THEOREM 1.6. *Under Assumptions A1-A2 and for all $\boldsymbol{\theta}$ satisfying Assumption A3′,*

$$\lim_{n\to\infty} R_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}})\big) - \mathbb{E}\big[L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}})\big)\big] = 0.$$

COROLLARY 1.2. *Under Assumptions A1-A2, for all $\boldsymbol{\theta}$ satisfying Assumption A3′ and for any estimator $\hat{\tau}_n \ge 0$ and $|\hat{\eta}_n| \le M_n$,*

I. $\lim_{n\to\infty} P\left\{ L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}})\big) \ge L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n, \hat{\tau}_n)\big) + \epsilon \right\} = 0$ *for any fixed* $\epsilon > 0$.

II. $\lim_{n\to\infty} R_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}})\big) - R_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n, \hat{\tau}_n)\big) \le 0.$

The EBML estimate of the hyperparameters are given by

$$\hat{\tau}_n^{\mathrm{ML}} = \operatorname*{arg\,min}_{\tau\in[0,\infty]} \frac{1}{n}\sum_{i=1}^{n} \left(\frac{(X_i - f(\tau))^2}{\tau + \nu_{p,i}} + \log(\tau + \nu_{p,i}) \right) \quad \text{and} \quad \hat{\eta}_n^{\mathrm{ML}} = f(\hat{\tau}_n^{\mathrm{ML}}) \ ,$$

where

$$f(\tau) = \left(\sum_{i=1}^{n} (\tau + \nu_{p,i})^{-1} X_i \right) \Big/ \left(\sum_{i=1}^{n} (\tau + \nu_{p,i})^{-1} \right) \ ,$$

and the method of moments (MM) estimates are roots of the following equations:

$$\tau = \frac{1}{n-1}\left(\sum_{i=1}^{n} \big(X_i - \eta\big)^2 - (1 - 1/n)\nu_{p,i} \right)_+ \text{ and } \eta = f(\tau).$$

In both the cases, the location estimate lies in $[-M_n, M_n]$. Also, unlike $(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}})$, the EBML and EBMM estimates of the hyperparameters do not depend on the shape of the loss functions $\{(b_i, h_i) : 1 \le i \le n\}$. So, the EBML and EBMM estimators $\hat{\boldsymbol{q}}(\hat{\eta}^{\mathrm{ML}}, \hat{\tau}^{\mathrm{ML}})$ and $\hat{\boldsymbol{q}}(\hat{\eta}^{\mathrm{MM}}, \hat{\tau}^{\mathrm{MM}})$ do not always agree with the asymptotically efficient ARE based estimator $\hat{\boldsymbol{q}}(\hat{\eta}^{\mathrm{D}}, \hat{\tau}^{\mathrm{D}})$ and cannot be generally asymptotically optimal in $\mathcal{S}$.

Results on Estimators in $\mathcal{S}^G$. Following (1.4), the class of estimators with shrinkage towards the Grand Mean ($\bar{\boldsymbol{X}}$) of the past observations is of the following form: for $i = 1, \ldots, n$,

$$\hat{q}_i^G(\tau) = \alpha_i(\tau) X_i + (1 - \alpha_i(\tau))\bar{\boldsymbol{X}} + (\nu_{f,i} + \alpha_i(\tau)\nu_{p,i})^{1/2}\, \Phi^{-1}(\tilde{b}_i)\ , \tag{1.13}$$

where τ varies over 0 to ∞, and $\alpha_i(\tau)$, and $\tilde{b}_i$ are defined just below Equation (1.4). For any fixed τ, unlike estimators in $\mathcal{S}$, $\hat{\boldsymbol{q}}^G(\tau)$ is no longer a coordinatewise independent rule. In Section 5, we develop an estimation strategy which estimates the loss of estimators in $\mathcal{S}^G$ uniformly well.

THEOREM 1.7. *Under Assumptions A1-A2, for all $\boldsymbol{\theta}$ satisfying Assumption A3′ and for all estimates $\hat{\boldsymbol{q}}^{\mathrm{G}}(\tau) \in \mathcal{S}^G$,*

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \mathbb{E}\left| \widehat{\mathrm{ARE}}_n^{\mathrm{G}}(\tau) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) \right| = 0\ .$$

We propose an estimate $\hat{\tau}_n^{\mathrm{ARE}^{\mathrm{G}}} = \arg\min_{\tau\in[0,\infty]} \widehat{\mathrm{ARE}}_n^{\mathrm{G}}(\tau)$ for the hyperparameter in this class and compare its asymptotic behavior with the corresponding oracle loss $\tau_n^{\mathsf{GOR}} = \arg\min_{\tau\in[0,\infty]} L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau))$. Like the other two classes, based on Theorem 1.7, here we also derive the asymptotic optimality of our proposed estimate in terms of both the predictive risk and loss.

THEOREM 1.8. *Under Assumptions A1-A2, for all $\boldsymbol{\theta}$ satisfying Assumption A3′*

(A) comparing with the oracle loss estimator, we have the following:

I. $\lim_{n\to\infty} \mathbb{P}\left\{ L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\hat{\tau}_n^{\mathrm{ARE}^{\mathrm{G}}})\big) \ge L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau_n^{\mathsf{GOR}})\big) + \epsilon \right\} = 0$ *for any fixed* $\epsilon > 0$.

II. $\lim_{n\to\infty} R_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\hat{\tau}_n^{\mathrm{ARE}^{\mathrm{G}}})\big) - \mathbb{E}\left[L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau_n^{\mathsf{GOR}})\big)\right] = 0.$

(B) for any estimate $\hat{\tau}_n \ge 0$ of the hyperparameter, we have the following:

I. $\lim_{n\to\infty} \mathbb{P}\left\{ L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\hat{\tau}_n^{\mathrm{ARE}^{\mathrm{G}}})\big) \ge L_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\hat{\tau}_n)\big) + \epsilon \right\} = 0$ *for any fixed* $\epsilon > 0$.

II. $\lim_{n\to\infty} R_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\hat{\tau}_n^{\mathrm{ARE}^{\mathrm{G}}})\big) - R_n\big(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\hat{\tau}_n)\big) \le 0\,.$

1.6. *Organization of the Paper.* In Section 2, we provide a detailed explanation of the results involving the class of estimators $\mathcal{S}^0$. Treating this class as the fundamental case, through the proof of Theorem 1.1, Section 2 explains the general

principle behind our asymptotic risk estimation methodology and the proof techniques used in this paper. The proofs of Theorems 1.2 and 1.3 and Corollary 1.1 are provided in Appendix A. Section 3 discusses the performance of our prediction methodology in simulation experiments. Section 4 and its associated Appendix B provide the proofs of Theorems 1.4, 1.5 and 1.6 and Corollary 1.2, which deal with estimators in class $\mathcal{S}$. The proofs of Theorems 1.7 and 1.8 involving class $\mathcal{S}^G$ are provided in Section 5 and Appendix C.

1.7. *Glossary.* In Table 1, we briefly list the notations that have been used repeatedly in the current paper. As a convention, multivariate vectors, expressions and estimates are represented in bold.

TABLE 1
List of important notations used in the current paper.

Notation	Description
n	number of products
i	product index
θ_i	*unknown* mean demand of product i
$\nu_{p,i}$	variance of the past demand of product i
$\nu_{f,i}$	variance of the future demand of product i
X_i	past demand data for product i with $X_i \sim N(\theta_i, \nu_{p,i})$
Y_i	future demand of product i with $Y_i \sim N(\theta_i, \nu_{f,i})$
b_i	per-unit lost sales cost associated with product i
h_i	per-unit holding cost associated with product i
$\tilde{b}_i$	the critical ratio $b_i/(b_i+h_i)$ of the lost sales and holding costs
$l_i(\theta_i, \hat{q}_i(\boldsymbol{x}))$	loss associated with product i under the policy $\hat{q}_i$ when $\boldsymbol{X}=\boldsymbol{x}$ is observed
$L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}})$	average loss over n products under the prediction policy $\hat{q}$
$r_i(\theta_i, \hat{q}_i)$	risk associated with product i under the prediction policy $\hat{q}_i$
$R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}})$	average risk of n products under the prediction policy $\hat{q}$
(η, τ)	hyperparameters for the prior distribution of θ_i, with $\theta_i \sim N(\eta, \tau)$
$\hat{q}_i^{\mathsf{Bayes}}(\eta, \tau)$	bayes estimate for $N(\eta, \tau)$ prior; see (1.4) for definition
$\alpha_i(\tau)$	shrinkage factor in our estimates, with $\alpha_i(\tau) = \tau/(\tau+\nu_{p,i})$ for all i
$\mathcal{S}^0$	class of shrinkage estimators $\hat{\boldsymbol{q}}(\tau)$ based on origin-centric priors
$\mathcal{S}^G$	class of shrinkage estimators $\hat{\boldsymbol{q}}^G(\tau)$ based on grand-mean centric priors
$\mathcal{S}$	class of data driven shrinkage estimators $\hat{\boldsymbol{q}}(\eta, \tau)$
$\widehat{\mathrm{ARE}}$	our proposed estimate of the risk function $R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))$ of estimators in $\mathcal{S}^0$
$\widehat{\mathrm{ARE}}^{\mathrm{G}}$	our proposed estimate of the risk function $R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^G(\tau))$ of estimators in $\mathcal{S}^G$
$\widehat{\mathrm{ARE}}^{\mathrm{D}}$	our proposed estimate of the risk function $R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta, \tau))$ of estimators in $\mathcal{S}$
τ_n^{OR}, τ_n^{GOR}	Oracle estimates of the hyperparameter τ for $\mathcal{S}^0$ and $\mathcal{S}^G$, respectively
$\hat{\tau}_n^{\mathrm{ARE}}$, $\hat{\tau}_n^{\mathrm{ARE^G}}$	ARE-based estimate of τ for $\mathcal{S}^0$ and $\mathcal{S}^G$, respectively
$(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}})$	oracle estimates of the hyperparameter in $\mathcal{S}$; ARE estimates are $\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}}$
$\hat{T}_i(X_i, \tau)$	coordinate-wise estimator used in our risk estimation method; see (1.10)
$\lambda_n(i)$	threshold parameter used in our risk estimation method; see (1.11)
$K_n(i)$	truncation parameter used in our risk estimation method; see (1.12)
$G(\omega, \beta)$	describes the newsvendor's predictive loss function; see (1.3)
$\mathcal{O}(\cdot)$, $o(\cdot)$	denote the Big O and the little-o mathematical notations, respectively

2. Proof of Theorem 1.1 and Explanation of the ARE Method. In this section, we provide a detailed explanation of the results on the estimators in $\mathcal{S}^0$. This case serves as a fundamental building block and contains all the essential

ingredients involved in the general risk estimation method. In subsequent sections, the procedure is extended to $\mathcal{S}$ and $\mathcal{S}^G$. We begin by laying out the proof of Theorem 1.1. The *decision theoretic optimality results* – Theorems 1.2 and 1.3 and Corollary 1.1 – follow easily from Theorem 1.1; their proofs are provided in Appendix A.

2.1. *Proof of Theorem 1.1.* The proof of Theorem 1.1 makes use of Lemma 2.1, which shows that, as the number of product increases, the loss function $L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))$ uniformly concentrates around its expected value, which is the risk $R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))$. To prove this result, we use the fact that the mean demand is bounded (Assumption A3) and apply the uniform SLLN argument (Newey and McFadden, 1994, Lemma 2.4) to establish the desired concentration. The detailed proof is in the Appendix A.

LEMMA 2.1 (Uniform Concentration of the Loss around the Risk). *Under Assumption A1, for any $\boldsymbol{\theta}$ obeying Assumption A3,*

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \mathbb{E}\big|R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))\big| = 0 \ .$$

Using Lemma 2.1 and Proposition 1.1, we can now give the proof of the main result.

Proof of Theorem 1.1. Let $\widehat{\mathrm{ARE}}_n(\boldsymbol{Z}, \tau)$ denote a randomized risk estimate before the Rao-Blackwellization step in Section 1.3. For any fixed τ, $\{\hat{T}_i(X_i, Z_i, \tau) : 1 \le i \le n\}$ are independent of each other, so the Bias-Variance decomposition yields

$$\begin{aligned} &\mathbb{E}\left[\big(R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}_n(\boldsymbol{Z}, \tau)\big)^2\right] \\ &\le A_n\left\{\left(\frac{1}{n}\sum_{i=1}^n \mathrm{Bias}(T_i(X_i, Z_i, \tau))\right)^2 + \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(T_i(X_i, Z_i, \tau))\right\} , \end{aligned} \tag{2.1}$$

where $A_n = \sup\{(b_i + h_i)^2(\nu_{f,i} + \alpha_i(\tau)\nu_{p,i}) : i = 1, \ldots, n\}$ and $\alpha_i(\tau) = \tau/(\tau + \nu_{p,i})$. By Assumption A2, $\sup_n A_n < \infty$. From Proposition 1.1, both terms on the right hand side above uniformly converge to 0 as $n \to \infty$. This shows that

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \mathbb{E}\big[\big(R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}_n(\boldsymbol{Z}, \tau)\big)^2\big] \ = \ 0 \ ,$$

where the expectation is over the distribution of $\boldsymbol{Z}$ and $\boldsymbol{X}$. As $\widehat{\mathrm{ARE}}_n(\tau) = \mathbb{E}[\widehat{\mathrm{ARE}}_n(\boldsymbol{Z}, \tau)|X]$, using Jensen's inequality for conditional expectation, we have $\mathbb{E}[(R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}_n(\boldsymbol{Z}, \tau))^2] \ge \mathbb{E}[(R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}_n(\tau))^2]$ for any n, $\boldsymbol{\theta}$ and τ. Thus,

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \mathbb{E}\big[\big(R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}_n(\tau)\big)^2\big] \ = \ 0 \ .$$

Again, by the triangle and Jensen inequalities,

$$\sup_{\tau\in[0,\infty]} \mathbb{E}\left|L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}_n(\tau)\right| \le \sup_{\tau\in[0,\infty]} \mathbb{E}\left|L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\tau)) - R_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\tau))\right| + \sqrt{\sup_{\tau\in[0,\infty]} \mathbb{E}\left[\left(R_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}_n(\tau)\right)^2\right]}\ .$$

The first term on the right hand side converges to zero by Lemma 2.1 and the second term follows from the above. This completes the proof. □

Thus, to complete the proof, it remains to establish Proposition 1.1, which shows that both the bias and variance converge to zero as the number of products increases. We undertake this in the next section. Understanding how the bias and variance is controlled will help the reader to appreciate the elaborate construction process of ARE estimates and our prescribed choices of the threshold parameter $\lambda_n(i)$ and truncation parameter $K_n(i)$.

2.2. *Proof of Proposition 1.1 Overview and Reduction to the Univariate Case.* In this section, we outline the overview of the proof techniques used to establish Proposition 1.1. It suffices to consider a generic univariate setting and consider each product individually. This will simplify our analysis considerably. In addition, we will make use of the following two results about the property of the loss function G. The proof of these lemmas are given in Appendix A.

LEMMA 2.2 (Formula for the Loss Function). *If $Y \sim N(\theta,\nu)$, then*

$$\mathbb{E}_\theta\left[b(Y-q)^+ + h(q-Y)^+\right] = (b+h)\sqrt{\nu}\, G\left((q-\theta)/\sqrt{\nu}\,,\,\tilde{b}\right)\ , \tag{2.2}$$

where $\tilde{b}$=b/(b+h) and for all $w\in\mathbb{R}$ and $\beta\in[0,1]$, $G(w,\beta)=\phi(w)+w\Phi(w)-\beta w$. Also, if θ is known, the loss $l(\theta,q)$ is minimized at $q=\theta+\sqrt{\nu}\Phi^{-1}(\tilde{b})$ and the minimum value is $(b+h)\sqrt{\nu}\phi(\Phi^{-1}(\tilde{b}))$.

The next lemma gives an explicit formula for the Bayes estimator and the corresponding Bayes risk in the univariate setting.

LEMMA 2.3 (Univariate Bayes Estimator). *Consider the univariate prediction problem where the past sales $X\sim N(\theta,\nu_p)$, the future demand $Y\sim N(\theta,\nu_f)$ and $\theta\sim N(\eta,\tau)$. Consider the problem of minimizing the integrated Bayes risk. Then,*

$$\min_q \int R(\theta,q)\pi_{(\eta,\tau)}(\theta\,|\,x)d\theta \;=\; (b+h)\sqrt{\nu_f+\alpha\nu_p}\;\phi(\Phi^{-1}(\tilde{b}))\ ,$$

where $\tilde{b}=b/(b+h)$, $\alpha=\tau/(\tau+\nu_p)$, and $\pi_{(\eta,\tau)}(\theta\mid x)$ is the posterior density of θ given $X=x$. Also, the Bayes estimate $\hat{q}^{\mathsf{Bayes}}(\eta,\tau)$ that achieves the above minimum is given by

$$\hat{q}^{\mathsf{Bayes}}(\eta,\tau) = \alpha x + (1-\alpha)\eta + \sqrt{\nu_f+\alpha\nu_p}\,\Phi^{-1}(\tilde{b})\ .$$

Finally, the risk $r(\theta, \hat{q}^{\mathsf{Bayes}}(\eta,\tau))$ *of the Bayes estimator is*

$$(b+h)\sqrt{\nu_f + \alpha^2\nu_p}\, G(c_\tau + d_\tau\,(\theta-\eta), \tilde{b})\ ,$$

where

$$c_\tau = \sqrt{(1+\alpha\nu_p)/(1+\alpha^2\nu_p)}\,\Phi^{-1}(\tilde{b}) \qquad \textit{and} \qquad d_\tau = -(1-\alpha)/\sqrt{\nu_f+\alpha^2\nu_p}.$$

By Lemma 2.2, note that the loss function is scalable in ν_f. Also by Lemma 2.3, we observe that the risk calculation depends only on the ratio ν_p/ν_f and scales with $b+h$. Thus, without loss of generality, henceforth we will assume that $\nu_f = 1$, $b+h=1$ and write $\nu = \nu_p$ and $\tilde{b} = b/(b+h) = b$. As a convention, for any number $\beta \in [0,1]$, we write $\bar{\beta} = 1-\beta$.

Reparametrization and some new notations. In order to prove the desired result, we will work with generic univariate risk estimation problems where $X \sim N(\theta,\nu)$ and $Y \sim N(\theta,1)$. Note that Assumption A2 requires that $\nu < 1/(4e)$. For ease of presentation, we restate and partially reformulate the univariate version of the methodology stated in Section 1.3. We conduct sample splitting by adding independent Gaussian noise Z:

$$U = X + \sqrt{\nu}Z, \quad V = X - \sqrt{\nu}Z.$$

Instead of $\tau \in [0,\infty]$, we reparameterize the problem using $\alpha = \tau/(\tau+\nu) \in [0,1]$. By Lemma 2.3 and the fact that $\nu_f = 1$ and $b+h=1$, the univariate risk function (with $\eta = 0$) is given by $\alpha \mapsto G(c_\alpha + d_\alpha\theta, b)$, where $b < 1$ and

$$c_\alpha = \Phi^{-1}(b)\sqrt{(1+\alpha\nu)/(1+\alpha^2\nu)} \qquad \text{and} \qquad d_\alpha = -\bar{\alpha}/\sqrt{1+\alpha^2\nu}\ .$$

Now, consider $U_\alpha = c_\alpha + d_\alpha U, V_\alpha = c_\alpha + d_\alpha V$ and $\theta_\alpha = c_\alpha + d_\alpha\theta$. By construction $(U_\alpha, V_\alpha) \sim N(\theta_\alpha, \theta_\alpha, 2\nu d_\alpha^2, 2\nu d_\alpha^2, 0)$ and $\alpha \mapsto G(\theta_\alpha, b)$ is estimated by the ARE estimator $\alpha \mapsto \hat{T}_{\alpha,n}(X,Z)$, where

$$\hat{T}_{\alpha,n}(X,Z) = -b\,U_\alpha\,\mathbb{I}_{\{V_\alpha < -\lambda_n\}} + \tilde{S}(U_\alpha)\,\mathbb{I}_{\{|V_\alpha| \le \lambda_n\}} + \bar{b}\,U_\alpha\,\mathbb{I}_{\{V_\alpha > \lambda_n\}}\ ,$$

where $\bar{b} = 1-b$, and the threshold is given $\lambda_n = \gamma\sqrt{2\log n}$, where γ is any positive number less than $\sqrt{2\nu}\left((1/\sqrt{4e\nu}) - 1\right) = (1/\sqrt{2e}) - \sqrt{2\nu}$, which is well-defined by Assumption A2 because $\nu < 1/(4e)$.

The estimator $\tilde{S}(U_\alpha)$ is the truncated Taylor series expansion of $G(\theta_\alpha, b)$, defined as follows. Let

$$K_n = 1 + \left\lceil e^2\left(\gamma + \sqrt{2\nu}\right)^2 (2\log n)\right\rceil\ .$$

Let $G_{K_n}(\theta_\alpha, b)$ denote the the K_n^{th} order Taylor series expansion of $G(\theta_\alpha, b)$. Let $S(U_\alpha)$ denote an unbiased estimate of $G_{K_n}(\theta_\alpha, b)$; that is,

$$\begin{aligned} S(U_\alpha) &= G(0,b) + G'(0,b)\,U_\alpha \\ &\quad + \phi(0)\sum_{l=0}^{K_n-2} \frac{(-1)^l H_l(0)}{(l+2)!}\left(\sqrt{2\nu d_\alpha^2}\right)^{l+2} H_{l+2}\left(\frac{U_\alpha}{\sqrt{2\nu d_\alpha^2}}\right)\ , \end{aligned} \tag{2.3}$$

and finally, we have $\tilde{S}(U_\alpha) = \text{sign}(S(U_\alpha)) \min\{|S(U_\alpha)|, n\}$, which is the truncated version of $S(U_\alpha)$. This completes the definition of the estimator $\hat{T}_{\alpha,n}(X, Z)$. This reparametrization allows us to deal with the stochasticity of the problem only through the random variables $\{U_\alpha, V_\alpha : \alpha \in [0,1]\}$ and saves us the inconvenience of dealing with the varied functionals of X and Z separately.

Proof Outline. We partition the univariate parameter space into 3 cases: `Case 1:` $|\theta_\alpha| \le \lambda_n/2$, `Case 2:` $\lambda_n/2 < |\theta_\alpha| \le \big(1 + \sqrt{2\nu}/\gamma\big)\lambda_n$ and `Case 3:` $|\theta_\alpha| > \big(1 + \sqrt{2\nu}/\gamma\big)\lambda_n$. We present a heuristic argument for considering such a decomposition. The following lemma, whose proof is provided in Appendix A, establishes a bound on the bias in different regimes.

LEMMA 2.4 (Bias Bounds). *There is an absolute constant c such that for all $b \in [0,1]$ and $\alpha \in [0,1]$,*

$$\text{I.} \quad |G(y,b) - G_{K_n}(y,b)| \le c\,\frac{n^{-(e^2-1)(\gamma+\sqrt{2\nu})^2}}{e^4(\gamma+\sqrt{2\nu})^2} \quad \text{for all} \quad |y| \le \big(1+\sqrt{2\nu}/\gamma\big)\lambda_n \ .$$

$$\text{II.} \quad |G(y,b) - \bar{b}y| \le \frac{e^{-y^2/2}}{y^2} \quad \text{for all } y > 0 \ .$$

$$\text{III.} \quad |G(y,b) - (-by)| \le \frac{e^{-y^2/2}}{y^2} \quad \text{for all } y < 0 \ .$$

So, we would like to use linear estimates when $|w|$ is large and $S(U_\alpha)$ otherwise. The choice of threshold λ_n is chosen such that this happens with high probability. As we have a normal model in Case 3, which includes unbounded parametric values, we will be mainly using the linear estimates of risk because when $|\theta_\alpha| \ge \big(1 + \sqrt{2\nu}/\gamma\big)\,\lambda_n$, the probability of selecting $\tilde{S}$ over the linear estimates is very low. Similarly, in Case 1, we will be mainly using $\tilde{S}$. Case 2 is the buffering zone where we may use either $\tilde{S}$ or the linear estimates.

We also need to control the variances of the 3 different kind of estimates used in $\hat{T}_{\alpha,n}(X, Z)$. While the variances of the linear estimators are easily controlled, we needed to pay special attention to control the variance of $S(U_\alpha)$. In the following lemma, we exhibit an upper bound on the quadratic growth of the estimator $S(U_\alpha)$. The choice of the truncation parameter K_n in $\tilde{S}(U_\alpha)$ was done is such a way that both its bias and squared growth are controlled at the desired limits.

LEMMA 2.5 (Variance Bounds). *For any $b \in [0,1]$,*

$$\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha| \le (1+\sqrt{2\nu}/\gamma)\lambda_n} \frac{\mathbb{E}_{\theta_\alpha}\left[S(U_\alpha)^2\right]}{n} = 0 \ ,$$

where the expection is over the distribution of U_α, which has $N(\theta_\alpha, 2\nu d_\alpha^2)$ distribution for all $\alpha \in [0,1]$.

Our proof also makes use of the following large deviation bounds.

LEMMA 2.6 (Large Deviation Bounds).

For `Case 1`,
$$\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha|\,\le\,\lambda_n/2} \lambda_n^2 \cdot \mathbb{P}_{\theta_\alpha}\{|V_\alpha| > \lambda_n\} = 0\ .$$

For `Case 2`,
$$\lim_{n\to\infty} \sup_{\alpha\,:\,\lambda_n/2\,<\,\theta_\alpha\,\le\,(1+\sqrt{2\nu}/\gamma)\lambda_n} |\theta_\alpha| \cdot \mathbb{P}_{\theta_\alpha}\{V_\alpha < -\lambda_n\} = 0\,.$$
$$\lim_{n\to\infty} \sup_{\alpha\,:\,-(1+\sqrt{2\nu}/\gamma)\lambda_n\,\le\,\theta_\alpha\,<\,-\lambda_n/2} |\theta_\alpha| \cdot \mathbb{P}_{\theta_\alpha}\{V_\alpha > \lambda_n\} = 0\,.$$

For `Case 3`,
$$\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha|\,>\,\left(1+\sqrt{2\nu}/\gamma\right)\lambda_n} n \cdot \mathbb{P}\{|V_\alpha| \le \lambda_n\} \quad = 0\,.$$
$$\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha|\,>\,\left(1+\sqrt{2\nu}/\gamma\right)\lambda_n} \theta_\alpha^2 \cdot \mathbb{P}\{|V_\alpha| \le \lambda_n\} \quad = 0\,.$$
$$\lim_{n\to\infty} \sup_{\alpha\,:\,\theta_\alpha\,>\,(1+\sqrt{2\nu}/\gamma)\lambda_n} \theta_\alpha^2 \cdot \mathbb{P}_{\theta_\alpha}\{V_\alpha < -\lambda_n\} \ = 0\,.$$
$$\lim_{n\to\infty} \sup_{\alpha\,:\,\theta_\alpha\,<\,-(1+\sqrt{2\nu}/\gamma)\lambda_n} \theta_\alpha^2 \cdot \mathbb{P}_{\theta_\alpha}\{V_\alpha > \lambda_n\} \ = 0\,.$$

The proofs of the above three lemmas are presented in Appendix A.

2.3. *Detailed Proof of Proposition 1.1.*

Bounding the Bias: As $\mathbb{E}[U_\alpha] = \theta_\alpha$, by definition $|\mathrm{Bias}_{\theta_\alpha}(\hat{T}_{\alpha,n})|$ equals
$$\left|\mathbb{E}[\tilde{S}(U_\alpha)] - G(\theta_\alpha, b)\right| \cdot \mathbb{P}\{|V_\alpha| \le \lambda_n\} + |\mathbb{E}[\tilde{S}(U_\alpha)] - G(\theta_\alpha, b)| \cdot \mathbb{P}\{|V_\alpha| > \lambda_n\}\ .$$

We will now show that each of the two terms on the RHS converges uniformly in α as n increases to infinity.

First Term: Consider θ_α in `Cases 1 and 2`; that is, $|\theta_\alpha| \le \left(1+\sqrt{2\nu}/\gamma\right)\lambda_n$. Since $\mathbb{E}S(U_\alpha) = G_{K_n}(\theta_\alpha, b)$, by our construction, we have that
$$|\mathbb{E}\tilde{S}(U_\alpha) - G(\theta_\alpha, b)| \le |\mathbb{E}\tilde{S}(U_\alpha) - \mathbb{E}S(U_\alpha)| + |G_{K_n}(\theta_\alpha, b) - G(\theta_\alpha, b)|,$$
and it follows from Lemma 2.4 that $\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha|\le\left(1+\sqrt{2\nu}/\gamma\right)\lambda_n} |G_{K_n}(\theta_\alpha, b) - G(\theta_\alpha, b)| = 0$. By Markov's Inequality,
$$|\mathbb{E}\tilde{S}(U_\alpha) - \mathbb{E}S(U_\alpha)| \le \mathbb{E}\left[|S(U_\alpha)|\,\mathbb{I}_{\{|S(U_\alpha)|\ge n\}}\right] \le \mathbb{E}\left[S^2(U_\alpha)\right]/n\ ,$$
which converges to zero uniformly in α as $n \to \infty$ by Lemma 2.5.

Now, consider `Case 3`, where $|\theta_\alpha| > \left(1+\sqrt{2\nu}/\gamma\right)\lambda_n$. By definition, $|\tilde{S}(U_\alpha)| \le n$, and by Lemma D.5, $G(\theta_\alpha, b) \le \phi(0) + \max\{\bar{b}, b\}|\theta_\alpha|$. From Lemma 2.6, we have that $\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha|\,>\,\left(1+\sqrt{2\nu}/\gamma\right)\lambda_n} \max\{n, \theta_\alpha^2\} \cdot \mathbb{P}\{|V_\alpha| \le \lambda_n\} = 0$, and thus,
$$\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha|\,>\,\left(1+\sqrt{2\nu}/\gamma\right)\lambda_n} \left|\mathbb{E}[\tilde{S}(U_\alpha)] - G(\theta_\alpha, b)\right| \cdot \mathbb{P}\{|V_\alpha| \le \lambda_n\} = 0\ .$$

So, in all three cases, the first term of the bias converges to zero.

Second Term: The second term in the bias formula is equal to

$$B_{\alpha,n} \equiv |\bar{b}\theta_\alpha - G(\theta_\alpha, b)| \cdot \mathbb{P}\{V_\alpha > \lambda_n\} + |G(\theta_\alpha, b) - (-b\theta_\alpha)| \cdot \mathbb{P}\{V_\alpha < -\lambda_n\} \ .$$

For θ_α in `Case 1` with $|\theta_\alpha| \le \lambda_n/2$, note that by Lemma D.5,

$$\max\{|\bar{b}\theta_\alpha - G(\theta_\alpha, b)| \ , \ |G(\theta_\alpha, b) - (-b\theta_\alpha)|\} \le |\theta_\alpha| + \phi(0) + |\theta_\alpha| \le \lambda_n + \phi(0) \ ,$$

and thus $B_{\alpha,n} \le (\lambda_n + \phi(0))\,\mathbb{P}\{|V_\alpha| > \lambda_n\}$. The desired result then follows from Lemma 2.6 for Case 1.

Now, consider θ_α in `Case 2`; that is, $\lambda_n/2 < |\theta_\alpha| \le (1+\sqrt{2\nu}/\gamma)\lambda_n$. We will assume that $\lambda_n/2 < \theta_\alpha \le (1+\sqrt{2\nu}/\gamma)\lambda_n$; the case $-(1+\sqrt{2\nu}/\gamma)\lambda_n < \theta_\alpha < -\lambda_n/2$ follows analogously. Since $\theta_\alpha > \lambda_n/2$, it follows from Lemma 2.4 that

$$|\bar{b}\theta_\alpha - G(\theta_\alpha, b)| \le e^{-\theta_\alpha^2/2}/\theta_\alpha^2 \le 4\, e^{-\lambda_n^2/8}/\lambda_n^2 = 4\, n^{-\gamma^2/4}/\lambda_n^2 \ .$$

Also, by Lemma D.5, $|G(\theta_\alpha, b) - (-b\theta_\alpha)| \le 2|\theta_\alpha| + \phi(0)$. So,

$$B_{\alpha,n} \le 4\, c\, n^{-\gamma^2/4}/\lambda_n^2 \ + \ (2|\theta_\alpha| + \phi(0))\mathbb{P}\{V_\alpha < -\lambda_n\} \ ,$$

and the desired result then follows from Lemma 2.6 for `Case 2`.

Now, consider θ_α in `Case 3`; that is, $|\theta_\alpha| > (1+\sqrt{2\nu}/\gamma)\lambda_n$. We will assume that $\theta_\alpha > (1+\sqrt{2\nu}/\gamma)\lambda_n$; the case $\theta_\alpha < -(1+\sqrt{2\nu}/\gamma)\lambda_n$ follows analogously. As before, it follows from Lemma 2.4 that

$$|\bar{b}\theta_\alpha - G(\theta_\alpha, b)| \le c\, e^{-(1+\sqrt{2\nu}/\gamma)^2\lambda_n^2/2}/\big((1+\sqrt{2\nu}/\gamma)^2\lambda_n^2\big) = c\, n^{-(\gamma+\sqrt{2\nu})^2}/\big((\gamma+\sqrt{2\nu})^2(2\log n)\big) \ .$$

By Lemma D.5, $|G(\theta_\alpha, b) - (-b\theta_\alpha)| \le 2|\theta_\alpha| + \phi(0)$. So,

$$B_{\alpha,n} \le \frac{c\, n^{-(\gamma+\sqrt{2\nu})^2}}{(\gamma+\sqrt{2\nu})^2(2\log n)} + (2|\theta_\alpha| + \phi(0))\mathbb{P}\{V_\alpha < -\lambda_n\} \ ,$$

and the desired result then follows from Lemma 2.6 for `Case 3`. Note that in `Case 3`, $|\theta_\alpha| \le \theta_\alpha^2$ for sufficiently large n. So, in all three cases, the first term of the bias converges to zero.

Bounding the Variance: According to the definition of $\hat{T}_{\alpha,n}$, it follows from Lemma D.10 that

$$\begin{aligned} &\mathrm{Var}_{\theta_\alpha}(\hat{T}_{\alpha,n}) \le 4\mathrm{Var}(A^1_{\alpha,n}) + 4\mathrm{Var}(A^2_{\alpha,n}) + 4\mathrm{Var}(A^3_{\alpha,n}), \text{ where} \\ &A^1_{\alpha,n} = \tilde{S}(U_\alpha)\mathbb{I}_{\{|V_\alpha|<\lambda_n\}}, \ A^2_{\alpha,n} = -bU_\alpha\mathbb{I}_{\{V_\alpha<-\lambda_n\}}, \text{ and } A^3_{\alpha,n} = \bar{b}U_\alpha\mathbb{I}_{\{V_\alpha>\lambda_n\}}. \end{aligned} \tag{2.4}$$

To establish the desired result, we will show that each term on the RHS is $o(n)$ uniformly in α; that is, for $i = 1, 2, 3$, $\lim_{n\to\infty} n^{-1} \sup_{\alpha\in[0,1]} \mathrm{Var}(A^i_{\alpha,n}) = 0$.

`Case 1:` $|\theta_\alpha| \le \lambda_n/2$. Since $\tilde{S}(U_\alpha) = \text{sign}(S(U_\alpha))\ \min\{|S(U_\alpha)|, n\}$, it follows from Lemma D.3 that $\text{Var}\left(A^1_{\alpha,n}\right) \le \mathbb{E}_{\theta_\alpha}\tilde{S}^2(U_\alpha) \le \mathbb{E}_{\theta_\alpha}S^2(U_\alpha) = o(n)$, where the last equality follows from Lemma 2.5. Again, by Lemma D.3,

$$\begin{aligned}
&\text{Var}(A^2_{\alpha,n}) + \text{Var}(A^3_{\alpha,n}) \\
&\le b^2\mathbb{E}\left[U^2_\alpha\right]\cdot\mathbb{P}\{V_\alpha < -\lambda_n\} + \bar{b}^2\left[U^2_\alpha\right]\cdot\mathbb{P}\{V_\alpha > \lambda_n\} \le \mathbb{E}\left[U^2_\alpha\right]\mathbb{P}\{|V_\alpha| > \lambda_n\} \\
&= \left(\theta^2_\alpha + 2\nu d^2_\alpha\right)\mathbb{P}\{|V_\alpha| > \lambda_n\} \le \left(\lambda^2_n/4 + 2\nu\right)\mathbb{P}\{|V_\alpha| > \lambda_n\}\ ,
\end{aligned}$$

where the equality follows from the definition of U_α. The desired result then follows from Lemma 2.6 for `Case 1`.

`Case 2:` $\lambda_n/2 < |\theta_\alpha| \le (1+\sqrt{2\nu}/\gamma)\lambda_n$. Suppose that $\lambda_n/2 < \theta_\alpha \le (1+\sqrt{2\nu}/\gamma)\lambda_n$; the proof for the case where $-(1+\sqrt{2\nu}/\gamma)\lambda_n \le \theta_\alpha < -\lambda_n/2$ is the same. By Lemma D.3,

$$\text{Var}(A^1_{\alpha,n}) \le \mathbb{E}\tilde{S}^2(U_\alpha) \le \mathbb{E}S^2(U_\alpha) = o(n)\ ,$$

where the equality follows from Lemma 2.5. By Lemma D.3,

$$\text{Var}(A^2_{\alpha,n}) \le b^2\,\mathbb{E}\left[U^2_\alpha\right]\mathbb{P}\{V_\alpha < -\lambda_n\} \le \left(2\nu + \theta^2_\alpha\right)\mathbb{P}\{V_\alpha < -\lambda_n\}\ .$$

For the range of θ_α in `Case 2`, $\theta_\alpha/n \to 0$ uniformly in α, and it follows that

$$\lim_{n\to\infty}\ \sup_{\alpha\,:\,\lambda_n/2\,<\,\theta_\alpha\,\le\,(1\,+\,\sqrt{2\nu}/\gamma)\lambda_n} n^{-1}\text{Var}(A^2_{\alpha,n}) = 0\,,$$

where the equality follows from Lemma 2.6 for `Case 2`. Note that $\text{Var}(A^3_{\alpha,n}) \le 4\mathbb{E}[b^2U^2_\alpha]\mathbb{P}\{V_\alpha < -\lambda_n\} \le 4\mathbb{E}[b^2U^2_\alpha] \le 4\left(2\nu + \theta^2_\alpha\right) = o(n)$ uniformly in α.

`Case 3:` $|\theta_\alpha| > (1+\sqrt{2\nu}/\gamma)\lambda_n$. Note that

$$\text{Var}_{\theta_\alpha}(A^1_{\alpha,n}) \le \mathbb{E}[\tilde{S}^2(U_\alpha)\mathbb{I}_{\{|V_\alpha|<\lambda_n\}}] \le n^2\mathbb{P}\{|V_\alpha| < \lambda_n\}\ ,$$

and by Lemma 2.6 for `Case 3`, $\lim_{n\to\infty}\sup_{\alpha:|\theta_\alpha|>(1+\sqrt{2\nu}/\gamma)\lambda_n}\text{Var}\theta_\alpha(A^1_{\alpha,n})/n = 0$.

Note that $\mathbb{E}[U_\alpha] = \theta_\alpha$ and $\text{Var}(U_\alpha) = 2\nu d^2_\alpha \le 2\nu$. By Lemma D.3,

$$\begin{aligned}
\text{Var}_{\theta_\alpha}(A^2_{\alpha,n}) &\le \mathbb{E}[U^2_\alpha]\,\mathbb{P}\{V_\alpha < -\lambda_n\} \le (2\nu + \theta^2_\alpha)\mathbb{P}\{V_\alpha < -\lambda_n\} \\
\text{Var}_{\theta_\alpha}(A^3_{\alpha,n}) &\le \text{Var}(\bar{b}U_\alpha) + (\mathbb{E}[\bar{b}U_\alpha])^2\mathbb{P}\{V_\alpha \le \lambda_n\} \le 2\nu + (2\nu + \theta^2_\alpha)\mathbb{P}\{V_\alpha \le \lambda_n\} \\
&= 2\nu + (2\nu + \theta^2_\alpha)\mathbb{P}\{|V_\alpha| \le \lambda_n\} + (2\nu + \theta^2_\alpha)\mathbb{P}\{V_\alpha < -\lambda_n\}\ ,
\end{aligned}$$

which implies that

$$\text{Var}_{\theta_\alpha}(A^2_{\alpha,n}) + \text{Var}_{\theta_\alpha}(A^3_{\alpha,n}) \le 2\nu + (2\nu+\theta^2_\alpha)\mathbb{P}\{|V_\alpha| \le \lambda_n\} + (2\nu+\theta^2_\alpha)\mathbb{P}\{V_\alpha < -\lambda_n\}\ .$$

Note that, by Lemma 2.6 for `Case 3`, both $\sup_{\alpha:|\theta_\alpha|>\left(1\,+\,\sqrt{2\nu}/\gamma\right)\lambda_n}\theta^2_\alpha\cdot\mathbb{P}\left\{|V_\alpha| \le \lambda_n\right\}$ and $\sup_{\alpha:\theta_\alpha>\left(1\,+\,\sqrt{2\nu}/\gamma\right)\lambda_n}\theta^2_\alpha\cdot\mathbb{P}\left\{V_\alpha < -\lambda_n\right\}$ converge to zero as n increases. Thus, we have that $\lim_{n\to\infty}\sup_{\alpha\,:\,|\theta_\alpha|\,>\,(1\,+\,\sqrt{2\nu}/\gamma)\lambda_n}\left(\text{Var}_{\theta_\alpha}(A^2_{\alpha,n}) + \text{Var}_{\theta_\alpha}(A^3_{\alpha,n})\right)/n = 0$, which is the desired result.

This completes the proof of Proposition 1.1. We end this section with a remark on the choice of threshold. The proof will work similarly for $\sqrt{2\log n}$ thresholds that are scalable with $\sqrt{\nu_{p,i}}$ and $|d_\alpha|$ for $1 \le i \le n$, $\alpha \in [0,1]$. Our choice λ_n being uniform over $\tau \in [0,\infty]$, however, yields a comparatively cleaner proof.

3. Simulation Experiments. In this section, we study the performances of our proposed estimators through numerical experiments. In the first example, we display a case where the performance of our proposed ARE-based estimate is close to that of the oracle estimator, but the traditional EBML and EBMM estimators perform poorly. It supports the arguments (provided below Corollaries 1.1 and 1.2) that as the formulae of the ML and MM estimates of the hyper parameters do not depend on the shape of the loss functions, they can be significantly different from the ARE-based estimates and hence sub-optimal. We calculate the inefficiency of an estimate τ of the shrinkage hyperparameter of members in $\mathcal{S}^0$ by comparing it with its corresponding Oracle risk-based estimator: $\tilde{\tau}_{OR} = \arg\min_{\tau\in[0,\infty]} R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))$. We define:

$$\text{Inefficency of } \hat{\tau} = \frac{R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau})) - R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tilde{\tau}_{OR}))}{\max_{\tau\geq 0} R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - \min_{\tau\geq 0} R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))} \times 100\,\%.$$

The measures for the other classes are defined analogously. In the other two examples, we study the performance of our proposed estimators as we vary the model parameters. Throughout this section, we set $\nu_{f,i} = 1$ and $b_i + h_1 = 1$ for all $i = 1, \ldots, n$. The R codes used for these simulation experiments can be downloaded from `http://www-bcf.usc.edu/~gourab/inventory-management/`.

3.1. *Example 1.* Here, we study a simple setup in a homoskedastic model where $\nu_{p,i} = 1/3$ for all $i = 1, \ldots, n$. As $\nu_{f,i} = 1$ for all i, it means that if we are observing monthly demand data, we are using 3 months of data to predict the future month's demand. We consider two different choices of n (a) $n = 20$, which yields comparatively low dimensional models, and (b) $n = 100$, which is large enough to expect our high-dimensional theory results to set in. We consider only two different values for the θ_i: $1/\sqrt{3}$ and $-3\sqrt{3}$. Also, we design the setup such that the lost sales cost b_i is related to the θ_i: when $\theta_i = 1/\sqrt{3}$, $b_i = 0.51$ and when $\theta_i = -3\sqrt{3}$, $b_i = 0.99$. For the case when $n = 20$, we consider $(\boldsymbol{\theta}, \boldsymbol{b})$ with 18 replicates of the (θ_i, b_i) pair of $(1/\sqrt{3}, 0.51)$ and 2 replicates of $(-3\sqrt{3}, 0.99)$. For $n = 100$, we have 90 replicates of the former and 10 replicates of the latter. Note that in both the cases, the mean of $\boldsymbol{\theta}$ across dimensions is 0.

In this homoskedastic setup, the MM and ML estimates of the hyperparameter are identical. In Table 2, we present their relative inefficiencies as well as that of the ARE with respect to the Oracle risk estimate. For computation of the ARE risk estimates, 5 Monte-Carlo simulations were used for the evaluation of the unconditional expectation in the Rao-Blackwellization step. In Table 1, based on 50 independent simulation experiments, we report the mean and standard deviation of the estimates as well as their inefficiency percentages. The EBML/EBMM perform very poorly in both cases. When $n = 100$, the ARE-based estimates are close to the Oracle risk-based estimates and are quite efficient. When $n = 20$, the ARE method is not as efficient as before but still performs remarkably better than the EBML/EBMM methods. The plots of the univariate risks of $\hat{q}_i(\tau)$ for the (θ_i, b_i) pairs $(1/\sqrt{3}, 0.51)$ and $(-3\sqrt{3}, 0.99)$ (as $\alpha_i = \tau/(\tau + \nu_{p,i})$ varies) are very different (see Figure 1). For the former, the oracle minimizer is at $\alpha_{OR} = 0.51$; that is,

$\tau_{OR} = 0.35$. For the latter, the oracle minimizer is at $\alpha_{OR} = 1$; that is, $\tau_{OR} = \infty$. The multivariate risk plot of our setup is different than those of the two univariate risk plots but is closer to the former than to the later. ARE approximates this multivariate risk function well and does a good job in estimating the shrinkage parameter. However, the ML/MM estimate of the hyperparameter is swayed by the extremity of fewer $(\theta_i, b_i) = (-3\sqrt{3}, 0.99)$ cases and fail to properly estimate the shrinkage parameter in the combined multivariate case.

TABLE 2

Comparison of the performances of ARE-*, MM- and ML-based estimates with the Oracle risk estimator in Example 1. The mean and standard deviation (in parentheses) across* 50 *independent simulation experiments are reported.*

METHODS	$n = 20$		$n = 100$	
	Inefficiency (%)	$\hat{\tau}$	Inefficiency (%)	$\hat{\tau}$
ARE	16.78 (30.42)	1.214 (4.823)	1.15 (2.57)	0.344 (0.079)
MM/ML	48.01 (3.55)	0.037 (0.006)	47.96 (2.01)	0.037 (0.003)
ORACLE	-	0.296 (0.000)	-	0.296 (0.000)

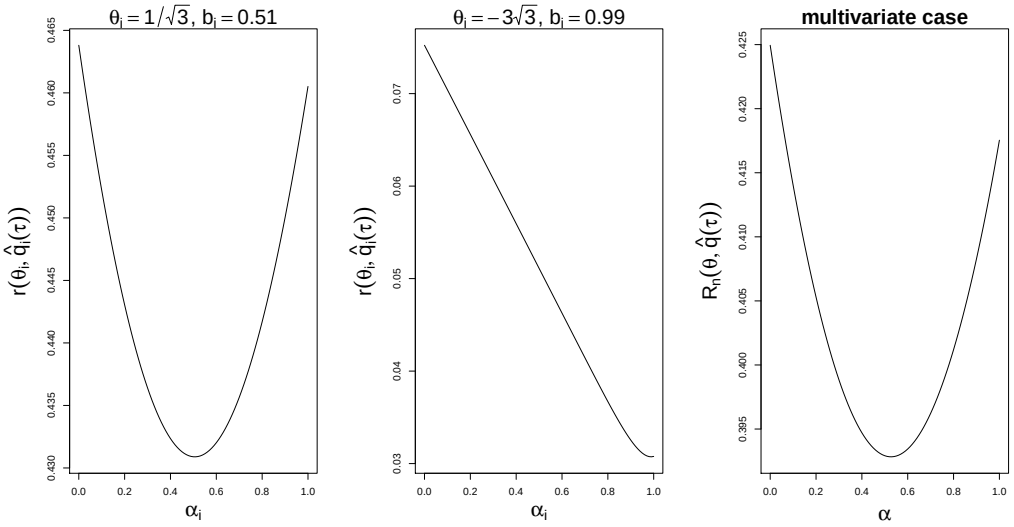

FIG 1. *From left to right we have the following: the plots of the univariate risks of $\hat{q}_i(\tau)$ for the (θ_i, b_i) pairs $(1/\sqrt{3}, 0.51)$ and $(-3\sqrt{3}, 0.99)$, respectively, as $\alpha_i = \tau/(\tau + \nu_{p,i})$ varies and the plot of the multivariate risk of $\hat{\boldsymbol{q}}(\tau)$ for the $(\boldsymbol{\theta}, \boldsymbol{b})$ choices described in Example 3.1.*

3.2. *Example 2.* We consider homoskedastic models with $\nu_{f,i} = 1$ and $\nu_{p,i} = \nu_p$ for all $i = 1, \ldots, n$. We vary ν_p to numerically test the performance of the ARE methodology when Assumption A2 of Section 1.2 is violated. We generate $\{\theta_i : i = 1, \ldots, n\}$ independently from $N(0, 1)$, and $\{b_i : i = 1, \ldots, n\}$ are generated uniformly from $[0.51, 0.99]$. Table 3 reports the mean and standard deviation (in

brackets) of the inefficiency percentages across 20 simulation experiments from each regime. We see that the ARE methodology does not work for larger values of the ratio ν_p/ν_f and starts performing reasonably when $\nu_p/\nu_f \leq 1/3$, which is quite higher than the prescribed theoretical bound in (1.6).

TABLE 3
Inefficiency (%) of ARE *estimators in Example 2 as the ratio* ν_p/ν_f *varies.*

ν_p/ν_f	$n = 20$	$n = 100$
1/1	75.34 (28.55)	88.88 (14.70)
1/2	31.70 (20.85)	27.81 (07.95)
1/3	19.21 (14.44)	12.91 (03.63)
1/4	06.93 (03.58)	07.43 (02.07)
1/5	05.56 (03.93)	04.36 (01.38)
1/6	04.07 (03.06)	03.06 (00.97)

3.3. *Example 3.* We now study the performance of our proposed ARE^{G} methodology in 6 heteroskedastic models, which are modified predictive versions of those used in Section 7 of Xie, Kou and Brown (2012). Here, $\{b_i : i = 1, \ldots, n\}$ are generated uniformly from $[0.51, 0.99]$ and $\nu_{f,i} = 1$ for all i. Also, based on Example 2, we impose the constraint $\max\{\nu_{p,i}/\nu_{f,i} : 1 \leq i \leq n\} \leq 1/3$. Next, we outline the 6 experimental setups by describing the parameters used in the predictive model of (1.1):

Case I. $\boldsymbol{\theta}$ are i.i.d. from Uniform(0,1), and $\nu_{p,i}$ are i.i.d. from Uniform(0.1,1/3).

Case II. $\boldsymbol{\theta}$ are i.i.d. from N(0,1), and $\nu_{p,i}$ are i.i.d. from Uniform(0.1,1/3).

Case III. Here, we bring in dependence between $\nu_{p,i}$ and $\boldsymbol{\theta}$. We generate $\{\nu_{p,i} : 1 \leq i \leq n\}$ independently from Uniform(0.1,1/3) and $\theta_i = 5\,\nu_{p,i}$ for $i = 1, \ldots, n$.

Case IV. Instead of uniform distribution in the above case, we now generate $\{\nu_{p,i} : 1 \leq i \leq n\}$ independently from Inv-χ^2_{10}, which is the conjugate distribution for normal variance.

Case V. This model reflects grouping in the data. We draw the past variances independently from the 2-point distribution $2^{-1}(\delta_{0.1} + \delta_{0.5})$, and the θ_i are drawn conditioned on the past variances:

$$(\theta_i|\nu_{p,i} = 0.1) \sim N(0, 0.1) \quad \text{and} \quad (\theta_i|\nu_{p,i} = 0.5) \sim N(0, 0.5).$$

Thus, there are two groups in the data.

Case VI. In this example, we assess the sensitivity in the performance of the ARE^{G} estimators to the normality assumption by allowing $\boldsymbol{X}$ to depart from the normal model of (1.1). We generate $\{\nu_{p,i} : 1 \leq i \leq n\}$ independently from Uniform(0.1,1/3) and $\theta_i = 5\,\nu_{p,i}$ for $i = 1, \ldots, n$. The past observations are generated independently from

$$X_i \sim \text{Uniform}\big(\theta_i - \sqrt{3\,\nu_{p,i}},\, \theta_i + \sqrt{3\,\nu_{p,i}}\big) \text{ for } i = 1, \ldots, n.$$

Table 4 reports the mean and standard deviation (in brackets) of the inefficiency percentages of our methodology in 20 simulation experiments from each of the 6 models. The $\mathrm{ARE}^{\mathrm{G}}$ estimator performs reasonably well across all 6 scenarios.

TABLE 4
Inefficiency (%) of $\mathrm{ARE}^{\mathrm{G}}$ *estimators in* 6 *different heteroskedastic models of Example 3.*

	$n = 20$	$n = 100$
Case I	02.79 (02.70)	01.81 (01.83)
Case II	12.90 (21.16)	11.31 (01.73)
Case III	13.90 (19.21)	07.84 (02.08)
Case IV	08.75 (14.26)	10.47 (20.65)
Case V	03.80 (04.32)	01.52 (03.13)
Case VI	06.20 (08.45)	08.74 (03.19)

4. Explanations and Proofs for Estimators in $\mathcal{S}$. We first describe the $\widehat{\mathrm{ARE}}^{\mathrm{D}}(\eta, \tau)$ risk estimation procedure. Note that, by Lemma 2.3, for any fixed $\eta \in \mathbb{R}$, the risk of estimators in $\mathcal{S}$ is related to risk of estimators in $\mathcal{S}^0$ as $R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta, \tau)) = R_n(\boldsymbol{\theta} - \eta, \hat{\boldsymbol{q}}(\tau))$. We rewrite the ARE risk estimate defined in (1.10) by explicitly denoting the dependence on $\boldsymbol{X}$ as

$$\widehat{\mathrm{ARE}}_n(\tau, \boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}(b_i + h_i)(\nu_{f,i} + \nu_{p,i}\alpha_i^2)^{1/2}\,\hat{T}_i(X_i, \tau). \tag{4.1}$$

The $\widehat{\mathrm{ARE}}^{\mathrm{D}}$ risk estimate is defined as $\widehat{\mathrm{ARE}}^{\mathrm{D}}_n(\eta, \tau, \boldsymbol{X}) = \mathrm{ARE}_n(\tau, \boldsymbol{X} - \eta)$. Henceforth, whenever we use the relation between $\widehat{\mathrm{ARE}}^{\mathrm{D}}$ and ARE, we will explicitly denote the dependence of the risk estimates on the data. Otherwise, we will stick to our earlier notation where the dependence on the data is kept implicit. We next prove Theorem 1.4.

4.1. *Proof of Theorem 1.4.* The proof follows from the following two lemmas. The first one shows that our proposed risk estimate does a good job in estimating the risk of estimators in $\mathcal{S}$. This lemma holds for all estimates $q(\eta, \tau)$ in $\mathcal{S}$ and does not need any restrictions on $\boldsymbol{\theta}$. The second lemma shows that the loss is uniformly close to the risk. It needs the restriction $|\eta| \le M_n$ on estimates in $\mathcal{S}$ and also the assumption A3′ on $\boldsymbol{\theta}$.

LEMMA 4.1. *Under Assumptions A1-A2, for all* $\boldsymbol{\theta}$,

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty],\,\eta\in\mathbb{R}} \mathbb{E}\left[\left(\widehat{\mathrm{ARE}}^{\mathrm{D}}_n(\eta, \tau) - R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta, \tau))\right)^2\right] = 0\ .$$

LEMMA 4.2. *Under Assumption A1, for all* $\boldsymbol{\theta}$ *satisfying Assumption A3′,*

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty],|\eta|\le M_n} \mathbb{E}\left|R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta, \tau)) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta, \tau))\right| = 0\ .$$

The proof of Lemma 4.1 is provided in Appendix B. *For the proof of Lemma 4.2*, we show uniform convergence over the set of location parameters $\{|\eta| \leq M_n\}$ by undertaking a moment-based approach. Here, we show that for any $\boldsymbol{\theta}$ obeying Assumption A3$'$

$$\sup_{\tau\in[0,\infty],|\eta|\leq M_n} \mathrm{Var}_{\boldsymbol{\theta}}(L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\eta,\tau)) \to 0 \text{ as } n\to\infty\ ,$$

from which the proof of the lemma follows easily. Now, note that, due to independence across coordinates, we have

$$\mathrm{Var}_{\boldsymbol{\theta}}\left(L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\eta,\tau)\right) = n^{-2}\sum_{i=1}^{n}\mathrm{Var}_{\theta_i}\left(l_i(\theta_i,\hat{q}_i(\eta,\tau))\right) \leq n^{-2}\sum_{i=1}^{n}\mathbb{E}_{\theta_i}\left[l_i^2(\theta_i,\hat{q}_i(\eta,\tau))\right].$$

By definition of the predictive loss, we have the following relation between the loss of estimators in $\mathcal{S}$ and $\mathcal{S}^0$: $\mathbb{E}_{\theta_i}[l_i^2(\theta_i,\hat{q}_i(\eta,\tau))] = \mathbb{E}_{\theta_i}[l_i^2(\theta_i-\alpha_i(\tau)\eta,\hat{q}_i(\tau))]$ and using the inequality in Equation (A.1) of the Appendix we see that it is dominated by $\mathcal{O}(1+\mathbb{E}[\theta_i-\alpha_i(\tau)\eta]^2) \leq \mathcal{O}(1+2\,\mathbb{E}_{\theta_i}[\theta_i^2+M_n^2])$ as $|\alpha_i(\tau)|\leq 1$ for any $\tau\in[0,\infty]$. Thus, we have

$$\mathrm{Var}_{\boldsymbol{\theta}}\left(L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\eta,\tau)\right) \leq \mathcal{O}\left(n^{-2}\sum_{i=1}^{n}\theta_i^2 + n^{-1}\mathbb{E}_{\boldsymbol{\theta}}[M_n^2]\right).$$

For any $\tau\in[0,\infty]$, $|\eta|\leq M_n$ and $\boldsymbol{\theta}$ satisfying assumption A3$'$, both the terms in the RHS above uniformly converge to 0, which completes the proof of Lemma 4.2. The following lemma (the proof of which follows immediately from Lemma A1 of Xie, Kou and Brown (2012)) is used for convergence of the second term.

LEMMA 4.3. *Under Assumption A3$'$, we have* $\lim_{n\to\infty} n^{-1}\mathbb{E}_{\boldsymbol{\theta}}\left[M_n^2\right] = 0$.

We next present the proof of the decision theoretic properties of our estimators.

4.2. *Proof of Theorem 1.5.* By construction, $\widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\hat{\eta}_n^{\mathrm{D}},\hat{\tau}_n^{\mathrm{D}}) \leq \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta_n^{\mathsf{DOR}},\tau_n^{\mathsf{DOR}})$. So, for any fixed $\epsilon>0$, we have: $\mathbb{P}\left\{L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}},\hat{\tau}_n^{\mathrm{D}})) \geq L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\eta_n^{\mathsf{DOR}},\tau_n^{\mathsf{DOR}})) + \epsilon\right\}$, which is bounded above by

$$\mathbb{P}\left\{A_n(\boldsymbol{\theta},\hat{\eta}_n^{\mathrm{D}},\hat{\tau}_n^{\mathrm{D}}) \geq B_n(\boldsymbol{\theta},\eta_n^{\mathsf{DOR}},\tau_n^{\mathsf{DOR}}) + \epsilon\right\}\ ,$$

where

$$A_n(\boldsymbol{\theta},\hat{\eta}_n^{\mathrm{D}},\hat{\tau}_n^{\mathrm{D}}) = L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}},\hat{\tau}_n^{\mathrm{D}})) - \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\hat{\eta}_n^{\mathrm{D}},\hat{\tau}_n^{\mathrm{D}})$$
$$B_n(\boldsymbol{\theta},\eta_n^{\mathsf{DOR}},\tau_n^{\mathsf{DOR}}) = L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\eta_n^{\mathsf{DOR}},\tau_n^{\mathsf{DOR}})) - \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta_n^{\mathsf{DOR}},\tau_n^{\mathsf{DOR}}).$$

Now, using Markov inequality, we have

$$\mathbb{P}\left\{A_n(\boldsymbol{\theta},\hat{\eta}_n^{\mathrm{D}},\hat{\tau}_n^{\mathrm{D}}) \geq B_n(\boldsymbol{\theta},\eta_n^{\mathsf{DOR}},\tau_n^{\mathsf{DOR}}) + \epsilon\right\} \leq \epsilon^{-1}\mathbb{E}|A_n(\boldsymbol{\theta},\hat{\eta}_n^{\mathrm{D}},\hat{\tau}_n^{\mathrm{D}}) - B_n(\boldsymbol{\theta},\eta_n^{\mathsf{DOR}},\tau_n^{\mathsf{DOR}})|,$$

which, again, by the triangle inequality is less than

$$2\epsilon^{-1}\sup_{\tau\in[0,\infty],|\eta|\leq M_n}\mathbb{E}|\widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta,\tau) - L_n(\boldsymbol{\theta},\hat{\boldsymbol{q}}(\eta,\tau))|.$$

By Theorem 1.4, it converges to 0 as $n\to\infty$, and we have the required result.

4.3. *Proof of Theorem 1.6.* We decompose the difference of the losses into 3 parts:

$$
\begin{aligned}
&L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}})) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}})) \\
&= \left(L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}})) - \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}}) \right) - \left(L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}})\} - \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}}) \right) \\
&\quad + \left(\widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}}) - \widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}}) \right).
\end{aligned}
$$

As the third term is less than 0, so $\mathbb{E}\left[L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\eta}_n^{\mathrm{D}}, \hat{\tau}_n^{\mathrm{D}})) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}}))\right]$ is bounded above by $2\sup_{\tau\in[0,\infty],|\eta|\le M_n} \mathbb{E}\big|\widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta,\tau) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta,\tau))\big|$, which converges to 0 by Theorem 1.4. Hence, the result follows.

4.4. *Proof of Corollary 1.1.* The results follow directly from Theorems 1.5 and 1.6 as $(\eta_n^{\mathsf{DOR}}, \tau_n^{\mathsf{DOR}})$ minimizes the loss $L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta,\tau))$ among the class $\mathcal{S}$.

5. Explanations and Proofs for Estimators in $\mathcal{S}^G$. By (1.3), the predictive loss an estimator $\hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)$ in $\mathcal{S}^G$ is given by $L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) = \frac{1}{n}\sum_{i=1}^n l_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau))$, where

$$
l_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau)) = \nu_{f,i}^{1/2}\,(b_i + h_i)\, G(\nu_{f,i}^{-1/2}(\hat{q}_i(\tau) + (1-\alpha_i)\bar{\boldsymbol{X}} - \theta_i), \tilde{b}).
$$

We define a surrogate of the loss by plugging in $\bar{\boldsymbol{\theta}}$ – the mean of the unknown parameter $\boldsymbol{\theta}$ in the place of $\bar{\boldsymbol{X}}$: $\tilde{L}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) = \frac{1}{n}\sum_{i=1}^n \tilde{l}_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau))$, where

$$
\tilde{l}_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau)) = \nu_{f,i}^{1/2}\,(b_i + h_i)\, G(\nu_{f,i}^{-1/2}(\hat{q}_i(\tau) + (1-\alpha_i)\bar{\boldsymbol{\theta}} - \theta_i), \tilde{b}).
$$

The following lemma, whose proof is provided in Appendix C, shows the surrogate loss is uniformly close to the actual predictive loss.

LEMMA 5.1. *For any $\boldsymbol{\theta} \in \mathbb{R}^n$ and $\hat{\boldsymbol{q}}^{\mathrm{G}}(\tau) \in \mathcal{S}^G$, we have*

$$
\lim_{n\to\infty} \mathbb{E}\left[\sup_{\tau\in[0,\infty]} \left|L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) - \tilde{L}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau))\right|\right] = 0\ .
$$

We define the associated surrogate risk by $\tilde{r}_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau)) = \mathbb{E}_{\boldsymbol{\theta}}\tilde{l}_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau))$. From Lemma 2.3, it follows that this surrogate risk is connected with the risk function of estimators in $\mathcal{S}$ as: $\tilde{r}_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau)) = r(\theta_i - \bar{\boldsymbol{\theta}}, \hat{q}(\tau))$. Thus, the associated multivariate surrogate risk $\tilde{R}_i(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) = \sum_{i=1}^n \tilde{r}_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau))$ equals $R_n(\theta - \bar{\boldsymbol{\theta}}, \hat{\boldsymbol{q}}(\tau))$. Also by Lemma 5.1, it follows that for any $\boldsymbol{\theta} \in \mathbb{R}^n$

$$
\lim_{n\to\infty} \mathbb{E}\left[\sup_{\tau\in[0,\infty]} \left|R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) - \tilde{R}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau))\right|\right] = 0\ . \tag{5.1}
$$

Now we will describe our proposed $\mathrm{ARE}^{\mathrm{G}}$ estimator. Explicitly denoting the dependence of the estimators on the data, for any fixed value of $\tau \in [0,\infty]$, we

define $\widehat{\mathrm{ARE}}_n^{\mathrm{G}}(\tau, \boldsymbol{X}) = \widehat{\mathrm{ARE}}_n(\tau, \boldsymbol{X} - \eta)|_{\eta=\bar{\boldsymbol{X}}}$. Note that $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$ are correlated, and $\boldsymbol{X} - \bar{\boldsymbol{X}}$ has a normal distribution with a non-diagonal covariance structure. However, we can still use the asymptotic risk estimation procedure described in Section 2 by just plugging in the value of $\bar{\boldsymbol{X}}$. We avoid the complications of incorporating the covariance structure in our calculations by cleverly using the concentration properties of $\bar{\boldsymbol{X}}$ around $\bar{\boldsymbol{\theta}}$. To explain this approach, we again define a surrogate to our $\mathrm{ARE}^{\mathrm{G}}$ estimator $\widehat{\mathrm{ARE}}_n(\tau, \boldsymbol{X}-\eta)|_{\eta=\bar{\boldsymbol{X}}} = \sum_{i=1}^n a_i \hat{\mathrm{T}}_i(X_i-\eta, \tau)|_{\eta=\bar{\boldsymbol{X}}}$ by

$$\widetilde{\mathrm{ARE}}_n(\tau, \boldsymbol{X} - \bar{\boldsymbol{\theta}}) = \sum_{i=1}^n a_i \, \widetilde{\mathrm{T}}_i(X_i - \bar{\boldsymbol{\theta}}, \tau) \ ,$$

where we plugin $\bar{\boldsymbol{\theta}}$ in the place of $\bar{\boldsymbol{X}}$. Here, $a_i = (b_i + h_i)\sqrt{\nu_{f,i} + \nu_{p,i}\alpha_i(\tau)^2}$. Note that $\widetilde{\mathrm{ARE}}$ and $\widetilde{\mathrm{T}}$ have the same functional form as $\widehat{\mathrm{ARE}}$ and $\hat{T}$, respectively, but with $\bar{\boldsymbol{X}}$ replaced by $\bar{\boldsymbol{\theta}}$ and so are not estimators. We now present the proof of Theorem 1.7.

Proof of Theorem 1.7. We will prove the theorem by establishing

(a) $\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \mathbb{E}\big|L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) - R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau))\big| = 0$ and,

(b) $\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \mathbb{E}\big|R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) - \widehat{\mathrm{ARE}}^{\mathrm{G}}(\tau)\big| = 0.$

For the proof of (a), based on (5.1) and Lemma 5.1, it suffices to show

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \mathbb{E}\big|\tilde{L}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) - \tilde{R}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau))\big| = 0 \ .$$

We prove it by showing that as $n \to \infty$, $\mathrm{Var}_{\boldsymbol{\theta}}(\tilde{L}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)))$ converges to 0 uniformly over τ for any $\boldsymbol{\theta}$ satisfying Assumption A3′. Again, as in the proof of Lemma 4.1, we have the bound

$$\mathrm{Var}_{\boldsymbol{\theta}}(\tilde{L}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau))) \le \mathcal{O}\bigg(\frac{1}{n^2}\sum_{i=1}^n \mathbb{E}_{\theta_i}(\theta_i - \alpha_i(\tau)\bar{\boldsymbol{\theta}})^2\bigg).$$

As $|\alpha_i(\tau)| \le 1$ for all $\tau \in [0, \infty]$, the RHS above is at most $\mathcal{O}(n^{-2}\sum_{i=1}^n \theta_i^2 + \bar{\boldsymbol{\theta}}^2/n)$. It converges to 0 as $n \to \infty$ for any $\boldsymbol{\theta}$ satisfying Assumption A3′.

Now for the proof of (b), using (5.1) as $n \to \infty$, we have

$$\begin{aligned}
&\sup_{\tau\in[0,\infty]} \mathbb{E}\big|R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) - \widehat{\mathrm{ARE}}^{\mathrm{G}}(\tau)\big| \to \sup_{\tau\in[0,\infty]} \mathbb{E}\big|\tilde{R}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}^{\mathrm{G}}(\tau, \boldsymbol{X})\big| \\
&= \sup_{\tau\in[0,\infty]} \mathbb{E}\big|R_n(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}, \hat{\boldsymbol{q}}(\tau)) - \widehat{\mathrm{ARE}}^{\mathrm{G}}(\tau, \boldsymbol{X})\big| \ ,
\end{aligned}$$

which is bounded above by the sum of $\sup_{\tau\in[0,\infty]} \mathbb{E}_{\boldsymbol{\theta}}|\widetilde{\mathrm{ARE}}_n(\tau, \boldsymbol{X} - \bar{\boldsymbol{\theta}}) - \mathrm{ARE}^{\mathrm{G}}(\tau)|$ and $\sup_{\tau\in[0,\infty]} \mathbb{E}_{\boldsymbol{\theta}}|R_n(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}, \hat{\boldsymbol{q}}(\tau)) - \widetilde{\mathrm{ARE}}_n(\tau, \boldsymbol{X} - \bar{\boldsymbol{\theta}})|$. Again, by Lemma 4.1, the

second term converges to 0 as $n \to \infty$. The first term is bounded above by

$$\sup_{\tau\in[0,\infty]} \frac{1}{n}\sum_{i=1}^{n} a_i \, \mathbb{E}_{\boldsymbol{\theta}} \left| (\bar{\boldsymbol{X}} - \bar{\boldsymbol{\theta}}) \cdot \left[\frac{\partial}{\partial \eta} \hat{T}_i(X_i - \eta, \tau) \right]_{\eta=\mu_i} \right| ,$$

where each $\{\mu_i : 1 \le i \le n\}$ lies between $\bar{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{X}}$. Using Cauchy-Schwarz inequality, the above term is less than

$$\lim_{n\to\infty} \sup_{\tau\in[0,\infty]} \frac{1}{n}\sum_{i=1}^{n} a_i \left\{ \mathbb{E}_{\boldsymbol{\theta}} (\bar{\boldsymbol{X}} - \bar{\boldsymbol{\theta}})^2 \cdot \mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \eta} \hat{T}_i(X_i - \eta, \tau) \right]^2_{\eta=\mu_i} \right\}^{1/2} = 0 .$$

As a_i are bounded by Assumptions A1 and A2, the asymptotic convergence above follows by using $\mathbb{E}_{\boldsymbol{\theta}} (\bar{\boldsymbol{X}} - \bar{\boldsymbol{\theta}})^2 = n^{-1}$ and the following lemma, whose proof is provided in Appendix C.

LEMMA 5.2. *For any $\boldsymbol{\theta} \in \mathbb{R}^n$ and μ_i lying in between $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{\theta}}$ for all $i = 1, \ldots, n$*

$$\lim_{n\to\infty} n^{-1} \left\{ \sup_{1\le i\le n} \sup_{\tau\in[0,\infty]} \mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \eta} \hat{T}_i(X_i - \eta, \tau) \right]^2_{\eta=\mu_i} \right\} = 0 .$$

This completes the proof of Theorem 1.7.

The proof of Theorem 1.8 follows similarly from the proofs of Theorems 1.5, 1.6 and Corollary 1.2 and is not presented here to avoid repetition.

6. Discussion. Here, we have developed an Empirical Bayes methodology for predicting the stocking levels that a seller of a large category of products needs to keep in its inventory. Our proposed method involves the calibration of the tuning parameters of skrinkage estimators by minimizing risk estimates that are adapted to the curvature of the newsvendor's loss functions. It produces asymptotically optimal stocking quantity estimates. Our risk estimation method and its proof techniques can also be used to construct optimal empirical Bayes predictive rules for piecewise linear and related asymmetric loss functions, where we do not have any natural unbiased risk estimate. In this paper, we have worked in a high-dimensional Gaussian model. Though normality transformations exist for a wide range of high-dimensional models (Brown, 2008), future works in extending the methodology to non-Gaussian models, particularly discrete setups, would be interesting. Noting that past sales are actually censored demand corresponding to the minimum of the demand and stocking quantity, incorporating the censoring effect in future models would be useful. Extending our Empirical Bayes approach from the one-period predictive setup to a multi-period dynamic inventory management setup (Lariviere and Porteus, 1999) would be another interesting future direction.

APPENDIX

APPENDIX A: PROOF DETAILS FOR ESTIMATORS IN THE CLASS $\mathcal{S}^0$ AND THE LEMMAS USED IN SECTION 2

A.1. Proof of Theorem 1.2. By construction $\widehat{\text{ARE}}_n(\hat{\tau}_n^{\text{ARE}}) \leq \widehat{\text{ARE}}_n(\tau_n^{\mathsf{OR}})$. So, for any fixed $\epsilon > 0$ we have:

$$\mathbb{P}\left\{L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\text{ARE}})) \geq L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau_n^{\mathsf{OR}})) + \epsilon\right\} \leq \mathbb{P}\left\{A_n(\boldsymbol{\theta}, \hat{\tau}_n^{\text{ARE}}) \geq B_n(\boldsymbol{\theta}, \tau_n^{\mathsf{OR}}) + \epsilon\right\}$$
$$\text{where } A_n(\boldsymbol{\theta}, \hat{\tau}_n^{\text{ARE}}) = L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\text{ARE}})) - \widehat{\text{ARE}}_n(\hat{\tau}_n^{\text{ARE}})$$
$$\text{and } B_n(\boldsymbol{\theta}, \tau_n^{\mathsf{OR}}) = L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau_n^{\mathsf{OR}})) - \widehat{\text{ARE}}_n(\tau_n^{\mathsf{OR}}).$$

Now, using Markov inequality we get:

$$\mathbb{P}\left\{A_n(\boldsymbol{\theta}, \hat{\tau}_n^{\text{ARE}}) \geq B_n(\boldsymbol{\theta}, \tau^{OR}) + \epsilon\right\} \leq \epsilon^{-1}\mathbb{E}|A_n(\boldsymbol{\theta}, \hat{\tau}_n^{\text{ARE}}) - B_n(\boldsymbol{\theta}, \tau_n^{\mathsf{OR}})|$$

which again is less than $2\epsilon^{-1}\sup_{\tau\in[0,\infty]}\mathbb{E}|\widehat{\text{ARE}}_n(\tau) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))|$. By Theorem 1.1, it converges to 0 as $n \to \infty$. Thus, we have the required result.

A.2. Proof of Theorem 1.3. We decompose the difference of the losses into the following 3 parts:

$$\begin{aligned}
&L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\text{ARE}})) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau_n^{\mathsf{OR}})) \\
&= \{L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\text{ARE}})) - \widehat{\text{ARE}}_n(\hat{\tau}_n^{\text{ARE}})\} - \{L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau_n^{\mathsf{OR}})\} - \widehat{\text{ARE}}_n(\tau_n^{\mathsf{OR}})) \\
&\quad + \{\widehat{\text{ARE}}_n(\hat{\tau}_n^{\text{ARE}}) - \widehat{\text{ARE}}_n(\tau_n^{\mathsf{OR}})\}.
\end{aligned}$$

Now, by construction the third term is less than 0 and so,

$$\mathbb{E}\left[L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\hat{\tau}_n^{\text{ARE}})) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau_n^{\mathsf{OR}}))\right] \leq 2\sup_{\tau\in[0,\infty]}\mathbb{E}\left|\widehat{\text{ARE}}_n(\tau) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))\right|$$

which converges to 0 by Theorem 1.1. Hence, the result follows.

A.3. Proof of Corollary 1.1. The results follows directly from Theorems 1.2 and 1.3 as τ_n^{OR} minimizes the loss $L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))$ among the class of all linear estimates with shrinkage towards the origin.

Next, we provide proofs of all the main lemmas used in this Section 2.3.

A.4. Proof of Lemma 2.1. Here, we will be proving something stronger than the stated result. We will prove the following:

$$\sup_{\tau\in[0,\infty]}\left|R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))\right| \to 0 \text{ in } L_1 \text{ as } n \to \infty.$$

In our proof we will use a version of the uniform SLLN (Newey and McFadden, 1994, Lemma 2.4). Based on form the loss functions in (1.3) and the form of the linear estimators in (1.5), we can reparametrize this problem with respect to $\tilde{\tau} = \tau/(\tau+1)$ instead of τ. The only τ dependent quantity in the expression of $\hat{q}_i(\tau)$ in (1.5) is α_i which is reparametrized to $\tilde{\tau}/(\tilde{\tau} + (1-\tilde{\tau})\nu_{p,i})$. As $\tilde{\tau} \in [0,1]$, the supremum here is actually over compact set. Also, $l_i(\theta_i, \hat{q}_i(\tilde{\tau})(x))$ is continuous at each $\tilde{\tau}$ for all most all x and θ. Also,

$$l(\theta, \hat{q}(\tilde{\tau})) = (b_i + h_i)\, {\nu_{f,i}}^{1/2}\, G\bigg(\frac{\alpha_i z_i + (\nu_{f,i} + \alpha_i \nu_{p,i})^{1/2}\Phi^{-1}(\tilde{b}_i) - \bar{\alpha}_i\theta_i}{{\nu_{f,i}}^{1/2}};\, \tilde{b}_i\bigg)$$

where and $z_i = x_i - \theta_i$ and has $N(0, \nu_{p,i})$ distribution. By Lemma D.5 we know $G(y, \tilde{b}) \le \phi(0) + (1-\tilde{b})|y|$ and we use $\alpha_i \in [0,1]$ to arrive at: for each θ_i and for all $\tilde{\tau} \in [0,1]$ we have,

(A.1)
$$l_i(\theta_i, \hat{q}_i(\tau)) \le (b_i + h_i)\big[\nu_{f,i}^{1/2}\phi(0) + (1-\tilde{b}_i)\big\{|z_i| + (\nu_{f,i} + \nu_{p,i})^{1/2}\Phi^{-1}(\tilde{b}_i) + |\theta_i|\big\}\big].$$

So, for any $\boldsymbol{\theta}$ and $\tau \in [0, \infty]$ we have:

(A.2)
$$L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) \le A_n\bigg(\phi(0) + n^{-1}\sum_{i=1}^{n}|\Phi^{-1}(\tilde{b}_i)|\bigg) + B_n\bigg(n^{-1}\sum_{i=1}^{n}|z_i| + n^{-1}\sum_{i=1}^{n}|\theta_i|\bigg)$$

where $A_n = \sup\{(b_i + h_i)\nu_{f,i}^{1/2} : i = 1, \ldots, n\}$ and $B_n = \sup\{b_i + h_i : i = 1, \ldots, n\}$. By Assumptions A1-A2 we have $\limsup_n A_n \le \infty$ and $\limsup_n B_n < \infty$. So, the expectation of the RHS in (A.2) is finite under Assumption A3. As all the conditions of Newey and McFadden (1994, Lemma 2.4) hold, we can apply the SLLN uniformly. So, the loss converge to the risk and we have:

$$\sup_{\tau\in[0,\infty]} \big|R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))\big| \to 0 \text{ in } P \text{ as } n \to \infty.$$

Now, noting that the upper bound in (A.2) is uniformly integrable as $\sum_{i=1}^{n}|z_i|$ is U.I. by the extra integrability condition (See Lemma D.6). So, $\sup_{\tau\in[0,\infty]}|R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau)) - L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\tau))\big|$ is U.I. and we also have L_1 convergence. Hence, the result follows.

A.5. Proof of Lemma 2.2. As both the L.H.S. and R.H.S. scale in $(b+h)$ without loss of generality we assume $b + h = 1$. We will first prove the result for $\nu = 1$ before proceeding to the general case. Noting that for all y and q, $(y-q)^+ = y - q + (q-y)^+$, we have:

$$b\,\mathbb{E}[Y-q]^+ + h\,\mathbb{E}[q-Y]^+ = b(\theta - q) + (b+h)\mathbb{E}[q - \theta - Z]^+$$

where Z is standard normal random variable. Direct calculation yields:

$$\begin{aligned}\mathbb{E}[q-\theta-Z]^+ &= \int_{-\infty}^{q-\theta}(q-\theta-x)\phi(x)dx \\ &= (q-\theta)\Phi(q-\theta) + \int_{-\infty}^{q-\theta} -x\phi(x)dx = (q-\theta)\Phi(q-\theta) + \phi(q-\theta)\ ,\end{aligned}$$

which gives the desired result. Also, note that in this case $\partial_w G(w,b) = \Phi(w) - b$. So, $G(w,b)$ is minimized at $\Phi^{-1}(b)$ and the minimum value is $\phi(\Phi^{-1}(b))$.

For general ν we rewrite the L.H.S. using $Y \stackrel{d}{=} \nu^{1/2}Z + \theta$ where Z is a standard normal random variable to obtain:

$$\nu^{1/2}\{b\,\mathbb{E}[Z - \nu^{1/2}(q-\theta)]^+ + h\,\mathbb{E}[\nu^{1/2}(q-\theta) - Z]^+\}.$$

Now, the result stated in the lemma follows by using the already proven result for the unit variance case.

A.6. Proof of Lemma 2.3. With out loss of generality we can assume that $b_i + h_i = 1$ as the univariate loss is just scaled by that factor. Now, the minimizer of the the Bayes risk $B_1(\eta,\tau)$ is given by

$$\hat{q}(\eta,\tau)(x) = \arg\min_{\hat{q}} \int l(\theta, \hat{q}(x))\pi(\theta|x)\,d\theta.$$

The posterior distribution $\pi(\theta|x) \sim N(\alpha x + \bar{\alpha}\eta, \alpha\nu_p)$. So, for any fixed x we have

$$\int l(\theta,\hat{q}(x))\pi(\theta|x)\,d\theta = \nu_f^{1/2}\,\mathbb{E}\left\{G\left(\frac{\hat{q}-T}{\sqrt{\nu_f}}, b\right)\right\}$$

where the expectation is over T which follows $N(\alpha x + \bar{\alpha}\eta, \alpha\nu_p)$. The above expectation equals $\nu_f^{1/2}\,\mathbb{E}\,G(\nu_f^{-1/2}\{\hat{q}(x) - (\alpha x + +\bar{\alpha}\eta + \alpha^{1/2}\nu_p^{1/2}Z)\}, b)$ where Z is a standard normal random variable. To evaluate the aforementioned expression we now use the identity in (2.2) with $Y \sim N(a(x), \nu_f)$ and $a(x) = \alpha x + \bar{\alpha}\eta - \hat{q}(x)$. Finally we get $\int l(\theta,\hat{q}(x))\pi(\theta|x)\,d\theta$ equals

$$\begin{aligned}&\mathbb{E}_{Z\sim N(0,1)}\big\{\mathbb{E}_{Y\sim N(a(x),\nu_f)}\big(b\,(Y + \alpha^{1/2}\nu_p^{1/2}Z)^+ + h\,(-Y - \alpha^{1/2}\nu_p^{1/2}Z)^+\big)\big\}\\ &= \mathbb{E}\big\{b\,\big(a(x) + (\nu_f + \alpha\nu_p)^{1/2}Z\big)^+ + h\big(-a(x) - (\nu_f+\alpha\nu_p)^{1/2}Z\big)^+\big\}.\end{aligned}$$

As $Y + \alpha^{1/2}\nu_p^{1/2}Z \sim N(-a(x), \nu_f + \alpha\nu_p)$ the above equality follows using $Y + (\alpha\nu_p)^{1/2}Z \stackrel{d}{=} a(x) + (\nu_f+\alpha\nu_p)^{1/2}Z$. Again, using change of variable, the problem can be ultimately reduced to finding the minimizer for:

$$(\nu_f + \alpha\nu_p)^{1/2}\big\{b\,\mathbb{E}\{Z - \tilde{a}(x)\}^+ + h\,\mathbb{E}\{\tilde{a}(x) - Z\}^+\big\}$$

where $\tilde{a}(x) = -a(x)(\nu_f+\alpha\nu_p)^{-1/2}$. By Lemma 2.2, it is minimized when $\tilde{a}(x) = \Phi^{-1}(b)$ which implies $\hat{q}_\alpha(x) = \alpha x + \bar{\alpha}\eta + (\nu_f+\alpha\nu_p)^{1/2}\Phi^{-1}(b)$. Also, the minimum value is $(\nu_f+\alpha\nu_p)^{1/2}\phi(\Phi^{-1}(b))$ which gives us the expression for the Bayes risk.

The risk of the Bayes estimate $\hat{q}(\eta,\tau)$ is given by:

$$b\cdot\mathbb{E}_\theta\big(Y - \alpha X - \bar{\alpha}\eta - (\nu_f+\alpha\nu_p)^{1/2}\Phi^{-1}(b)\big)^+ + h\cdot\mathbb{E}_\theta\big(\alpha X + \bar{\alpha}\eta + (\nu_f+\alpha\nu_p)^{1/2}\Phi^{-1}(b) - Y\big)^+$$

where $X \sim N(\theta,\nu_p)$, $Y \sim N(\theta,\nu_f)$ and given θ, $X \perp Y$. And so, the above equals,

$$(\nu_f + \alpha^2\nu_p)^{1/2}\left\{b\cdot\mathbb{E}_0\big(Z - J\big)^+ + h\mathbb{E}_0\big(Z-J\big)^-\right\} = (\nu_f+\alpha^2\nu_p)^{1/2}G(J,b)$$

where $J = (\nu_f+\alpha^2\nu_p)^{-1/2}\{-\bar{\alpha}(\theta-\eta) + (\nu_f+\alpha\nu_p)^{1/2}\Phi^{-1}(b)\}$. This completes the proof.

A.7. Proof of Lemma 2.4. By Taylor's Theorem,

$$|G_{K_n}(y,b) - G(y,b)| = \frac{\phi(\zeta)\,|H_{K_n-1}(\zeta)|\,|\zeta|^{K_n+1}}{(K_n+1)!}\ ,$$

where ζ lies between 0 and y. Noting that $\phi(\zeta) \le 1$ for all ζ. By Lemma D.1, there exists an absolute constant c such that:

$$|H_{K_n-1}(\zeta)| \le c\, e^{\zeta^2/4}(K_n-1)!(K_n-1)^{-1/3}\,\{(K_n-1)/e\}^{-(K_n-1)/2}$$

which provide us with the following error bound:

(A.3)

$$\begin{aligned}|G_{K_n}(y,b) - G(y,b)| &\le c\,\frac{e^{y^2/4}}{(K_n+1)!} \times \frac{(K_n-1)!}{(K_n-1)^{1/3}\,((K_n-1)/e)^{(K_n-1)/2}} \times |y|^{K_n+1} \\ &= c\left(\frac{e\,y^2}{K_n-1}\right)^{(K_n-1)/2} \frac{e^{y^2/4}\,y^2}{(K_n-1)^{1/3}(K_n+1)K_n}.\end{aligned}$$

By definition of K_n just before Equation (2.3), $K_n - 1 \ge e^2(\gamma + \sqrt{2\nu})^2(2\log n)$. Since $|y| \le (1+\sqrt{2\nu}/\gamma)\lambda_n = (\gamma+\sqrt{2\nu})\sqrt{2\log n}$,

$$\begin{aligned}\left(\frac{e\,y^2}{K_n-1}\right)^{(K_n-1)/2} &\le e^{-(K_n-1)/2} \le e^{-e^2(\gamma+\sqrt{2\nu})^2\log n} = n^{-e^2\,(\gamma+\sqrt{2\nu})^2} \\ \frac{y^2}{(K_n-1)^{1/3}(K_n+1)K_n} &\le \frac{y^2}{(K_n-1)^2} \le \frac{1}{e^4(\gamma+\sqrt{2\nu})^2(2\log n)} \\ e^{y^2/4} &\le e^{(\gamma+\sqrt{2\nu})^2(\log n)/2} = n^{(\gamma+\sqrt{2\nu})^2/2} \ \le\ n^{(\gamma+\sqrt{2\nu})^2}\ ,\end{aligned}$$

which implies that

$$\sup_{|y|\le(1+\sqrt{2\nu}/\gamma)\lambda_n} |G_{K_n}(y,b) - G(y,b)| \le c\,\frac{n^{-(e^2-1)(\gamma+\sqrt{2\nu})^2}}{e^4(\gamma+\sqrt{2\nu})^2}\ ,$$

which is the desired result.

The second part follows because $G(y,b) = \phi(y) - y\tilde{\Phi}(y) + \bar{b}y$, and thus,

$$|G(y,b) - \bar{b}y| = |\phi(y) - y\tilde{\Phi}(y)| \le \frac{\phi(y)}{y^2} \le \frac{e^{-y^2/2}}{y^2}\ ,$$

where the first inequality follows from Lemma D.4. For the proof of the third statement, note that for $y < 0$,

$$G(y,b) = \phi(y) - y\tilde{\Phi}(y) + \bar{b}y = \phi(-y) - (-y)\tilde{\Phi}(-y) - by$$

and we can then apply Lemma D.4 as before because $-y$ is now positive.

A.8. Proof of Lemma 2.5. By Lemma D.8, $\mathbb{E}\left[S^2(U_\alpha)\right]$ is bounded above by

$$\left(G(0,b)+|G'(0,b)|\sqrt{\mathbb{E}U_\alpha^2}+\sum_{l=0}^{K_n}\frac{|H_l(0)|}{(l+2)!}(2\nu d_\alpha^2)^{(l+2)/2}\sqrt{\mathbb{E}\,H_{l+2}^2\left(\frac{U_\alpha}{\sqrt{2\nu d_\alpha^2}}\right)}\right)^2$$

We will now bound each of the term in the above expression. Note that $G(0,b)=\phi(0)$, $G'(0,b)=\frac{1}{2}-b$, and $\mathbb{E}U_\alpha^2=2\nu d_\alpha^2+\theta_\alpha^2$. So, the first two terms are $o(\sqrt{n})$ as $n\to\infty$. Thus, it suffices to show that the last term is also $o(\sqrt{n})$. Let $m_\alpha=2e\nu d_\alpha^2$. Then, by Lemma D.2, we have that for all $l\geq 0$

$$\begin{aligned}(2\nu d_\alpha^2)^{(l+2)/2}\sqrt{\mathbb{E}\,H_{l+2}^2\left(\frac{U_\alpha}{\sqrt{2\nu d_\alpha^2}}\right)} &= (2\nu d_\alpha^2)^{(l+2)/2}(l+2)^{(l+2)/2}\left(1+\frac{\theta_\alpha^2}{2\nu d_\alpha^2(l+2)}\right)^{(l+2)/2}\\ &= \left(\frac{l+2}{e}\right)^{(l+2)/2}\left(m_\alpha+\frac{\theta_\alpha^2 m_\alpha}{2\nu d_\alpha^2(l+2)}\right)^{(l+2)/2}\\ &\leq \left(\frac{l+2}{e}\right)^{(l+2)/2}\left(1+\frac{\theta_\alpha^2 m_\alpha}{2\nu d_\alpha^2(l+2)}\right)^{(l+2)/2}\\ &\leq \left(\frac{l+2}{e}\right)^{(l+2)/2} e^{\theta_\alpha^2 m_\alpha/(4\nu d_\alpha^2)}\,,\end{aligned}$$

where the first inequality follows from $m_\alpha\leq 2e\nu<1$ because $|d_\alpha|\leq 1$ and Assumption A2 implies that $\nu<1/(4e)$. The final inequality follows Lemma D.7. Since $|\theta_\alpha|\leq(1+\sqrt{2\nu}/\gamma)\lambda_n=(\gamma+\sqrt{2\nu})\sqrt{2\log n}$ and $m_\alpha/(2\nu d_\alpha^2)=e$,

$$\frac{\theta_\alpha^2 m_\alpha}{4\nu d_\alpha^2}\leq e(\gamma+\sqrt{2\nu})^2\log n=n^{e(\gamma+\sqrt{2\nu})^2}$$

Using the above abound and Lemma D.1, it follows that there exists an absolute constant c such that

$$\begin{aligned}&\sum_{l=0}^{K_n}\frac{|H_l(0)|}{(l+2)!}(2\nu d_\alpha^2)^{(l+2)/2}\sqrt{\mathbb{E}\,H_{l+2}^2\left(\frac{U_\alpha}{\sqrt{2\nu d_\alpha^2}}\right)}\\ &\quad\leq n^{e(\gamma+\sqrt{2\nu})^2}\sum_{l=0}^{K_n}\frac{|H_l(0)|\,(l+2)^{(l+2)/2}}{(l+2)!\,e^{(l+2)/2}}\\ &\quad\leq c\,n^{e(\gamma+\sqrt{2\nu})^2}\sum_{l=1}^{K_n}\frac{(l+2)^{(l+2)/2}}{(l+2)!\,e^{(l+2)/2}}\times\frac{l!}{l^{1/3}\left(\frac{l}{e}\right)^{l/2}}\\ &\quad\leq\ c\,n^{e(\gamma+\sqrt{2\nu})^2}\sum_{l=1}^{K_n}\frac{1}{l^{4/3}}\,,\end{aligned}$$

where the last inequality follows from Lemma D.7 because

$$\frac{(l+2)^{(l+2)/2}}{(l+2)!\,e^{(l+2)/2}}\times\frac{l!}{l^{1/3}\left(\frac{l}{e}\right)^{l/2}}=\frac{1}{e}\left(\frac{l+2}{l}\right)^{l/2}\frac{1}{(l+1)l^{1/3}}\leq\frac{1}{(l+1)l^{1/3}}\leq\frac{1}{l^{4/3}}$$

Note that $\sum_{l=1}^{\infty}\frac{1}{l^{4/3}}<\infty$. Also, by our definition, $0<\gamma<(1/\sqrt{2e})-\sqrt{2\nu}$, which implies that $e(\gamma+\sqrt{2\nu})^2<1/2$, so $n^{e(\gamma+\sqrt{2\nu})^2}=o(\sqrt{n})$, which completes the proof.

A.9. Proof of Lemma 2.6. Since $V_\alpha = \theta_\alpha + \sqrt{2\nu d_\alpha^2} Z$ and $d_\alpha^2 \le 1$, for **Case 1** where $|\theta_\alpha| \le \lambda_n/2$,

$$\begin{aligned}\mathbb{P}\{|V_\alpha| > \lambda_n\} &\le 2\mathbb{P}\left\{\lambda_n/2 + \sqrt{2\nu d_\alpha^2} Z > \lambda_n\right\} \;\le\; 2\mathbb{P}\left\{Z > \lambda_n/(2\sqrt{2\nu})\right\} \\ &= 2\tilde{\Phi}\left(\gamma\sqrt{\log n/(4\nu)}\right) \le \frac{2\phi\left(\gamma\sqrt{\log n/(4\nu)}\right)}{\gamma\sqrt{\log n/(4\nu)}} \le \frac{2n^{-\gamma^2/(8\nu)}}{\gamma\sqrt{\log n/(4\nu)}} \,,\end{aligned}$$

where the next to last inequality follows from Lemma D.4. Thus,

$$\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha|\,\le\,\lambda_n/2} \lambda_n^2 \cdot \mathbb{P}\{|V_\alpha| > \lambda_n\} = 0 \,,$$

which is the desired result.

For Case 2, we will assume that $\lambda_n/2 < \theta_\alpha \le (1 + \sqrt{2\nu}/\gamma)\lambda_n$; the proof for the other case is the same by symmetry. Since $V_\alpha = \theta_\alpha + \sqrt{2\nu d_\alpha^2} Z$ and $d_\alpha^2 \le 1$,

$$\begin{aligned}\mathbb{P}\{V_\alpha < -\lambda_n\} &\le 2\mathbb{P}\left\{\lambda_n/2 + \sqrt{2\nu d_\alpha^2} Z < -\lambda_n\right\} \;\le\; \mathbb{P}\left\{Z < -3\lambda_n/(2\sqrt{2\nu})\right\} \\ &\le \frac{\phi\left(\gamma\sqrt{9\log n/(4\nu)}\right)}{\gamma\sqrt{9\log n/(4\nu)}} \le \frac{n^{-9\gamma^2/(8\nu)}}{\gamma\sqrt{9\log n/(4\nu)}} \,,\end{aligned}$$

and since $\theta_\alpha \le (1 + \sqrt{2\nu}/\gamma)\lambda_n = (\gamma + \sqrt{2\nu})\sqrt{2\log n}$, we have that

$$\lim_{n\to\infty} \sup_{\alpha\,:\,\lambda_n/2\,<\,\theta_\alpha\,\le\,(1+\sqrt{2\nu}/\gamma)\lambda_n} |\theta_\alpha| \cdot \mathbb{P}\{V_\alpha < -\lambda_n\} = 0 \,,$$

which is the desires result.

For Case 3, suppose that $\theta_\alpha > \left(1 + \sqrt{2\nu}/\gamma\right)\lambda_n$; the proof for the other case is the same. Since $V_\alpha = \theta_\alpha + \sqrt{2\nu d_\alpha^2} Z$ and $d_\alpha^2 \le 1$,

$$\begin{aligned}\mathbb{P}\left\{|V_\alpha| \le \lambda_n\right\} &\le \mathbb{P}\left\{V_\alpha \le \lambda_n\right\} \le \mathbb{P}\left\{\left(1 + \sqrt{2\nu}/\gamma\right)\lambda_n + \sqrt{2\nu d_\alpha^2} Z \le \lambda_n\right\} \\ &\le \mathbb{P}\left\{Z \le -(\sqrt{2\nu}/\gamma)\lambda_n/(\sqrt{2\nu})\right\} \le \frac{\phi\left(\lambda_n/\gamma\right)}{\lambda_n/\gamma} = \frac{n^{-1}}{\sqrt{2\log n}} \,,\end{aligned}$$

which implies that $\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha|\,>\,\left(1+\sqrt{2\nu}/\gamma\right)\lambda_n} n \cdot \mathbb{P}\left\{|V_\alpha| \le \lambda_n\right\} \;=\; 0$. Also,

$$\begin{aligned}\mathbb{P}\left\{|V_\alpha| \le \lambda_n\right\} &\le \mathbb{P}\left\{V_\alpha \le \lambda_n\right\} \le \mathbb{P}\left\{\theta_\alpha + \sqrt{2\nu d_\alpha^2} Z \le \lambda_n\right\} \le \mathbb{P}\left\{Z \le -(\theta_\alpha - \lambda_n)/(\sqrt{2\nu})\right\} \\ &\le \frac{\phi\left((\theta_\alpha - \lambda_n)/(\sqrt{2\nu})\right)}{(\theta_\alpha - \lambda_n)/(\sqrt{2\nu})} = \frac{e^{-\theta_\alpha^2\left(1-\frac{\lambda_n}{\theta_\alpha}\right)^2/(4\nu)}}{\theta_\alpha\left(1 - \frac{\lambda_n}{\theta_\alpha}\right)/(\sqrt{2\nu})} \le \frac{e^{-\theta_\alpha^2/(2(\gamma+\sqrt{2\nu})^2)}}{\theta_\alpha/(\gamma + \sqrt{2\nu})} \,,\end{aligned}$$

where the last inequality follows from the fact that $\theta_\alpha > \left(1 + \sqrt{2\nu}/\gamma\right)\lambda_n$, which implies that $1 > 1 - (\lambda_n/\theta_\alpha) > \sqrt{2\nu}/(\gamma + \sqrt{2\nu})$. Note that for any $a > 0$, $\max_{x\ge 0} x e^{-ax} = 1/(ea)$, which implies that

$$\theta_\alpha^2 \cdot \mathbb{P}\left\{|V_\alpha| \le \lambda_n\right\} \le \frac{\theta_\alpha^2 e^{-\theta_\alpha^2/(2(\gamma+\sqrt{2\nu})^2)}}{\theta_\alpha/(\gamma + \sqrt{2\nu})} \le \frac{2(\gamma + \sqrt{2\nu})^2/e}{\theta_\alpha/(\gamma + \sqrt{2\nu})} = \frac{2(\gamma + \sqrt{2\nu})^3/e}{\left(1 + \sqrt{2\nu}/\gamma\right)\lambda_n} \,,$$

which implies that $\lim_{n\to\infty} \sup_{\alpha\,:\,|\theta_\alpha| \,>\, \left(1+\sqrt{2\nu}/\gamma\right)\lambda_n} \theta_\alpha^2 \cdot \mathbb{P}\left\{|V_\alpha| \le \lambda_n\right\} \;=\; 0$, which is the desired result. To complete the proof for Case 3, we will show that

$$\lim_{n\to\infty} \sup_{\alpha\,:\,\theta_\alpha \,>\, (1+\sqrt{2\nu}/\gamma)\lambda_n} \theta_\alpha^2 \;\cdot\; \mathbb{P}\{V_\alpha < -\lambda_n\} = 0\ .$$

This follows immediately from the above analysis as $\mathbb{P}\{V_\alpha < -\lambda_n\} \le \mathbb{P}\{V_\alpha < \lambda_n\}$, and we have just shown that $\lim_{n\to\infty} \sup_{\alpha\,:\,\theta_\alpha \,>\, (1+\sqrt{2\nu}/\gamma)\lambda_n} \theta_\alpha^2 \cdot \mathbb{P}\left\{V_\alpha \le \lambda_n\right\} = 0$.

APPENDIX B: PROOF DETAILS FOR ESTIMATORS IN THE CLASS $\mathcal{S}$ AND THE LEMMAS USED IN SECTION 4

B.1. Proof of Lemma 4.1. Using the relation between $\widehat{\mathrm{ARE}}$ and $\widehat{\mathrm{ARE}}^{\mathrm{D}}$ it follows that

$$\mathbb{E}\big[\big(\widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta,\tau) - R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta,\tau))\big)^2\big] = \mathbb{E}\big[\big(\widehat{\mathrm{ARE}}_n(\tau, \boldsymbol{X}-\eta) - R_n(\boldsymbol{\theta}-\eta, \hat{\boldsymbol{q}}(\tau))\big)^2\big]\ .$$

Now, similarly as in the proof of Theorem 1.1, following the Bias-Variance decomposition and the argument below (2.1) we upper bound the RHS above by

$$\begin{aligned} &A_n\left\{\left(\frac{1}{n}\sum_{i=1}^n \mathrm{Bias}_{\theta_i}(T_i(X_i-\eta, Z_i, \tau))\right)^2 + \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_{\theta_i}(T_i(X_i-\eta, Z_i, \tau))\right\} \\ &= A_n\left\{\left(\frac{1}{n}\sum_{i=1}^n \mathrm{Bias}_{\theta_i-\eta}(T_i(X_i, Z_i, \tau))\right)^2 + \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_{\theta_i-\eta}(T_i(X_i, Z_i, \tau))\right\} \end{aligned}$$

where $\boldsymbol{Z} = \{Z_1, \ldots, Z_n\}$ follows $N(0, I_n)$; $T_i(X_i-\eta, Z_i, \tau)$ are randomized rules defined in Section 1.3; $A_n = \sup\{(b_i+h_i)^2(\nu_{f,i} + \alpha_i\nu_{p,i}) : i = 1, \ldots, n\}$ which by Assumption A2 satisfies $\sup_n A_n \le \infty$. Now, using Lemma 1.1 we get the required result: $\sup_{\tau\in[0,\infty], \eta\in\mathbb{R}} \mathbb{E}\big(\widehat{\mathrm{ARE}}_n^{\mathrm{D}}(\eta,\tau) - R_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}(\eta,\tau))\big)^2 \to 0$ as $n\to\infty$.

APPENDIX C: PROOF DETAILS FOR ESTIMATORS IN THE CLASS $\mathcal{S}^G$ AND THE LEMMAS USED IN SECTION 5

C.1. Proof of Lemma 5.1. By a first order Taylor series expansion we have:

$$\tilde{l}_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau)) = l_i(\theta_i, \hat{q}_i^{\mathrm{G}}(\tau)) + a_i\,(\bar{\boldsymbol{X}} - \bar{\boldsymbol{\theta}})\left[\frac{\partial}{\partial\eta} G(\nu_{f,i}^{-1/2}(\hat{q}_i(\tau) + (1-\alpha_i)\eta - \theta_i), \tilde{b})\right]_{\eta=\mu_i}$$

where μ_i lies between $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{\theta}}$ and $a_i = \nu_{f,i}^{1/2}\,(b_i+h_i)$. Again, based on the definition of G from Equation (1.3) we have for any $\tau \ge 0$ and any $\eta \in \mathbb{R}$:

$$\left|\frac{\partial}{\partial\eta} G\big(\nu_{f,i}^{-1/2}(\hat{q}_i(\tau) + (1-\alpha_i)\eta - \theta_i), \tilde{b}\big)\right| \le \nu_{f,i}^{-1/2} \text{ for all } i = 1, \ldots, n.$$

Thus, for the multivariate versions we have:

$$\sup_{\tau\in[0,\infty]} |L_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau)) - \tilde{L}_n(\boldsymbol{\theta}, \hat{\boldsymbol{q}}^{\mathrm{G}}(\tau))| \le |\bar{\boldsymbol{X}} - \bar{\boldsymbol{\theta}}| \cdot \frac{1}{n}\sum_{i=1}^{n}(b_i + h_i)$$

which converges to 0 in L^1 as $\bar{\boldsymbol{X}} \sim N(\bar{\boldsymbol{\theta}}, n^{-1})$ and $n^{-1}\sum_{i=1}^{n}(b_i + h_i)$ is bounded by Assumption A1. This completes the proof.

C.2. Proof of Lemma 5.2. From the description of the ARE procedure in Section 1.3, recall that, $\hat{T}_i(X_i - \eta, \tau) = \mathbb{E}\{\tilde{T}_i(X_i - \eta, \boldsymbol{Z}, \tau)\}$ where the expectation is over $\boldsymbol{Z}$ which is independent of X and follows $N(0, I_n)$ distribution. And,

$$\hat{T}_i(X_i - \eta, Z_i, \tau) = \begin{cases} -\tilde{b}_i\, U_i(\eta,\tau) & \text{if } V_i(\eta,\tau) < -\lambda_n(i) \\ \tilde{S}_i(U_i(\eta,\tau)) & \text{if } |V_i(\eta,\tau)| \le \lambda_n(i) \\ (1-\tilde{b}_i)\, U_i(\eta,\tau) & \text{if } V_i(\eta,\tau) > \lambda_n(i) \end{cases} \quad \text{for } i = 1,\ldots,n$$

where the threshold parameter defined in (1.11) and

$$U_i(\eta,\tau) = c_i(\tau) + d_i(\tau)(X_i - \eta + \nu_{p,i}^{1/2} Z_i), V_i(\eta,\tau) = c_i(\tau) + d_i(\tau)(X_i - \eta - \nu_{p,i}^{1/2} Z_i)$$

with $|d_i(\tau)|$ is less than $\nu_{f,i}^{-1/2}$. $\tilde{S}_i(U_i(\eta,\tau))$ is a truncated version of

$$\begin{aligned} S_i(U_i(\eta,\tau)) =& G(0,\tilde{b}_i) + G'(0,\tilde{b}_i)U_i(\eta,\tau) \\ &+ \phi(0)\sum_{k=0}^{K-2} \frac{(-1)^k H_k(0)}{(k+2)!}\big(2\nu_{p,i}d_i^2(\tau)\big)^{\frac{k+2}{2}} H_{k+2}\bigg(\frac{U_i(\eta,\tau)}{\sqrt{2\nu_{p,i}d_i^2(\tau)}}\bigg). \end{aligned}$$

So, the derivative exists almost everywhere and for all $i = 1,\ldots,n$:

$$\text{(C.1)} \qquad \left|\frac{\partial}{\partial\eta}\hat{T}_i(X_i - \eta, Z_i, \tau)\right| \le \begin{cases} \nu_{f,i}^{-1/2}|\tilde{S}_i'(U_i(\eta,\tau))| & \text{if } |V_i(\eta,\tau)| < \lambda_n(i) \\ \nu_{f,i}^{-1/2} & \text{if } |V_i(\eta,\tau)| > \lambda_n(i) \end{cases}.$$

Noting that for Hermite polynomials of order k the derivative satisfies: $H_k'(x) = kH_{k-1}(x)$, we have $\tilde{S}_i'(U_i(\eta,\tau))$ exempting the two discontinuity points is either 0 or given by:

(C.2)

$$\nu_{f,i}^{-1/2}\bigg\{G'(0,\tilde{b}_i) + \phi(0)\sum_{k=0}^{K-2}\frac{(-1)^k H_k(0)}{(k+1)!}\big(2\nu_{p,i}d_i^2(\tau)\big)^{\frac{k+1}{2}} H_{k+1}\bigg(\frac{U_i(\eta,\tau)}{\sqrt{2\nu_{p,i}d_i^2(\tau))}}\bigg)\bigg\}.$$

Define, $\theta_i(\eta,\tau) = c_i(\tau) + d_i(\tau)(\theta_i - \eta)$ and $\Lambda_i(\eta,\tau) = (1 + \sqrt{2\nu_{p,i}}/\gamma_i)\lambda_n(i)$. Now, by exactly following the proof technique used in Lemma 2.5, it can be shown that:

$$\text{(C.3)} \qquad \sup_{\tau\ge 0} \sup_{|\theta_i(\eta,\tau)|\le\Lambda_i(\eta,\tau)} \mathbb{E}_{\theta_i(\eta,\tau)}\big\{\tilde{S}_i'(U_i(\eta,\tau))\big\}^2 = o(n) \text{ as } n\to\infty,$$

where γ_i is defined below Equation (1.11) and the expectation is over the distribution of $U_i(\eta,\tau)$ which follows $N(\theta_i(\eta,\tau), 2d_i^2(\tau)\nu_{p,i})$. So, by (C.1) for all values of η, τ and θ_i such that $|\theta_i(\eta,\tau)| \le \Lambda_i(\eta,\tau)$ we have:

$$\mathbb{E}_{\theta_i}\left(\frac{\partial}{\partial\eta}\hat{T}_i(X_i-\eta,Z_i,\tau)\right)^2 = o(n)$$

where the expectation is over the joint distribution of X_i and Z_i.

We now concentrate on all values of η, τ and θ_i such that $|\theta_i(\eta,\tau)| > \Lambda_i(\eta,\tau)$. For this note that for all large n, $|\tilde{S}_i'(U_i(\eta,\tau)| \le n$ a.e. It follows as: $\tilde{S}_i(U_i(\eta,\tau)$ is truncated above $\pm n$ and so by definition of derivative we have $|\tilde{S}_i'(U_i(\eta,\tau)| \le n$ for all $|U_i(\eta,\tau)| \ge 1$; and when $|U_i(\eta,\tau)| < 1$, using uniform approximation bounds (Szegö, 1939) on the Hermite polynomials in the expression (C.2), we have $|\tilde{S}_i'(U_i(\eta,\tau)| \le n$. So, using (C.1) for all values of η,τ and θ_i such that $|\theta_i(\eta,\tau)| > \Lambda_i(\eta,\tau)$ we have the following upper bound:

$$\mathbb{E}_{\theta_i}\left(\frac{\partial}{\partial\eta}\hat{T}_i(X_i-\eta,Z_i,\tau)\right)^2 \le \nu_{f,i}^{-1}\left\{n^2P\big(|V_i(\eta,\tau)| < \lambda_n(i)\big) + P\big(|V_i(\eta,\tau)| > \lambda_n(i)\big)\right\}$$

where $V_i(\eta,\tau)$ has $N(\theta_i(\eta,\tau), 2d_i^2(\tau)\nu_{p,i})$ distribution. Also from Lemma 2.6 we know that: $\sup_{|\theta_i(\eta,\tau)|>\Lambda_i(\eta,\tau)} n^2P(|V_i(\eta,\tau)| < \lambda_n(i)) = o(n)$ which produces the desired bound for $|\theta_i(\eta,\tau)| > \Lambda_i(\eta,\tau)$. Thus, we have:

$$\sup_i \sup_{\tau\in[0,\infty]} \mathbb{E}_{\theta_i}\left(\frac{\partial}{\partial\eta}\hat{T}_i(X_i-\eta,Z_i,\tau)\right)^2 = o(n).$$

As $\hat{T}_i(X_i-\eta,\tau) = \mathbb{E}\{\hat{T}_i(X_i-\eta,\boldsymbol{Z},\tau)\}$, by Jensen's inequality we have:

$$\mathbb{E}_{\theta_i}\left(\frac{\partial}{\partial\eta}\hat{T}_i(X_i-\eta,Z_i,\tau)\right)^2 \ge \mathbb{E}_{\theta_i}\left\{\mathbb{E}\left(\frac{\partial}{\partial\eta}\hat{T}_i(X_i-\eta,Z_i,\tau)\middle| X_i\right)\right\}^2$$

and the result of the lemma in terms of $\hat{T}_i(X_i-\eta,\tau)$ follows.

APPENDIX D: AUXILIARY LEMMAS

The following lemma provides an upper bound on Hermite polynomial.

LEMMA D.1. *There is an absolute constant c such that for all $k \ge 1$ and $x \in \mathbb{R}$,*

$$|H_k(x)| \le c\, e^{x^2/4}\frac{k!}{k^{1/3}\,(k/e)^{k/2}}.$$

Proof of Lemma D.1

Krasikov (2004) shows that for $k \ge 6$,

$$(2k)^{1/6}2^k \max_{x\in\mathbb{R}}(H_k(x))^2e^{-x^2/2} \;\le\; \frac{2}{3}C_k\exp\left(\frac{15}{8}\left(1+\frac{12}{4(2k)^{1/3}-9}\right)\right)\;,$$

where

$$C_k = \begin{cases} \frac{2k\sqrt{4k-2}\ (k!)^2}{\sqrt{8k^2-8k+3}\ ((k/2)!)^2}, & \text{if} \quad k \text{ is even}, \\ \frac{\sqrt{16k^2-16k+6}\ k!(k-1)!}{\sqrt{2k-1}\ ((k-1)/2)!^2}, & \text{if} \quad k \text{ is odd}, \end{cases}$$

Note that for $k \geq 6$, $\frac{2}{3}e^{\frac{15}{8}\left(1+\frac{12}{4(2k)^{1/3}-9}\right)}$ is decreasing in k and

$$\begin{aligned} C_k &\leq \begin{cases} 2\sqrt{k} \times \frac{(k!)^2}{((k/2)!)^2}, & \text{if} \quad k \text{ is even}, \\ 4\sqrt{k} \times \frac{k!(k-1)!}{((k-1)/2)!^2}, & \text{if} \quad k \text{ is odd}, \end{cases} \\ &\leq \begin{cases} 2\sqrt{k} \times \frac{(k!)^2}{\left(\sqrt{2\pi}\left(\frac{k}{2}\right)^{(k+1)/2}e^{-k/2}\right)^2}, & \text{if} \quad k \text{ is even}, \\ 4\sqrt{k} \times \frac{(k!)^2}{k\left(\sqrt{2\pi}\left(\frac{k-1}{2}\right)^{k/2}e^{-(k-1)/2}\right)^2}, & \text{if} \quad k \text{ is odd}, \end{cases} \\ &= \begin{cases} 2\sqrt{k} \times \frac{(k!)^2}{\frac{k}{2}(2\pi)\left(\left(\frac{k}{2}\right)^{k/2}e^{-k/2}\right)^2}, & \text{if} \quad k \text{ is even}, \\ 4\sqrt{k} \times \frac{(k!)^2}{ke(2\pi)\left(\left(\frac{k}{2}\right)^{k/2}\times\left(1-\frac{1}{k}\right)^{k/2}\times e^{-k/2}\right)^2}, & \text{if} \quad k \text{ is odd}, \end{cases} \\ &= \begin{cases} 4\sqrt{k} \times \frac{(k!)^2}{k(2\pi)\left(\left(\frac{k}{2}\right)^{k/2}e^{-k/2}\right)^2}, & \text{if} \quad k \text{ is even}, \\ 4\sqrt{k} \times \frac{(k!)^2}{ke(2\pi)\left(1-\frac{1}{k}\right)^{k}\left(\left(\frac{k}{2}\right)^{k/2}e^{-k/2}\right)^2}, & \text{if} \quad k \text{ is odd}, \end{cases} \end{aligned}$$

where the second equality follows from Sterling's bound. It is easy to verify that $(1-\frac{1}{k})^k$ is increasing in k and approaches $\frac{1}{e}$ as k approaches infinity. Putting everything together, we conclude that there is an absolute constant a_1 such that for all $k \geq 6$ and $x \in \mathbb{R}$,

$$|H_k(x)| \leq a_1 e^{x^2/4}\frac{k!}{k^{1/3}\,2^{k/2}\left(\frac{k}{2e}\right)^{k/2}} = a_1 e^{x^2/4}\frac{k!}{k^{1/3}\left(\frac{k}{e}\right)^{k/2}}$$

Since the results hold for $k \geq 6$, it is easy to verify that it also holds for all $k \geq 1$, by choosing appropriately large constant a_1.

LEMMA D.2. *If $X \sim N(\theta, 1)$ then*

$$\mathbb{E}H_k^2(X) \leq k^k(1+\theta^2/k)^k.$$

Proof. See Lemma 3 of Cai et al. (2011).

LEMMA D.3. *If Y and I_A are independent random variables then:*

- $\mathrm{Var}(YI_A) = \mathrm{Var}(Y)P(A) + (\mathbb{E}[Y])^2 P(A)P(A^c)$
- $\mathrm{Var}(YI_A) \leq \mathbb{E}[Y^2]\, P(A)$
- $\mathrm{Var}(YI_A) \leq \mathrm{Var}(Y) + (\mathbb{E}[Y])^2 P(A^c)$

Proof. Using the independence between Y and I_A we have $\mathrm{Var}(YI_A)$ equals $\mathbb{E}[Y^2]P(A) - (\mathbb{E}[Y]P(A))^2 = \mathrm{Var}(Y)P(A) + (\mathbb{E}[Y])^2 P(A) - (\mathbb{E}[Y]P(A))^2$ from which we have the identity in the lemma. The inequalities immediately follow from it.

LEMMA D.4. *Mills Ratio and Gaussian Tails: For any $a > 0$ we have:*

$$-a^{-3}\phi(a) \leq \tilde{\Phi}(a) - a^{-1}\phi(a) \leq 0.$$

Proof. See Exercise 8.1, Chapter 8 in Johnstone (2013).

LEMMA D.5. *For any $w \in \mathbb{R}$ and $b \in (0,1)$, let $G(w,b)$ be defined as in Equation (1.3). Then, $G(w,b) \leq \phi(0) + \max\{1-b\,,\, b\}|w|$.*

PROOF. By definition $G(w,b) = \phi(w) + w\Phi(w) - bw$. Since $\phi(w) \leq \phi(0)$ for all w, the result follows. □

LEMMA D.6. *Extra Integrability condition. If family $\{X_t : t \in T\}$ is such that $\sup_{t\in T} \mathbb{E}|X_t|^{1+\delta} < \infty$ for some $\delta > 0$ then $\{X_t : t \in T\}$ is uniformly integrable.*

Proof. See Billingsley (2008).

LEMMA D.7. *For any fixed $m > 0$ we have*

$$\left(1 + \frac{m}{k}\right)^k \leq e^m \text{ for all } k \geq 1$$

Proof. We know that for any $x > 0$, $\log(1+x) \leq x$ and taking logarithm and dividing both sides by m the statement in the lemma reduces to $\log(1 + m/k) \leq m/k$.

The following well known random variable lemmas have been used in our proofs.

LEMMA D.8. *For random variables $W_1, \ldots, W_n$ we have:*

$$\mathbb{E}\left[\left(\sum_{i=1}^n W_i\right)^2\right] \leq \left(\sum_{i=1}^n \sqrt{\mathbb{E}(W_i^2)}\right)^2$$

LEMMA D.9. *For any random variable X and $\lambda > 0$, we have*

$$\mathbb{E}\{XI\{X \geq \lambda\}\} \leq |\lambda|^{-1}\mathbb{E}X^2.$$

LEMMA D.10. *For any finite $l \geq 1$ we have*

$$\mathrm{Var}\Big(\sum_{i=1}^l X_i\Big) \leq 2^{l-1}\sum_{i=1}^l \mathrm{Var}(X_i).$$

REFERENCES

Aitchison, J. and Dunsmore, I. R. (1976). Statistical Prediction Analysis. *Bulletin of the American Mathematical Society* **82** 683–688.

Arrow, K. J., Harris, T. and Marschak, J. (1951). Optimal inventory policy. *Econometrica* **19** 250–272.

Azoury, K. S. (1985). Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science* **31** 1150–1160.

Berger, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *The Annals of Statistics* **4** 223–226.

Billingsley, P. (2008). *Probability and Measure*. John Wiley & Sons, New York.

Brown, L. D. (1975). Estimation with incompletely specified loss functions (the case of several location parameters). *Journal of the American Statistical Association* **70** 417–427.

Brown, L. D. (2008). In-season prediction of batting averages: a field test of empirical Bayes and Bayes methodologies. *Ann. Appl. Stat.* **2** 113–152.

Cai, T. T., Low, M. G. et al. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *The Annals of Statistics* **39** 1012–1041.

Chang, S. H. and Fyffe, D. E. (1971). Estimation of forecast errors for seasonal style-goods sales. *Management Science* **18** B89–B96.

Chihara, T. S. (2011). *An Introduction to Orthogonal Polynomials*. Dover, New York.

Choi, T. (2012). Handbook of Newsvendor Problems: Models, Extensions and Applications, International Series in Operations Research & Management Science, Vol. 176.

DasGupta, A. and Sinha, B. K. (1999). A new general interpretation of the Stein estimate and how it adapts: Applications. *Journal of Statistical Planning and Inference* **75** 247 - 268.

Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *The Annals of Statistics* **13** 1581–1591.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* **90** 1200–1224.

Edgeworth, F. Y. (1888). The mathematical theory of banking. *Journal of the Royal Statistical Society* **51** 113–127.

Efron, B. and Morris, C. (1973a). Combining possibly related estimation problems. *Journal of the Royal Statistical Society. Series B (Methodological)* **35** 379–421.

Efron, B. and Morris, C. (1973b). Stein's estimation rule and its competitors—an empirical Bayes approach. *Journal of the American Statistical Association* **68** 117–130.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association* **70** 311–319.

Geisser, S. (1993). *Predictive Inference. Monographs on Statistics and Applied Probability* **55**. Chapman and Hall, New York.

George, E. I., Liang, F. and Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *The Annals of Statistics* **34** 78–91.

George, E. I. and Strawderman, W. E. (2012). A tribute to Charles Stein. *Stat. Sci.* **27** 1–2.

Good, I. J. (1980). Some history of the hierarchical Bayesian methodology. *Trabajos de estadística y de investigación operativa* **31** 489–519.

Greenshtein, E. and Ritov, Y. (2009). Asymptotic efficiency of simple decisions for the compound decision problem. In *Optimality: The Third Erich L. Lehmann Symposium* **57** 266–275.

Hoffmann, K. (2000). Stein estimation—A review. *Statistical Papers* **41** 127-158.

Iglehart, D. L. (1964). The dynamic inventory problem with unknown demand distribution. *Management Science* **10** 429–440.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley symposium on Mathematical Statistics and Probability* **1** 361–379.

Johnstone, I. M. (2013). Gaussian Estimation: Sequence and Wavelet Models. Version: 11 June, 2013. Available at `"http://www-stat.stanford.edu/~imj"`.

Karlin, S. (1960). Dynamic inventory policy with varying stochastic demands. *Management Science* **6** 231–258.

Karlin, S. and Scarf, H. (1958). Inventory models of the Arrow-Harris-Marschak type with time lag. In *Studies in the Mathematical Theory of Inventory and Production* Stanford University

Press.
Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society* 33–50.
Krasikov, I. (2004). New bounds on the Hermite polynomials. *arXiv preprint math/0401310*.
Lariviere, M. A. and Porteus, E. L. (1999). Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* **45** 346–363.
Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation* **31**. Springer Science & Business Media.
Lindley, D. (1962). Discussion of the paper by Stein. *J. Roy. Statist. Soc. Ser. B* **24** 265–296.
Lovejoy, W. S. (1990). Myopic policies for some inventory models with uncertain demand distributions. *Management Science* **36** 724–738.
Mallows, C. L. (1973). Some comments on C p. *Technometrics* **15** 661–675.
Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association* **78** 47–55.
Mukherjee, G. and Johnstone, I. M. (2015). Exact minimax estimation of the predictive density in sparse Gaussian models. *The Annals of Statistics* **43** 937–961.
Murray, G. R. and Silver, E. A. (1966). A Bayesian analysis of the style goods inventory problem. *Management Science* **12** 785–797.
Newey, W. K. and McFadden, D. (1994). Chapter 36: Large sample estimation and hypothesis testing. (R. F. Engle and D. L. McFadden, eds.). *Handbook of Econometrics* **4** 2111–2245. Elsevier, Edition 1.
Press, S. J. (2009). *Subjective and objective Bayesian statistics: principles, models, and applications* **590**. John Wiley & Sons.
Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *The Annals of Mathematical Statistics* **35** 1–20.
Robbins, H. (1985). Asymptotically subminimax solutions of compound statistical decision problems. In *Herbert Robbins Selected Papers* 7–24. Springer.
Scarf, H. (1959). Bayes solution to the statistical inventory problem. *Annals of Mathematical Statistics* **30** 490–508.
Scarf, H. (1960). Some remarks on Bayes solutions to the inventory problem. *Naval Research Logistics Quarterly* **7** 591–596.
Stein, C. M. (1962). Confidence sets for the mean of a multivariate normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)* **24** 265–296.
Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics* **9** 1135–1151.
Steinwart, I. and Christmann, A. (2011). Estimating conditional quantiles with the help of the pinball loss. *Bernoulli* **17** 211–225.
Stigler, S. M. (1990). The 1988 Neyman memorial lecture: a Galtonian perspective on shrinkage estimators. *Statistical Science* **5** 147–155.
Szegö, G. (1939). *Orthogonal Polynomials* **23**. American Mathematical Soc.
Thangavelu, S. (1993). *Lectures on Hermite and Laguerre Expansions* **42**. Princeton Uni. Press.
Xie, X., Kou, S. and Brown, L. D. (2012). SURE estimates for a heteroscedastic hierarchical model. *Journal of the American Statistical Association* **107** 1465–1479.
Xie, X., Kou, S. and Brown, L. D. (2015). Optimal shrinkage estimation of mean parameters in family of distributions with quadratic variance. *Journal of the American Statistical Association (to appear)* **45**.
Zhang, C.-H. (2003). Compound decision theory and empirical Bayes methods: invited paper. *Ann. Statist.* **31** 379–390.

Address of the First and Third authors
3670 Trousdale Parkway,
401 Bridge Hall,
University of Southern California,
Los Angeles, CA 90089-0809
E-mail: gourab@usc.edu
rusmevic@usc.edu

Address of the Second author
Department of Statistics
University of Pennsylvania
400 Jon M. Huntsman Hall
3730 Walnut Street
Philadelphia, PA 19104
E-mail: lbrown@wharton.upenn.edu