Asymptotic Expansion and Bounds for the Bias of Empirical Tail Value-at-Risk

Nadezhda Gribkova [*,1,2], Jianxi Su [†,3,4], and Mengqi Wang [‡,4,5]

[1]Saint Petersburg State University, Saint Petersburg, 199034 Russia

[2]Emperor Alexander I Saint Petersburg State Transport University, Saint Petersburg, 190031 Russia

[3]Purdue University, West Lafayette, Indiana 47907, United States

[4]RISC Foundation, Toronto, Ontario M2N 7E9, Canada

[5]Western University, London, Ontario N6A 5B7, Canada

January 16, 2026

Abstract

Tail Value-at-Risk (TVaR) is a widely adopted risk measure playing a critically important role in both academic research and industry practice in insurance. In data applications, TVaR is often estimated using the empirical method, owing to its simplicity and nonparametric nature. The empirical TVaR has been explicitly advocated by regulatory authorities as a standard approach for computing TVaR. However, prior literature has pointed out that the empirical TVaR estimator is negatively biased, which can lead to a systemic underestimation of risk in finite-sample applications. This paper aims to deepen the understanding of the bias of the empirical TVaR estimator in two dimensions: its magnitude as well as the key distributional and structural determinants driving the severity of the bias. To this end, we derive a leading-term approximation for the bias based on its asymptotic expansion. The closed-form expression associated with the leading-term approximation enables us to obtain analytical insights into the structural properties governing the bias of the empirical TVaR estimator. To account for the discrepancy between the leading-term approximation and the true bias, we further derive an explicit upper bound for the bias. We validate the proposed bias analysis framework via simulations and demonstrate its practical relevance using real data.

Keywords: Expected shortfall, bootstrap, finite-sample bias, heavy-tailed distributions, trimmed means

Dedication:

This research was initiated by Ričardas Zitikis (1963–2025) in early 2025. This article is dedicated to his memory.

[*]e-mail: n.gribkova@spbu.ru

[†]Corresponding author; e-mail: jianxi@purdue.edu

[‡]e-mail: mwan259@uwo.ca

1 Introduction

The measurement of risk plays a pivotal role across a wide spectrum of applications in finance and insurance, including but not limited to, determining risk-loaded prices, assessing capital adequacy, supporting regulatory compliance, measuring risk-based performance, and guiding portfolio optimization. Among many risk measures employed in these contexts, Tail Value-at-Risk (TVaR), which is also referred to as AVaR, CVaR, or expected shortfall, has gained particular prominence among academics and professionals, especially following its adoption as the replacement for Value-at-Risk (VaR) as the standard risk measure within recent regulatory frameworks (e.g., BCBS, 2016, 2019). In detail, consider the risk position of a generic one-period model, denoted by X, where positive values represent losses and negative values represent surpluses. Let F be the cumulative distribution function (CDF) of X. The left-continuous inversion of F is given by

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : \ F(x) \geq u\}, \qquad u \in (0,1).$$

For a given probability level $p \in (0,1)$, the TVaR of X at level p is defined, provided the integral is finite, as

$$T_p = \frac{1}{1-p} \int_p^1 F^{-1}(u) \, \mathrm{d}u,$$

which can be interpreted as the average of the upper tail quantiles of X.

Given the practical prominence of TVaR, a large and continually growing body of literature has been generated. A wide range of its economically grounded properties have been extensively studied, including its adherence to the coherence axioms (Artzner et al., 1999), comonotonic additivity (Kusuoka, 2001), and its representation as a Choquet integral (Schmeidler, 1989) and as the solution to an expected loss minimization problem (Rockafellar and Uryasev, 2002). We also refer readers to Wang and Zitikis (2021) for an axiomatic characterization of TVaR, which also includes a comprehensive review of recent theoretical developments related to TVaR.

While different from, yet closely related to, the aforementioned literature on studying the mathematical and economic properties of TVaR, the primary objective of this paper pertains to the estimation of TVaR. In applied settings, the choice of estimation method is often guided by practical considerations. Practitioners often favor approaches that are simple to implement, transparent to communicate, and whose performance does not rely on a specific set of parametric assumptions. One such widely adopted method is the empirical, or plug-in, estimator of TVaR:

$$\widehat{T}_{n,p} = \frac{1}{1-p} \int_p^1 \widehat{F}_n^{-1}(u) \, \mathrm{d}u, \tag{1.1}$$

where $\widehat{F}_n^{-1}$ denotes the inversion of the empirical CDF $\widehat{F}_n$, based on $n \in \mathbb{N}$ independent and identically distributed (iid) observations $X_1, \ldots, X_n$ drawn from the distribution of X. In particular, the empirical CDF is computed via

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}},$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function, and its inversion can be calculated as

$$\widehat{F}_n^{-1}(u) = \inf\{x \in \mathbb{R} : \ \widehat{F}_n(x) \geq u\} = \begin{cases} X_{nu:n}, & \text{if } nu \text{ is an integer;} \\ X_{[nu]+1:n}, & \text{otherwise,} \end{cases} \qquad u \in (0,1),$$

where $[\cdot]$ represents the greatest integer function, and $X_{1:n} \leq \cdots \leq X_{n:n}$ are the order statistics of the samples $X_1, \ldots, X_n$. Throughout the rest of this paper, empirical TVaR refers to the estimator in (1.1), unless specified otherwise. This empirical estimator has been recommended by the European Banking Authority as the standard formula for computing TVaR (EBA, 2020).

The large-sample properties of empirical TVaR $\widehat{T}_{n,p}$ have been thoroughly studied in the related literature (e.g., Brazauskas et al., 2008; Wei and Zitikis, 2023). More general results based on the empirical tail conditional allocation have been developed in Gribkova et al. (2022a,b). In particular, it has been shown that $\widehat{T}_{n,p}$ in (1.1) is consistent. Thereby, empirical TVaR $\widehat{T}_{n,p}$ ultimately provides an accurate evaluation of the true TVaR as the sample size tends to infinity.

That said, in real-world applications, data are finite and sometimes not very large. In such cases, empirical TVaR in (1.1) has been found to exhibit a negative bias (Gribkova et al., 2026). More precisely, existing literature has established the existence and sign of this bias, but not its magnitude. Consequently, two possible scenarios emerge. If the bias is small even for moderately sized samples and diminishes rapidly as the sample size increases, then empirical TVaR can still be used with confidence. On the other hand, if the bias is substantial and persistent, caution is warranted, and bias reduction techniques should be employed. Otherwise, the inherent negative bias may lead to a systematic underestimation of the riskiness of a financial position as measured by TVaR.

The goal of our paper is to examine the magnitude of the negative bias in empirical TVaR $\widehat{T}_{n,p}$, thereby providing insights into the two potential scenarios discussed above. While numerical methods such as jackknife (Quenouille, 1956; Tukey, 1958) and bootstrap (Efron and Tibshirani, 1994) are commonly used to estimate the bias of an estimator for a given dataset, the resulting bias estimates are inherently tied to the specific sample at hand. As such, these methods are limited in their ability to provide a structural understanding of how the distributional characteristics of data

impact the extent of the bias. Moreover, they offer little insight into how the bias evolves as the sample size increases.

In this paper, we take an analytical route to analyze the bias of empirical TVaR in (1.1). Specifically, following the initial conjecture the same authors of the present paper made in equation (30) of Gribkova et al. (2026), this paper formally establishes the asymptotic expansion of the bias under detailed conditions and with a rigorous proof. We then use the leading term of this expansion as a practical approximation for analyzing the bias. It is important to emphasize that, although the bias of the empirical TVaR estimator vanishes asymptotically due to its consistency property, the proposed leading-term approximation still offers useful practical insights, even in finite-sample settings. This approach is analogous to using the leading term of a Taylor expansion to analyze the behavior of a complicated function. Compared with numerical resampling techniques, the closed-form expression obtained from the leading-term approximation provides explicit insight into the distributional characteristics that drive the magnitude of the bias, as well as the rate at which the bias diminishes as the sample size grows.

The aforementioned leading-term approximation, while informative, does not generally coincide exactly with the finite-sample bias. Thereby, we complement it with an explicit upper bound for the negative bias. Our numerical experiments indicate that, with an appropriate choice of tuning parameters, the proposed upper bound can achieve a reasonable balance between tightness and conservativeness when assessing the bias of the empirical TVaR.

Our proposed bias analysis framework carries significant practical relevance. Beyond providing analytical insights into the structural properties of the bias of empirical TVaR, the framework delivers tractable and satisfactory approximations of the bias magnitude across a range of probability levels and sample sizes, without requiring repeated simulation or resampling procedures commonly used in numerical approaches. Understanding these relationships is particularly useful for practitioners when designing a risk assessment procedure, where trade-offs between data availability constraints and feasible or desirable choices of probability levels in TVaR applications must be balanced. Moreover, as we will see in the numerical experiments, simulation-based bias evaluation methods such as the bootstrap may, due to sampling variability, occasionally produce improper bias estimates contradicting the theoretical property that the bias of empirical TVaR is always negative. In contrast, the proposed bias analysis framework is deterministic and therefore does not suffer from this drawback. When a risk analyst concerns the extent of bias, the bias evaluation framework developed in this paper can be readily employed to implement bias reduction.

Finally, we note that, in addition to the empirical method considered in this paper, other commonly used approaches for estimating TVaR include parametric methods (e.g., Chen et al., 2023; Brazauskas and Kleefeld, 2009; Lee and Lin, 2010), kernel-based methods (e.g., Bolance

et al., 2008; Chen, 2008; Yu et al., 2010), and methods originated from Extreme Value Theory (e.g., Goegebeur et al., 2022; Hua and Joe, 2011; Necir et al., 2010). We also refer the reader to Nadarajah et al. (2014) for a systematic review of different methods for estimating TVaR. It is important to note taht the objective of this paper is not to advocate for the superiority of one estimation method over another for computing TVaR. Rather, given the widespread use of the empirical TVaR estimator in (1.1), this paper contributes to deepening the fundamental understanding of its properties, limitations, and explores how it may be improved when necessary.

The remainder of this paper is organized as follows. Before presenting the main results, Section 2 compares the bias of the empirical TVaR estimator in (1.1) with that of an alternative empirical estimator based on the conditional-mean formulation, which clarifies why our study focuses on the empirical TVaR estimator defined as the tail average of the quantile function. Section 3 then presents the proposed explicit bias analysis framework. Section 4 assesses the performance of the proposed framework through an extensive simulation study. In Section 5, the practical relevance of the proposed bias analysis framework is illustrated using the Danish fire loss data. Section 6 concludes the paper. To facilitate readability, all technical proofs are relegated to Appendix A.

2 A commentary on empirical Tail Conditional Expectation

Before presenting the proposed bias analysis framework, we devote this section to commenting on the comparison of empirical TVaR $\widehat{T}_{n,p}$ in (1.1), from the perspective of bias, with another empirical estimator based on the conditional-mean formulation.

In what follows, for notational convenience, let us shorthand the $(p \times 100)\%$-th left percentile of F by

$$\xi_p := F^{-1}(p), \qquad p \in (0,1).$$

The following relationship holds:

$$T_p = \frac{1}{1-p}\int_p^1 F^{-1}(u)\,\mathrm{d}u = \frac{1}{1-p}\Big[\mathbf{E}\Big(X\,\mathbb{1}_{\{X>\xi_p\}}\Big) + \xi_p \times \Big(F(\xi_p) - p\Big)\Big].$$

Moreover, when the quantile function F^{-1} is continuous around p, it holds that $F(\xi_p) = p$, and the above expression simplifies to

$$T_p = \mathbf{E}(X \,|\, X > \xi_p),$$

where the right-hand side corresponds to the Tail Conditional Expectation (TCE) risk functional.

In this case, TVaR can be estimated using the empirical TCE estimator, defined as

$$\widehat{\mathrm{TCE}}_{n,p} = \frac{1}{n-[np]} \sum_{i=[np]+1}^{n} X_{i:n}. \tag{2.1}$$

We argue that the empirical TVaR estimator of interest, namely $\widehat{T}_{n,p}$ in (1.1), exhibits a smaller magnitude of negative bias (i.e., a smaller absolute bias) than $\widehat{\mathrm{TCE}}_{n,p}$ in (2.1), regardless of whether the distribution F is continuous or not. Therefore, from the perspective of bias minimization, $\widehat{T}_{n,p}$ in (1.1) should be preferred over $\widehat{\mathrm{TCE}}_{n,p}$ in (2.1) for computing TVaR. To see the above claim, let us rewrite empirical TVaR in terms of the following more computational friendly form:

$$\begin{aligned}
\widehat{T}_{n,p} = \frac{1}{1-p}\int_p^1 \widehat{F}_n^{-1}(u)\,\mathrm{d}u &= \frac{1}{1-p}\left[\int_p^{\frac{[np]+1}{n}} \widehat{F}_n^{-1}(u)\,\mathrm{d}u + \int_{\frac{[np]+1}{n}}^1 \widehat{F}_n^{-1}(u)\,\mathrm{d}u\right] \\
&= \frac{[np]+1-np}{(1-p)n} X_{[np]+1:n} + \frac{1}{(1-p)n}\sum_{i=[np]+2}^{n} X_{i:n} \\
&= -\frac{np-[np]}{(1-p)n} X_{[np]+1:n} + \frac{1}{(1-p)n}\sum_{i=[np]+1}^{n} X_{i:n} \\
&= -\frac{np-[np]}{(1-p)n} X_{[np]+1:n} + \frac{n-[np]}{(1-p)n}\widehat{\mathrm{TCE}}_{n,p}.
\end{aligned} \tag{2.2}$$

Accordingly, we have

$$\begin{aligned}
\widehat{\mathrm{TCE}}_{n,p} &= \frac{(1-p)n}{n-[np]}\widehat{T}_{n,p} + \frac{np-[np]}{n-[np]} X_{[np]+1:n} \\
&= \widehat{T}_{n,p} - \frac{np-[np]}{n-[np]}\Big(\widehat{T}_{n,p} - X_{[np]+1:n}\Big).
\end{aligned} \tag{2.3}$$

Recall that the bias of $\widehat{T}_{n,p}$ is always negative (Gribkova et al., 2026). Moreover, note that

$$\widehat{T}_{n,p} = \frac{1}{1-p}\int_p^1 \widehat{F}_n^{-1}(u)\,\mathrm{d}u \geq \frac{1}{1-p}\int_p^1 \widehat{F}_n^{-1}(p)\,\mathrm{d}u = \widehat{F}_n^{-1}(p).$$

If np is not an integer, then $\widehat{F}_n^{-1}(p) = X_{[np]+1:n}$, hence $\widehat{T}_{n,p} \geq X_{[np]+1:n}$. Thereby, the subtraction term in (2.3) is always non-negative, which introduces an additional negative bias to the estimator $\widehat{\mathrm{TCE}}_{n,p}$ compared to $\widehat{T}_{n,p}$, unless np is an integer, in which case this additional bias is zero. In other words, the empirical estimator of interest $\widehat{T}_{n,p}$ always outperforms $\widehat{\mathrm{TCE}}_{n,p}$ in the sense of having a smaller, or at the very least an equal, negative bias. This observation justifies our focus on $\widehat{T}_{n,p}$ as the estimator of interest in this paper.

3 Main results

We aim to study the bias of empirical TVaR $\widehat{T}_{n,p}$ in (1.1) under minimal assumptions, so that the results can be broadly applicable across a wide range of settings. Nevertheless, certain finiteness conditions are still required to guarantee that the problem under study is well-posed. The first condition ensures that the risk measure of interest, T_p, is well-defined: For a given $p \in (0,1)$

$$\int_p^1 F^{-1}(u)\,\mathrm{d}u < \infty. \tag{C$_1$}$$

In addition, we need the mean of the empirical TVaR estimator, and thus its bias, to exist. This is ensured by the following condition: For a given $p \in (0,1)$,

$$\mathbf{E}\Big(|X|^{\varepsilon}\,\mathbb{1}_{\{X\le\xi_p\}}\Big) < \infty, \text{ for some } \varepsilon > 0. \tag{C$_2$}$$

It is worth noting that condition (C$_2$) is very mild. In particular, it may fail only in the presence of a sufficiently heavy left tail. In the context of risk management, loss models are typically assumed to have positive support or at the very least lower bounded. In these cases, condition (C$_2$) can be automatically satisfied.

To understand why conditions (C$_1$) and (C$_2$) together imply the finiteness of the bias, we shall examine the difference between $\widehat{T}_{n,p}$ and T_p. To this end, we need to introduce some additional notations. For $i = 1,\dots,n$, let W_i denote the winsorized counterpart of X_i with respect to the threshold ξ_p:

$$W_i = \max(X_i,\,\xi_p) = \begin{cases} \xi_p, & X_i \le \xi_p; \\ X_i, & X_i > \xi_p. \end{cases}$$

Note that $W_1,\dots,W_n$ are iid random variables. We denote their common mean by

$$\mu_W = \mathbf{E}(W_1).$$

Moreover, let $N_p := \mathrm{card}\{i \,:\, X_i \le \xi_p\}$ denote the number of iid samples $X_1,\dots,X_n$ that fall below or at the threshold ξ_p.

The following lemma spells out $(\widehat{T}_{n,p} - T_p)$ explicitly in terms of a sum of iid centered winsorized random variables, accompanied by a non-positive remainder term. At the outset, let us define

$$a \wedge b := \min(a,b); \quad a \vee b := \max(a,b); \quad \mathrm{sign}(a) := \begin{cases} \frac{a}{|a|}, & a \neq 0, \\ 0, & a = 0. \end{cases} \tag{3.1}$$

Lemma 3.1. *Suppose that condition* (C_1) *holds, then we have*

$$\widehat{T}_{n,p} - T_p = \frac{1}{n(1-p)} \sum_{i=1}^{n} \big(W_i - \mu_W\big) - R_n, \tag{3.2}$$

where

$$R_n = \frac{np - [np]}{n(1-p)}(X_{[np]+1:n} - \xi_p) - \frac{\text{sign}(N_p - [np])}{n(1-p)} \sum_{i=([np]\wedge N_p)+1}^{[np]\vee N_p} (X_{i:n} - \xi_p). \tag{3.3}$$

Furthermore,

$$R_n \geq 0. \tag{3.4}$$

From here on, let us denote the bias of empirical TVaR $\widehat{T}_{n,p}$ by

$$B_n = \mathbf{E}\big(\widehat{T}_{n,p}\big) - T_p.$$

From the representation (3.2) in Lemma 3.1, it follows that

$$B_n = -\mathbf{E}(R_n). \tag{3.5}$$

Therefore, the finiteness of the bias B_n is equivalent to the finiteness of $\mathbf{E}(R_n)$, or equivalently, to the finiteness of the expectations of the order statistics involved in the expression of R_n in (3.3). By Proposition 2 of Stigler (1974), conditions (C_1) and (C_2) together ensure the existence of moments of central order statistics of any order, and in particular, the finiteness of the first moment, which is sufficient for our purposes in the current context.

Remark 3.1. *By the relationships noted in* (3.4) *in Lemma 3.1 as well as in* (3.5)*, it follows that the bias B_n is non-positive, which is in agreement with the result established in Gribkova et al. (2026).*

We are now in a position to present one of the main results of this paper, which can offer analytical insights into the structural properties of the bias B_n through the leading term of its asymptotic expansion.

Theorem 3.1. *For a fixed $p \in (0,1)$, suppose that conditions* (C_1) *and* (C_2) *hold. Further, assume that the probability density function (PDF) of X, denoted by f, exists and is continuous in a neighborhood of the point $\xi_p = F^{-1}(p)$, and that $f(\xi_p) > 0$. Then, the following relationship holds*

for the bias of empirical TVaR $\widehat{T}_{n,p}$*:*

$$B_n = \ell_n + o\big(n^{-1}\big), \qquad \ell_n = -\frac{p}{2nf(\xi_p)}. \tag{3.6}$$

Remark 3.2. *In equation (30) of Gribkova et al. (2026), the same authors as in the present paper conjectured the asymptotic expansion in (3.6), motivated by heuristic reasoning derived from Edgeworth-type expansions for the distribution of normalized trimmed means studied in Gribkova and Helmers (2006, 2007). That conjecture was stated without a complete specification of the required conditions nor a rigorous proof. Theorem 3.1 of the present paper formally confirms the correctness of this initial conjecture by establishing the expansion under a complete and explicit set of conditions.*

Thanks to the explicit expression established in Theorem 3.1, several analytical insights emerge immediately, which are summarized in the succeeding remarks.

Remark 3.3. *The asymptotic expansion in (3.6) shows that the negative bias decreases at the rate of* n^{-1} *as the sample size increases. It also confirms that the empirical TVaR estimator is asymptotically unbiased.*

Remark 3.4. *For a fixed sample size* n*, the leading term* ℓ_n *in (3.6) is proportional to* $p/f(\xi_p)$*. Assuming that* f *decreases monotonically along the far right tail, which is typically observed in empirical (insurance) data, then the function* $p \mapsto p/f(\xi_p)$ *is increasing for sufficiently large values of* p*. This implies that the negative bias becomes more pronounced as the probability level* p *increases. Intuitively, increasing* p *reduces the size of the tail region and leads to fewer observations above the VaR threshold* ξ_p*, so the estimation of the tail mean becomes more difficult and its finite-sample bias becomes larger.*

Remark 3.5. *When* n *and* p *are fixed, the magnitude of the leading term* ℓ_n *is inversely related to* $f(\xi_p)$*. In risk management practice, TVaR is commonly evaluated at high probability levels. When a data distribution becomes more heavy-tailed, its associated probability density tends to spread farther into the extreme region. As a result, the density around a fixed high quantile becomes smaller, or equivalently* $f(\xi_p)$ *tends to decrease. While this opposite relationship between tail heaviness and* $f(\xi_p)$ *may not hold universally for all distribution families, it is indeed the case for models commonly used in risk modeling practice. For instance, the Pareto distribution considered in our succeeding simulation study in Section 4 statisfies this relationship. Consequently, heavier tails lead to a larger value of the term* $1/f(\xi_p)$*, and thus a more pronounced negative bias of the empirical TVaR* $\widehat{T}_{n,p}$*.*

In addition, we note that the relationship between the leading-term bias of empirical TVaR and $f(\xi_p)$ stands in stark contrast with kernel-based TVaR estimators, whose leading bias term is

increasing with $f(\xi_p)$ rather than decreasing (e.g., Chen, 2008; Hand et al., 2010). The reasoning behind this observation may be that a smaller density at ξ_p implies fewer observations and larger spacing between upper order statistics in the neighborhood surrounding ξ_p. As a result, an empirical estimator is less capable of accurately locating the true quantile threshold, which leads to a more pronounced downward bias. However, for kernel methods, a kernel estimator smooths the loss distribution in a neighborhood of the quantile level and therefore assigns positive weight to observations just below ξ_p. These sub-threshold observations are smaller than the true tail losses, and their inclusion pulls the TVaR estimate downward. When $f(\xi_p)$ is larger, a greater amount of probability mass accumulates near ξ_p, amplifying this smoothing-induced downward pull and can increase the magnitude of the negative bias.

We note that in practical applications, the quantile ξ_p and the PDF f are unknown. To evaluate the leading term ℓ_n in (3.6), one may estimate ξ_p using the empirical quantile and estimate $f(\xi_p)$ using a kernel density estimator or another nonparametric method, such as local likelihood density estimation, local polynomial density estimation, or an orthogonal series estimator, and so forth.

Thus far, Theorem 3.1 identifies the leading term in the asymptotic expansion of the bias B_n of the empirical TVaR. Although informative, the leading term ℓ_n need not match the finite-sample bias exactly. To strengthen our analysis of the bias B_n, we next complement the leading-term approximation in Theorem 3.1 with an explicit upper bound for $|B_n|$, which yields a more conservative evaluation of the bias.

Theorem 3.2. *For a fixed $p \in (0,1)$, suppose that conditions (C_1) and (C_2) are satisfied. Assume also that the inversion F^{-1} is locally Hölder continuous of order $\gamma \in (0,1]$ at the point p, that is, there exists a neighborhood $\mathbb{U}_p$ of the point p and a constant $c_\gamma > 0$ such that*

$$|F^{-1}(u) - F^{-1}(p)| \le c_\gamma \, |u-p|^\gamma, \quad u \in \mathbb{U}_p. \tag{3.7}$$

Then, for any $\delta > 0$, there exists n_0 such that the bias satisfies

$$-B_n \;\le\; \frac{C\, p^{\frac{1+\gamma}{2}}}{n^{\frac{1+\gamma}{2}} (1-p)^{\frac{1-\gamma}{2}}}, \qquad n \ge n_0, \tag{3.8}$$

where $C = c_\gamma \, (1+\delta)$.

Remark 3.6. *The local Hölder continuity constant c_γ in (3.7) quantifies the local smoothness of the quantile function F^{-1} near p. A larger value of c_γ indicates greater steepness or variability of F^{-1} in a neighborhood of p. Hence, a larger value of c_γ implies that even small fluctuations in the sample order statistics may translate into notable errors in the empirical TVaR estimate. Consequently, a larger upper bound is needed to accommodate the increased sensitivity of the negative bias, which*

explains why the bias bound in (3.8) is monotonically increasing in c_γ when the other parameters are held constant.

Moreover, since the derivative of the upper bound with respect to c_γ is proportional to $n^{-(1+\gamma)/2}$, its sensitivity to c_γ decreases as the sample size n grows. This implies that for small samples, an accurate assessment of c_γ has a much stronger impact on correctly implementing the upper bound than it does for large samples. We will later discuss how to estimate c_γ in practice.

Remark 3.7. *In the choice of $C = c_\gamma(1+\delta)$, the slack parameter $\delta > 0$ is not a structural parameter of the data distribution but an arbitrarily small inflation factor introduced to ensure that the upper bound (3.8) holds uniformly for all sufficiently large n. A smaller δ yields a tighter upper bound in (3.8), but the sample size n may then be larger before the inequality is guaranteed.*

As discussed in Remark 3.6, implementing the negative-bias bound in Theorem 3.2 requires an appropriate evaluation of the local Hölder continuity constant c_γ. First, let us consider an ideal scenario where the underlying data distribution F is known. For almost all commonly used continuous parametric distributions with a positive and continuous density at ξ_p, the Hölder exponent of F^{-1} near p satisfies $\gamma = 1$. To calculate c_1, namely the local Hölder constant c_γ when $\gamma = 1$, note that for a small interval $\mathbb{X}_{\xi_p,h} = [\xi_p - h, \xi_p + h]$, $h > 0$, the mean value theorem implies that for all u sufficiently close to p,

$$|F^{-1}(u) - F^{-1}(p)| \leq \Big(\sup_{x \in \mathbb{X}_{\xi_p,h}} \frac{1}{f(x)} \Big) |u - p|.$$

Thereby, a local Hölder (or Lipschitz in this case of $\gamma = 1$) constant can be computed via

$$c_1 = \Big(\inf_{x \in \mathbb{X}_{\xi_p,h}} f(x) \Big)^{-1}. \tag{3.9}$$

In a more realistic scenario where the data distribution is unknown and only data are available, a natural estimator of c_1 is obtained by replacing f with a local density estimator near $\widehat{\xi}_p$, the empirical estimate of the $(p \times 100)$%-th quantile. Let $\widehat{f}(x)$ denote such an estimator (e.g., kernel density estimator) and let $\mathbb{X}_{\widehat{\xi}_p,h} = [\widehat{\xi}_p - h, \widehat{\xi}_p + h]$ be a small data-driven neighborhood. Then an empirical Lipschitz constant can be

$$\widehat{c}_1 = \Big(\inf_{x \in \mathbb{X}_{\widehat{\xi}_p,h}} \widehat{f}(x) \Big)^{-1}. \tag{3.10}$$

4 Simulation analysis

The simulation analysis in this section serves two complementary purposes. The first is to assess how well the explicit leading bias identified in Theorem 3.1 reflects the behavior of the finite-sample

bias across varying sample sizes, probability levels, and tail behaviors. The second is to evaluate the effectiveness of the leading-term approximation from Theorem 3.1 and the finite-sample bound from Theorem 3.2 in approximating the true bias of empirical TVaR $\widehat{T}_{n,p}$.

Throughout this simulation study, we assume that the risk random variable X follows a Pareto distribution with CDF:

$$F(x) = 1 - x^{-\alpha}, \qquad x > 1, \tag{4.1}$$

where $\alpha > 0$ denotes the shape parameter. The smaller the value of α, the heavier the tail of the distribution of X. The mean of X is finite when $\alpha > 1$, and so is the TVaR T_p for any $p \in (0, 1)$. This is the case of interest throughout this section.

4.1 Leading-term approximation

Let us begin by examining the leading-term approximation of the bias of empirical TVaR. Under the model given in (4.1), the mean of the i-th order statistic $X_{i:n}$ based on an iid sample $X_1, \ldots, X_n$ from X can be computed explicitly as, for $i < n + 1 - \alpha^{-1}$,

$$\mathbf{E}\big(X_{i:n}\big) = \frac{\Gamma(n+1)\,\Gamma\big(n-i+1-\alpha^{-1}\big)}{\Gamma(n-i+1)\,\Gamma\big(n+1-\alpha^{-1}\big)},$$

where $\Gamma(\cdot)$ represents the gamma function. Consequently, using the representation in the third line of (2.2), the exact bias of empirical TVaR can be derived explicitly as

$$\begin{aligned} B_n &= \mathbf{E}\Big[-\frac{np - [np]}{(1-p)n}\, X_{[np]+1:n} + \frac{1}{(1-p)n} \sum_{i=[np]+1}^{n} X_{i:n} \Big] - T_p \\ &= \frac{\Gamma(n+1)}{(1-p)\,n\,\Gamma\big(n+1-\alpha^{-1}\big)} \left[-(np - [np])\, \frac{\Gamma\big([np]+1-\alpha^{-1}\big)}{\Gamma([np]+1)} + \sum_{i=[np]+1}^{n} \frac{\Gamma\big(i-\alpha^{-1}\big)}{\Gamma(i)} \right] - T_p, \end{aligned}$$

where the true TVaR of (4.1) is given by

$$T_p = \frac{\alpha}{\alpha - 1}(1-p)^{-1/\alpha}, \qquad p \in (0, 1).$$

We consider the following baseline parameter setting: $n = 500$, $p = 95\%$, and $\alpha = 3$. We then perturb each parameter individually to examine how the leading-term bias ℓ_n varies with n, p, and α, and to verify that the resulting behavior of the bias of empirical TVaR aligns with the analytical insights established in Section 3. Two versions of the leading-term bias ℓ_n are evaluated. The first assumes that the data-generating model (4.1) is known, in which case ℓ_n can be computed exactly.

The second scenario considers a more realistic situation in which only the simulated samples are available. In this case, the kernel density estimator with a Gaussian kernel is used to estimate the unknown PDF appearing in ℓ_n. Because of the sampling variability inherent in the simulated data, each configuration of the perturbed parameters is replicated $m = 100$ times to assess variability in the resulting empirical leading-term bias estimates.

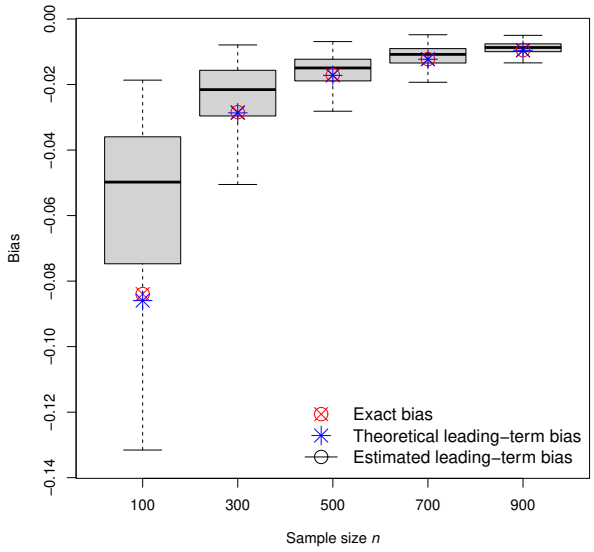

Figure 4.1: Comparison of the exact TVaR bias (red circle-cross markers), the theoretical leading-term approximation (blue star markers), and the estimated leading-term approximation (boxplots), with fixed parameters $p = 95\%$, $\alpha = 5$, and varying sample sizes $n \in \{100, 300, 500, 700, 900\}$.

Figure 4.1 presents a comparison of the exact TVaR bias with the leading-term bias ℓ_n derived in Theorem 3.1 with varying sample sizes. In the figure, the boxplots are constructed from $m = 100$ simulated datasets of size n. As shown, when the data-generating distribution is assumed to be known, the exact bias and its leading-term approximation agree very closely. Among the sample sizes considered, a visible yet minor discrepancy is observed only when $n = 100$. This indicates that, at least for the Pareto model, the residual term in the asymptotic expansion (3.6) is negligible even at small sample sizes. When the distribution is assumed to be unknown and $f(\xi_p)$ must be estimated from data, the performance of the leading-term approximation deteriorates due to the limited accuracy of the kernel-based estimation of $f(\xi_p)$. For smaller samples (e.g., $n = 100$ or 300), the exact bias B_n can deviate substantially from the estimated leading-term bias, occasionally falling near or outside the whiskers of the boxplots. However, as the sample size increases, the estimation of $f(\xi_p)$ becomes more accurate and stable. Thereby, for larger sample sizes (e.g., $n = 900$), the leading-term approximation based on the kernel density estimate tracks the exact bias of empirical TVaR very closely.

Next, we examine the impact of the probability level p on both the exact bias B_n and the leading-term bias ℓ_n. As p increases, fewer observations fall beyond the quantile threshold ξ_p, thereby reducing the amount of data available for estimating TVaR. Consequently, the negative bias of empirical TVaR estimator becomes more pronounced. This pattern observed in Figure 4.2 is consistent with the analytical insights discussed in Remark 3.4. When the data-generating distribution is treated to be unknown, increasing p also deteriorates the accuracy of the kernel density estimator for $f(\xi_p)$, leading to a larger discrepancy between the estimated leading-term bias and the exact bias. Nevertheless, the theoretical leading-term bias tracks the exact bias very closely across all probability levels considered. This suggests that the discrepancy observed in the estimated leading-term bias is mainly attributable to the estimation error in the kernel density estimator, rather than caused by the asymptotic expansion in (3.6).

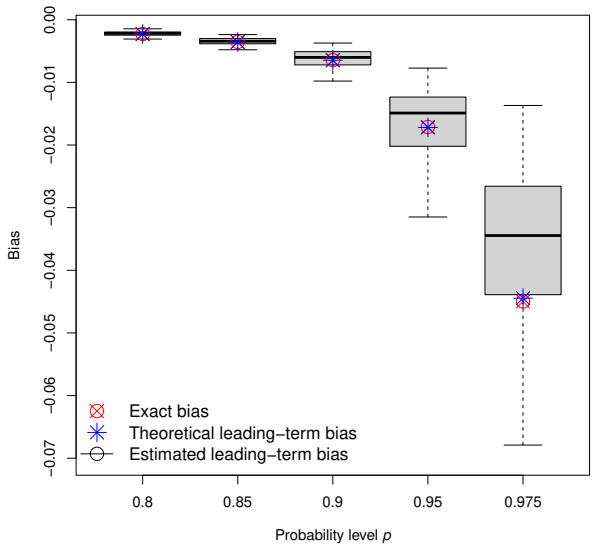

Figure 4.2: Comparison of the exact TVaR bias (red circle-cross markers), the theoretical leading-term approximation (blue star markers), and the estimated leading-term approximation (boxplots), with fixed parameters $n = 500$, $\alpha = 5$, and varying probability levels $p \in \{80, 85, 90, 95, 97.5\}\%$.

Further, we study the impact of the tail heaviness of the data distribution in (4.1), controlled by the shape parameter α, on the bias of empirical TVaR B_n and its leading-term approximation ℓ_n. The results are summarized in Figure 4.3. As shown, smaller values of α, corresponding to heavier tails, lead to a larger magnitude of the negative bias for both the exact TVaR and its leading-term approximation. This observation is consistent with the analytical insights discussed in Remark 3.5. In particular, under data model in (4.1), elementary algebraic calculations yield

$$f(\xi_p) = \alpha\,(1-p)^{(\alpha+1)/\alpha},$$

which is increasing in $\alpha > 0$ for a fixed $p \in (0, 1)$. Thus, smaller values of α, thus heavier tails, lead to smaller values of $f(\xi_p)$, which in turn amplify the negative leading-term bias. This behavior of ℓ_n matches that of the exact bias B_n across all values of α considered. We also observe from Figure 4.3 that heavier tails impair the accuracy of estimating $f(\xi_p)$ via kernel density estimation, thus resulting in a more noticeable discrepancy between the estimated leading-term bias and the exact bias when α is small.

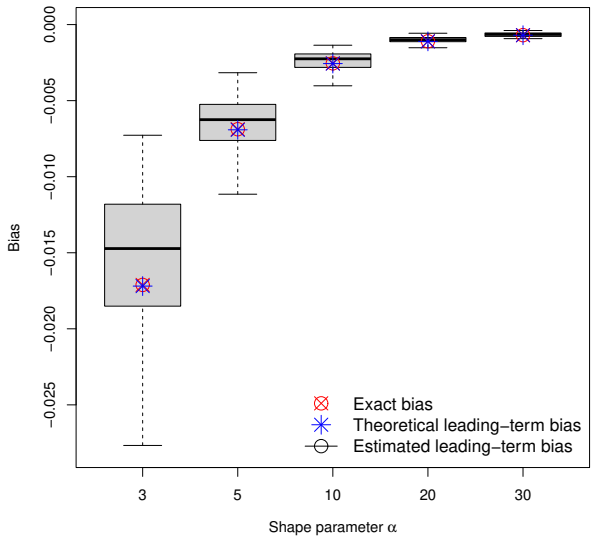

Figure 4.3: Comparison of the exact TVaR bias (red circle-cross markers), the theoretical leading-term approximation (blue star markers), and the estimated leading-term approximation (box-plots), with fixed parameters $n = 500$, $p = 95\%$, and varying values of the shape parameter $\alpha \in \{3, 5, 10, 20, 30\}$.

4.2 Negative-bias upper bound

We now turn to examining how well the negative-bias upper bound in Theorem 3.2 captures the true bias of empirical TVaR. To implement the bound, it is necessary to specify two tuning parameters. The first one is the slack parameter δ, which controls the trade-off between the tightness of the bound and the sample size required for its validity in finite samples (see Remark 3.7 for a detailed discussion). The second one is the locality parameter h, which determines the local Lipschitz constant through (3.9). Although introduced from different perspectives, note that these two tuning parameters in fact play similar roles in practice. Specifically, a larger value of the locality parameter h enlarges the interval over which the Lipschitz constant is evaluated, resulting in a larger estimate of c_1 and, consequently, a more conservative upper bound. This effect parallels the role of the slack parameter δ discussed in Remark 3.7. For this reason, throughout the analysis in

this subsection, we fix the locality parameter $h = 0.05$ and focus on examining the sensitivity of the upper bound with respect to different choices of the slack parameter δ.

Suppose that the data-generating model (4.1) is known. For a given locality parameter h, the local Lipschitz constant defined in (3.9) admits the following closed-form expression:

$$c_1 = \left(\inf_{x \in \mathbb{X}_{\xi_p, h}} f(x) \right)^{-1} = \frac{(\xi_p + h)^{\alpha+1}}{\alpha} = \frac{\big((1-p)^{-1/\alpha} + h\big)^{\alpha+1}}{\alpha},$$

where the second equality follows from the fact that the PDF of the Pareto distribution in (4.1) is strictly decreasing. We vary the sample size n to examine whether the negative-bias upper bound remains valid for small sample sizes and, if so, how tight the bound is across different values of n. The comparison results are summarized in Figure 4.4.

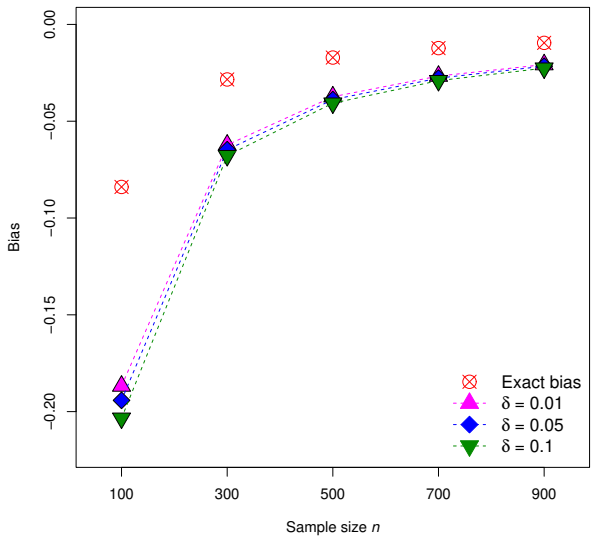

Figure 4.4: Comparison of the exact TVaR bias with the theoretical negative-bias upper bound, with fixed parameters $p = 95\%$, $\alpha = 3$, and $h = 0.05$, across varying sample sizes $n \in \{100, 300, 500, 700, 900\}$ and different choices of the slack parameter $\delta \in \{0.01, 0.05, 0.1\}$.

As can be observed in Figure 4.4, the derived upper bound always covers the true negative bias across all values of the slack parameter δ and sample sizes n considered. For a fixed δ, the bound captures the decreasing trend of the exact negative bias as the sample size n increases. For a fixed n, larger values of the slack parameter δ lead to a more conservative bound and thus a larger discrepancy from the exact bias, which is consistent with the discussion in Remark 3.7. Interestingly, the gap between the exact bias and the bound is substantially larger for smaller sample sizes and lessens noticeably as n increases. This behavior can be attributed to the presence

of positive residual terms in the derived upper bound under the Pareto model (4.1), which reduce as the sample size grows. As a result, the bound becomes increasingly tight in larger samples.

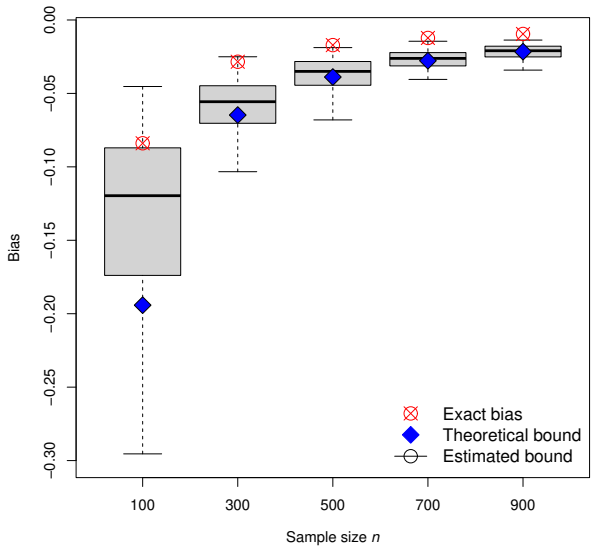

Figure 4.5: Comparison of the exact TVaR bias (red circle-cross markers), the theoretical negative-bias upper bound (blue diamond markers), and the estimated negative-bias upper bound (boxplots), with fixed parameters $p = 95\%$, $\alpha = 3$, $\delta = 0.05$, and $h = 0.05$, across varying sample sizes $n \in \{100, 300, 500, 700, 900\}$.

In another scenario where the data-generating model (4.1) is treated as unknown, the kernel-based estimator (3.10) is employed to estimate the Lipschitz constant required for computing the negative-bias upper bound. Similar to numerical experiments conducted in the previous subsection, each configuration is repeated $m = 100$ times to account for the sampling variability inherent in the simulated data. The resulting performance of the bound relative to the exact negative bias is summarized in Figure 4.5. Several observations can be made. First, the boxplots corresponding to the estimated upper bound always cover the theoretical upper bound. When the sample size is small, the kernel density estimator more frequently misestimates the PDF in the tail region, leading to a noticeable discrepancy between the center of the boxplots and the theoretical upper bound. As a result, the estimated upper bound may occasionally fail to capture the exact bias. As the sample size increases, the kernel density estimator becomes more accurate and stable. As a result, the boxplots become narrower, and their centers converges toward the theoretical upper bound. In this example, we observe that once the sample size exceeds 500, the estimated upper bound consistently covers the true bias, which indicates improved finite-sample reliability of the negative-bias upper bound as the sample size grows.

5 Real-data illustration

We now consider a real-data illustration of the theoretical results obtained in this current paper. Our real data example is based on the Danish fire loss dataset (McNeil, 1997). The data consist of 2 167 individual fire insurance losses recorded between 1980 and 1990, which have been inflation-adjusted to 1985 price levels and are reported in millions of Danish kroner. Prior studies of the dataset have documented that its distribution exhibits a heavy right tail, with a tail index of approximately 1.4 (e.g., McNeil, 1997; Resnick, 1997). As demonstrated by the simulation experiments in Section 4, heavy-tailed patterns can induce an increased bias in $\widehat{T}_{n,p}$, particularly at high probability levels. This motivates us to examine the magnitude of the bias of the estimator.

Although the tail of the distribution of the Danish fire data is relatively heavy, as indicated by the estimated tail index of approximately 1.4, the distribution still has a finite first moment. Consequently, Condition (C_1) is satisfied. Moreover, since the loss data are strictly positive, Condition (C_2) also holds. These conditions together justify the validity of using the leading-term bias approximation in Theorem 3.1, as well as the corresponding upper bound in Theorem 3.2, to analyze the bias of empirical TVaR.

In the absence of knowledge of the true underlying loss distribution, unlike in the simulation study, the exact bias is not available for direct comparison. To address this, we compare the proposed leading-term approximation and upper bound against estimates obtained using a classical numerical approach, such as the bootstrap method. Specifically, we employ the nonparametric Efron bootstrap with $l = 1\,000$ resample sets, each of size $n = 2\,167$, which is the same as the size of the original dataset. To account for the variability inherent in the bootstrap procedure, we repeat the bootstrap bias estimation $m = 100$ times.

Figure 5.1 depicts a comparison of bias estimates for the empirical TVaR (1.1) obtained using different methods, including the leading-term approximation and the negative-bias upper bound derived in Theorems 3.1 and 3.2, respectively, as well as the bootstrap method. We observe that, due to resampling variability, the bootstrap procedure may occasionally produce positive bias estimates. Such estimates contradict the theoretical result that the bias of empirical TVaR B_n is always negative (Gribkova et al., 2026); also see discussion in Remark 3.1. For the two proposed approximations, owing to their analytical nature, they do not suffer from this drawback induced by resampling variation, and therefore provide bias assessments that are always consistent with the negativity property.

At lower probability levels, the leading-term approximation matches closely with the means of the bootstrap estimates. At higher probability levels, the leading-term approximation exhibits more pronounced discrepancies from the mean of the bootstrap estimates. At the same time, the bootstrap estimates themselves display increased variability, which is caused by the greater sampling

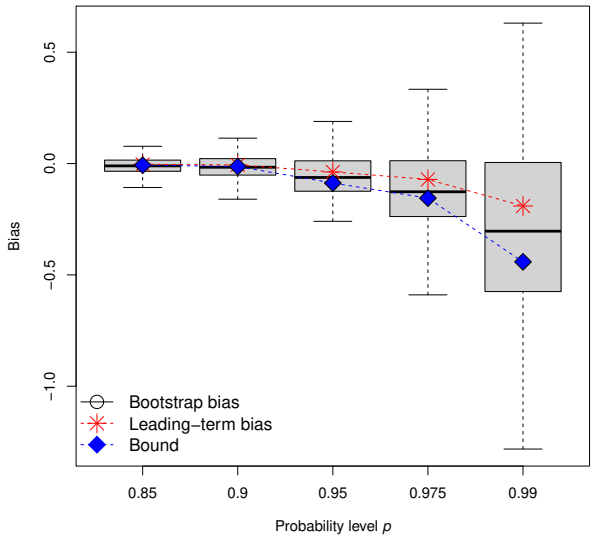

Figure 5.1: A comparison of bias estimates for the empirical TVaR based on the Danish fire loss dataset, obtained via the leading-term approximation (red star markers), the negative-bias upper bound (blue diamond markers), and the bootstrap method (boxplots).

uncertainty associated with observations in the extreme tail region that drive the estimation of TVaR. Moreover, across all probability levels considered, the proposed negative-bias upper bound consistently covers both the leading-term approximation and the means of the bootstrap bias estimates, while remaining reasonably close to the lower edge of the bootstrap boxplot whiskers. This behavior indicates a satisfactory performance of the negative-bias upper bound in terms of achieving a practical balance between tightness and conservativeness.

Compared with the bootstrap approach which is a numerical method, the proposed analytical methods including the leading-term bias approximation and the negative-bias upper bound, are more convenient to implement, as they do not rely on repeated resampling. In addition to the analytical insights already discussed in Section 3, these methods can provide practitioners with an efficient and transparent framework for assessing the significance of bias. In particular, they support informed decision-making regarding whether bias-reduction strategies are warranted or whether additional data collection is desirable.

For instance, in this data example, $\widehat{T}_{n,p}$ yields a TVaR estimate of approximately 24 at the 95% probability level, while the corresponding bias of the empirical TVaR is estimated to be around -0.1, representing less than 0.05% of the TVaR value. This is perhaps an affirmative finding, as such a negligible level of bias indicates that empirical TVaR can be used with confidence even without applying any bias correction. For more sophisticated users seeking to further reduce the bias, the leading-term approximation ℓ_n can be readily applied for bias reduction. Moreover, when

TVaR is used for economic capital calculation, a conservative analyst may adopt the negative-bias upper bound to incorporate an additional buffer that accounts for the bias inherent in the TVaR estimation process.

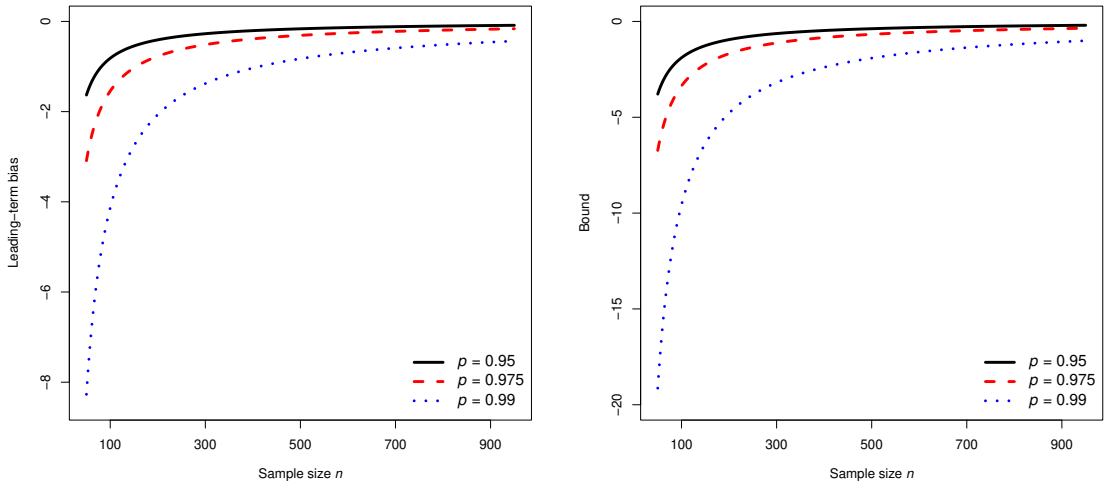

Figure 5.2: Leading-term bias (left panel) and negative-bias upper bound (right panel) of the empirical TVaR as functions of the sample size n, evaluated at probability levels $p \in \{95, 97.5, 99\}\%$.

For analysts interested in understanding how different choices of sample size and probability level affect the accuracy of TVaR estimation, the derived leading-term approximation and the negative-bias upper bound provide useful tools for assessing these relationships. In Figure 5.2, we first use the original Danish fire loss data to estimate $f(\xi_p)$ and the Lipschitz constant c_1, which are required for implementing the leading-term approximation ℓ_n in (3.6) and the negative-bias upper bound in (3.8). We then plot the resulting bias versus sample-size curves across different probability levels. These curves can offer insights into whether the marginal benefit of acquiring additional data is meaningful, or whether adjusting the probability level used in the calculation of TVaR is more appropriate given the current sample size.

6 Conclusions

TVaR has been widely adopted by practitioners and regulators as a conservative risk measure for capturing tail risk. In practical applications, TVaR is typically estimated using the empirical method due to its nonparametric nature and ease of implementation. It is known that the empirical TVaR estimator exhibits a negative finite-sample bias (Gribkova et al., 2026). However, despite its practical importance, the magnitude of this bias and its dependence on the distributional characteristics of the data and the sample size remain underexplored.

This paper addresses this gap by developing two analytical approaches for evaluating the bias of the empirical TVaR. The first approach is based on the leading term in the asymptotic expansion of the bias, while the second approach establishes an explicit upper bound on the negative bias. These results yield explicit analytical insights into how data distributional properties, sample sizes, and probability levels jointly determine the magnitude of the bias. Simulation studies demonstrate that the proposed methods perform satisfactorily in approximating the exact bias, even in moderately small sample settings. We further apply the proposed approaches to study the bias involved in TVaR estimation using the Danish fire loss data, and demonstrate how the resulting insights can inform prospective study design and risk assessment procedures, especially when balancing data availability constraints against feasible and desirable choices of probability levels in TVaR applications.

Acknowledgment

Mengqi Wang's research has been supported by the NSERC Alliance-MITACS Accelerate grant (ALLRP 580632-22) entitled "New Order of Risk Management: Theory and Applications in the Era of Systemic Risk" from the Natural Sciences and Engineering Research Council (NSERC) of Canada and the national research organization Mathematics of Information Technology and Complex Systems (MITACS) of Canada.

References

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

BCBS (2016). Minimum capital requirements for market risk. Technical report, Basel Committee on Banking Supervision, Bank for International Settlements, Basel, Switzerland. Available at https://www.bis.org/bcbs/publ/d352.htm.

BCBS (2019). Minimum capital requirements for market risk. Technical report, Basel Committee on Banking Supervision, Bank for International Settlements, Basel, Switzerland. Available at https://www.bis.org/bcbs/publ/d457.htm.

Bolance, C., Guillen, M., Pelican, E., and Vernic, R. (2008). Skewed bivariate models and nonparametric estimation for the CTE risk measure. *Insurance: Mathematics and Economics*, 43(3):386–393.

Brazauskas, V., Jones, B., Puri, M., and Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138:3590–3604.

Brazauskas, V. and Kleefeld, A. (2009). Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insurance: Mathematics and Economics*, 45(3):424–435.

Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, 6(1):87–107.

Chen, Y., Song, Q., and Su, J. (2023). On a class of multivariate mixtures of gamma distributions: Actuarial applications and estimation via stochastic gradient methods. *Variance*, 16(1):1–17.

EBA (2020). Final draft regulatory technical standards on the calculation of the stress scenario risk measure under article 325bk(3) of regulation (EU) no 575/2013 (CRR2). Technical report, European Banking Authority, Paris, France.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC, Boca Raton.

Goegebeur, Y., Guillou, A., Pedersen, T., and Qin, J. (2022). Extreme-value based estimation of the conditional tail moment with application to reinsurance rating. *Insurance: Mathematics and Economics*, 107:102–122.

Gribkova, N. (1995). Bounds for absolute moments of order statistics. In Skorokhod, A. and Borovskikh, Y., editors, *Exploring Stochastic Laws*, pages 129–134. de Gruyter, Berlin.

Gribkova, N. and Helmers, R. (2006). The empirical Edgeworth expansion for a studentized trimmed mean. *Mathematical Methods of Statistics*, 15:61–87.

Gribkova, N. and Helmers, R. (2007). On the Edgeworth expansion and the M out of N bootstrap accuracy for a studentized trimmed mean. *Mathematical Methods of Statistics*, 16:142–176.

Gribkova, N., Su, J., and Zitikis, R. (2022a). Empirical tail conditional allocation and its consistency under minimal assumptions. *Annals of the Institute of Statistical Mathematics*, 74(4):713–735.

Gribkova, N., Su, J., and Zitikis, R. (2022b). Inference for the tail conditional allocation: Large sample properties, insurance risk assessment, and compound sums of concomitants. *Insurance: Mathematics and Economics*, 107:199–222.

Gribkova, N. V., Wang, M., and Zitikis, R. (2026). Fundamentals of non-parametric statistical inference for integrated quantiles. *Risk Sciences*, 2:100026.

Hand, K. Y., Ally, A. K., Yang, S., and David, J. (2010). Kernel quantile based estimation of expected shortfall. *Journal of Risk*, 12(4):15–32.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Hua, L. and Joe, H. (2011). Second order regular variation and conditional tail expectation of multiple risks. *Insurance: Mathematics and Economics*, 49(3):537–546.

Kusuoka, S. (2001). On law invariant coherent risk measures. In *Advances in Mathematical Economics*, pages 83–95. Springer, Tokyo.

Lee, S. C. and Lin, X. S. (2010). Modeling and evaluating insurance losses via mixtures of erlang distributions. *North American Actuarial Journal*, 14(1):107–130.

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin: The Journal of the IAA*, 27(1):117–137.

Nadarajah, S., Zhang, B., and Chan, S. (2014). Estimation methods for expected shortfall. *Quantitative Finance*, 14(2):271–291.

Necir, A., Rassoul, A., and Zitikis, R. (2010). Estimating the conditional tail expectation in the case of heavy-tailed losses. *Journal of Probability and Statistics*, 2010(1):596839.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4):353–360.

Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin: The Journal of the IAA*, 27(1):139–151.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, pages 571–587.

Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Stigler, S. (1974). Linear functions of order statistics with smooth weight functions. *Annals of Statistics*, 2:676–693.

Tukey, J. (1958). Bias and confidence in not quite large samples. *The Annals of Mathematical Statistics*, 29:614.

Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the expected shortfall. *Management Science*, 67(3):1413–1429.

Wei, Y. and Zitikis, R. (2023). Assessing the difference between integrated quantiles and integrated cumulative distribution functions. *Insurance: Mathematics and Economics*, 111:163–172.

Yu, K., Ally, A. K., Yang, S., and Hand, D. J. (2010). Kernel quantile-based estimation of expected shortfall. *Journal of Risk*, 12(4):15–32.

Appendix A Technical proofs

Let us introduce some additional notation to facilitate the succeeding technical developments. Recall that the winsorized version of X_i with respect to the threshold ξ_p is denoted by $W_i = \max(X_i, \xi_p)$, $i = 1, \dots, n$. The random variables $W_1, \dots, W_n$ are iid. The common CDF of $W_1, \dots, W_n$ can be computed via

$$F_W(x) = \begin{cases} 0, & x < \xi_p, \\ F(x), & x \geq \xi_p, \end{cases}$$

and the corresponding quantile function is given by

$$F_W^{-1}(u) = \max(\xi_p, F^{-1}(u)) = \begin{cases} \xi_p, & u \leq p, \\ F^{-1}(u), & u > p. \end{cases}$$

The common mean of $W_1, \dots, W_n$, shorthanded by $\mu_W = \mathbf{E}(W_1)$, can be expressed as

$$\mu_W = \int_0^1 F_W^{-1}(u)\,\mathrm{d}u = p\,\xi_p + \int_p^1 F^{-1}(u)\,\mathrm{d}u. \tag{A.1}$$

Moreover, let $W_{1:n} \leq \cdots \leq W_{n:n}$ denote the order statistics based on the random variables $W_1, \dots, W_n$. For $i = 1, \dots, n$, it holds that

$$W_{i:n} = \begin{cases} \xi_p, & i \leq N_p, \\ X_{i:n}, & i > N_p. \end{cases}$$

Proof of Lemma 3.1. Recall the representation given in the third line of (2.2). Multiplying both sides by $(1-p)$, we get:

$$(1-p)\widehat{T}_{n,p} = -\frac{np - [np]}{n} X_{[np]+1:n} + \frac{1}{n} \sum_{i=[np]+1}^{n} X_{i:n}.$$

Consider the average of $W_1, \ldots, W_n$. We have

$$\frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n}\sum_{i=1}^{n} W_{i:n} = \frac{N_p}{n}\xi_p + \frac{1}{n}\sum_{i=N_p+1}^{n} X_{i:n},$$

because $W_{i:n} = X_{i:n}$ for all $i > N_p$. Taking into account (A.1), we find

$$\begin{aligned}(1-p)\big(\widehat{T}_{n,p} - T_p\big) - \frac{1}{n}\sum_{i=1}^{n}(W_i - \mu_W) &= -\frac{np - [np]}{n} X_{[np]+1:n} + \frac{1}{n}\sum_{i=[np]+1}^{n} X_{i:n} - \frac{N_p}{n}\xi_p - \frac{1}{n}\sum_{i=N_p+1}^{n} X_{i:n} \\ &\quad - \int_p^1 F^{-1}(u)\,\mathrm{d}u + \underbrace{p\,\xi_p + \int_p^1 F^{-1}(u)\,\mathrm{d}u}_{\mu_W} \\ &= -\frac{np - [np]}{n}\big(X_{[np]+1:n} - \xi_p\big) - \frac{N_p - [np]}{n}\xi_p + S_n, \qquad \text{(A.2)}\end{aligned}$$

where

$$S_n = \frac{1}{n}\sum_{i=[np]+1}^{n} X_{i:n} - \frac{1}{n}\sum_{i=N_p+1}^{n} X_{i:n} = \frac{\operatorname{sign}(N_p - [np])}{n}\sum_{i=([np]\wedge N_p)+1}^{[np]\vee N_p} X_{i:n}. \qquad \text{(A.3)}$$

Combining (A.2) and (A.3), we obtain

$$\begin{aligned}&\widehat{T}_p - T_p - \frac{1}{n(1-p)}\sum_{i=1}^{n}(W_i - \mu_W) \\ &\qquad = -\frac{np - [np]}{n(1-p)}\big(X_{[np]+1:n} - \xi_p\big) + \frac{\operatorname{sign}(N_p - [np])}{n(1-p)}\sum_{i=([np]\wedge N_p)+1}^{[np]\vee N_p}(X_{i:n} - \xi_p) \\ &\qquad = -R_n,\end{aligned}$$

which implies (3.2) and (3.3).

Next, we prove (3.4). There are three cases: (i) $[np] = N_p$; (ii) $[np] < N_p$; and (iii) $[np] > N_p$. For case (i), we have $X_{[np]+1:n} - \xi_p > 0$, hence the first term of R_n is not negative, whereas its second term is zero because, according to the definition of the sign function in (3.1), $\operatorname{sign}(N_p - [np]) = 0$.

For case (ii), we have $X_{i:n} \le \xi_p$ for $i = [np]+1, \ldots, N_p$, and hence,

$$\begin{aligned}R_n &= \frac{np - [np]}{n(1-p)}\big(X_{[np]+1:n} - \xi_p\big) - \frac{1}{n(1-p)}\sum_{i=[np]+1}^{N_p}(X_{i:n} - \xi_p) \\ &= \frac{np - [np] - 1}{n(1-p)}\big(X_{[np]+1:n} - \xi_p\big) - \frac{\mathbb{1}_{\{N_p > [np]+1\}}}{n(1-p)}\sum_{i=[np]+2}^{N_p}(X_{i:n} - \xi_p) \ge 0.\end{aligned}$$

Here, it is noteworthy that the indicator $\mathbb{1}_{\{N_p>[np]+1\}}$ in front of the second sum is needed because if $N_p = [np] + 1$, then the second sum disappears.

For case (iii), we have $\text{sign}(N_p - [np]) = -1$, and $X_{i:n} > \xi_p$ for all $i > N_p$. Hence,

$$R_n = \frac{np - [np]}{n(1-p)}\big(X_{[np]+1:n} - \xi_p\big) + \frac{1}{n(1-p)} \sum_{i=N_p+1}^{[np]} (X_{i:n} - \xi_p) \geq 0.$$

The proof for this assertion is now finished. □

To establish the main theorems in Section 3, the following assertion serves as an important auxiliary result. Specifically, let k and ε be arbitrary positive numbers, and set $\rho := k/\varepsilon$. The following lemma suggests an upper bound estimate for the absolute moment of the i-th order statistic, $\mathbf{E}|X_{i:n}|^k$, for $\rho \leq i \leq n - \rho + 1$, in terms of the absolute moment of the underlying distribution, $\mathbf{E}|X|^\varepsilon$. The proof of the lemma can be found in Gribkova (1995).

Lemma A.1 (Gribkova, 1995). *For all $n \geq 2\rho + 1$ and all i such that $\rho \leq i \leq n - \rho + 1$, the following inequality holds:*

$$\mathbf{E}|X_{i:n}|^k \leq C(\rho) \left\{ \mathbf{E}|X|^\varepsilon \, h\Big(\frac{i}{n+1}\Big) \right\}^\rho, \tag{A.4}$$

where the function $h : (0,1) \to (0,\infty)$ is given by $h(u) = [u(1-u)]^{-1}$, $u \in (0,1)$, and the constant $C(\rho)$ may be taken as $C(\rho) = 2\sqrt{\rho}\, \exp\big(\rho + 7/6\big)$.

To facilitate the presentation of the proofs of Theorems 3.1 and 3.2, some additional notations are needed. Let $U_1, \ldots, U_n$ be independent $[0,1]$-uniform random variables, and $U_{1:n} \leq \cdots \leq U_{n:n}$ represent their order statistics. For a fixed $p \in (0,1)$, define the binomial random variable $M_p = \text{card}\{i : U_i \leq p\}$, where $\text{card}\{\cdot\}$ denotes the cardinality of a set. Then, since the joint distribution of the order statistics $X_{1:n}, \ldots, X_{n:n}$ is identical to that of $F^{-1}(U_{1:n}), \ldots, F^{-1}(U_{n:n})$, the remainder term R_n in (3.3) can be written as

$$R_n \stackrel{d}{=} R_{n,1} + R_{n,2}, \tag{A.5}$$

where

$$\begin{aligned} R_{n,1} :=& \frac{np - [np]}{n(1-p)}\big(F^{-1}(U_{[np]+1:n}) - F^{-1}(p)\big), \\ R_{n,2} :=& -\frac{\text{sign}(M_p - [np])}{n(1-p)} \sum_{i=([np]\wedge M_p)+1}^{[np]\vee M_p} \big(F^{-1}(U_{i:n}) - F^{-1}(p)\big). \end{aligned} \tag{A.6}$$

Accordingly, we have

$$-B_n = \mathbf{E}(R_n) = \mathbf{E}(R_{n,1}) + \mathbf{E}(R_{n,2}). \tag{A.7}$$

Further, let $\mathbb{U}_p \subset (0,1)$ denote a neighborhood of the point $p \in (0,1)$ in which the quantile function F^{-1} is continuously differentiable (for the proof of Theorem 3.1), or Hölder continuous of order $\gamma \in (0,1]$ (for the proof of Theorem 3.2). In both cases, the following inequality holds (e.g., Shorack and Wellner, 1986, pp. 453–454):

$$\mathbf{P}\big(U_{[np]+1:n} \notin \mathbb{U}_p\big) \leq e^{-c\,n}, \tag{A.8}$$

for some constant $c > 0$. The inequality above will be used repeatedly throughout the proofs of the two theorems.

Proof of Theorem 3.1. Recall the relationship noted in (A.7). In what follows, we will prove that

$$\mathbf{E}(R_{n,1}) = o\big(n^{-1}\big) \tag{A.9}$$

and that

$$\mathbf{E}(R_{n,2}) = \frac{p}{2nf(\xi_p)} + o\big(n^{-1}\big), \tag{A.10}$$

which together yield the desired asymptotic expansion in (3.6).

To do so, define the following shorthand notation: $g(u) := \frac{\mathrm{d}}{\mathrm{d}u}\big(F^{-1}(u)\big) = 1/f\big(F^{-1}(u)\big)$ for $u \in \mathbb{U}_p$. We fix $\epsilon_1 > 0$ and $\delta > 0$ such that $I_\delta := (p-\delta, p+\delta) \subset \mathbb{U}_p$ and $|g(u) - g(v)| \leq \epsilon_1$ for all $u, v \in I_\delta$. To prove (A.9), we write

$$\mathbf{E}(R_{n,1}) = \mathbf{E}_1 + \mathbf{E}_2, \tag{A.11}$$

where

$$\mathbf{E}_1 := \mathbf{E}\big[R_{n,1}\mathbb{1}_{I_\delta}(U_{[np]+1:n})\big], \qquad \mathbf{E}_2 := \mathbf{E}\big[R_{n,1}\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\big].$$

Hereafter, $\mathbb{1}_D(\cdot)$ denotes the indicator function of the set D. Let us first study $\mathbf{E}_1$. The following relationships hold:

$$\begin{aligned}
\mathbf{E}_1 &= \frac{np - [np]}{n(1-p)}\mathbf{E}\Big[g(\zeta)\big(U_{[np]+1:n} - p\big)\mathbb{1}_{I_\delta}(U_{[np]+1:n})\Big] \\
&= \frac{np - [np]}{n(1-p)}\mathbf{E}\Big[\Big(g(p)\big(U_{[np]+1:n} - p\big) + \big(g(\zeta) - g(p)\big)\big(U_{[np]+1:n} - p\big)\Big)\mathbb{1}_{I_\delta}(U_{[np]+1:n})\Big],
\end{aligned} \tag{A.12}$$

where ζ is a point lying between p and $U_{[np]+1:n}$.

Moreover, consider the following decomposition:

$$\mathbf{E}\Big[\Big(g(p)\big(U_{[np]+1:n} - p\big) + \big(g(\zeta) - g(p)\big)\big(U_{[np]+1:n} - p\big)\Big)\mathbb{1}_{I_\delta}(U_{[np]+1:n})\Big] = e_1 + e_2 + e_3. \tag{A.13}$$

For the first component in the above decomposition, we have

$$e_1 := g(p)\mathbf{E}\big(U_{[np]+1:n} - p\big) = g(p)\left(\frac{[np]+1}{n+1} - p\right),$$

thus it holds that

$$|e_1| \le g(p)\,\frac{\max(p, 1-p)}{n+1}. \tag{A.14}$$

For the second component which is given by

$$e_2 := -g(p)\mathbf{E}\Big[\big(U_{[np]+1:n} - p\big)\big(1 - \mathbb{1}_{I_\delta}(U_{[np]+1:n})\big)\Big],$$

we have

$$|e_2| \le g(p)\mathbf{E}\Big[1 - \mathbb{1}_{I_\delta}(U_{[np]+1:n})\Big] = g(p)\mathbf{P}\Big(U_{[np]+1:n} \notin I_\delta\Big) \le g(p)\exp(-c\,n), \tag{A.15}$$

where the rightmost inequality holds due to (A.8). Here and throughout, c denotes a positive constant that does not depend on n but may vary from line to line.

Now, consider the third component in the decomposition (A.13), which is given by

$$e_3 := \mathbf{E}\Big[\big(g(\zeta) - g(p)\big)\big(U_{[np]+1:n} - p\big)\mathbb{1}_{I_\delta}\big(U_{[np]+1:n}\big)\Big]. \tag{A.16}$$

By the choice of ϵ_1 and δ noted at the beginning of this proof, and using well-known formulas for the variance of order statistics (see, e.g., Shorack and Wellner, 1986), we can obtain

$$|e_3| \le \epsilon_1\,\mathbf{E}\big|U_{[np]+1:n} - p\big| \le \epsilon_1\left(\frac{p(1-p)}{n}\big(1 + O(n^{-1})\big)\right)^{1/2}. \tag{A.17}$$

Combining (A.12)–(A.17) then yields

$$\mathbf{E}_1 = o(n^{-1}). \tag{A.18}$$

Next, we turn to $\mathbf{E}_2$ in (A.11), which can be expressed as

$$\frac{np - [np]}{n(1-p)}\mathbf{E}\Big[\big(F^{-1}(U_{[np]+1:n}) - F^{-1}(p)\big)\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\Big].$$

Let us consider the following decomposition:

$$\mathbf{E}\Big[\big(F^{-1}(U_{[np]+1:n}) - F^{-1}(p)\big)\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\Big] = e_4 + e_5, \tag{A.19}$$

where

$$e_4 := \mathbf{E}\big[F^{-1}(U_{[np]+1:n})\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\big],$$
$$e_5 := -F^{-1}(p)\mathbf{E}\big[\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\big].$$

For e_4, it holds that

$$\begin{aligned}
|e_4| &\le \mathbf{E}\big[\big|F^{-1}(U_{[np]+1:n})\big|\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\big] \\
&\le \big[\mathbf{E}\big|F^{-1}(U_{[np]+1:n})\big|^{1+\nu}\big]^{\frac{1}{1+\nu}}\big[\mathbf{E}\big(\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\big)\big]^{1-\frac{1}{1+\nu}} \\
&\le \underbrace{C(\rho)^{\frac{1}{1+\nu}}}_{C_1(\rho,\nu)}\left[\left(\mathbf{E}\big|X\big|^{\varepsilon}\frac{(n+1)^2}{([np]+1)(n-[np])}\right)^{\frac{1+\nu}{\varepsilon}}\right]^{\frac{1}{1+\nu}}\Big[\mathbf{P}\Big(U_{[np]+1:n}\notin I_\delta\Big)\Big]^{1-\frac{1}{1+\nu}} \\
&\le C_1(\rho,\nu)\left(\mathbf{E}\big|X\big|^{\varepsilon}\frac{(n+1)^2}{([np]+1)(n-[np])}\right)^{\frac{1}{\varepsilon}}\exp\left(-n\,\frac{c\nu}{1+\nu}\right) \\
&= C_1(\rho,\nu)\left(\frac{\mathbf{E}\big|X\big|^{\varepsilon}}{p(1-p)}\Big(1+O(n^{-1})\Big)\right)^{\frac{1}{\varepsilon}}\exp\left(-n\,\frac{c\nu}{1+\nu}\right). \qquad \text{(A.20)}
\end{aligned}$$

In the above calculations, the second inequality follows from the Hölder inequality with an arbitrary $\nu > 0$, while the third inequality applies the inequality in (A.4) with $\rho = (1+\nu)/\varepsilon$.

For c_5, we get

$$|e_5| = \big|F^{-1}(p)\big|\,\mathbf{P}\Big(U_{[np]+1:n}\notin I_\delta\Big) \le \big|F^{-1}(p)\big|\,\exp(-c\,n). \qquad \text{(A.21)}$$

Combining (A.19)–(A.21) then implies

$$\mathbf{E}_2 = o(n^{-1}). \qquad \text{(A.22)}$$

Collectively, the relationship in (A.9) holds because of (A.18) and (A.22).

So far, we have established (A.9). Now, we proceed to proving (A.10), which is more demanding. Recall that ϵ_1 and δ are defined and fixed at the beginning of the proof of this theorem. We note that if $U_{[np]+1:n} \in I_\delta$, then $U_{i:n} \in I_\delta$ for all $i = ([np]\wedge M_p)+1,\ldots,[np]\vee M_p$. Indeed, let $U_{[np]+1:n} \in I_\delta$. If $M_p > [np]$, then $U_{[np]+1:n} \le U_{M_p:n} \le p$ and $(U_{[np]+1:n}, U_{M_p:n}) \subset I_\delta$. If $M_p < [np]$, then $p < U_{M_p+1:n} \le U_{[np]+1:n}$, and hence, $(U_{M_p+1:n}, U_{[np]:n}) \subset I_\delta$. If $M_p = [np]$, then $R_{n,2}$ is zero. In passing, we note that, whenever $M_p \neq [np]$, the following inequality holds:

$$\max_{([np]\wedge M_p)+1\le i\le([np]\vee M_p)}\Big|F^{-1}(U_{i:n}) - F^{-1}(p)\Big| \le \Big|F^{-1}(U_{[np]:n}) - F^{-1}(p)\Big|. \qquad \text{(A.23)}$$

This inequality will be useful later.

We evaluate $\mathbf{E}(R_{n,2})$ in (A.10) by decomposing it as

$$\mathbf{E}(R_{n,2}) = \mathbf{E}_3 + \mathbf{E}_4, \tag{A.24}$$

where

$$\mathbf{E}_3 := \mathbf{E}\big[R_{n,2}\mathbb{1}_{I_\delta}(U_{[np]+1:n})\big], \qquad \mathbf{E}_4 := \mathbf{E}\big[R_{n,2}\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\big].$$

We first treat $\mathbf{E}_3$. Consider $R_{n,2}$ on the event $\{U_{[np]+1:n} \in I_\delta\}$, on which $\mathbb{1}_{I_\delta}(U_{[np]+1:n}) = 1$. On this event, we have

$$\begin{aligned} R_{n,2} = &-\frac{\operatorname{sign}(M_p - [np])}{n(1-p)} \sum_{i=([np]\wedge M_p)+1}^{[np]\vee M_p} g(p)\big(U_{i:n} - p\big) \\ &- \frac{\operatorname{sign}(M_p - [np])}{n(1-p)} \sum_{i=([np]\wedge M_p)+1}^{[np]\vee M_p} \big(g(\zeta_i) - g(p)\big)\big(U_{i:n} - p\big) =: R'_{n,2} + R''_{n,2}, \end{aligned} \tag{A.25}$$

where each ζ_i lies between p and $U_{i:n}$, $R'_{n,2}$ and $R''_{n,2}$ denote the first and second summation terms in (A.25), respectively. Accordingly, $\mathbf{E}_3$ can be written as

$$\begin{aligned} \mathbf{E}_3 &= \mathbf{E}\big[\big(R'_{n,2} + R''_{n,2}\big)\mathbb{1}_{I_\delta}(U_{[np]+1:n})\big] \\ &= \mathbf{E}\big(R'_{n,2}\big) - \mathbf{E}\big[R'_{n,2}\big(1 - \mathbb{1}_{I_\delta}(U_{[np]+1:n})\big)\big] + \mathbf{E}\big[R''_{n,2}\mathbb{1}_{I_\delta}(U_{[np]+1:n})\big] \\ &=: \mathbf{E}'_3 + e_6 + e_7. \end{aligned} \tag{A.26}$$

We claim that $|e_6 + e_7| = o\big(n^{-1}\big)$. To see this, for e_6, we have

$$\begin{aligned} |e_6| &\le \frac{g(p)}{n(1-p)}\mathbf{E}\Big(\big|M_p - [np]\big|\,\big(1 - \mathbb{1}_{I_\delta}(U_{[np]+1:n})\big)\Big) \\ &\le \frac{g(p)}{n(1-p)}\Big(\mathbf{E}(|M_p - [np]|)^2\mathbf{P}(U_{[np]+1:n} \notin I_\delta)\Big)^{1/2} \le \frac{g(p)p^{1/2}}{(n(1-p))^{1/2}}\exp(-c\,n). \end{aligned}$$

For e_7, we get

$$\begin{aligned} |e_7| &\le \frac{\sup_{u\in I_\delta}|g(u) - g(p)|}{n(1-p)}\mathbf{E}\Big(\big|M_p - [np]\big|\,\big|U_{[np]:n} - p\big|\Big) \\ &\le \frac{\epsilon_1}{n(1-p)}\Big(\mathbf{E}\big(M_p - [np]\big)^2\,\mathbf{E}\big(U_{[np]:n} - p\big)^2\Big)^{1/2} \\ &= \frac{\epsilon_1\big(np(1-p)\big)^{1/2}}{n(1-p)}\left(\frac{p(1-p)}{n}\big(1 + O(n^{-1})\big)\right)^{1/2} = \frac{\epsilon_1 p}{n} + O\big(n^{-2}\big), \end{aligned}$$

where we used the inequality in (A.23) and the Cauchy–Schwarz inequality to establish the desired

relationships. Since $\epsilon_1 > 0$ is arbitrary, it follows that $e_7 = o(n^{-1})$.

We now proceed to evaluating the first term in (A.26). We note that the following relationships hold:

$$\mathbf{E}_3' = \mathbf{E}\big(R_{n,2}'\big) = -\frac{g(p)}{n(1-p)}\mathbf{E}\big(\Sigma_n\big), \tag{A.27}$$

where

$$\begin{aligned}
\Sigma_n &:= \operatorname{sign}(M_p - [np]) \sum_{i=([np]\wedge M_p)+1}^{[np]\vee M_p} \big(U_{i:n} - p\big) = \Sigma_n' + \Sigma_n'',\\
\Sigma_n' &:= \mathbb{1}_{\{M_p > [np]\}} \sum_{i=[np]+1}^{M_p} \big(U_{i:n} - p\big) = T_{1,n} - \mathbb{1}_{\{M_p > [np]\}}(M_p - [np])\big(p - U_{M_p:n}\big),\\
\Sigma_n'' &:= -\mathbb{1}_{\{M_p < [np]\}} \sum_{i=M_p+1}^{[np]} \big(U_{i:n} - p\big) = T_{2,n} - \mathbb{1}_{\{M_p < [np]\}}([np] - M_p)\big(U_{M_p+1:n} - p\big),
\end{aligned}$$

with

$$T_{1,n} := -\mathbb{1}_{\{M_p > [np]\}} \sum_{i=[np]+1}^{M_p} \big(U_{M_p:n} - U_{i:n}\big), \quad T_{2,n} := -\mathbb{1}_{\{M_p < [np]\}} \sum_{i=M_p+1}^{[np]} \big(U_{i:n} - U_{M_p+1:n}\big).$$

Moreover, note that if $M_p = [np] + 1$, then $T_{1,n} = 0$. Therefore, we can slightly simplify $T_{1,n}$ into

$$T_{1,n} = -\mathbb{1}_{\{M_p > [np]+1\}} \sum_{i=[np]+1}^{M_p-1} \big(U_{M_p:n} - U_{i:n}\big) = -\mathbb{1}_{\{M_p > [np]+1\}} \sum_{i=[np]+1}^{M_p-1} \sum_{j=i+1}^{M_p} (U_{j:n} - U_{j-1:n}).$$

Similarly, if $M_p = [np] - 1$, then $T_{2,n} = 0$. Thus, $T_{2,n}$ can be simplified into

$$T_{2,n} = -\mathbb{1}_{\{M_p < [np]-1\}} \sum_{i=M_p+2}^{[np]} \big(U_{i:n} - U_{M_p+1:n}\big) = -\mathbb{1}_{\{M_p < [np]-1\}} \sum_{i=M_p+2}^{[np]} \sum_{j=M_p+2}^{i} (U_{j:n} - U_{j-1:n}).$$

Let $s_{j,n} := U_{j:n} - U_{j-1:n}$, $j = 1, \ldots, n+1$, which correspond to the spacings of the uniform order statistics on $[0,1]$, with $U_{0:n} = 0$ and $U_{n+1:n} = 1$. Then we can rewrite $T_{1,n}$ and $T_{2,n}$ as

$$\begin{aligned}
T_{1,n} = -\mathbb{1}_{\{M_p > [np]+1\}} \sum_{i=[np]+1}^{M_p-1} \sum_{j=i+1}^{M_p} s_{j,n} &= -\mathbb{1}_{\{M_p > [np]+1\}} \sum_{i=[np]+1}^{M_p-1} \left(\sum_{j=i+1}^{M_p} s_{j,n} - \frac{M_p - i}{n+1} \right)\\
&\quad - \mathbb{1}_{\{M_p > [np]+1\}} \frac{(M_p - [np] - 1)(M_p - [np])}{2(n+1)},\\
T_{2,n} = -\mathbb{1}_{\{M_p < [np]-1\}} \sum_{i=M_p+2}^{[np]} \sum_{j=M_p+2}^{i} s_{j,n} &= -\mathbb{1}_{\{M_p < [np]-1\}} \sum_{i=M_p+2}^{[np]} \left(\sum_{j=M_p+2}^{i} s_{j,n} - \frac{i - M_p - 1}{n+1} \right)
\end{aligned}$$

$$- \mathbb{1}_{\{M_p<[np]-1\}} \frac{([np]-M_p-1)([np]-M_p)}{2(n+1)}.$$

In addition, we observe that

$$-\mathbb{1}_{\{M_p>[np]+1\}} \frac{(M_p-[np]-1)(M_p-[np])}{2(n+1)} - \mathbb{1}_{\{M_p<[np]-1\}} \frac{([np]-M_p-1)([np]-M_p)}{2(n+1)} \tag{A.28}$$

can be written equivalently as

$$-\frac{(|M_p-[np]|-1)|M_p-[np]|}{2(n+1)} =: V_{p,n}. \tag{A.29}$$

Indeed, the expression in (A.29) equals zero when $M_p = [np]$, $M_p = [np]+1$, or $M_p = [np]-1$, which matches exactly the cases excluded by the indicator functions involved in (A.28).

Collectively, we can conclude

$$\Sigma_n = V_{p,n} + r_{1,n} + r_{2,n} + r_{3,n}, \tag{A.30}$$

where

$$\begin{aligned}
r_{1,n} &:= -\mathbb{1}_{\{M_p>[np]\}}(M_p-[np])\big(p-U_{M_p:n}\big) - \mathbb{1}_{\{M_p<[np]\}}([np]-M_p)\big(U_{M_p+1:n}-p\big),\\
r_{2,n} &:= -\mathbb{1}_{\{M_p>[np]+1\}} \sum_{i=[np]+1}^{M_p-1} \left(\sum_{j=i+1}^{M_p} s_{j,n} - \frac{M_p-i}{n+1} \right),\\
r_{3,n} &:= -\mathbb{1}_{\{M_p<[np]-1\}} \sum_{i=M_p+2}^{[np]} \left(\sum_{j=M_p+2}^{i} s_{j,n} - \frac{i-M_p-1}{n+1} \right).
\end{aligned}$$

For the first term on the right-hand side of (A.30), we have

$$\begin{aligned}
\mathbf{E}\big(V_{p,n}\big) &= -\frac{1}{2(n+1)}\mathbf{E}\Big((M_p-[np])^2 - |M_p-[np]|\Big)\\
&= -\frac{1}{2(n+1)}\mathbf{E}\Big((M_p-np)^2 + (np-[np])(M_p-[np]+M_p-np) - |M_p-[np]|\Big)\\
&= -\frac{np(1-p)}{2(n+1)} + \delta_n,
\end{aligned} \tag{A.31}$$

where

$$\delta_n := -\frac{1}{2(n+1)}\mathbf{E}\Big((np-[np])(M_p-[np]+M_p-np) - |M_p-[np]|\Big).$$

Since $np - [np] < 1$ and $|M_p - [np]| < |M_p - np| + 1$, it follows that

$$|\delta_n| < \frac{3\mathbf{E}|M_p - np| + 2}{2(n+1)} < \frac{3\big(\mathbf{E}(M_p - np)^2\big)^{1/2} + 2}{2n} = \frac{3\big(p(1-p)\big)^{1/2}}{2n^{1/2}} < \frac{3}{4n^{1/2}} < \frac{1}{n^{1/2}}. \tag{A.32}$$

Next we develop bounds for $r_{k,n}$, $k = 1, 2, 3$, appearing on the right-hand side of (A.30). For $r_{1,n}$, we have

$$\begin{aligned} |r_{1,n}| &\leq |M_p - [np]\,|(p - U_{M_p:n}) + |M_p - [np]\,|(U_{M_p+1:n} - p) = |M_p - [np]\,|(U_{M_p+1:n} - U_{M_p:n}) \\ &\leq \big(|M_p - np| + 1\big)(U_{M_p+1:n} - U_{M_p:n}) \overset{d}{=} \big(|M_p - np| + 1\big)U_{1:n}. \end{aligned} \tag{A.33}$$

where, in the last step, we used the fact that $(U_{M_p+1:n} - U_{M_p:n}) = s_{M_p+1,n} \overset{d}{=} s_{1,n} = U_{1:n}$; that is, the uniform spacings are identically distributed and exchangeable random variables (see, Shorack and Wellner, 1986, p. 721). The relationships derived in (A.33) imply that

$$\begin{aligned} |\mathbf{E}(r_{1,n})| &\leq \mathbf{E}\big(|M_p - np|\,U_{1:n}\big) + \mathbf{E}U_{1:n} \leq \Big(np(1-p)\mathbf{E}\big(U_{1:n}\big)^2\Big)^{1/2} + n^{-1} \\ &= \left(\frac{2np(1-p)}{(n+2)(n+1)}\right)^{1/2} + n^{-1} < (2n)^{-1/2} + n^{-1} < 2n^{-1/2}. \end{aligned} \tag{A.34}$$

We now turn to the terms $r_{2,n}$ and $r_{3,n}$ involved in (A.30). Let us set $k = i - [np]$, then $r_{2,n}$ can be rewritten as

$$\begin{aligned} r_{2,n} &= -\mathbb{1}_{\{M_p > [np]+1\}} \sum_{k=1}^{M_p - [np] - 1} \left(\sum_{j=[np]+k+1}^{M_p} s_{j,n} - \frac{M_p - [np] - k}{n+1} \right) \\ &\overset{d}{=} -\mathbb{1}_{\{M_p > [np]+1\}} \sum_{k=1}^{M_p - [np] - 1} \left(\sum_{j=1}^{M_p - [np] - k} s_{j,n} - \frac{M_p - [np] - k}{n+1} \right) \\ &= -\mathbb{1}_{\{M_p > [np]+1\}} \sum_{k=1}^{M_p - [np] - 1} \left(U_{M_p - [np] - k:n} - \frac{M_p - [np] - k}{n+1} \right) \\ &= -\mathbb{1}_{\{M_p > [np]+1\}} \sum_{i=1}^{M_p - [np] - 1} \left(U_{i:n} - \frac{i}{n+1} \right), \end{aligned} \tag{A.35}$$

where in the second line, we used again the fact that the uniform spacings are identically distributed and exchangeable random variables.

Similarly, setting $k = i - M_p - 1$, we can obtain the following relationships for $r_{3,n}$:

$$r_{3,n} = -\mathbb{1}_{\{M_p < [np]-1\}} \sum_{k=1}^{[np] - M_p - 1} \left(\sum_{j=M_p+2}^{k+M_p+1} s_{j,n} - \frac{k}{n+1} \right)$$

$$\stackrel{d}{=} -\mathbb{1}_{\{M_p<[np]-1\}} \sum_{k=1}^{[np]-M_p-1} \left(\sum_{j=1}^{k} s_{j,n} - \frac{k}{n+1} \right)$$
$$= -\mathbb{1}_{\{M_p<[np]-1\}} \sum_{k=1}^{[np]-M_p-1} \left(U_{k:n} - \frac{k}{n+1} \right). \tag{A.36}$$

Then combining (A.35) and (A.36) yields

$$\begin{aligned}
&|\mathbf{E}(r_{2,n}+r_{3,n})| \le \mathbf{E}|r_{2,n}| + \mathbf{E}|r_{3,n}| \\
\le &\mathbf{E}\left(\mathbb{1}_{\{M_p>[np]+1\}} \sum_{i=1}^{M_p-[np]-1} \left| U_{i:n} - \frac{i}{n+1} \right| \right) + \mathbf{E}\left(\mathbb{1}_{\{M_p<[np]-1\}} \sum_{i=1}^{[np]-M_p-1} \left| U_{i:n} - \frac{i}{n+1} \right| \right) \\
= &\mathbf{E}\left(\mathbb{1}_{\{M_p>[np]+1\}} \sum_{i=1}^{|M_p-[np]-1|} \left| U_{i:n} - \frac{i}{n+1} \right| + \mathbb{1}_{\{M_p<[np]-1\}} \sum_{i=1}^{|[np]-M_p-1|} \left| U_{i:n} - \frac{i}{n+1} \right| \right) \\
\le &\mathbf{E}\left(\sum_{i=1}^{|M_p-[np]|+1} \left| U_{i:n} - \frac{i}{n+1} \right| \right).
\end{aligned} \tag{A.37}$$

Let $A > 1$ be arbitrary. By Hoeffding's inequality (Hoeffding, 1963, Theorem 1), we know

$$\mathbf{P}\left(|M_p - np| \ge \sqrt{A\, n \log n} \right) \le 2\exp\left(-2A \log n\right) = 2n^{-2A} < 2n^{-2}$$

for all sufficiently large n such that $0 < \sqrt{A \log n / n} < p$. Therefore, the right-hand side of (A.37) can be bounded by

$$\begin{aligned}
&\sum_{i=1}^{\lceil \sqrt{A\, n \log n} \rceil} \mathbf{E}\left| U_{i:n} - \frac{i}{n+1} \right| + \mathbf{E}\left(\mathbb{1}_{\{|M_p - np| \ge \sqrt{A\, n \log n}\}} \sum_{i=1}^{|M_p-[np]|+1} \left| U_{i:n} - \frac{i}{n+1} \right| \right) \\
\le &\sum_{i=1}^{\lceil \sqrt{A\, n \log n} \rceil} \left(\frac{i(n-i+1)}{(n+1)^3} \right)^{1/2} + n\, \mathbf{P}\left(|M_p - np| \ge \sqrt{A\, n \log n} \right) \\
\le &\sum_{i=1}^{\lceil \sqrt{A\, n \log n} \rceil} \left(\frac{i}{(n+1)^2} \right)^{1/2} + 2n^{-1} \le \frac{\lceil \sqrt{A\, n \log n} \rceil}{n+1} (\lceil \sqrt{A\, n \log n} \rceil)^{1/2} + 2n^{-1} \\
= &\frac{(A\, n \log n)^{3/4}}{n} + O(n^{-1}) = O\left(\frac{(\log n)^{3/4}}{n^{1/4}} \right),
\end{aligned} \tag{A.38}$$

where $\lceil \cdot \rceil$ denotes the ceiling function.

Collecting the relationships established in (A.27), (A.30)–(A.31) and the estimates (A.32),

(A.34), and (A.38), we readily obtain

$$\mathbf{E}_3' = -\frac{g(p)}{n(1-p)}\mathbf{E}\left[V_{p,n} + r_{1,n} + r_{2,n} + r_{3,n}\right] = \frac{g(p)}{n(1-p)}\left[\frac{np(1-p)}{2(n+1)} + O\left(\frac{(\log n)^{3/4}}{n^{1/4}}\right)\right]$$
$$= \frac{g(p)p}{2(n+1)} + O\left(\frac{(\log n)^{3/4}}{n^{5/4}}\right) = \frac{p}{2nf(\xi_p)} + o\big(n^{-1}\big).$$

To complete the proof, it remains to show that the term $\mathbf{E}_4$ in (A.24) satisfies

$$\mathbf{E}_4 = o\big(n^{-1}\big). \tag{A.39}$$

To this end, the following bound is established:

$$\begin{aligned}
|\mathbf{E}_4| &\le \mathbf{E}\big[|R_{n,2}|\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\big] \\
&= \frac{1}{n(1-p)}\mathbf{E}\left[\left|\sum_{i=([np]\wedge M_p)+1}^{[np]\vee M_p}\big(F^{-1}(U_{i:n}) - F^{-1}(p)\big)\right|\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\right] \\
&\le \frac{1}{1-p}\mathbf{E}\left[\left|F^{-1}(U_{[np]:n}) - F^{-1}(p)\right|\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\right] \\
&\le \frac{1}{1-p}\mathbf{E}\left[\left|F^{-1}(U_{[np]:n})\right|\mathbb{1}_{[0,1]\setminus I_\delta}(U_{[np]+1:n})\right] + \frac{\left|F^{-1}(p)\right|}{1-p}\mathbf{P}\Big(U_{[np]+1:n} \notin I_\delta\Big),
\end{aligned} \tag{A.40}$$

where, in particular, the inequality in (A.23) was used in the third step. The first term on the right-hand side of (A.40) has already been bounded in (A.20). Consequently, for an arbitrary $\nu > 0$, the right-hand side of (A.40) is bounded above by

$$\frac{C_1(\rho,\nu)}{1-p}\left(\frac{\mathbf{E}|X|^{\varepsilon}}{p(1-p)}\big(1 + O(n^{-1})\big)\right)^{\frac{1}{\varepsilon}}\exp\left(-n\,\frac{c\nu}{1+\nu}\right) + \frac{\left|F^{-1}(p)\right|}{1-p}\exp(-c\,n),$$

where $C_1(\rho,\nu) > 0$ is a constant depending only on ν and $\rho = (1+\nu)/\varepsilon$, as defined in (A.20). Since both terms in the above expression decay exponentially fast in n, this bound implies (A.39).

The proof of Theorem 3.1 is finally completed. □

Proof of Theorem 3.2. Recall that, from (A.5)–(A.7), we have

$$-B_n = \mathbf{E}(R_n) = \mathbf{E}(R_{n,1}) + \mathbf{E}(R_{n,2}). \tag{A.41}$$

We shall first show that $\mathbf{E}(R_{n,1})$ is of negligible order relative to the right-hand side of (3.8). To

this end, we write

$$\mathbf{E}(R_{n,1}) = \mathbf{E}_1 + \mathbf{E}_2,$$

where

$$\mathbf{E}_1 := \mathbf{E}\big(R_{n,1}\mathbb{1}_{\mathbb{U}_p}(U_{[np]+1:n})\big), \qquad \mathbf{E}_2 := \mathbf{E}\big(R_{n,1}\mathbb{1}_{[0,1]\setminus\mathbb{U}_p}(U_{[np]+1:n})\big).$$

Here, $\mathbb{U}_p$ denotes the neighborhood of p on which F^{-1} is locally Hölder continuous of order $\gamma \in (0,1]$ with constant c_γ, as specified in (3.7).

For $\mathbf{E}_1$, we get

$$\begin{aligned}|\mathbf{E}_1| &\le \frac{c_\gamma}{n(1-p)}\mathbf{E}|U_{[np]+1:n} - p|^\gamma \le \frac{c_\gamma}{n(1-p)}\Big(\mathbf{E}(U_{[np]+1:n} - p)^2\Big)^{\gamma/2} \\ &= \frac{c_\gamma}{n(1-p)}\Big(\frac{p(1-p)}{n}(1+o(1))\Big)^{\gamma/2} = \frac{c_\gamma p^{\gamma/2}}{n^{1+\gamma/2}(1-p)^{1-\gamma/2}}\big(1+o(1)\big).\end{aligned}$$

Therefore, $|\mathbf{E}_1| = O(n^{-1-\gamma/2}) = o(n^{-1})$, which is negligible relative to the bound in (3.8).

For $\mathbf{E}_2$, we have

$$\begin{aligned}|\mathbf{E}_2| &= \frac{np - [np]}{n(1-p)}\left|\mathbf{E}\Big(\big(F^{-1}(U_{[np]+1:n}) - F^{-1}(p)\big)\mathbb{1}_{[0,1]\setminus\mathbb{U}_p}(U_{[np]+1:n})\Big)\right| \\ &\le \frac{1}{n(1-p)}\left(\mathbf{E}\Big|F^{-1}(U_{[np]+1:n})\mathbb{1}_{[0,1]\setminus\mathbb{U}_p}(U_{[np]+1:n})\Big| + |F^{-1}(p)|\mathbf{P}\big(U_{[np]+1:n} \notin \mathbb{U}_p\big)\right).\end{aligned}$$

Arguing as in the estimation of e_4 in (A.20), it follows that for any $\nu > 0$,

$$|\mathbf{E}_2| \le \frac{1}{n(1-p)}\left(C_1(\rho,\nu)\left(\frac{\mathbf{E}|X|^\varepsilon}{p(1-p)}\big(1+O(n^{-1})\big)\right)^{\frac{1}{\varepsilon}}\exp\left(-n\,\frac{c\nu}{1+\nu}\right) + |F^{-1}(p)|e^{-c\,n}\right),$$

where $\rho = (1+\nu)/\varepsilon$. Since both terms in the above expression decay exponentially fast in n, we can conclude that $\mathbf{E}_2 = o(n^{-1})$, and hence $\mathbf{E}_2$ is negligible relative to the bound in (3.8).

It remains to treat $\mathbf{E}(R_{n,2})$ in (A.41). We decompose it as

$$\mathbf{E}(R_{n,2}) = \mathbf{E}_3 + \mathbf{E}_4,$$

where

$$\mathbf{E}_3 := \mathbf{E}\big(R_{n,2}\mathbb{1}_{\mathbb{U}_p}(U_{[np]+1:n})\big), \qquad \mathbf{E}_4 := \mathbf{E}\big(R_{n,2}\mathbb{1}_{[0,1]\setminus\mathbb{U}_p}(U_{[np]+1:n})\big).$$

We first consider the term $\mathbf{E}_3$. Note that all terms in the sum defining $R_{n,2}$ in (A.6) have the same sign, either all positive or all negative, so the absolute value of the sum equals the sum of the

absolute values. By the local Hölder continuity of F^{-1}, we can obtain

$$|\mathbf{E}_3| \le \frac{c_\gamma}{n(1-p)}\mathbf{E}\left(\sum_{i=([np]\wedge M_p)+1}^{[np]\vee M_p} \big|U_{i:n}-p\big|^\gamma\right). \tag{A.42}$$

Moreover, by (A.23), the largest term in the sum in (A.42) is bounded above by $|U_{[np]:n}-p|^\gamma$. Consequently, the right-hand side of (A.42) is bounded by

$$\begin{aligned}
&\frac{c_\gamma}{n(1-p)}\mathbf{E}\Big(|M_p-[np]||U_{[np]:n}-p|^\gamma\Big)\\
\le&\frac{c_\gamma}{n(1-p)}\mathbf{E}\Big((|M_p-np|+1)|U_{[np]:n}-p|^\gamma\Big)\\
\le&\frac{c_\gamma}{n(1-p)}\Big[\mathbf{E}\big(|M_p-np|+1\big)^2\,\mathbf{E}|U_{[np]:n}-p|^{2\gamma}\Big]^{1/2}\\
\le&\frac{c_\gamma}{n(1-p)}\Big[np(1-p)+2\mathbf{E}|M_p-np|+1\Big]^{1/2}\Big[\big(\mathbf{E}(U_{[np]:n}-p)^2\big)^{2\gamma/2}\Big]^{1/2}\\
\le&\frac{c_\gamma}{n(1-p)}\Big[np(1-p)+2(np(1-p))^{1/2}+1\Big]^{1/2}\left[\frac{p(1-p)}{n}\big(1+O(n^{-1})\big)\right]^{\gamma/2}\\
=&\frac{c_\gamma}{n(1-p)}\Big[np(1-p)\big(1+O(n^{-1/2})\big)\Big]^{1/2}\left[\frac{p(1-p)}{n}\big(1+O(n^{-1})\big)\right]^{\gamma/2}\\
=&\frac{c_\gamma}{n(1-p)}\Big[np(1-p)\Big]^{1/2}\big(1+O(n^{-1/2})\big)\left[\frac{p(1-p)}{n}\right]^{\gamma/2}\big(1+O(n^{-1})\big)\\
=&\frac{c_\gamma p^{\frac{1+\gamma}{2}}}{n^{\frac{1+\gamma}{2}}(1-p)^{\frac{1-\gamma}{2}}}\big(1+O(n^{-1/2})\big).
\end{aligned}\tag{A.43}$$

The above result leads exactly to the expression of the bound (3.8) in Theorem 3.2.

To complete the proof, it remains to show that the term $\mathbf{E}_4$ is of negligible order relative to the bound on the right-hand side of (A.43). Proceeding as before, but now taking into account the indicator $\mathbb{1}_{[0,1]\setminus\mathbb{U}_p}(U_{[np]+1:n})$, we obtain

$$\begin{aligned}
|\mathbf{E}_4| &\le \frac{1}{n(1-p)}\mathbf{E}\Big(|M_p-[np]|\big|F^{-1}(U_{[np]:n})-F^{-1}(p)\big|\mathbb{1}_{[0,1]\setminus\mathbb{U}_p}(U_{[np]+1:n})\Big)\\
&\le \frac{1}{1-p}\mathbf{E}\Big(\big|F^{-1}(U_{[np]:n})-F^{-1}(p)\big|\mathbb{1}_{[0,1]\setminus\mathbb{U}_p}(U_{[np]+1:n})\Big)\\
&\le \frac{1}{1-p}\mathbf{E}\Big(\big|F^{-1}(U_{[np]:n})\big|\mathbb{1}_{[0,1]\setminus\mathbb{U}_p}(U_{[np]+1:n})\Big)+\frac{\big|F^{-1}(p)\big|}{1-p}\mathbf{P}\big(U_{[np]+1:n}\notin\mathbb{U}_p\big)\\
&\le \frac{C_1(\rho,\nu)}{1-p}\left(\frac{\mathbf{E}|X|^\varepsilon}{p(1-p)}\Big(1+O(n^{-1})\Big)\right)^{\frac{1}{\varepsilon}}\exp\left(-n\,\frac{c\nu}{1+\nu}\right)+\frac{\big|F^{-1}(p)\big|}{1-p}\exp(-c\,n),
\end{aligned}$$

for each $\nu>0$ and $\rho=(1+\nu)/\varepsilon$; see a similar derivation in (A.20). This implies that $\mathbf{E}_4$ is of exponentially small order and therefore negligible relative to the bound in (A.43). This completes

the proof of Theorem 3.2. □