Advances in Derivative Estimation:

Volk-Makarewicz, W.M.

2014

document version
Publisher's PDF, also known as Version of record

Chapter 4

A Distributional approach for Quantile Derivative Estimation

4.1 Introduction

Let $Z \in \mathbb{R}$ be a continuous random variable that has distribution function F and density function f, and let $\mathbf{Z} = (Z_i : 1 \leq i \leq m)$ be a sequence of m independent and identically distributed (i.i.d.) copies of Z. The random variable Z is influenced by a controllable intrinsic parameter $\theta \in \Theta$ that, for simplicity, has a domain on an interval $\Theta = (a, b) \subset \mathbb{R}$. In applications $Z = h(X)$ represents a performance function of a discrete-event model with an input random vector $X \in \mathbb{R}^J$ and Borel measurable map h, and the influence of θ is through this vector.

Earlier work on the derivative estimation of quantiles is based on both pathwise differentiation, by Hong and co-authors [57], [71], [32], [58], and distributional differentiation, [55], [56], methods. Let $q_\alpha(\theta)$ denote the quantile function of Z at level α, $\alpha \in (0,1)$, and let $Z_{\lceil \alpha m \rceil : m}$, $1 \leq \lceil \alpha m \rceil \leq m$, denote the corresponding order statistic, [4], that estimates and converges almost surely to $q_\alpha(\theta)$ as $m \to \infty$. Quantiles and order statistics have been previously discussed in Sections 1.3.1 and 2.1 respectively, and concepts from either of these areas will only be brought up when needed.

The first result was [57], where the quantile derivative is derived by the differentiation of a random variable $Z(\theta) = h(X(\theta))$, contingent on the value of this random variable to be equal to the quantile, i.e., yielding the estimator

$$\frac{\partial}{\partial \theta} \hat{q}_\alpha(\theta) \;=\; \mathbb{E}\left[\frac{\partial}{\partial \theta} h(X(\theta)) \,\middle|\, h(X(\theta)) = q_\alpha(\theta) \right].$$

Estimating the quantile via the order statistic $Z_{\lceil \alpha m \rceil : m}$, Infinitesimal Perturbation Analysis (IPA) is employed to differentiate the random variable with the realization associated the order statistic. This is permitted since for continuous random variables, where $Z \sim F_\theta$, and F_θ^{-1} is the inverse distribution function of Z, $Z_{\lceil \alpha m \rceil : m} = F_\theta^{-1}(U_{\lceil \alpha m \rceil : m})$, and $U_{\lceil \alpha m \rceil : m}$ is an order statistic based on m i.i.d. uniform-$(0,1)$ samples, independent of θ. Differentiating the order statistic for intrinsic parameters is equivalent to differentiating F_θ^{-1} and the order of the order statistic sequence is preserved for all $\omega \in \Omega$. This derivative estimator is a precise method to attain the quantile sensitivity due to (2.11) in Section 2.4.1. This is also observed to be the most precise and expedient ranked-data gradient estimation method, Section 2.5. In [59], the same authors discerned a similar derivative estimator via IPA for the parameter derivative of the conditional value-at-risk.

In [71], the authors improved two deficiencies present in the first paper: a rate of convergence of only $O(n^{-1/3})$, and, given a fixed computational budget, the ranked-data estimator produced irregular RMS Error and coverage probability estimates. The authors proposed a kernel-weighted averaging of the IPA derivatives where the kernel function is centred at $q_\alpha(\theta)$. Though kernel estimation does introduce bias, the authors showed that the extent of bias is negligible and improved the rate of convergence to $O(n^{-2/5})$. This renders the kernel estimator to be a interesting alternative method to attain the quantile sensitivity from readily known statistical methods.

However, there are a couple of caveats. The first is that the IPA derivative methods requires the map h to the piece-wise differentiable w.r.t. state variable x and to be Lipschitz continuous w.r.t. the parameter θ, Section 1.2.1.1. Secondly, in applications the IPA estimator may not exist with the required convergence properties. For instance, consider the stationary waiting time in a (non)-exponential queueing network. Except for some special cases, pathwise differentiation of the stationary waiting time is not applicable (a general result is [37], which provides an IPA estimator based on regenerative analysis). Or, as a second example, consider the maintenance problem studied in [50] where no IPA estimator exists for this model. Thirdly, we comment on the implementation of kernel estimation methods. There are many considerations, however, we focus on two of these. Firstly, following [46], which used a simpler estimation scheme than in [72], the choice of kernel is important to ensure the optimal bandwidth parameter will yield an estimate with a small RMS Error to the quantity that is estimated. Secondly, the efficacy of kernel estimation can be sensitive to the choice of the bandwidth parameter.

Alternatively, the quantile sensitivity can be estimated directly from the representation of the quantile for continuous random variables: $F_\theta(q_\alpha(\theta)) = \alpha$. Implicit differentiation w.r.t. θ of this equation yields

$$\frac{\partial}{\partial\theta} q_\alpha(\theta) \;=\; -\frac{\frac{\partial}{\partial\theta} F_\theta(q_\alpha(\theta))}{f_\theta(q_\alpha(\theta))}. \tag{4.1}$$

This computation is provided in Section 1.3.1. We observe that the above expression is a distributional representation of the pathwise derivative, [100], presented in Equation (2.12). In Fu, Hu, and Hong [32], the authors separately wrote the parameter derivative and the inverse of the density function in Equation (4.1) as expectations and attained a pathwise derivative by utilizing the structure of the underlying model, using conditioning. This estimation method is called Smoothed Perturbation Analysis (SPA), Section 1.2.1.2. For this derivative estimator, the standard rate of convergence for simulation is attained, $O(n^{-1/2})$, whilst requiring weaker set of assumptions. The conditioning of random variables in this method is used to "smooth-out" discontinuities and allows the pathwise derivative to be applied under weaker conditions. For instance, differentiation of the map h and Lipschitz continuity w.r.t. θ is not required. However, the disadvantage of SPA is the requirement of manipulating the performance function to be amenable to pathwise differentiation. Even for models that are relatively simple, the computational complexity makes this method impractical, e.g., the aforementioned maintenance problem in [50].

Instead, this chapter presents a distributional method, following the representation in Equation (4.1) to estimate the quantile sensitivity that is computed from the sequence $\mathbf{Z}$ and the distribution of Z and is independent of the choice the performance function. This estimation method is applicable to non-smooth cost functions, requires no conditioning, and in the most general variant of our estimator, the map h only needs to be Borel. In the case that F_θ is analytically tractable, the parameter derivative $\partial_\theta F_\theta$ is attained via distributional differentiation and this gradient estimator is a single-run estimator. However, if $\partial_\theta F_\theta$ fails to be attainable, $\partial_\theta F_\theta$ is estimated empirically via Measure Valued Differentiation (MVD). Though this method has the disadvantage of requiring additional random variable generations, MVD has comparable precision to IPA.

The chapter is organized as follows. In Section 4.2, we present the quantile sensitivity estimators and explain further the statistical concepts. In Section 4.3, the mathematical analysis of the derivative estimator in the case when $\partial_\theta F_\theta$ is analytically tractable is supplied. Section 4.4 repeats the mathematical analysis when $\partial_\theta F_\theta$ is estimated. Applications to queues are discussed in Section 4.5 and applications to financial risk management are presented in Section 4.6. As an application, in Section 4.6.2.1 we provide the sensitivity analysis of the cash-flow of options defined on the Variance Gamma process. Though IPA estimators were proposed by Fu in [28], no proof of unbiasedness was shown.

4.2 Quantile Sensitivity Analysis

The form of the quantile sensitivity estimator follows from Equation (4.1), in which the estimation of the parameter derivative $\partial_\theta F_\theta(q_\alpha(\theta))$ and the quantile density function $1/f_\theta(q_\alpha(\theta))$ are separately ascertained. Estimation of the parameter derivative will be discussed presently. To estimate the quantile density function we utilize the spacing limit for i.i.d. realizations from Pyke, [86]:

$$m\left(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil-1:m}\right) \overset{d}{\to} \frac{E}{f_\theta(q_\alpha(\theta))}, \tag{4.2}$$

as $m \to \infty$, where $\overset{d}{\to}$ denotes convergence in distribution and E is an Exponential random variable with mean one. Hence, we can attain an estimate for the quantile density only from the order statistic sequence. This spacing sequence is an rudimentary version of a nearest neighbour estimator, [73]. This density estimator is discussed in Section 4.3.2.

The combination of the almost sure convergence of $Z_{\lceil \alpha m\rceil:m}$ towards the quantile together with the limit of spacings in Equation (4.2) suggests the following result for the quantile sensitivity

$$\lim_{m\to\infty} \mathbb{E}_\theta\left[-m\left(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil-1:m}\right)\frac{\partial}{\partial\theta}F_\theta(Z_{\lceil \alpha m\rceil:m})\right] = \frac{\partial}{\partial\theta}q_\alpha(\theta). \tag{4.3}$$

The mathematical analysis for this derivative estimator follows Equation (4.3), and has two main aspects: the statistical and the distributional differentiation aspect. For the *statistical analysis* we will first provide sufficient conditions for the above limit to hold, which already yields an asymptotically unbiased estimator for $\partial_\theta q_\alpha(\theta)$. Provided that Equation (4.3) holds, taking averages over i.i.d. realizations of

$$-m\left(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil-1:m}\right)\frac{\partial}{\partial\theta}F_\theta(Z_{\lceil \alpha m\rceil:m}), \tag{4.4}$$

then yields a strongly consistent estimator for $\partial_\theta q_\alpha(\theta)$. As we will show in this chapter, confidence intervals for $\partial_\theta q_\alpha(\theta)$ can be established as well. The other aspect in Equation (4.3) is how to evaluate the *distributional derivative* $\partial_\theta F_\theta(Z_{\lceil \alpha m\rceil:m})$. If the model is relatively simple, the parameter derivative may be analytically tractable, either by determining F_θ exactly or by modifying the performance expression to write F_θ as a distribution function in terms of the input random variable X. Otherwise, efficient simulation of $\partial_\theta F_\theta$ has to be addressed. To simulate $\partial_\theta F_\theta$, we will apply Measure-Valued Differentiation to operationalize $\partial_\theta F_\theta(Z_{\lceil \alpha m\rceil:m})$ for estimation. In particular, if $\partial_\theta F_\theta$ exists, then under rather weak conditions the parameter derivative can be written as $\partial_\theta F_\theta = c_\theta(F_\theta^+ - F_\theta^-)$, with constant $c_\theta > 0$ and distribution functions $F_\theta^\pm$. Inserting the above difference expression for $\partial_\theta F_\theta$ into Equation (4.3), we arrive at the estimator

$$-mc_\theta\left(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil-1:m}\right)\left(F_\theta^+(Z_{\lceil \alpha m\rceil:m}) - F_\theta^-(Z_{\lceil \alpha m\rceil:m})\right). \tag{4.5}$$

The above estimators are single-run estimators as no additional simulations other than sampling the order statistics are required.

Denote by $D(m,k)$ the sample average over k realizations of one of the estimators in Equations (4.4) and (4.5). Let n denote the computational budget, i.e., n is total number of realizations of Z that can be used for estimating $\partial_\theta q_\alpha(\theta)$. If we take $m = n$, and $k = 1$, the estimator yields the most accurate point estimate for $\partial_\theta q_\alpha(\theta)$, though no statistical assessment on the quality of the estimator can be made. Therefore, the computational budget is typically divided into parts, namely $n = mk$, where m is the number of realizations of Z assigned to an estimator and k is the number of independent replications the sensitivity estimator is attained from. Letting m and k tend to infinity simultaneously, the limit with respect to m procures an arbitrarily close approximation to $\partial_\theta q_\alpha(\theta)$ and the limit with respect to k allows for constructing confidence intervals for the estimator. Details will be provided in the statistical analysis sections of this chapter.

Now suppose that F_θ is not known or computationally intractable. In this case we will split the estimation process into two parts. The first part will use i.i.d. replications of the spacing variable on the LHS of Equation (4.1), which from Pyke's result yields a strongly consistent estimator for the value for the quantile density function $1/f_\theta(q_\alpha(\theta))$. Alongside with estimating the quantile density we use $Z_{\lceil \alpha m\rceil:m}$ as an estimator for $q_\alpha(\theta)$. In the second part of the procedure, we will use an MVD estimator to estimate $\partial_\theta F_\theta(q_\alpha(\theta))$. Typically, the distribution of Z is attained from the measure-valued derivative of the input random vector $X = (X_1,\dots,X_J)$, and $Z = h(X)$. In this case, we will resort to the product rule of weak differentiation [52] to derive unbiased derivative estimator(s) for $\partial_\theta F_\theta$ at $q_\alpha(\theta)$. Instead, if θ, for instance, is only a parameter of the distribution X_1, denoted by $F_{\theta,1}$, and h as well as X_2 to X_J are independent of θ, then, provided that $F_{\theta,1}$ is weakly differentiable with weak derivative $(c_\theta, F^+_{\theta,1}, F^-_{\theta,1})$, it holds under fairly general conditions that

$$\frac{\partial}{\partial\theta}F_\theta(z) = c_\theta \mathbb{E}_\theta\left[1\left\{h(X_1^+,X_2,\dots,X_J)\le z\right\} - 1\left\{h(X_1^-,X_2,\dots,X_J)\le z\right\}\right],$$

in which $X_1^\pm$ are distributed according to $F^\pm_{\theta,1}$. Let $Z^\pm = h(X_1^\pm,X_2,\dots,X_J)$ be the weak derivatives of Z. The overall derivative estimator is then either the expression

$$\begin{aligned}\frac{\partial}{\partial\theta}q_\alpha(\theta) \;=\; \lim_{m\to\infty}\Big(c_\theta\mathbb{E}_\theta\big[-m\big(Z_{\lceil\alpha m\rceil:m} - Z_{\lceil\alpha m\rceil-1:m}\big)\\ \cdot\big(1\{Z^+\le Z_{\lceil\alpha m\rceil:m}\} - 1\{Z^-\le Z_{\lceil\alpha m\rceil:m}\}\big)\big]\Big),\end{aligned} \tag{4.6}$$

or, alternatively,

$$\frac{\partial}{\partial\theta} q_\alpha(\theta) = \lim_{m\to\infty} \Big(c_\theta \mathbb{E}_\theta\big[-m\big(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil-1:m}\big)\big] \\ \times \mathbb{E}_\theta\big[\, 1\{Z^+ \le Z_{\lceil \alpha m\rceil:m}\} - 1\{Z^- \le Z_{\lceil \alpha m\rceil:m}\}\big]\Big). \tag{4.7}$$

The estimators in Equation (4.6) and Equation (4.7) are evaluated only from the realizations of the random variable $\mathbf{Z} = (Z_i : 1 \le i \le m)$ and the corresponding sequences for the weak derivatives. The drawback of this method is that the rate of convergence has least upper bound of $O(n^{-1/3})$ given a computational budget of n, asymptotically reached from below. The result is presented in Section 4.4.3. In the computation of (4.6) and (4.7), we split up the simulation budget in computing the order statistics sequence and the generation of the weak derivatives. Specifically, $m \times k$ observations are used for obtaining k i.i.d. replications of the spacing estimator for the quantile density function as well as the quantile estimator, and l i.i.d. samples of $Z^\pm$ for estimating the distributional derivative.

Remark 4.1. *Rather than MVD, the Finite Difference (FD) method, Section 1.2.1.3 is a general propose approach to gradient estimation that can be applied to attain the parameter derivative $\partial_\theta F_\theta(q_\alpha(\theta))$. Although easy to use it has the main drawback of producing biased estimates and the constant to attain the optimal size as a function of the number of realizations requires supplemental analysis. Consequently, FD is not considered as a competitive method in the gradient estimation literature.*

4.3 Statistical Analysis

In this section we provide the analysis of the statistical properties of our estimator. We will first provide an analysis for the estimator in Equation (4.3) in Section 4.3.1. In Section 4.3.2 we discuss possible extensions.

4.3.1 Main Analysis

Below are the main assumptions required for our analysis:

(C1) $\mathbf{Z} = (Z_i)_{i=1}^m$ is a sequence of i.i.d. continuous random variables taking values in some state space S, where $S \subset \mathbb{R}$, with distribution function F_θ, density function f_θ, such that $\mathbb{E}_\theta[Z_1^{2+\delta}] < \infty$ for some $\delta > 0$.

(C2) There exists an open neighbourhood $B(\alpha)$ of $q_\alpha(\theta)$, such that

(i) the density $f_\theta(x)$ is continuous and strictly positive within this neighbourhood, i.e.,

$$\inf_{x \in B(\alpha)} f_\theta(x) > 0,$$

(ii) the derivative of the density $f_\theta(x)$ w.r.t. x is bounded on $B(\alpha)$, i.e.,

$$\sup_{x \in B(\alpha)} \left| \frac{\partial}{\partial x} f_\theta(x) \right| < \infty.$$

(C3) The distribution function $F_\theta(x)$ is differentiable w.r.t. θ, and the derivative is continuous as a mapping of x.

The first statistical property of the derivative estimator that we derive is asymptotic unbiasedness. For the proof we use the fact that uniform convergence, proceeding from the uniform integrability of spacings, Appendix A.1, together with convergence in distribution implies convergence in the mean. For the proof we need the following property of the derivative of the distribution function:

(C4) It holds that

$$\sup_{x \in \mathbb{R}} \left| \frac{\partial}{\partial \theta} F_\theta(x) \right| < \infty.$$

We note that if the support of f_θ is $S \subset \mathbb{R}$, we extend f_θ onto $\mathbb{R}$ by stating that $f_\theta(x) = 0$ on the complement set S^c.

As will show in Lemma 4.4 in the section on distributional differentiation, Assumption **(C4)** holds under rather weak conditions on F_θ. We now state the first main theorem.

Theorem 4.1. *Let* **(C1)** *to* **(C4)** *hold, then*

$$\lim_{m \to \infty} \mathbb{E}_\theta \left[-m \big(Z_{\lceil \alpha m \rceil : m} - Z_{\lceil \alpha m \rceil - 1 : m} \big) \frac{\partial}{\partial \theta} F_\theta(Z_{\lceil \alpha m \rceil : m}) \right] = \frac{\partial}{\partial \theta} q_\alpha(\theta).$$

Proof: By **(C3)** it follows that

$$\mathbb{E}_\theta \left[m^2 \big(Z_{\lceil \alpha m \rceil : m} - Z_{\lceil \alpha m \rceil - 1 : m} \big)^2 \left(\frac{\partial}{\partial \theta} F_\theta(Z_{\lceil \alpha m \rceil : m}) \right)^2 \right]$$

is well defined. Using the fact that $(\partial_\theta F_\theta(Z_{\lceil \alpha m \rceil : m}))^2$ is bounded from Assumption **(C4)**), we may, by Assumption **(C1)** together with **(C2)**, argue similarly as in the proof of Lemma A.2 to show that

$$\sup_m \left| \mathbb{E}_\theta \left[m^2 \big(Z_{\lceil \alpha m \rceil : m} - Z_{\lceil \alpha m \rceil - 1 : m} \big)^2 \left(\frac{\partial}{\partial \theta} F_\theta(Z_{\lceil \alpha m \rceil : m}) \right)^2 \right] \right| < \infty.$$

This implies uniform integrability of the sequence $(m(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil:m}) \times \partial_\theta F_\theta(Z_{\lceil \alpha m\rceil:m}))$. By Assumption **(C1)**, $\partial_\theta F_\theta(Z_{\lceil \alpha m\rceil:m})$ converges almost surely toward the deterministic value $\partial_\theta F_\theta(q_\alpha(\theta))$. From Slutsky's Theorem it then follows that $m(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil:m})\partial_\theta F_\theta(Z_{\lceil \alpha m\rceil:m})$ converges weakly. This together with the fact that this sequence is uniformly integrable, establishes convergence in the mean. □

Let

$$D(m) = -m\big(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil-1:m}\big)\frac{\partial}{\partial\theta}F_\theta(Z_{\lceil \alpha m\rceil:m})$$

and denote by $D_i(m)$, for $1 \le i \le k$, a realization of $D(m)$. Then, $D(m)$ is the estimator for $\partial_\theta q_\alpha(\theta)$ introduced in Equation (4.3), and we denote by $D^\alpha(m,k)$ the sample average over k realizations $D_i(m)$, $1 \le i \le k$, based on an i.i.d. sample of m random variables $\mathbf{Z}$:

$$D^\alpha(m,k) = \frac{1}{k}\sum_{i=1}^{k} D_i(m).$$

By asymptotic unbiasedness determined in Theorem 4.1, letting k first tend to infinity and then m, yields

$$\lim_{m\to\infty}\lim_{k\to\infty} D^\alpha(m,k) = \frac{\partial}{\partial\theta}q_\alpha(\theta),$$

which establishes strong consistency of the estimator for a particular, and impractical, order of limit. The next lemma proves that $D^\alpha(m,k)$ is a strongly consistent estimator for $\partial_\theta q_\alpha(\theta)$ as $m := m_k$ tends to infinity as k approaches infinity.

Lemma 4.1. *Suppose Assumptions* **(C1)** *to* **(C4)** *hold. Then for any nondecreasing sequence* $(m_k : k \ge 1)$ *such that* m_k *tends to infinity as* k *tends to infinity*

$$\lim_{k\to\infty} D^\alpha(m_k,k) = \frac{\partial}{\partial\theta}q_\alpha(\theta),$$

with probability one.

Proof: The proof is an application of the Strong Law of Large Numbers result put forward in Theorem 6.6 in [82], which covers the arrangement of a sequence of random variables as a triangular array in the result. This theorem states that if $(X_i(k) : k, i \ge 1)$ is a collection of independent zero mean random variables, where for each k, $X_i(k)$ are identically distributed for $i \geq 1$, and $(a_k : k \geq 1)$ is a sequence such that $a_k \uparrow \infty$ as $k \to \infty$, then, provided that

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}\left[|X_i(k)|^p\right]}{a_k^p} < \infty$$

for some $1 \leq p \leq 2$, the expression

$$\frac{1}{a_k} \sum_{i=1}^{k} X_i(k) \overset{a.s.}{\to} 0.$$

In the following we will show that the conditions of the above theorem are met. Let (m_k) be a nondecreasing sequence of integers dependent on k such that $m_k \to \infty$ as $k \to \infty$. For each $k \geq 1$, let $(D_i(m_k) : 1 \leq i \leq k)$ be a sequence of independent random variables depicting our per sample estimators. We let $\mu(m_k) = \mathbb{E}[D_1(m_k)]$ and apply the above limit result to the sequence $X_i(k) = D_i(m_k) - \mu(m_k)$. By construction $\mathbb{E}[D_1(m_k) - \mu(m_k)] = 0$. We set $a_k = k$, $p = 1+\gamma$, in which $a_k^p = k^{1+\gamma}$.

To simplify the notation, we let in the following $m = m_k$. Using the Minkowski inequality, it holds

$$\left(\mathbb{E}_\theta\left[\left(|D_1(m)| + |\mu(m)|\right)^{1+\gamma}\right]\right)^{\frac{1}{1+\gamma}} \leq \mathbb{E}_\theta^{\frac{1}{1+\gamma}}\left[|D_1(m)|^{1+\gamma}\right] + |\mu(m)|,$$

which implies

$$\begin{aligned}\mathbb{E}_\theta\left[\left|D_1(m) - \mu(m)\right|^{1+\gamma}\right] &\leq \mathbb{E}\left[\left(|D_1(m)| + |\mu(m)|\right)^{1+\gamma}\right] \\ &\leq \left(\mathbb{E}^{\frac{1}{1+\gamma}}\left[|D_1(m)|^{1+\gamma}\right] + |\mu(m)|\right)^{1+\gamma}. \end{aligned} \tag{4.8}$$

We define $A_1 := \sup_{y\in\mathbb{R}} |\partial_\theta F_\theta(y)|$ as the finite supremum of the θ-derivative of the distribution function, and c_m as given in Lemma A.2. Using the transformation from Equation (A.2) from the same Appendix, for m sufficiently large we obtain

$$Z_{\lceil\alpha m\rceil:m} - Z_{\lceil\alpha m\rceil-1:m} \leq \frac{1}{c_m}\left(U_{\lceil\alpha m\rceil:m} - U_{\lceil\alpha m\rceil-1:m}\right).$$

and consequently,

$$\begin{aligned}|\mu(m)| &= \left|\mathbb{E}\left[-m\left(Z_{\lceil\alpha m\rceil:m} - Z_{\lceil\alpha m\rceil-1:m}\right)\frac{\partial}{\partial\theta}F_\theta(Z_{\lceil\alpha m\rceil:m})\right]\right| \\ &\leq A_1\left|\mathbb{E}\left[m\left(Z_{\lceil\alpha m\rceil:m} - Z_{\lceil\alpha m\rceil-1:m}\right)\right]\right| \\ &\leq \frac{A_1}{c_m}\mathbb{E}\left[m\left(U_{\lceil\alpha m\rceil:m} - U_{\lceil\alpha m\rceil-1:m}\right)\right] \leq \frac{A_1}{c_m},\end{aligned}$$

where we use Equation (A.8) from Appendix A.2. By the same arguments, for m sufficiently large we have

$$\mathbb{E}^{\frac{1}{1+\gamma}}\left[|D_1(m)|^{1+\gamma}\right] \leq \frac{A_1}{c_m}\,\mathbb{E}^{\frac{1}{1+\gamma}}\left[m^{1+\gamma}\big(U_{\lceil\alpha m\rceil:m} - U_{\lceil\alpha m\rceil-1:m}\big)^{1+\gamma}\right].$$

For $0 < \gamma \leq 1$, the Lyapanov inequality implies

$$\mathbb{E}^{\frac{1}{1+\gamma}}\left[m^{1+\gamma}\big(U_{\lceil\alpha m\rceil:m} - U_{\lceil\alpha m\rceil-1:m}\big)^{1+\gamma}\right] \leq \mathbb{E}^{\frac{1}{2}}\left[m^2\left(U_{\lceil\alpha m\rceil:m} - U_{\lceil\alpha m\rceil-1:m}\right)^2\right] \leq 2^{\frac{1}{2}},$$

since

$$\mathbb{E}\left[m^2\big(U_{\lceil\alpha m\rceil:m} - U_{\lceil\alpha m\rceil-1:m}\big)^2\right] = \frac{2m^2}{(m+1)(m+2)} \leq 2.$$

Note that $\lim_{m\to\infty} c_m = f_\theta(q_\alpha(\theta))$ and we may choose A finite and sufficiently large such that

$$\mathbb{E}\big[|D_1(m) - \mu(m)|^{1+\gamma}\big] \leq \left(2^{\frac{1}{2}}+1\right)^2 \frac{A^2}{f_\theta^2(q_\alpha(\theta))},$$

for all m, and, replacing m by m_k, the series

$$\sum_{k=1}^{\infty} k^{-(1+\gamma)}\,\mathbb{E}\big[|D_1(m_k) - \mu(m_k)|^{1+\gamma}\big] \leq \left(2^{\frac{1}{2}}+1\right)^2 \frac{A^2}{f_\theta^2(q_\alpha(\theta))}\sum_{k=1}^{\infty}\frac{1}{k^{1+\gamma}} < \infty$$

as the harmonic series is convergent for $\gamma > 0$. Hence, we have proved that

$$\frac{1}{k}\sum_{i=1}^{k} X_i(k) = \frac{1}{k}\sum_{i=1}^{k}\big(D_i(m_k) - \mu(m_k)\big) = D^\alpha(m_k,k) - \mu(m_k) \overset{a.s.}{\to} 0$$

as k tends to infinity. Since, by Theorem 4.1, $\lim_{k\to\infty}\mu(m_k) = \partial_\theta q_\alpha(\theta)$, we conclude from

$$D^\alpha(m_k,k) - \frac{\partial}{\partial\theta}q_\alpha(\theta) = \big(D^\alpha(m_k,k) - \mu(m_k)\big) + \left(\mu(m_k) - \frac{\partial}{\partial\theta}q_\alpha(\theta)\right)$$

that

$$\lim_{k\to\infty} D^\alpha(m_k,k) - \frac{\partial}{\partial\theta}q_\alpha(\theta) \overset{a.s.}{\to} 0,$$

for any sequence m_k that tends to infinity for k to infinity, proving strong consistency. □

We can delve further in our asymptotic result, Theorem 4.1, and determine the extent of biasedness for finite samples between our estimator and the quantile sensitivity. This result is also needed for the Central Limit Theorem, which will be provided in Theorem 4.3 later on.

Theorem 4.2. *Under Assumptions* **(C1)** *to* **(C3)** *it holds that for any* $k \geq 1$ *that*

$$\left|\mathbb{E}_\theta\left[D^\alpha(m,k) - \frac{\partial}{\partial\theta}q_\alpha(\theta)\right]\right| = O\left(\frac{1}{m}\right).$$

Proof: To simplify the notation, let $Y_1 = Z_{\lceil\alpha m\rceil-1:m}$, $Y_2 = Z_{\lceil\alpha m\rceil:m}$, and

$$W_1 = Y_2 - Y_1 \quad \text{and} \quad W_2 = F_\theta^+(Y_2) - F_\theta^-(Y_2).$$

With this notation, our estimator reads

$$\mathbb{E}_\theta\big[D^\alpha(m,k)\big] = \mathbb{E}_\theta[-mc_\theta W_1 W_2]\,. \tag{4.9}$$

We now expand W_1, W_2 via a Taylor polynomial approximation, heeding the method in [18], pp. 84.

We first deal with W_1. We have assumed that **(C1)** and **(C2)** hold, and we may take $r_m = (2/f_\theta(q_\alpha(\theta)))\cdot(\ln(m)/m)^{1/2}$ from Lemma A.1, so that for m sufficiently large there exists a neighbourhood $B_{r_m}(\alpha)$ of $q_\alpha(\theta)$ such that $B_{r_m}(\alpha) \subset B(\alpha)$ and the bound in Equation (A.1) applies. By means of the mapping $F_\theta(x)$, a set $B \subset S$ corresponds to a pre-image B^{-1} on the unit interval, and we define

$$B^{-1} = \{F_\theta(x) : x \in B\} \subset [0,1].$$

By Assumption **(C2)** it now holds that

$$c_{1,m} := \sup_{y\in(B_{r_m}(\alpha))^{-1}} \left|\frac{\partial^2}{\partial y^2}F_\theta^{-1}(y)\right| = \sup_{y\in B_{r_m}(\alpha)}\left|-\frac{1}{f_\theta^3(y)}\frac{\partial}{\partial y}f_\theta(y)\right| < \infty,$$

in which we have used the fact that $\partial_x F_\theta^{-1}(x) = 1/f_\theta(F_\theta^{-1}(x))$, which follows from the fact that $F_\theta(x)$ is monotone and continuous. Let $(U_{l:m} : 1 \leq l \leq m)$ be the order statistic sequence of i.i.d. uniform $(0,1)$ random variables. Writing $Y_i = F_\theta^-(T_i)$, for $i = 1,2$, where $T_1 = U_{\lceil\alpha m\rceil-1:m}$, $T_2 = U_{\lceil\alpha m\rceil:m}$, we can expand Y_i around α as follows:

$$Y_i = F_\theta^{-1}(\alpha) + \frac{\partial}{\partial x}F_\theta^{-1}(\alpha)(T_i - \alpha) + \frac{1}{2}\frac{\partial^2}{\partial x^2}F_\theta^{-1}(\xi_i)(T_{i,m} - \alpha)^2,$$

with $\xi_i \in (B_{r_m}(\alpha))^{-1}$, for $i = 1,2$. As the sequence of sets $(B_{r_m}(\alpha))^{-1}$ is decreasing with limit $\{\alpha\}$ this implies, as $m \to \infty$

$$\lim_{m\to\infty} c_{1,m} := c_1 = \frac{1}{f_\theta^3(q_\alpha(\theta))}\left|\frac{\partial}{\partial y}f_\theta(q_\alpha(\theta))\right|.$$

Hence, for m sufficiently large

$$\begin{aligned} W_1 &\approx \left(F_\theta^{-1}(\alpha) + \frac{\partial}{\partial x}F_\theta^{-1}(\alpha)(T_2-\alpha) + c_1(T_2-\alpha)^2\right) \\ &\quad - \left(F_\theta^{-1}(\alpha) + \frac{\partial}{\partial x}F_\theta^{-1}(\alpha)(T_{1,m}-\alpha) + c_1(T_1-\alpha)^2\right) \\ &= \frac{\partial}{\partial x}F_\theta^{-1}(\alpha)(T_{2,m}-T_{1,m}) + c_1(T_{2,m}-T_{1,m})(T_{2,m}+T_{1,m}-2\alpha). \end{aligned}$$

We now turn to the term W_2. We use a linear Taylor expansion with a Lagrangian remainder around $x = q_\alpha(\theta)$, providing

$$\begin{aligned} F_\theta^+(Y_2) &= F_\theta^+(q_\alpha(\theta)) + \frac{\partial}{\partial x}F_\theta^+(\eta_1)(Y_2 - q_\alpha(\theta)) \\ &= F_\theta^+(q_\alpha(\theta)) + f_\theta^+(\eta_1)(Y_2 - q_\alpha(\theta)), \end{aligned}$$

for $\eta_1 \in B_{r_m}(\alpha)$, and $f_\theta^+(x)$ denoting the density of $F_\theta^+(x)$. In the same vein

$$F_\theta^-(Y_2) = F_\theta^-(q_\alpha(\theta)) + f_\theta^-(\eta_2)(Y_2 - q_\alpha(\theta)),$$

for $\eta_2 \in B_{r_m}(\alpha)$, and $f_\theta^-(x)$ denoting the density of $F_\theta^-(x)$. Letting

$$c_{2,m} = \sup_{y\in B_{r_m}(\alpha)} \max\{f_\theta^+(y), f_\theta^-(y)\} < \infty,$$

it follows that $(B_{r_m}(\alpha))$, $m \in \mathbb{N}$, is also a decreasing sequence with limit $\{q_\alpha(\theta)\}$, that $c_{2,m}$ converges to $c_2 := \max\{f_\theta^+(q_\alpha(\theta)), f_\theta^-(q_\alpha(\theta))\}$. Hence, for m sufficiently large

$$W_2 \approx F_\theta^+(q_\alpha(\theta)) - F_\theta^-(q_\alpha(\theta)) + c_2(Y_2 - \alpha).$$

We define

$$c_{3,m} = \sup_{y\in(B_{r_m}(\alpha))^{-1}} \left|\frac{\partial}{\partial y}F_\theta^{-1}(y)\right| = \sup_{y\in(B_{r_m}(\alpha))^{-1}} \frac{1}{f_\theta(F_\theta^{-1}(y))},$$

in which we have used the result $\partial_x F_\theta^{-1}(x) = 1/f_\theta(F_\theta^{-1}(x))$ again. Utilizing a Taylor series approximation, for m sufficiently large Y_2 is approximated simply to

$$Y_2 - q_\alpha(\theta) \approx c_3(T_2 - \alpha).$$

Since the sequence $(B_{r_m}(\alpha))^{-1}$ monotone decreases to the atom set $\{\alpha\}$ as $m \to \infty$, this implies

$$\lim_{m\to\infty} c_{3,m} := c_3 = \frac{1}{f_\theta(q_\alpha(\theta))}.$$

Note that by Assumption **(C2)** it holds that $c_2, c_3 < \infty$. Inserting the above approximation for $Y_2 - q_\alpha(\theta)$ into the Taylor polynomial for W_2 we arrive at

$$W_{2,m} \approx F_\theta^+(q_\alpha(\theta)) - F_\theta^-(q_\alpha(\theta)) + c_2 c_3 (T_2 - \alpha),$$

for m sufficiently large. The previously introduced constants c_l, $l = 1,2,3$, are only important in the sense that they represent a finite limit.

From the definition for the parameter derivative of the quantile, Equation (4.1), in combination with the formula $\partial_\theta F_\theta(q_\alpha(\theta)) = c_\theta(F_\theta^+(q_\alpha(\theta)) - F_\theta^-(q_\alpha(\theta)))$, we obtain that

$$\frac{\partial}{\partial\theta} q_\alpha(\theta) = c_\theta \left(F_\theta^+(q_\alpha(\theta)) - F_\theta^-(q_\alpha(\theta)\right) \frac{\partial}{\partial x} F_\theta^{-1}(\alpha).$$

We again make use of the observation $\partial_x F_\theta^{-1}(\alpha) = f_\theta(F_\theta^{-1}(\alpha)) = f_\theta(q_\alpha(\theta))$. Collecting the above approximations and inserting them into Equation (4.9) yields

$$\begin{aligned}
\mathbb{E}_\theta[D^\alpha(m,k)] \approx\, & m q_\alpha(\theta)\, \mathbb{E}[T_2 - T_1] \\
& - m \left(\frac{\partial}{\partial x} F_\theta(q_\alpha(\theta))\right) c_1\, \mathbb{E}[(T_2 - T_1)(T_2 + T_1 - 2\alpha)] \\
& - m c_\theta \left(\frac{\partial}{\partial x} F_\theta^{-1}(\alpha)\right) c_2 c_3\, \mathbb{E}[(T_2 - T_1)(T_2 - \alpha)] \\
& - m c_\theta c_1 c_2 c_3\, \mathbb{E}[(T_2 - T_1)(T_2 - \alpha)(T_2 + T_1 - 2\alpha)]\,.
\end{aligned}$$

Since $\mathbb{E}[T_{2,m} - T_{1,m}] = 1/(m+1)$, Equation (A.8) from Appendix A.2,

$$m \frac{\partial}{\partial\theta} q_\alpha(\theta)\, \mathbb{E}[T_2 - T_1] \approx \frac{\partial}{\partial\theta} q_\alpha(\theta),$$

for m sufficiently large. We now attain

$$\begin{aligned}
\mathbb{E}_\theta[D^\alpha(m,k)] - \frac{\partial}{\partial\theta} q_\alpha(\theta) \approx & - m \left(\frac{\partial}{\partial\theta} F_\theta(q_\alpha(\theta))\right) c_1\, \mathbb{E}[(T_2 - T_1)(T_2 + T_1 - 2\alpha)] \\
& - m c_\theta \left(\frac{\partial}{\partial\theta} F_\theta^{-1}(\alpha)\right) c_2 c_3\, \mathbb{E}[(T_2 - T_1)(T_2 - \alpha)] \\
& - m c_\theta c_1 c_2 c_3\, \mathbb{E}[(T_2 - T_1)(T_2 - \alpha)(T_2 + T_1 - 2\alpha)]\,. \quad (4.10)
\end{aligned}$$

The results from the remaining uniform order statistics expectations are provided in Equations (A.11), (A.12), and (A.15) via the Maple 15 programming package. Inserting these expectations into Equation (4.10) will show that $\mathbb{E}_\theta[D^\alpha(m,k)] - \partial_\theta q_\alpha(\theta)$ behaves for sufficiently large m as a sum of terms that are at most of order $O(m^{-1})$, proving the claim. □

For the Central Limit Theorem result of $D^\alpha(m,k)$, we require the result from Lyapanov, [96], that is sufficiently flexible to deal with a sequence of random variables in the form of the triangular array. The advantage of utilizing this Central Limit Theorem is that the only additional condition, compared to a sequence of i.i.d. random variables, is the finiteness of $E_\theta[Z^{2+\delta}]$ for some $\delta > 0$, Assumption **(C1)**.

Theorem 4.3 (Central Limit Theorem)**.** *Suppose that Assumptions* **(C1)** *to* **(C4)** *holds for any nondecreasing sequence* $\{m_k : k \geq 1\}$ *such that* $k^{1/2}/m_k \to 0$ *as* $(m_k, k) \to (\infty,\infty)$. *Then*

$$\frac{D^\alpha(m_k,k) - \frac{\partial}{\partial\theta}q_\alpha(\theta)}{(\mathrm{Var}_\theta(D^\alpha(m_k,k)))^{1/2}} \xrightarrow{d} N(0,1). \tag{4.11}$$

where $N(0,1)$ *is the standard normal distribution.*

Proof: The left-hand side of (4.11) can be written as

$$\frac{D^\alpha(m_k,k) - \frac{\partial}{\partial\theta}q_\alpha(\theta)}{(\mathrm{Var}_\theta(D^\alpha(m_k,k)))^{1/2}} = \frac{D^\alpha(m_k,k) - \mathbb{E}_\theta[D^\alpha(m_k,k)]}{(\mathrm{Var}_\theta(D^\alpha(m_k,k)))^{1/2}} + \frac{\mathbb{E}_\theta[D^\alpha(m_k,k)] - \frac{\partial}{\partial\theta}q_\alpha(\theta)}{(\mathrm{Var}_\theta(D^\alpha(m_k,k)))^{1/2}}. \tag{4.12}$$

Demonstration of the convergence to the limiting distribution, given the assumptions, occurs in two parts. Firstly, need to prove that the first term on the RHS in Equation (4.12), for any nondecreasing sequence $\{m_k\}$, converges to the standard normal distribution as both $k \to \infty$ and $m_k \to \infty$. Secondly, given the additional requirement $k^{1/2}/m_k \to 0$ as both k, $m_k \to \infty$, the second terms tends to zero almost surely. Once these two results are derived, the proof is completed via Slutsky's Theorem.

Part(i): Let $k \geq 1$ and $\{m_k : k \geq 1\}$ be a nondecreasing sequence of integers such that as $k \to \infty$, $m_k \to \infty$. To simplify notation, we define $m := m_k$. For each $k \geq 1$ let $(D_i(m), 1 \leq i \leq k)$ be a sequence of independent random variables depicting our per sample estimators. Let $\mu(m) = \mathbb{E}_\theta[D_1(m)]$, and $\sigma^2(m) = \mathrm{Var}_\theta(D_1(m))$ denote the mean and variance of these estimators, and define $B(k)$ to be the standard deviation of the sum of the per sample estimators:

$$B(k) = \left(\sum_{i=1}^{k} \sigma^2(m)\right)^{1/2} = \sqrt{k}\sigma(m), \tag{4.13}$$

for $k \geq 1$. Then due to Lemma A.2, $\sigma(m)$ hence $B(k)$ is bounded as a function of m.

In the following we will show that for some $\delta > 0$, the following Lyapanov-type condition is satisfied:

$$\frac{1}{B^{2+\delta}(k)} \mathbb{E}_\theta\left[\sum_{i=1}^{k} \left|D_i(m) - \mu(m)\right|^{2+\delta}\right] \to 0$$

as $k \to \infty$. This in turn will imply the Central Limit Theorem result. By the i.i.d. assumption together with Equation (4.13), this is equivalent to

$$\frac{1}{k^{1+\frac{\delta}{2}}} \mathbb{E}_\theta\left[\left|D_1(m) - \mu(m)\right|^{2+\delta}\right] \to 0,$$

in which this condition is satisfied if that for all m

$$\mathbb{E}_\theta\left[\left|D_1(m) - \mu(m)\right|^{2+\delta}\right] < C, \tag{4.14}$$

for some finite constant C. Following the line of argument leading to Equation (4.8), using the Minkowski inequality, the above expectation is bounded above by

$$\mathbb{E}_\theta\left[\left|D_1(m) - \mu(m)\right|^{2+\delta}\right] \leq \left(\mathbb{E}_\theta^{\frac{1}{2+\delta}}\left[|D_1(m)|^{2+\delta}\right] + |\mu(m)|\right)^{2+\delta}.$$

Therefore, we only need to show that the $2+\delta$ central moment of $D_1(m)$ is finite, leading us to the exact condition that Hong in [57] had assumed, and which we will prove for our per sample estimator. By Assumption **(C4)**, the θ-derivative of the distribution function is bounded, i.e., $\sup_{x\in\mathbb{R}} |\partial_\theta F_\theta(x)| \leq A_1 < \infty$, which yields:

$$\mathbb{E}_\theta\left[|D_1(m)|^{2+\delta}\right] \leq \mathbb{E}_\theta\left[A_1^{2+\delta} m^{2+\delta} \left(Z_{\lceil\alpha m\rceil:m} - Z_{\lceil\alpha m\rceil-1:m}\right)^{2+\delta}\right].$$

In addition, by Assumption **(C2)** the density $f_\theta > 0$ in a neighbourhood $B(\alpha)$ containing the quantile $q_\alpha(\theta)$. Especially, we denote c_m, as defined in Lemma A.2, to be the infimum of the density function in the neighbourhood $B(\alpha)$. Following Equation (A.2) in the same lemma,

$$Z_{\lceil\alpha m\rceil:m} - Z_{\lceil\alpha m\rceil-1:m} \leq \frac{1}{c_m}\left(U_{\lceil\alpha m\rceil:m} - U_{\lceil\alpha m\rceil-1:m}\right).$$

and so

$$\mathbb{E}_\theta^{\frac{1}{2+\delta}}\left[|D_i(m)|^{2+\delta}\right] \leq \frac{A_1}{c_m} \mathbb{E}^{\frac{1}{2+\delta}}\left[m^{2+\delta}\left(U_{\lceil\alpha m\rceil:m} - U_{\lceil\alpha m\rceil-1:m}\right)^{2+\delta}\right]. \tag{4.15}$$

This uniform spacing is bounded via the Lyapanov inequality, in which

$$\mathbb{E}^{\frac{1}{2+\delta}}\left[m^{2+\delta}\left(U_{\lceil\alpha m\rceil:m}-U_{\lceil\alpha m\rceil-1:m}\right)^{2+\delta}\right] \le \mathbb{E}^{\frac{1}{3}}\left[m^{3}\left(U_{\lceil\alpha m\rceil:m}-U_{\lceil\alpha m\rceil-1:m}\right)^{3}\right] \le 6^{\frac{1}{3}}.$$

This last inequality is obtained by computation, where

$$\mathbb{E}\left[m^{3}\left(U_{\lceil\alpha m\rceil:m}-U_{\lceil\alpha m\rceil-1:m}\right)^{3}\right] = \frac{6m^3}{(m+1)(m+2)(m+3)} \le 6.$$

The insertion of this upper bound into (4.15) then yields

$$\mathbb{E}_\theta^{\frac{1}{2+\delta}}\left[|D_1(m)|^{2+\delta}\right] \le 6^{\frac{1}{3}}\frac{A_1}{c_m}.$$

Since $c_m \to f_\theta(q_\alpha(\theta))$ as $m\to\infty$, together with the assumption $\mathbb{E}_\theta[Z^{2+\delta}]<\infty$, there exists a constant $A<\infty$ such that the $2+\delta$ moment of our per sample estimator is bounded by

$$\mathbb{E}_\theta^{\frac{1}{2+\delta}}\left[|D_1(m)|^{2+\delta}\right] \le 6^{\frac{1}{3}}\frac{A}{f_\theta(q_\alpha(\theta))}.$$

From this upper bound together with the result that $\lim_{m\to\infty}\mu(m) = \partial_\theta q_\alpha(\theta)$ from Theorem 4.1, we obtain Equation (4.14). Recalling $m := m_k$, this concludes the proof of the Central Limit Theorem for the first term of the RHS in (4.12), i.e., the expression, as $k\to\infty$:

$$\frac{D^\alpha(m_k,k)-\mathbb{E}_\theta[D^\alpha(m_k,k)]}{(\mathrm{Var}_\theta(D^\alpha(m_k,k)))^{1/2}} \xrightarrow{d} N(0,1).$$

converges weakly.

Part(ii): For the second term on the RHS of Equation (4.12), the variance of the estimator satisfies

$$\mathrm{Var}_\theta\left(D^\alpha(m_k,k)\right) = \frac{1}{k}\mathrm{Var}_\theta\left(D_1(m_k)\right).$$

By Theorem 4.2, $|\mathbb{E}_\theta[D^\alpha(m_k,k)]-q_\alpha(\theta)| = O(m_k^{-1})$, which means that $|\mathbb{E}_\theta[D^\alpha(m_k,k)]-q_\alpha(\theta)|$ behaves asymptotically like L/m_k for some finite constant L. From these two observations, the second term is bounded above by

$$\left|\frac{\mathbb{E}_\theta[D^\alpha(m_k,k)]-\frac{\partial}{\partial\theta}q_\alpha(\theta)}{(\mathrm{Var}_\theta(D^\alpha(m_k,k)))^{1/2}}\right| \le \frac{L}{(\mathrm{Var}_\theta(D_1(m_k)))^{1/2}}\frac{k^{1/2}}{m_k}. \tag{4.16}$$

Given our requirement that $k^{1/2}/m_k \to 0$ as $k, m_k \to \infty$, Equation (4.16) tends to zero and our proof is complete. □

In order to construct confidence intervals for $\partial_\theta q_\alpha(\theta)$, the variance of the estimator $D^\alpha(m,k)$ has to be estimated as well. Let

$$S^\alpha(m,k) = \frac{1}{k-1}\left(\sum_{i=1}^{k}(D_i(m))^2 - \frac{1}{k}\left(\sum_{i=1}^{k}D_i(m)\right)^2\right)$$

denote our estimator for $\mathrm{Var}_\theta(D_i(m))$. The following lemma provides a sufficient condition for strong consistency of the above estimator.

Lemma 4.2. *Suppose Assumption* **(C1)** *holds, then*

$$\lim_{k\to\infty} S^\alpha(m,k) = \mathrm{Var}_\theta(D_i(m))$$

with probability one.

Proof: This proof is another application of the Strong Law of Large Numbers (SLLN) result used in Lemma 4.1,[82]. We begin with, simplifying the notation with $m := m_k$:

$$S^\alpha(m,k) = \frac{k}{k-1}\left(\frac{1}{k}\sum_{i=1}^{k}(D_i(m))^2 - \frac{1}{k^2}\left(\sum_{i=1}^{k}D_i(m)\right)^2\right). \tag{4.17}$$

From Theorem 4.3, $\mathbb{E}_\theta[(D^\alpha(m,k))^{2+\delta}] < \infty$ for some $\delta > 0$ and all m. Combining this result, with the SLLN and the Continuity Mapping Theorem, where we use the mapping $g(x) = x^2$, we attain

$$\lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}(D_i(m))^2 = \mathbb{E}_\theta\big[(D^\alpha(m,k))^2\big]$$

and

$$\lim_{k\to\infty}\frac{1}{k^2}\left(\sum_{i=1}^{k}D_i(m)\right)^2 = \mathbb{E}_\theta^2\big[D^\alpha(m,k)\big], \tag{4.18}$$

with probability one. Combining Equation (4.18) with (4.17), and the algebra of limits for almost sure convergence, see, for example, [45], we obtain the result

$$\lim_{k\to\infty} S^\alpha(m,k) = \lim_{k\to\infty}\left(\mathbb{E}_\theta\left[\left(D^\alpha(m,k)\right)^2\right] - \mathbb{E}_\theta^2\left[D^\alpha(m,k)\right]\right) = \mathrm{Var}_\theta(D^\alpha(m,k))$$

with probability one. □

With the above statistical analysis we can construct a two-sided confidence interval for $\partial_\theta q_\alpha(\theta)$, where we assume that the conditions put forward in Theorem 4.3 are adhered. Let $\beta \in (0,1)$ denote the confidence level and let q_γ, $\gamma \in (0,1)$, denote the quantile of the standard t-distribution with $k-1$ degrees of freedom. Then it holds asymptotically that

$$\frac{\partial}{\partial\theta} q_\alpha(\theta) \in \left(D^\alpha(m,k) - \frac{q_{1-\beta/2}}{\sqrt{k}} S^\alpha(m,k), D^\alpha(m,k) + \frac{q_{1-\beta/2}}{\sqrt{k}} S^\alpha(m,k) \right)$$

with probability of at least $1-\beta$.

Given k samples of i.i.d. sequences of length m, the computational budget n for the derivative estimator in either Equation (4.4) is $n = km$. If we increase m at the expense of k, we reduce the bias and conversely, if k increases, we reduce the variance of our estimator. Following the prescription of Hong, [57], we denote $m = n^{1/3+\delta}$ and $k = n^{2/3-\delta}$. Since $k^{1/2}/m \to 0$ as $m, k \to \infty$ from Theorem 4.3, the parameter δ is constrained in the interval $0 < \delta < 2/3$. If $m = k = n^{1/2}$ ($\delta = 1/4$), the rate of convergence of this derivative estimator is $O(n^{-1/4})$ with bias $O(n^{-1/2})$. The maximum rate of convergence for the derivative estimator when $\partial_\theta F_\theta$ is analytically tractable is when $\delta \downarrow 0$ and the rate of convergence is asymptotically below $O(n^{-1/3})$, equally to the order statistic derivative estimator from Hong [57]. The corresponding bias is $O(n^{-1/3})$.

4.3.2 Extensions

The generalized spacing presented by Yu, [108], [109], improves the limiting spacing result from Pyke, [86], Equation (4.2), by improving the mode of convergence to almost sure convergence. This generalization, as a nearest neighbour estimator, requires a non-decreasing sequence (k_m) such that $k_m/m \to 0$ as m tends to infinity. Given an i.i.d. collection of random variables $\mathbf{Z} = (Z_i : 1 \leq i \leq m)$, and $z \in \mathbb{R}$, the density estimator is based on the difference of the order statistic values lying $2k_m$ steps apart. This density estimator estimates $f_\theta(z)$ in the region z where half of the order statistics is less than the argument and the other half greater. Specifically, let

$$f_{\theta,m}(z) = \begin{cases} \frac{2k_m}{m(Z_{2k_m+j:m} - Z_{1+j:m})} & \text{if } z \in [Z_{k_m+j:m}, Z_{k_m+j+1:m}) \text{ for } j = 0,1,\ldots,m-2k_m; \\ 0 & \text{if } z < Z_{k_m:m} \text{ or } z \geq Z_{m-k_m+1:m}. \end{cases} \tag{4.19}$$

Then it holds that

$$\lim_{m\to\infty} f_{\theta,m}(z) = f_\theta(z), \tag{4.20}$$

with probability one. Intuitively, the almost sure result is equivalent to a Strong Law of Large Numbers result where the expression is a sum of spacings which are approximately independent for large m.

Elaborating on Equation (4.20), a strongly consistent estimator for $\partial_\theta q_\alpha(\theta)$ can be attained. The precise statement is presented in the following lemma.

Lemma 4.3. *Let* **(C1)** *to* **(C4)** *hold. Let $f_{\theta,m}$ be defined as in Equation (4.19) and assume that*

(i) $f_\theta(z)$ is uniformly continuous as a mapping in z on $\mathbb{R}$,

(ii) the sequence (k_m) is such that $\lim_{m\to\infty} k_m/m = 0$ and,

$$\sum_{m=1}^{\infty} e^{-ck_m} < \infty$$

for all $c > 0$, then

$$\lim_{m\to\infty} -\frac{\frac{\partial}{\partial\theta}F_\theta(Z_{\lceil \alpha m\rceil:m})}{f_{\theta,m}(Z_{\lceil \alpha m\rceil:m})} = \frac{\partial}{\partial\theta}q_\alpha(\theta),$$

with probability one.

Proof: The limiting result for the order statistic, $Z_{\lceil \alpha m\rceil:m} \overset{a.s.}{\to} q_\alpha(\theta)$ for i.i.d. random variables, [4], together with continuity of $\partial_\theta F_\theta$ implies that almost surely $\partial_\theta F_\theta(Z_{\lceil \alpha m\rceil:m}) \overset{a.s.}{\to} \partial_\theta F_\theta(q_\alpha(\theta))$ as $m \to \infty$.

As for the denominator, note that

$$\lim_{m\to\infty} f_{\theta,m}(Z_{\lceil \alpha m\rceil:m}) = \lim_{m\to\infty} f_{\theta,m}(q_\alpha(\theta)) + \lim_{m\to\infty} \left(f_{\theta,m}(Z_{\lceil \alpha m\rceil:m}) - f_{\theta,m}(q_\alpha(\theta))\right).$$

Provided that the conditions put forward in the lemma hold, the limit in (4.20) holds uniformly, [110], and with probability one it holds that

$$\lim_{m\to\infty} \sup_z \left| f_{\theta,m}(z) - f_\theta(z)\right| = 0.$$

Consequently, for m sufficiently large, the expression $|f_{\theta,m}(Z_{\lceil \alpha m\rceil:m}) - f_{\theta,m}(q_\alpha(\theta))|$ becomes arbitrarily close to $|f_\theta(Z_{\lceil \alpha m\rceil:m}) - f_\theta(q_\alpha(\theta))|$ and, together with the Mean Value Theorem and Lemma A.1, we conclude that $f_{\theta,m}(Z_{\lceil \alpha m\rceil:m}) - f_{\theta,m}(q_\alpha(\theta)) = o(1)$. Hence,

$$\lim_{m\to\infty} f_{\theta,m}(Z_{\lceil \alpha m\rceil:m}) = f_\theta(q_\alpha(\theta))$$

with probability one, and the proof follows from Slutsky's Theorem. □

Remark 4.2. *The optimal rate of convergence is given in [110], [111], and depends on the further requirements on the density function f_θ. For example, if $f_\theta(x)$ is Lipschitz continuous function w.r.t. x, then the optimal sequence is* $k_m = \lceil m^{4/7}/(\log(m))^{3/7}\rceil$.

While, on the one side, the strongly consistent estimator in Lemma 4.1 is preferable to the estimator in Equation (4.4), which is based on the weak limit in (4.2), there is no result available on the extent of the asymptotic bias, see Theorem 4.2. The extent of asymptotic bias is necessary to establish a Central Limit Theorem for this derivative estimator. In addition, uniform continuity of $f_\theta(x)$ as a function of x as required in Lemma 4.1 is hard to check for complex models.

4.4 Distributional Differentiation

This subsection is a visitation of the Measure Differentiation method needed in this chapter. A more detailed discussion of the theory of the Measure Valued Differentiation method, [83], derivative estimation method is supplied in Section 1.2.2.2. Examples of a measure-valued derivative for the exponential and normal distribution together with a simulation method to generate the measure-valued derivative samples is given respectively in Sections 1.2.4.1 and 1.2.4.2.

4.4.1 Basic Definitions

In terms of a probability measure $\mathbb{P}_\theta$, $\theta \in \Theta$, the distribution function $F_\theta(z) = \mathbb{P}_\theta((-\infty, z])$ can be alternatively written as an expectation of an indicator mapping

$$F_\theta(z) = \int_{\mathbb{R}} 1\{x \leq z\}\, \mathbb{P}_\theta(dx).$$

The term $1\{x \leq z\} \in \mathfrak{B}$, $z \in \mathbb{R}$, denotes the indicator mapping, which is an element of the function space of bounded Borel mappings. In terms of the continuous random variable Z, this probability measure $\mathbb{P}_\theta$ has particular structure. Specifically, if Z is a continuous random variable, the induced probability measure $\mathbb{P}_\theta$ of Z is absolutely continuous w.r.t. the Lebesgue measure λ and $f_\theta = d\mathbb{P}_\theta/d\lambda$ is the Radon-Nikodym derivative.

In this section we revisit the notion of the distributional differentiation w.r.t. a family of density functions $\{f_\theta\}, \theta \in \Theta$, w.r.t. the collection of Borel and bounded functions $\mathfrak{B}$. This section is the first step in establishing the needed weak derivative concepts for the statistical analysis when $\partial_\theta F_\theta$ is expressed via MVD. An introduction into distributional differentiation is presented in Section 1.2.2.

A density function (Radon-Nikodym derivative) f_θ, $\theta \in \Theta$, is $\mathfrak{B}$-differentiable if for all $h \in \mathfrak{B}$:

$$\frac{d}{d\theta}\int_{\mathbb{R}} h(x) f_\theta(x)\, dx = \int_{\mathbb{R}} h(x) f'_\theta(x) dx, \tag{4.21}$$

in which $f'_\theta(x) = \partial_\theta f_\theta(x)$ is a "signed" density function, Equation (1.16). As $\mathfrak{B} \supset \mathfrak{C}_b$, the collection of continuous and bounded functions, f'_θ is unique up to a set of Lebesgue measure zero. The measure-valued derivative rewriting of the above expression arises from the Hahn-Jordan decomposition, [62], with non-unique density functions f^+_θ, f^-_θ and pre-factor $c_\theta > 0$:

$$\frac{d}{d\theta}\int_{\mathbb{R}} h(x) f'_\theta(x)\, dx = c_\theta \left(\int_{\mathbb{R}} h(x) f^+_\theta(x)\, dx - \int_{\mathbb{R}} h(x) f^-_\theta(x)\, dx \right). \tag{4.22}$$

The corresponding measure-valued derivative triple for the distribution function F_θ is then $(c_\theta, F^+_\theta, F^-_\theta)$. In terms of random variables, the equivalent expression to (4.22), with $h \in \mathfrak{B}$, is:

$$\frac{d}{d\theta}\mathbb{E}_\theta[h(Z)] = c_\theta\big(\mathbb{E}_\theta\big[h(Z^+)\big] - \mathbb{E}_\theta[h(Z^-)]\big).$$

The measure-valued derivative triple for the random variable Z is then (c_θ, Z^+, Z^-).

For the statistical analysis in Theorem 4.1 we assumed that $\partial_\theta F_\theta(z)$ is bounded as a mapping of z. The following lemma shows that Assumption **(C4)** is always satisfied provided the corresponding random variable, respectively distribution, is $\mathfrak{B}$-differentiable.

Lemma 4.4. *Let Z, with support $S \subset \mathbb{R}$, denote the random variable associated with the probability measure $\mathbb{P}_\theta$ and let $F_\theta(z) = \mathbb{P}_\theta((-\infty, z])$, $z \in \mathbb{R}$, be the corresponding distribution function. In addition, let $\mathfrak{B}$ denote the function space of bounded Borel mappings. If Z is $\mathfrak{B}$-differentiable, then $\partial_\theta F_\theta$ exists and*

$$\sup_{z\in\mathbb{R}} \left| \frac{\partial}{\partial\theta} F_\theta(z) \right| < \infty.$$

Moreover, $\partial_\theta F_\theta(x)$ is continuous as a mapping in z.

Proof: Let Z have probability measure $\mathbb{P}_\theta$ with $\mathfrak{B}$-derivative $(c_\theta, \mathbb{P}^+_\theta, \mathbb{P}^-_\theta)$. Since all indicator mappings are contained in $\mathfrak{B}$, then as a consequence from Equation (4.22)

$$\begin{aligned}
\frac{\partial}{\partial\theta} F_\theta(z) &= \int_{\mathbb{R}} 1\{x \le z\}\, \mathbb{P}'_\theta(x)\, dx \\
&= c_\theta \left(\int_{\mathbb{R}} 1\{x \le z\}\, \mathbb{P}^+_\theta(x)\, dx - \int_{\mathbb{R}} 1\{x \le z\}\, \mathbb{P}^-_\theta(x)\, dx \right) \\
&= c_\theta \left(F^+_\theta(z) - F^-_\theta(z) \right),
\end{aligned}$$

and the boundedness follows from the observation that $|F_\theta^+(z) - F_\theta^-(z)| \leq 1$ for any z. The reason that $\partial_\theta F_\theta(z)$ is continuous as a function of z stems from the result that $F_\theta(z)$ is $\mathfrak{B}$-differentiable, implying strong differentiability, [52], of F_θ and thus implying continuity of $\partial_\theta F_\theta$. □

Remark 4.3. *In Lemma 4.4 the same result holds verbatim for any probability measure $\mathbb{P}_\theta$ that is differentiable w.r.t. a Banach base $(\mathfrak{D}, v)$, where $v \geq 1$ and $\mathfrak{B} \subset \mathfrak{D}_v$, see Section 1.2.2.*

4.4.2 The Inverse-Transformation Approach and Extensions for Finite Product Measures

Let $Z = h(X)$ be written as a transformation of a random vector $X = (X_1, \ldots, X_J) \in \mathbb{R}^J$ via a Borel mapping h. In this section we verify $\mathfrak{B}$-differentiability of $Z = h(X)$ when the constituent elements of the random vector X are also $\mathfrak{B}$-differentiable. This result is built in two stages, and it will be helpful for establishing Assumption **(C3)** in applications.

Lemma 4.5. *Assume that $X \in \mathbb{R}$ is $\mathfrak{B}$ differentiable with measure-valued derivative (c, X^+, X^-). Let $h : \mathbb{R} \mapsto \mathbb{R}$ be a real-valued Borel measurable mapping defined on the state space of X. Then $Z = h(X)$ is $\mathfrak{B}$-differentiable with measure-valued derivative (c, Z^+, Z^-), with $Z^+ = h(X^+)$ and $Z^- = h(X^-)$. In addition the distribution function of Z is differentiable w.r.t. θ.*

Proof: For any $g \in \mathfrak{B}$ it holds that $g \circ h$, with $(g \circ h)(x) = g(h(x))$ belongs to $\mathfrak{B}$. By computation, it holds for any $g \in \mathfrak{B}$ that

$$\begin{aligned}\frac{d}{d\theta}\mathbb{E}_\theta[g(Z)] = \frac{d}{d\theta}\mathbb{E}_\theta[(g \circ h)(X)] &= c\left(\mathbb{E}_\theta[(g \circ h)(X^+)] - \mathbb{E}_\theta[(g \circ h)(X^-)]\right) \\ &= c\left(\mathbb{E}_\theta[g(h(X^+))] - \mathbb{E}_\theta[g(h(X^-))]\right),\end{aligned}$$

which shows that $(c, h(X^+), h(X^-))$ is an instance of an $\mathfrak{B}$-derivative of $Z = h(X)$. Differentiability of the distribution function Z then follows from Lemma 4.4, which concludes the proof. □

For the random vector X, let $\mathbb{P}_{\theta,i}$, $1 \leq i \leq J$, be the probability measure induced by the random variable $X_i \in \mathbb{R}$, and let $F_{\theta,i}$ denote the corresponding distribution function. Without loss of generality the corresponding density function $f_{\theta,i}$ has support on the entire real line, by defining $f_{\theta,i} = 0$ on S^c if necessary. Applying the product rule of weak differentiation leads to an unbiased derivative estimator in this case. The precise statement is provided in the following lemma.

Lemma 4.6. *For $1 \le i \le J$, suppose that the random variable X_i is $\mathfrak{B}$-differentiable w.r.t. θ and let (c_i, X_i^+, X_i^-) denote a version of the $\mathfrak{B}$-derivative. Let $h : \mathbb{R}^J \mapsto \mathbb{R}$ be a Borel mapping. Then $Z = h(X_1, \ldots, X_J)$ is $\mathfrak{B}$-differentiable with $\mathfrak{B}$-derivative (c, Z^+, Z^-), which is given by*

$$Z^+ = h(X_1, \ldots, X_{\rho-1}, X_\rho^+, X_{\rho+1}, \ldots, X_J), \quad Z^- = h(X_1, \ldots, X_{\rho-1}, X_\rho^-, X_{\rho+1}, \ldots, X_J),$$

where ρ is uniformly distributed on $\{1, \ldots, J\}$ independent of everything else, and $c = J c_\rho$.

Proof: Applying Theorem 6.1 in [52] it readily follows that Z is $\mathfrak{B}$-differentiable. The particular form of the weak derivative stems from the randomization principle, Corollary 4 in [48]. □

The estimator in Lemma 4.6 requires re-simulation, which might render the estimator inefficient. In the following we will show that in case that h is invertible, we can rewrite this estimator in single run form by conditioning w.r.t. all but one random variable. Let $x = (x_1, \ldots, x_J)$, and denote, for $1 \le i \le J$, $x_{\bar{i}}$ as the removal of element x_i from the vector x, i.e., $x_{\bar{i}} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_J)$. In this conditioning, for $1 \le i \le J$, we fix $X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, \cdot, X_{i+1} = x_{i+1}, \ldots, X_J = x_J$, and define the inverse image of $h(x_1, \ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_J)$ as follows

$$\{y \in \mathbb{R} : h(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_J) \le z\} = h^{I,i}(x_{\bar{i}})(z).$$

Hence, for $i = 1$,

$$\begin{aligned}\mathbb{P}(Z \le z \,|\, X_2 = x_2, \ldots, X_j = x_J) &= \mathbb{P}\big(X_1 \in h^{I,1}(x_{\bar{1}})(z) \,\big|\, X_2 = x_2, \ldots, X_J = x_J\big) \\ &= \mathbb{E}\big[\mathbb{P}_{\theta,1}\big(h^{I,1}(x_{\bar{1}})(z)\big) \,\big|\, X_2 = x_2, \ldots, X_J = x_J\big].\end{aligned}$$

We call the inverse image *simple* if $h^{I,i}(x_{\bar{i}})(z)$ yields sets of type $(-\infty, w]$ for any value of z and $x_{\bar{i}}$. This then implies that $h^{I,i}$ is Borel measurable. If the inverse image is simple, we let $h^{-1,i}(x_{\bar{i}})(z) = w$ if $(-\infty, w] = h^{I,i}(x_{\bar{i}})(z)$, and the above RHS probability can be written as

$$= \mathbb{E}\big[F_{\theta,1}\big(h^{-1,1}(x_{\bar{1}})(z)\big) \big| X_2 = x_2, \ldots, X_J = x_J\big]. \tag{4.23}$$

As has been noted in [59], the inverse of h may not be available in closed form or the evaluation of it may be numerically infeasible, for instance the portfolio credit risk example in [59], also Section 2.3.2.2.

In the next lemma we will present the single run version of the estimator in Lemma 4.6. For $1 \le i \le J$, we denote analogously by $X_{\bar{i}} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_J)$ the random vector X with element X_i removed.

Lemma 4.7. *Let $F_{\theta,i}$ denote the distribution function for X_i, for $1 \le j \le J$. If h is simple, then it holds under the conditions put forward in Lemma 4.6 that*

$$\frac{\partial}{\partial\theta} F_\theta(z) = J\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta} F_{\theta,\rho}\left(h^{-1,\rho}_{X_{\bar\rho}}(z)\right)\right],$$

for ρ being uniformly distributed on $\{1,\dots,J\}$ independent of everything else.

Proof: By Lemma 4.6 it holds that

$$\begin{aligned}\frac{\partial}{\partial\theta} F_\theta(z) &= J\mathbb{E}_\theta\big[c_\rho\left(1\{h(X_1,\dots,X_{\rho-1},X_{\rho^+},X_{\rho+1},\dots,X_J)\le z\}\right.\\ &\qquad\left. - 1\{h(X_1,\dots,X_{\rho-1},X_{\rho^-},X_{\rho+1},\dots,X_J)\le z\}\right)\big]\\ &= J\mathbb{E}\left[c_\rho\mathbb{E}_\theta\left[1\left\{X^+_\rho \le h^{-1,\rho}(X_{\bar\rho})(z)\right\} - 1\left\{X^-_\rho \le h^{-1,\rho}(X_{\bar\rho})(z)\right\}\middle| X_{\bar\rho}\right]\right]\\ &= J\mathbb{E}\left[\frac{\partial}{\partial\theta} F_{\theta,\rho}\left(h^{-1,\rho}_{X_{\bar\rho}}(z)\right)\right],\end{aligned}$$

which proves the claim. □

4.4.3 Incorporating Distributional Derivative Estimators

If $\partial_\theta F_\theta$ is not available in a closed form, we resort to estimating $\partial_\theta F_\theta$ through re-simulation. We first address the estimator put forward in Equation (4.7). The derivative estimator in Equation (4.6) yields similar outcomes. Let

$$\mathrm{distr}^\alpha_l(z) = \frac{c}{l}\sum_{i=1}^{l}\left(1\{Z^+_i \le z\} - 1\{Z^-_i \le z\}\right)$$

be the unbiased estimator of $\partial_\theta F_\theta(z)$ based on l i.i.d. replications of (Z^+, Z^-), for instance via Lemma 4.6. Furthermore, given m, k let $\mathbf{Z}^k = (\mathbf{Z}(j) : 1 \le j \le k)$ be an i.i.d. collection of k sequences of $\mathbf{Z} = (Z_i : 1 \le i \le m)$. In addition, let

$$\mathrm{dens}^\alpha_{m,k} := \mathrm{dens}^\alpha_{m,k}(\mathbf{Z}^k) = \frac{m}{k}\sum_{i=1}^{k}\left(Z_{\lceil\alpha m\rceil:m}(i) - Z_{\lceil\alpha m\rceil:m}(i)\right)$$

be an unbiased estimator for $1/f_\theta(q_\alpha(\theta))$.

Lemma 4.8. *Under Assumptions* **(C1)** *to* **(C3)** *it holds for $k \ge 1$ that*

$$\left|\mathbb{E}_\theta\left[\mathrm{dens}^\alpha_{m,k} - \frac{1}{f_\theta(q_\alpha(\theta))}\right]\right| = O\left(\frac{1}{m}\right).$$

Proof: Replacing $\partial_\theta F_\theta$ by the constant one in the proof of Theorem 4.2, the result readily occurs. □

The following lemma provides by means of re-simulation an extension of Theorem 4.2 to the case that $\partial_\theta F_\theta$ is not attainable in a closed form.

Lemma 4.9. *Under Assumptions* **(C1)** *to* **(C3)** *it holds that for all* $k, l \geq 1$ *that*

$$\left|\mathbb{E}_\theta\left[-\mathrm{dens}^\alpha_{m,k} \times \mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m}) - \frac{\partial}{\partial\theta} q_\alpha(\theta)\right]\right| = O\left(\frac{1}{m}\right).$$

Proof: By definition,

$$\begin{aligned}
\mathbb{E}_\theta\left[D^\alpha(m,k) - \frac{\partial}{\partial\theta} q_\alpha(\theta)\right] &= \mathbb{E}_\theta\left[-\mathrm{dens}^\alpha_{m,k} \times \frac{\partial}{\partial\theta} F_\theta(Z_{\lceil \alpha m \rceil : m}) - \frac{\partial}{\partial\theta} q_\alpha(\theta)\right] \\
&= \mathbb{E}_\theta\Big[-\mathrm{dens}^\alpha_{m,k} \times \mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m}) \\
&\quad - \mathrm{dens}^\alpha_{m,k} \times \left(\frac{\partial}{\partial\theta} F_\theta(Z_{\lceil \alpha m \rceil : m}) - \mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m})\right) - \frac{\partial}{\partial\theta} q_\alpha(\theta)\Big] \\
&= \mathbb{E}_\theta\left[-\mathrm{dens}^\alpha_{m,k} \times \mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m}) - \frac{\partial}{\partial\theta} q_\alpha(\theta)\right] \\
&\quad - \mathbb{E}_\theta\left[\mathrm{dens}^\alpha_{m,k} \times \left(\frac{\partial}{\partial\theta} F_\theta(Z_{\lceil \alpha m \rceil : m}) - \mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m})\right)\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
&\left|\mathbb{E}_\theta\left[-\mathrm{dens}^\alpha_{m,k} \times \mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m}) - \frac{\partial}{\partial\theta} q_\alpha(\theta)\right]\right| \\
&\quad \leq \left|\mathbb{E}_\theta\left[D^\alpha(m,k) - \frac{\partial}{\partial\theta} q_\alpha(\theta)\right]\right| + \left|\mathbb{E}\left[\mathrm{dens}^\alpha_{m,k} \times \left(\frac{\partial}{\partial\theta} F_\theta(Z_{\lceil \alpha m \rceil : m}) - \mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m})\right)\right]\right|.
\end{aligned}$$

The rate of convergence of $\mathbb{E}_\theta[D^\alpha(m,k) - \partial_\theta q_\alpha(\theta)]$ is put forward in Theorem 4.2. Since $\mathrm{distr}^\alpha_l(z)$ is an unbiased estimator for $\partial_\theta F_\theta(z)$, there is no bias introduced by switching from $\partial_\theta F_\theta(Z_{\lceil \alpha m \rceil : m})$ to $\mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m})$, i.e.,

$$\mathbb{E}\left[\mathrm{dens}^\alpha_{m,k} \times \left(\frac{\partial}{\partial\theta} F_\theta(Z_{\lceil \alpha m \rceil : m}) - \mathrm{distr}^\alpha_l(Z_{\lceil \alpha m \rceil : m})\right)\right] = 0.$$

□

Following the same line of argument as for the proof of the Central Limit Theorem for estimator $D^\alpha(m,k)$ in Theorem 4.3, we arrive at the following result:

Lemma 4.10 (Central Limit Theorem)**.** *Suppose that Assumptions* **(C1)** *to* **(C4)** *hold and suppose that* $k^{1/2}/m \to 0$ *as* $k, m \to \infty$*. If both* $\mathbb{E}_\theta[(Z^+)^{2+\delta}] < \infty$ *and*

$\mathbb{E}_\theta[(Z^-)^{2+\delta}] < \infty$ *for some* $\delta > 0$*, then*

$$\frac{\text{dens}^\alpha_{m,k} \times \text{distr}^\alpha_l(Z_{\lceil \alpha m\rceil:m}) - \frac{\partial}{\partial\theta} q_\alpha(\theta)}{\left(\text{Var}_\theta\left(\text{dens}^\alpha_{m,k} \times \text{distr}^\alpha_l(Z_{\lceil \alpha m\rceil:m})\right)\right)^{1/2}} \xrightarrow{d} N(0,1)$$

as $k, m, l \to \infty$*, where* $Z_{\lceil \alpha m\rceil:m}$ *is sampled independently of* $\mathbf{Z}^k$*.*

The rate of convergence for the standard deviation of the re-simulation estimator differs from the $O(k^{-1/2})$ result from the single run estimator, Equation (4.4). The precise rate is derived below

Lemma 4.11. *Suppose that Assumptions* **(C1)** *to* **(C4)** *hold, then as* $m \to \infty$

$$\text{Var}_\theta\left(-\text{dens}^\alpha_{m,k} \times \text{distr}^\alpha_l(Z_{\lceil \alpha m\rceil:m})\right) = O\left(\frac{k+l}{kl}\right)$$

where $Z_{\lceil \alpha m\rceil:m}$ *is sampled independently of* $\mathbf{Z}^k$*.*

Proof: By mutual independence of the samples that constitute the quantile density and parameter derivative estimators:

$$\begin{aligned}
&\text{Var}_\theta\left(-\text{dens}^\alpha_{m,k} \times \text{distr}^\alpha_l(Z_{\lceil \alpha m\rceil:m})\right) \\
&= \mathbb{E}\left[\left(\text{dens}^\alpha_{m,k}\right)^2\right] \mathbb{E}_\theta\left[\left(\text{distr}^\alpha_l\left(Z_{\lceil \alpha m\rceil:m}\right)\right)^2\right] - \mathbb{E}^2\left[\text{dens}^\alpha_{m,k}\right] \mathbb{E}^2_\theta\left[\text{distr}^\alpha_l\left(Z_{\lceil \alpha m\rceil:m}\right)\right].
\end{aligned}$$

This is our starting point. We continue by considering each of the four parts, beginning with $\text{dens}^\alpha_{m,k}$ and $(\text{dens}^\alpha_{m,k})^2$. Since $\mathbf{Z}^k$ is composed of i.i.d. sequences

$$\begin{aligned}
\mathbb{E}\left[\text{dens}^\alpha_{m,k}\right] &= \mathbb{E}\left[m\left(Z_{\lceil \alpha m\rceil:m}(1) - Z_{\lceil \alpha m\rceil-1:m}(1)\right)\right], \quad \text{and} \\
\mathbb{E}\left[\left(\text{dens}^\alpha_{m,k}\right)^2\right] &= \frac{1}{k}\mathbb{E}\left[m^2\left(Z_{\lceil \alpha m\rceil:m}(1) - Z_{\lceil \alpha m\rceil-1:m}(1)\right)^2\right] \\
&\quad + \left(1 - \frac{1}{k}\right)\mathbb{E}^2\left[m\left(Z_{\lceil \alpha m\rceil:m}(1) - Z_{\lceil \alpha m\rceil-1:m}(1)\right)\right].
\end{aligned}$$

We denote the spacing term by $A_m := m(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil-1:m})$, which reduces the notation in the above expressions to

$$\begin{aligned}
\mathbb{E}\left[\text{dens}^\alpha_{m,k}\right] &= \mathbb{E}[A_m], \quad \text{and} \\
\mathbb{E}\left[\left(\text{dens}^\alpha_{m,k}\right)^2\right] &= \frac{1}{k}\mathbb{E}[A_m^2] + \left(1 - \frac{1}{k}\right)\mathbb{E}^2[A_m].
\end{aligned}$$

For the parameter derivative estimator, simplification of distr_l^α and $(\mathrm{distr}_l^\alpha)^2$ is slightly more complicated. We begin by employing the i.i.d. property of the random pair sequence $((Z_i^+, Z_i^-) : 1 \le i \le l)$ to the mean of this estimator:

$$\mathbb{E}_\theta\big[\mathrm{distr}_l^\alpha\big(Z_{\lceil \alpha m\rceil:m}\big)\big] = c\,\mathbb{E}_\theta\big[1\{Z_1^+ \le Z_{\lceil \alpha m\rceil:m}\} - 1\{Z_1^- \le Z_{\lceil \alpha m\rceil:m}\}\big]$$

Like in the preceding derivation, we denote by $\mathbf{Z}(k+1)$ the sequence of random variables that form the order statistic. By conditioning w.r.t. this sequence the above expression is further simplified to

$$\begin{aligned}
&= c\,\mathbb{E}\big[\mathbb{E}_\theta\big[1\{Z_1^+ \le Z_{\lceil \alpha m\rceil:m}\} - 1\{Z_1^- \le Z_{\lceil \alpha m\rceil:m}\}\big|\mathbf{Z}(k+1)\big]\big]\\
&= c\,\mathbb{E}\big[F_\theta^+(Z_{\lceil \alpha m\rceil:m}) - F_\theta^-(Z_{\lceil \alpha m\rceil:m})\big].
\end{aligned}$$

The reduction for the second moment follows similar arguments. Below, we define $F_\theta^\pm$ to be the bivariate distribution of (Z^+, Z^-).

$$\begin{aligned}
\mathbb{E}_\theta\Big[\big(\mathrm{distr}_l^\alpha\big(Z_{\lceil \alpha m\rceil:m}\big)\big)^2\Big] &= \frac{c^2}{l}\mathbb{E}_\theta\Big[\big(1\{Z_1^+ \le Z_{\lceil \alpha m\rceil:m}\} - 1\{Z_1^- \le Z_{\lceil \alpha m\rceil:m}\}\big)^2\Big]\\
&\quad + c^2\left(1-\frac{1}{l}\right)\mathbb{E}_\theta\big[\big(1\{Z_1^+ \le Z_{\lceil \alpha m\rceil:m}\} - 1\{Z_1^- \le Z_{\lceil \alpha m\rceil:m}\}\big)\\
&\qquad\qquad \cdot\big(1\{Z_2^+ \le Z_{\lceil \alpha m\rceil:m}\} - 1\{Z_2^- \le Z_{\lceil \alpha m\rceil:m}\}\big)\big]\\
&= \frac{c^2}{l}\mathbb{E}_\theta\big[F_\theta^+(Z_{\lceil \alpha m\rceil:m}) + F_\theta^-(Z_{\lceil \alpha m\rceil:m}) - 2F_\theta^\pm(Z_{\lceil \alpha m\rceil:m}, Z_{\lceil \alpha m\rceil:m})\big]\\
&\quad + c^2\left(1-\frac{1}{l}\right)\mathbb{E}_\theta\Big[\big(F_\theta^+(Z_{\lceil \alpha m\rceil:m}) - F_\theta^-(Z_{\lceil \alpha m\rceil:m})\big)^2\Big].
\end{aligned}$$

Similar to the quantile density estimator, we define

$$\begin{aligned}
B_{1,m} &:= F_\theta^+(Z_{\lceil \alpha m\rceil:m}) - F_\theta^-(Z_{\lceil \alpha m\rceil:m}), \quad \text{and}\\
B_{2,m} &:= F_\theta^+(Z_{\lceil \alpha m\rceil:m}) + F_\theta^-(Z_{\lceil \alpha m\rceil:m}) - 2F_\theta^\pm(Z_{\lceil \alpha m\rceil:m}, Z_{\lceil \alpha m\rceil:m}),
\end{aligned}$$

and given these definitions, the first and second moments of distr_l^α are concisely written as

$$\begin{aligned}
\mathbb{E}_\theta\big[\mathrm{distr}_l^\alpha\big] &= c\,\mathbb{E}[B_{1,m}], \quad \text{and}\\
\mathbb{E}_\theta\Big[\big(\mathrm{distr}_l^\alpha\big)^2\Big] &= \frac{c^2}{l}\mathbb{E}[B_{2,m}] + c^2\left(1-\frac{1}{l}\right)\mathbb{E}[B_{1,m}^2].
\end{aligned}$$

Combining all of these above results together into the variance expression, we have extracted the necessary terms for our claim

$$\begin{aligned}
&\mathrm{Var}_\theta\Big(-\mathrm{dens}_{m,k}^\alpha \times \mathrm{distr}_l^\alpha(Z_{\lceil \alpha m\rceil:m})\Big)\\
&= c^2\left(\frac{1}{k}\mathbb{E}[A_m^2] + \left(1-\frac{1}{k}\right)\mathbb{E}^2[A_m]\right)\left(\frac{1}{l}\mathbb{E}[B_{2,m}] + \left(1-\frac{1}{l}\right)\mathbb{E}[B_{1,m}^2]\right) - c^2\mathbb{E}^2[A_m]\mathbb{E}^2[B_{1,m}].
\end{aligned}$$

As the spacing and distribution terms are $O(1)$, the leading order behaviour is unaffected by m. Therefore, we approximate A_m by its spacing limit $A = f_\theta(q_\alpha(\theta))$, where E is an exponential mean one random variable. In addition, we approximate $B_{1,m}$, $B_{2,m}$ with B_1, B_2 respectively by replacing the order statistic in the argument of the respective distribution functions with $q_\alpha(\theta)$. From Postscript I of Section 3.4.1 and Lemma A.1 the distribution function $F_\theta^\pm$ is also continuous around $q_\alpha(\theta)$, indicating that $B_{2,m}$ can be arbitrarily close to B_2 for sufficiently large m. As a result, the terms B_1 and B_2 are now deterministic. After computing the moments of the exponential random variable and a little algebra, the leading term of the variance is equal to

$$\begin{aligned} &= \frac{c^2}{kl f_\theta^2(q_\alpha(\theta))}\big(k(B_2 - B_1^2) + lB_1^2 + (B_2 - B_1^2)\big) \\ &\leq \frac{c^2 \max\{B_2 - B_1^2, B_1^2\}}{f_\theta^2(q_\alpha(\theta))}\left(\frac{k+l}{kl}\right) + O\left(\frac{1}{kl}\right) \end{aligned}$$

and the maximum is replaced by the minimum for the lower bound. As the $O((kl)^{-1})$ term is sub-leading, the claim is proved. □

Assuming that $Z = h(X_1, \ldots, X_J)$, where X_i, for $1 \leq i \leq J$, has distribution function $F_{\theta,i}$, we can alternatively estimate the distributional derivative term, following Equation (4.23) and Lemma 4.7, through

$$\frac{J}{l}\sum_{j=1}^{l} \frac{\partial}{\partial\theta} F_{\theta,\rho}\big(h^{-1,\rho}(X_{\bar{\rho}}(j))(z)\big), \tag{4.24}$$

where $(X(j) : 1 \leq j \leq l)$, with $X(j) = (X_1(j), \ldots, X_J(j))$, are l i.i.d. copies of the random vector X. The term $X_{\bar{i}}(j)$, $1 \leq i \leq J$ denotes the removal of X_i from random vector $X(j)$.

We conclude this section with a discussion of the derivative estimator in Equation (4.7). For this estimator, a confidence interval can be obtained by choosing α_1, α_2, such that $1 - \alpha = (1 - \alpha_1)(1 - \alpha_2)$, where α denotes the level of the overall confidence interval, and the estimated confidence intervals for $\mathbb{E}_\theta[\mathrm{dens}_{m,k}^\alpha]$ of level α_1 and for $\mathbb{E}[\mathrm{distr}_l^\alpha(Z_{\lceil \alpha m\rceil:m})]$ of level α_2, separately.

To estimate the quantile sensitivity according to Equation (4.7), we require mk generations to form k replications of the order statistic, composed from m random variables, to estimate the statistic $\mathrm{dens}_{m,k}^\alpha$. In addition, we require $m + 2l$ random variables to estimate the distribution functions $F_\theta^+(Z_{\lceil \alpha m\rceil:m})$, $F_\theta^-(Z_{\lceil \alpha m\rceil:m})$, namely, m random variables to independently estimate the order statistic $Z_{\lceil \alpha m\rceil:m}$ from $\mathrm{dens}_{m,k}^\alpha$ and l generations each of the measure-valued derivative random variables Z^+ and Z^-. Therefore, the computational budget is

$n = mk+m+2l$. As argued in Section 6 of Chapter III in [3], the relation between m, k, l, should be such that the standard deviation dominates the bias such that the confidence intervals are meaningful. From (4.11), the standard deviation of this estimator is $O(((k+l)/kl)^{1/2})$. The rate of convergence is maximized when $k = l$. The bias from Lemma 4.9 is $O(m^{-1})$. Setting $k = l$ together with requirement $k^{1/2}/m \to 0$ as $k, m \to \infty$, automatically the standard deviation is greater than the bias, i.e.,

$$\begin{aligned}\frac{\text{Bias}}{\text{Std. Dev.}} &= \frac{1}{m}\left(\frac{k^2}{2k-1}\right)^{1/2}\\ &= \frac{1}{2^{1/2}}\frac{k^{1/2}}{m}.\end{aligned}$$

With $k = l$, the computation budget becomes $n = mk + m + 2k$, and taking $m = p^{1/3+\delta}$, and $k = p^{2/3-\delta}$, with $p < n$ and $0 < \delta < 2/3$, the computational budget is given by the polynomial $n = p + p^{1/3+\delta} + 2p^{2/3-\delta}$. Solving n approximately from this equation yields $p \approx O(n)$ (the exact term is less). Again, the rate of convergence is maximized when $\delta \downarrow 0$. From $p \approx O(n)$ and assuming a small value for δ, we attain $k \approx O(n^{2/3})$, and $m \approx O(n^{1/3})$. The rate of convergence has approximate order of $O(n^{-1/3})$ though less overall due to the sub-leading terms. The bias is also of approximate order $O(n^{-1/3})$.

Instead, if the additional computational budget is preferred to be minimized them $l = m$ is an appropriate choice and the computational budget becomes $n = km+3m$. Parameterising $m = p^{1/2+\delta}$ and $k = p^{1/2-\delta}$, and given the condition $k^{1/2}/m \to 0$ as $m, k \to \infty$, in this case the rate of convergence is approximately $O(n^{-1/4})$ and the bias is approximately $O(n^{-1/2})$.

As for the alternative representation of the parameter derivative, Equation (4.24), we assume that each element for the random vector X has a distribution function $F_{\theta,i}$, $1 \le i \le J$. To estimate $\partial_\theta F_\theta$ we require $m+l$ random variables to be generated. To recognize this, as noted, we define a unit of the computational budget to be a generation of the J elements of random variable Z, or equivalently $Z^\pm$. This implies an expense of $1/J$ to generate one random variable. To generate the conditioned components $X_{\bar{\rho}}$ once, we require a budget of $(J-1)/J$ to generate the elements of random vector of length $J-1$ and $1/J$ to generate the index variable ρ. Then the total computational budget to generate $X_{\bar{\rho}}$ once is one and for the l replications, l. This together with m random variables to generate the order statistic $Z_{\lceil \alpha m\rceil:m}$, the total computational budget is $m+l$. Including the km generations to generate $\text{dens}^{\alpha}_{m,k}$, independent of the estimate of $\partial_\theta F_\theta$, the overall budget for this variation of the re-simulation estimator is $n = mk + m + l$. Taking $k = l$, and following the same argument, and restriction

$k^{1/2}/m \to 0$ as $m, k \to \infty$ the maximum rate of convergence is asymptotically below $O(n^{-1/3})$ with bias $O(n^{-1/3})$.

Remark 4.4. *For the estimator put forward in Lemma 4.1, let*

$$\text{dens}^{\alpha}_{m,k} = \frac{m\left(Z_{\lceil \alpha m\rceil + k_m : m} - Z_{\lceil \alpha m\rceil - k_m + 1 : m}\right)}{2k_m}$$

be the estimator for $1/f_\theta(Z_{\lceil \alpha m\rceil : m})$ as provided in Equation (4.23). Then the overall estimator becomes $-\text{dens}^{\alpha}_{m,k} \times \text{distr}^{\alpha}_{l}(Z_{\lceil \alpha m\rceil : m})$, where the bias is reduced by letting $m \to \infty$ and variance is reduced by letting $l \to \infty$. Unfortunately, no analytical results are available on the rate the bias is reduced through the choice of m, and there no guidelines are available for choosing m in relation to l.

4.5 Application to Queues

Consider a stationary tandem queue with Poisson-λ arrival stream consisting of two queues with infinite buffer capacity. Jobs arrive from the outside to server 1 and are being served with exponential service rate μ_1. From server 1 they continue to server 2 where they are served with exponential service rate μ_2. The jobs leave the system once the service at station 2 is completed. For the sake of simplicity, we let $\mu = \mu_1 = \mu_2$, and we assume that the tandem queue is stable, i.e., $\lambda < \mu$. By Burke's theorem [11], it is known that the departure process at the first server is a Poisson-λ process and additionally, from [87], it is known that the sojourn time from server 1 is independent of server 2. With these results it can be shown that the distribution of the sojourn time of a job, i.e., the total time elapsed between entering and leaving the system, is the convolution of two exponential distributions with rate $\mu - \lambda$, i.e., an Erlang distribution with shape 2 and rate $\mu - \lambda$.

Let q_α denote the α-quantile of the stationary sojourn time. In the following we apply our estimator to estimate the sensitivity of q_α with respect to, for example, μ. Even though the distribution of the stationary sojourn time can be attained in a closed form, the α-quantile cannot be expressed in a closed form and has to be solved numerically. Let F^D_μ denote the distribution function of the stationary sojourn time, denoted by Z, and since this distribution is of Erlang type, the distribution function is given as follows

$$F^D_\mu(z) = 1 - e^{-(\mu-\lambda)z} - (\mu - \lambda)ze^{-(\mu-\lambda)z}, \; z \geq 0.$$

Following the line of thought from the MVD derivative estimation in Section

1.2.4.1, Z is $\mathfrak{B}$-differentiable as Equation (4.21) is satisfied and therefore

$$\frac{\partial}{\partial \mu} F_\mu^D(z) = (\mu - \lambda) z^2 e^{-(\mu-\lambda)z}, \; z \geq 0.$$

Inserting the above expression into Equation (4.4) yields

$$D(m) = -m(\mu - \lambda)\left(Z_{\lceil \alpha m \rceil : m} - Z_{\lceil \alpha m \rceil - 1 : m}\right)\left(Z_{\lceil \alpha m \rceil : m}\right)^2 e^{-(\mu-\lambda) Z_{\lceil \alpha m \rceil : m}}$$

as an estimator, where the order statistic is attained from an i.i.d. sample $\mathbf{Z}$ of m stationary sojourn times.

Next, we consider the excess sojourn time $D1\{D \geq d\}$ for some fixed value d. For example, in service systems waiting is often only considered to be relevant if it exceeds some pre-specified threshold value d. We are again interested in the sensitivity w.r.t. the common service time parameter μ. Let $h(x) = x1\{x \geq d\}$ and note that $h(x) = y$ implies $x = y$ for $y > d$. Hence, under the reasonable assumption that $q_\alpha(\mu) > d$, we can apply the estimator in (4.4), which yields

$$D(m) = -m(\mu - \lambda)\left(Z_{\lceil \alpha m \rceil : m} - Z_{\lceil \alpha m \rceil - 1 : m}\right)\left(Z_{\lceil \alpha m \rceil : m}\right)^2 e^{-(\mu-\lambda) Z_{\lceil \alpha m \rceil : m}} 1\{Z_{\lceil \alpha m \rceil : m} > d\}$$

as a derivative estimator. The order statistic is obtained from an i.i.d. sample $\mathbf{Z}$ of m stationary sojourn times. For this mapping, $h(x)$ fails to be Lipschitz continuous and, therefore, IPA cannot be applied.

In the following we show that Assumptions **(C1)** to **(C4)** hold for the above examples. Since the distribution of the stationary sojourn time in the tandem queue is explicitly known and has finite second moment, sampling $\mathbf{Z}$ from this distribution is feasible and Assumption **(C1)** is thus satisfied. We now turn to Assumption **(C2)**. Since the support of the sojourn time distribution is $\mathbb{R}^+$, $f_\theta(x) > 0$ on every ball $B(\alpha)$, and the first part of **(C2)** is satisfied. Let $f_\mu^D(x)$ denote the density function of the stationary sojourn time, which is given by

$$f_\mu^D(x) = (\mu - \lambda)^2 x e^{-(\mu-\lambda)x},$$

for $x \geq 0$. The remaining conditions for **(C2)** follow from

$$\frac{\partial}{\partial \mu} f_\mu^D(x) = \left(2(\mu - \lambda)x - (\mu - \lambda)^2 x^2\right) e^{-(\mu-\lambda)x},$$

which is bounded since $x^2 e^{-(\mu-\lambda)x}$ is bounded on $\mathbb{R}^+$.

As we have already argued that the stationary waiting is $\mathfrak{B}$-differentiable, Assumptions **(C3)** and **(C4)** follow from applying Lemma 4.4. For the excess waiting time, we resort to Lemma 4.5 to verify that $h(X) = X1\{X \geq d\}$ is $\mathfrak{B}$-differentiable.

In addition, the above line of argument can be extended to more general feed-forward exponential queueing networks.

4.6 Application to Financial Risk Management

In this section we will discuss the estimation of sensitivities relating to the value-at-risk (VaR) of market risk[1] factors. Market risk is defined as the extent of loss in the value of financial assets in a financial firm's portfolio. These assets include stocks, bonds, as well as forwards, derivatives, and options contracts relating to interest rates, foreign exchange, and commodities. Following the definition in [20], the VaR is the α-quantile of the loss distribution of the asset value of a portfolio over a prescribe period t.

For $t \geq 0$, let $v(w, H_t)$ denote the value of an asset portfolio comprised of N contracts $w = (w_i : 1 \leq i \leq N)$ of assets $H_t = (H_{i,t})$, in which $H_{i,t} := H_i(S(t))$ are comprised from J underlying prices $S(t) = (S_j(t))$, $1 \leq j \leq J$. Note that a stock may appear in more than one asset, or more than one commodity may de comprised in a derivative contract. Then the VaR, denoted by $q_{\alpha,t}(w, H_t)$, is the α-quantile of the distribution of the loss $-(v(w, H_t) - v(w, H_0))$, which describes the maximal amount by which the portfolio can fall short at time t of its present value with probability α. From a regulatory standpoint, the VaR is calculated at the $\alpha = 0.99$ level over a period of $t = 10$ trading days.

Parameter derivatives with respect to macro- and microeconomic variables such as an interest rate, or the annual after tax cash flow of a company provide an understanding of the dependence of the VaR on these factors. This is of particular interest in scenarios relating to periods of financial crisis. In addition, the estimate of sensitivities are needed in stochastic optimization aimed at minimizing the VaR.

This section provides two basic examples with respect to market risk, especially to stocks, to highlight how the presented techniques can be utilized to evaluate sensitivities of the VaR. The first example in Section 4.6.2 studies VaR sensitivity of the cash flow of a single option contract, composed of one or multiple assets. Section 4.6.3 discusses VaR sensitivities of a stock portfolio. The price of a stock in this section is either modelled via Geometric Brownian motion or the Variance Gamma model. These stochastic processes are briefly explained in Section 4.6.1.

4.6.1 The Underlying Financial Models

In the following we introduce the Geometric Brownian motion and the Variance Gamma model as descriptions of a security price path. Geometric Brownian motion is discussed in Section 4.6.1.1, and the Variance Gamma model is introduced in Section 4.6.1.2.

[1] The VaR can also be used to incorporate credit risk due to creditworthiness of counterparties.

4.6.1.1 Geometric Brownian Motion

Geometric Brownian motion is a simplified model in depicting the future price of a stock within a financial market. This representation of stock price is used in the famous Black-Scholes-Merton model, [81], in which the stock price in this instance is modelled under the "risk-neutral" measure of investor preferences. Let $S(t)$ at time $t \geq 0$, with initial price $S(0) = s_0$ almost surely, denote the price of the stock that pays no dividend. For the purposes of risk-management, we are interested in the distributed according to the probability measure observed under market conditions. With $\mu \in \mathbb{R}$ being the instantaneous mean return of the stock and $\sigma > 0$ denoting the implied volatility, the evolution of the price of the stock at time $t \geq 0$ is given by

$$S(t) = S(0)e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}W},$$

with W being a standard normal random variable. Alternatively, let $X_{a(t),b(t)}$ denote a normal random variable with mean

$$a(t) = \ln(S(0)) + \left(\mu - \frac{\sigma^2}{2}\right)t$$

and standard deviation

$$b(t) = \sigma\sqrt{t},$$

then,

$$S(t) = e^{X(t)},$$

for $t \geq 0$. Note that the above presentation is a simplified discussion of a geometric Brownian motion. Indeed, the standard approach for obtaining $S(t)$ is to model the time evolution of the stock price, denoted by $\{\hat{S}(t) : t \geq 0\}$, by means of the stochastic exponential of a Brownian motion process. The above definition of $S(t)$ is such that the marginal distribution of $\{\hat{S}(t) : t \geq 0\}$ at time t is equal to the distribution of $S(t)$.

While Geometric Brownian motion is widely used, it fails to represent the market properly and more accurate models that build on Geometric Brownian motion have been developed in the literature. In the subsequent section we will discuss one of these extensions which is the Variance Gamma model, [75], [76]. The Variance Gamma model is obtained from substituting the time parameter of a Brownian motion describing the continuously compounded return process with a Gamma process as a subordinator. In other words, the value of a stock

$S(t)$ at time t is obtained by evaluating the value of a Brownian motion at time $\tau(t)$, where $\tau(t)$ is the value of a Gamma process at time t. A Gamma process is a stochastic process in which the increments of $\tau(t)$ are independently gamma distributed and the Variance Gamma process denotes the Gamma process subordination of a Brownian motion. The Variance Gamma model has been introduced, though not a complete model in the area of option pricing, as it better mimics the return of an actual stock as it allows for skewness, leptokurticity, and its sample path appears as a series of jumps, which reflects the change in security value due to successive trades on the security.

4.6.1.2 The Variance Gamma Model

In defining the Variance Gamma process, we first give a precise definition of Gamma process $(\tau(t) : t \geq 0)$ that acts as a subordinator for this model. Let $\gamma(a,b)$ denote the gamma distribution with shape parameter a and scale parameter b. In the case $a \in \mathbb{N}$, $\gamma(a,b)$ represents the distribution of the sum of a independent exponential mean $1/b$ random variables. Let ν denote the variance-rate of the time change, then $\tau(t) - \tau(s) = \tau(t-s)$, $0 \leq s < t$, is $\gamma(t/\nu, 1/\nu)$ distributed, i.e., the mean value of $\tau(t)$ is t and the variance of $\tau(t)$ is given by νt.

The Variance Gamma process is determined through choosing appropriate values for its parameters $(\mu, \sigma, \nu, \kappa)$, where μ is, like in Geometric Brownian motion, the instantaneous mean return of the stock, σ denotes the analogous implied volatility, ν is the parameter determining the Gamma process that influences the kurtosis of the stock price process, and κ is an artificial parameter that introduces asymmetry in the model. With mean $\mathbb{E}[\tau(t)] = t$, the Gamma process represents a business 'time' of which each event represents a 'trade' during the trading time of the security. The Gamma process is composed of a countable number of very small positive jumps over any given time interval. Specifically, $\tau(t) < t$ represents trading of the stock in a market that is relatively quiet, and $\tau(t) > t$ represents trading of the stock in a market that is relatively active.

Because this process has more parameters than Geometric Brownian motion, the Variance Gamma process encompasses more features of the security price process. In the literature, efficient statistical estimators exist to compute the parameters $(\mu, \sigma, \nu, \kappa)$, which makes the VG-process interesting from a practical point of view, [75]. Similar to Section 4.6.1.1, we simplify the presentation of the Variance Gamma process as we are only interested in the marginal distribution of the process at some time point t. The following construction of the Variance Gamma process follows [28].

Let $(W(t): t \geq 0)$ denote the Wiener process, the

$$\tilde{Y}(\tau(t)) = \kappa\tau(t) + \sigma W(\tau(t))$$

yields a Variance Gamma process. The price of the asset at time t in the Variance Gamma model is given by $S(0) = s_0$ and

$$S(t) = s_0 \exp\big((\mu+\omega)t + \tilde{Y}(t)\big),$$

where

$$\omega = \frac{1}{\nu}\ln\left(1 - \kappa\nu - \frac{\sigma^2\nu}{2}\right). \tag{4.25}$$

As with Geometric Brownian motion we are only interested in the marginal distribution of $(S(t): t \geq 0)$ at time t. Therefore, we fix t and set

$$\begin{aligned} \alpha(y) := \alpha_t(y) &= \ln s_0 + (\mu+\omega)t + \kappa y \quad \text{and,} \\ \beta(y) := \beta_t(y) &= \sigma\sqrt{y}, \end{aligned}$$

and we let $Y(y) = N(\alpha(y), \beta(y))$ be a normal random variable with mean $\alpha(y)$ and standard deviation $\beta(y)$. Then, the price of the asset at time t in the Variance Gamma model is in distribution equal to

$$S(t) = e^{Y(\tau(t))}.$$

and consequently $Y(\tau(t))$ for fixed $t > 0$ is a normal mean-variance mixture with a gamma mixing distribution. We can write the density function of $S(t)$ in the Variance Gamma model at time $t > 0$, denoted by $\phi^{\text{VG}}(\cdot, t)$, as

$$\phi^{\text{VG}}(x,t) = \int_0^\infty \phi_{\alpha(y),\beta(y)}(x)\, \frac{y^{\frac{t}{\nu}-1} e^{-\frac{y}{\nu}}}{\Gamma\left(\frac{t}{\nu}\right)\nu^{\frac{t}{\nu}}}\, dy,$$

where $\phi_{\mu,\sigma}(x)$ denotes the density function of the normal distribution with mean μ and standard deviation σ. By Fubini's Theorem, the distribution function of $S(t)$ in the VG-model at time $t > 0$ is written as

$$\Phi^{\text{VG}}(x,t) = \int_0^\infty \Phi_{\alpha(y),\beta(y)}(x)\, \frac{y^{\frac{t}{\nu}-1} e^{-\frac{y}{\nu}}}{\Gamma\left(\frac{t}{\nu}\right)\nu^{\frac{t}{\nu}}}\, dy,$$

where $\Phi_{\mu,\sigma}(x)$ denotes the normal distribution function with mean μ and standard deviation σ. In this compilation of market risk examples we will represent the security price process by the Variance Gamma model for the cash-flow of the call option, Section 4.6.2. In this section we we will derive the quantile sensitivity with respect to the initial price s_0 and the implied volatility σ. Both parameters are only present in the normal distribution and hence do not affect the gamma process.

For a parameter θ that presents either in the mean or the standard deviation of the normal distribution function, the measure-valued derivative of the VG-model can be written as follows

$$\frac{\partial}{\partial\theta}\Phi^{\mathrm{VG}}(x,t) = \int_0^\infty \left(\frac{\partial}{\partial\mu}\Phi_{\alpha(y),\beta(y)}(x)\,\frac{\partial}{\partial\theta}\alpha(y) + \frac{\partial}{\partial\sigma}\Phi_{\alpha(y),\beta(y)}(x)\,\frac{\partial}{\partial\theta}\beta(y)\right)\frac{y^{\frac{t}{\nu}-1}e^{-\frac{y}{\nu}}}{\Gamma\left(\frac{t}{\nu}\right)\nu^{\frac{t}{\nu}}}\,dy, \tag{4.26}$$

in which $\partial_\mu\Phi_{\alpha(y),\beta(y)}$ denotes the weak derivative of the normal distribution w.r.t. the mean and $\partial_\sigma\Phi_{\alpha(y),\beta(y)}$ denotes the weak derivative of the normal distribution w.r.t. the standard deviation. The details are provided in the MVD component of Section 1.2.4.2. The interchange of derivative and integral in Equation (4.26) is permitted as the density function of a gamma distribution is bounded, and the normal distribution function is $\mathfrak{B}$-differentiable with respect to both μ and σ.

For the calculations in the subsequent sections, we are interested in the present value, at $t = 0$, of the sensitivities of these portfolios. For this, we denote r to be a short-term interest rate based from a very liquid asset. At a fixed time t, the VaR of a financial portfolio is then discounted by a factor e^{-rt} to represent the present value.

4.6.2 VaR of a Single Option Contract

In this section we first analyse sensitivities of the VaR of a vanilla call option in Section 4.6.2.1. Section 4.6.2.2 is devoted to the Rainbow option.

4.6.2.1 A Vanilla Call Option

In this section we first determine quantile sensitivities of a vanilla call option with respect to the initial price as well as to the implied volatility. With r denoting the short-term interest rate, a purchaser of a call option receives at expiration time t is given by

$$H_t := H(S(t)) = e^{-rt}\max\{S(t) - K, 0\},$$

where $K > 0$ denotes the strike price, i.e., the price at which the stock can be purchased at the expiration date. Alternatively, we denote the vanilla call option payoff via $h(X_{a(t),b(t)})$, in which

$$h(x) = e^{-rt}\max\{e^x - K, 0\},$$

The pay-off mapping $H(x)$ is continuous throughout $\mathbb{R}$ and differentiable on $\mathbb{R}\setminus\{K\}$. As $S(t) \neq K$ with probability one, the conditions for IPA are met and the IPA quantile estimator is applicable.

The α-Quantile Delta Suppose that we are interested in the derivative of the α-quantile of $H(S(t))$ with respect to the initial stock price $S(0) = s_0$. In other words, we are interested in the "α-quantile Delta." We first consider the sock price process via Geometric Brownian motion. In light of the representation $H(S(t)) = h(X(t))$, the main task in applying our estimator is to compute the derivative of the distribution function $X(t)$ w.r.t $S(0)$. Denoting μ and σ to be the mean and standard deviation of a normal distribution function we integrate Equation (1.32) in Section 1.2.4.2 to attain

$$\frac{\partial}{\partial\mu}\Phi_{\mu,\sigma}(x) = -\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad x \in \mathbb{R}.$$

We apply Equation (1.37) with $\theta = s_0$, together with the above expression, yielding

$$\frac{\partial}{\partial S_0}\bigg|_{S(0)=s_0}\Phi_{a(t),b(t)}(x) = -\frac{1}{s_0}\frac{1}{\sqrt{2\pi}b(t)}\exp\left(-\frac{1}{2}\left(\frac{x-a(t)}{b(t)}\right)^2\right),$$

for $x \in \mathbb{R}$. The mapping $h(x)$ has an inverse for $h(x) > 0$ and is written as

$$h^{-1}(x) = \ln\left(xe^{rt} + K\right), \quad x > 0.$$

Inserting the expression into $\partial_{S(0)}\Phi_{a(t),b(t)}(x)$, the parameter derivative component of the derivative estimator, Equation (4.4), yields the single-run estimator

$$D_i(m) = m\left(Z_{\lceil\alpha m\rceil:m} - Z_{\lceil\alpha m\rceil-1:m}\right)\frac{1}{\sqrt{2\pi}b(t)s_0}\exp\left(-\frac{1}{2}\left(\frac{h^{-1}(Z_{\lceil\alpha m\rceil:m}) - a(t)}{b(t)}\right)^2\right).$$

For the Variance Gamma model, using integral result 3.471.9 in Gradshteyn & Rhyzik, [44], the partial derivative w.r.t. $S(0) = s_0$ is of the form

$$\begin{aligned}
&\frac{\partial}{\partial S(0)}\bigg|_{S(0)=s_0}\Phi^{\mathrm{VG}}(x,t) \\
&= -\frac{1}{s_0}\int_0^\infty \frac{1}{\sqrt{2\pi}\beta(y)}\exp\left(-\frac{1}{2}\left(\frac{x-\alpha(y)}{\beta(y)}\right)^2\right)\frac{y^{\frac{t}{\nu}-1}e^{-\frac{y}{\nu}}}{\Gamma\left(\frac{t}{\nu}\right)\nu^{\frac{t}{\nu}}}\,dy \\
&= -\frac{2}{\sqrt{2\pi}\sigma s_0\Gamma(\frac{t}{\nu})\nu^{\frac{t}{\nu}}}\exp\left(\frac{\kappa}{\sigma^2}\eta(x)\right)\left(\frac{\eta^2(x)}{\kappa^2+\frac{2\sigma^2}{\nu}}\right)^{\frac{t}{2\nu}-\frac{1}{4}} K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{\eta^2(x)\left(\kappa^2+\frac{2\sigma^2}{\nu}\right)}\right),
\end{aligned}$$

with

$$\eta(x) = x - \ln s_0 - (\mu + \omega)\,t, \tag{4.27}$$

in which ω is defined in (4.25), and K_n is the modified Bessel function of the second kind of order n. Consequently, the consequent estimator in the form of Equation (4.4) is

$$\frac{2}{\sqrt{2\pi}\sigma s_0 \Gamma(\frac{t}{\nu})\nu^{\frac{t}{\nu}}} m\big(Z_{\lceil \alpha m\rceil:m} - Z_{\lceil \alpha m\rceil -1:m}\big) \exp\Big(\frac{\kappa}{\sigma^2}\eta(h^{-1}(Z_{\lceil \alpha m\rceil:m}))\Big)$$
$$\cdot \left(\frac{\eta^2(h^{-1}(Z_{\lceil \alpha m\rceil:m}))}{\kappa^2 + \frac{2\sigma^2}{\nu}}\right)^{\frac{t}{2\nu}-\frac{1}{4}} K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{\eta^2(h^{-1}(Z_{\lceil \alpha m\rceil:m}))\left(\kappa^2 + \frac{2\sigma^2}{\nu}\right)}\right).$$

Provided that Assumptions **(C1)** to **(C4)** hold, we have a strongly consistent estimator and confidence intervals for the quantile sensitivity. The order statistic is obtained from **Z** which is a sample of m i.i.d. copies of $H(S(t)) = h(X(t))$, respectively $h(Y(\tau(t)))$.

We now verify the assumptions for this α-quantile derivative estimator w.r.t. both pricing models. Assumption **(C1)** is verified because the moment generating function is finite for the normal distribution and by sampling **Z** is an i.i.d. vector. For the Variance Gamma model, a similar line of argument holds for $S(t)$, though we require the more stringent constraint $(2+\delta)\kappa\nu + (2+\delta)^2\sigma^2\nu/2 < 1$ for some $\delta > 0$ for the expectation $\mathbb{E}_\theta[Z^{2+\delta}]$ to be finite. The proof that Assumption **(C2)** is satisfied for both the Geometric Brownian motion and the Variance Gamma model is postponed to the Appendix. We now turn to Assumption **(C3)** and **(C4)**. As h is a Borel mapping, $\mathfrak{B}$-differentiability of the normal and variance gamma distribution implies $\mathfrak{B}$-differentiability of the distribution of the respective option value. Assumption **(C3)** and **(C4)** is then verified for both models after applying Lemmas 4.4 and 4.5.

The α-Quantile Vega Let us consider the same situation but with the parameter being the implied volatility, σ. In parlance, this is considered the α-quantile Vega. To attain the estimator in Equation (4.4) we require the parameter derivative $\partial_\sigma \Phi_{\mu,\sigma}$. Integrating Equation (1.33) from Section 1.2.4.2 and after a cancellation of error function terms, the resulting expression is very similar

$$\frac{\partial}{\partial\sigma}\Phi_{\mu,\sigma}(x) = -\frac{1}{\sqrt{2\pi}}\frac{x-\mu}{\sigma^2}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad x \in \mathbb{R}.$$

Letting $\theta = \sigma$, for Geometric Brownian motion, it follows from Equation (1.37) that

$$\frac{\partial}{\partial\sigma}\Phi_{a(t),b(t)}(x) = -\sigma t\frac{\partial}{\partial\mu}\Phi_{a(t),b(t)}(x) + \sqrt{t}\frac{\partial}{\partial\sigma}\Phi_{a(t),b(t)}(x)$$
$$= \sqrt{t}\left(1 - \frac{x-a(t)}{b^2(t)}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right), \tag{4.28}$$

for $x \in \mathbb{R}$. Setting $\partial_\theta F_\theta(x) = \partial_\sigma \Phi_{a(t),b(t)}(x)$, as provided in (4.28), would then yield the single-run derivative estimator.

We will, however, not choose Equation (4.4) as our estimator but we rather will illustrate the application of the general estimator put forward in (4.6). Inserting the representation for the measure-valued derivatives attained in Equations (1.35) and (1.36) from Section 1.2.4.2 into (4.28) yields

$$\frac{\partial}{\partial\sigma}\phi_{a(t),b(t)}(x) = -\frac{\sqrt{t}}{\sqrt{2\pi}}\left(r^+_{a(t),b(t)}(x) - r^-_{a(t),b(t)}(x)\right) + \frac{1}{\sigma}\left(m_{a(t),b(t)}(x) - \phi_{a(t),b(t)}(x)\right)$$
$$= (c_1 + c_2)\Bigg(\frac{c_1}{c_1+c_2}r^+_{a(t),b(t)}(x) + \frac{c_2}{c_1+c_2}m_{a(t),b(t)}(x)$$
$$- \frac{c_1}{c_1+c_2}r^+_{a(t),b(t)}(x) - \frac{c_2}{c_1+c_2}\phi_{a(t),b(t)}(x)\Bigg) \tag{4.29}$$

for $x \in \mathbb{R}$, where

$$c_1 = \frac{\sqrt{t}}{\sqrt{2\pi}} \quad \text{and} \quad c_2 = \sqrt{t}.$$

We denote $R^+(a(t), b(t)) \sim r^+_{a(t),b(t)}$, $R^-(a(t), b(t)) \sim r^-_{a(t),b(t)}$, being the random variables associated with the respective shifted Rayleigh distribution, and $M(a(t), b(t)) \sim m_{a(t),b(t)}$ being the Double-Maxwell random variable. We define $X^+(t)$ according to

$$X^+(t) = \frac{c_1}{c_1+c_2}R^+(a(t), b(t)) + \frac{c_2}{c_1+c_2}M(a(t), b(t)),$$

i.e., with probability $c_1/(c_1 + c_2)$ let $X^+(t)$ be distributed according to a shifted Rayleigh distribution and with probability $c_2/(c_1 + c_2)$ let $X^+(t)$ follow a Double-Maxwell distribution. In the same vein, let $X^-(t)$ be defined as

$$X^-(t) = \frac{c_1}{c_1+c_2}R^-(a(t), b(t)) + \frac{c_2}{c_1+c_2}N(a(t), b(t)),$$

i.e., with probability $c_1/(c_1 + c_2)$ let $X^-(t)$ be distributed according to a shifted negative Rayleigh distribution and with probability $c_2/(c_1 + c_2)$ let $X^-(t)$ follow a normal distribution. With this in mind, the derivative estimator in Equation (4.6) becomes

$$-m(c_1+c_2)\left(Z_{\lceil\alpha m\rceil:m}-Z_{\lceil\alpha m\rceil:m}\right)\left(1\{h(X^+(t))\le Z_{\lceil\alpha m\rceil:m}\}-1\{h(X^-(t))\le Z_{\lceil\alpha m\rceil:m}\}\right).$$

For the Variance Gamma model, either estimator in Equation (4.4) or (4.6) is determined via the respective depictions of the normal distribution. In determining the measure-valued derivative triple for the re-simulation estimator, there is also the additional complication with the pre-factor depending on τ. This will be discussed shortly.

We can attain a closed form expression for $\partial_\sigma\Phi^{\mathrm{VG}}(x)$ and hence the single-run estimator following Equation (4.26) with $\eta(x)$ defined as in (4.27). After some calculation, with the result of the same integral, 3.471.9 in Gradshteyn & Rhyzik, [44], we get

$$\begin{aligned}
&\frac{\partial}{\partial\sigma}\Phi^{\mathrm{VG}}(x,t)\\
&=\int_0^\infty\frac{1}{\sqrt{2\pi}\beta(y)}\left(\frac{\sigma t}{1-\kappa\nu-\frac{\sigma^2}{2}\nu}-\sqrt{y}\frac{x-\alpha(y)}{\beta(y)}\right)\exp\left(-\frac{1}{2}\left(\frac{x-\alpha(y)}{\beta(y)}\right)^2\right)\frac{y^{\frac{t}{\nu}-1}e^{-\frac{y}{\nu}}}{\Gamma\left(\frac{t}{\nu}\right)\nu^{\frac{t}{\nu}}}\,dy\\
&=\frac{2\exp\left(\frac{\kappa}{\sigma^2}\eta(x)\right)}{\sqrt{2\pi}\sigma\Gamma\left(\frac{t}{\nu}\right)\nu^{\frac{t}{\nu}}}\left(\left(\sigma te^{-\nu\omega}-\frac{\eta(x)}{\sigma}\right)\left(\frac{\eta^2(x)}{\kappa^2+\frac{2\sigma^2}{\nu}}\right)^{\frac{t}{2\nu}-\frac{1}{4}}K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{\eta^2(x)\left(\kappa^2+\frac{2\sigma^2}{\nu}\right)}\right)\right.\\
&\quad\left.+\frac{\kappa}{\sigma}\left(\frac{\eta^2(x)}{\kappa^2+\frac{2\sigma^2}{\nu}}\right)^{\frac{t}{2\nu}+\frac{1}{4}}K_{\frac{t}{\nu}+\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{\eta^2(x)\left(\kappa^2+\frac{2\sigma^2}{\nu}\right)}\right)\right).
\end{aligned}$$

The resulting derivative estimator where $h^{-1}(x)$ for $x>0$ defined previously, is then written as

$$\begin{aligned}
&-\frac{2\exp\left(\frac{\kappa}{\sigma^2}\eta(h^{-1}(Z_{\lceil\alpha m\rceil:m}))\right)}{\sqrt{2\pi}\sigma\Gamma\left(\frac{t}{\nu}\right)\nu^{\frac{t}{\nu}}}m\left(Z_{\lceil\alpha m\rceil:m}-Z_{\lceil\alpha m\rceil:m}\right)\left(\left(\sigma te^{-\nu\omega}-\frac{\eta(h^{-1}(Z_{\lceil\alpha m\rceil:m}))}{\sigma}\right)\right.\\
&\cdot\left(\frac{\eta^2(h^{-1}(Z_{\lceil\alpha m\rceil:m}))}{\kappa^2+\frac{2\sigma^2}{\nu}}\right)^{\frac{t}{2\nu}-\frac{1}{4}}K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{\eta^2(h^{-1}(Z_{\lceil\alpha m\rceil:m}))\left(\kappa^2+\frac{2\sigma^2}{\nu}\right)}\right)\\
&\left.+\frac{\kappa}{\sigma}\left(\frac{\eta^2(h^{-1}(Z_{\lceil\alpha m\rceil:m}))}{\kappa^2+\frac{2\sigma^2}{\nu}}\right)^{\frac{t}{2\nu}+\frac{1}{4}}K_{\frac{t}{\nu}+\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{\eta^2(h^{-1}(Z_{\lceil\alpha m\rceil:m}))\left(\kappa^2+\frac{2\sigma^2}{\nu}\right)}\right)\right).
\end{aligned}$$

Following Equation (4.26), the re-simulation estimator for the derivative w.r.t. σ for Variance Gamma model is attained following from the measure-valued derivative of $\partial_\sigma \phi_{\alpha(y),\beta(y)}$. This leads to similar measure-valued derivative distributions as presented in Equation (4.29):

$$
\begin{aligned}
&\frac{\partial}{\partial\sigma}\phi^{\mathrm{VG}}(x,t) \\
&\quad= \int_0^\infty \left(-\frac{\sigma t}{1-\kappa\nu-\frac{\sigma^2}{2}\nu}\right)\frac{y^{\frac{t}{\nu}-1}e^{-\frac{y}{\nu}}}{\Gamma\left(\frac{t}{\nu}\right)\nu^{\frac{t}{\nu}}}\,dy\frac{\partial}{\partial\mu}\phi_{\alpha(y),\beta(y)}(x)+\sqrt{y}\frac{\partial}{\partial\sigma}\phi_{\alpha(y),\beta(y)}(x) \\
&\quad= \frac{te^{-\nu\omega}}{\sqrt{2\pi y}}\left(r^-_{\alpha(y),\beta(y)}(x)-r^+_{\alpha(y),\beta(y)}(x)\right)+\frac{1}{\sigma}\left(m_{\alpha(y),\beta(y)}(x)-\phi_{\alpha(y),\beta(y)}(x)\right) \\
&\quad= (d_1(y)+d_2)\left(\frac{d_1(y)}{d_1(y)+d_2}r^-_{\alpha(y),\beta(y)}(x)+\frac{d_2}{d_1(y)+d_2}m_{\alpha(y),\beta(y)}(x)\right. \\
&\qquad\qquad\left.-\frac{d_1(y)}{d_1(y)+d_2}r^+_{\alpha(y),\beta(y)}(x)-\frac{d_2}{d_1(y)+d_2}\phi_{\alpha(y),\beta(y)}(x)\right)
\end{aligned}
$$

for $x \in \mathbb{R}$. The pre-factors are given by

$$
d_1(y) = \frac{te^{-\nu\omega}}{\sqrt{2\pi y}} \quad \text{and} \quad d_2 = \frac{1}{\sigma},
$$

where in particular d_1 depends on a gamma random variable. For the measure-valued derivative $\partial_\sigma \phi^{\mathrm{VG}}(x,t)$ itself, the state variable y is replaced with the gamma distributed random variable $\tau(t) \sim \gamma(t/\nu, 1/\nu)$.

To generate the measure-valued derivative densities, following the same notation, we denote $R^+(\alpha(y),\beta(y)) \sim r^+_{\alpha(y),\beta(y)}$, $R^-(\alpha(y),\beta(y)) \sim r^-_{\alpha(y),\beta(y)}$, as the shifted-Rayleigh random variables. In addition, $M(\alpha(y),\beta(y)) \sim m_{\alpha(y),\beta(y)}$ denotes the Double-Maxwell random variable and $N(\alpha(y),\beta(y)) \sim \phi_{\alpha(y),\beta(y)}$ denotes a generation from the normal distribution. The measure-valued derivative random variables $Y^\pm(\tau(t))$ for the Variance Gamma model are gamma distributed mean-variance mixtures with structure analogous to Geometric Brownian motion. We define $Y^+(\tau(t))$ as

$$
Y^+(\tau(t)) = \frac{d_1(\tau(t))}{d_1(\tau(t))+d_2}R^-\big(\alpha(\tau(t)),\beta(\tau(t))\big)+\frac{d_2}{d_1(\tau(t))+d_2}M\big(\alpha(\tau(t)),\beta(\tau(t))\big), \tag{4.30}
$$

that is with probability $d_1(\tau(t))/(d_1(\tau(t))+d_2)$, $Y^+(\tau(t))$ is a shifted Rayleigh mixture with gamma mixing distribution on the negative half-line and with probability $d_2/(d_1(\tau(t))+d_2)$, $Y^+(\tau(t))$ is distributed as a Double-Maxwell mixture distribution. Similarly, $Y^-(\tau(t))$ is described via

$$
Y^-(\tau(t)) = \frac{d_1(\tau(t))}{d_1(\tau(t))+d_2}R^+\big(\alpha(\tau(t)),\beta(\tau(t))\big)+\frac{d_2}{d_1(\tau(t))+d_2}N\big(\alpha(\tau(t)),\beta(\tau(t))\big), \tag{4.31}
$$

i.e., as a shifted Rayleigh mixture distribution on the positive half-line with contingent probability $d_1(\tau(t))/(d_1(\tau(t))+d_2)$, and as a normal mean-variance mixture random variable with probability $d_2/(d_1(\tau(t))+d_2)$, the general estimator. The derivative re-simulation estimator, Equation (4.6), is then

$$-m(d_1(\tau(t))+d_2)\left(Z_{\lceil\alpha m\rceil:m}-Z_{\lceil\alpha m\rceil-1:m}\right) \cdot \left(1\{h_1(Y^+(\tau(t)))\le Z_{\lceil\alpha m\rceil:m}\}-1\{h_1(Y^-(\tau(t)))\le Z_{\lceil\alpha m\rceil:m}\}\right).$$

The order statistic within both estimators are attained from **Z**, a sample of m i.i.d. copies of either $H_1(S(t)) = h_1(X(t))$ or $H_1(S(t)) = h_1(Y(\tau(t)))$. This verifies Assumption **(C1)**. Assumptions **(C3)** and **(C4)** are confirmed identically as in the α-Quantile Delta case, applying Lemmas 4.4 and 4.5. For Assumption **(C2)** we refer this verification to Appendix A.3.

Remark 4.5. *The above analysis shows that $Y^+(\tau(t))$ and $Y^-(\tau(t))$ constructed in Equations (4.30) and (4.31) is a weak derivative of the Variance Gamma process. This is an important result in its own right in derivative estimation, as this is the first unbiased estimator for the derivative of the Variance Gamma process. Unbiased estimation of Malliavin Greeks for the Variance Gamma process is provided in [64], [65].*

4.6.2.2 Rainbow Options

Rainbow Options are contingent financial claims that are based on multiple stocks. Let $S(t) = (S_1(t), S_2(t))$, with $t \ge 0$, denote the joint price vector of two securities, where the price of an individual stock is depicted by Geometric Brownian motion

$$S_i(t) = s_0(i)\exp\left(\left(r-\frac{\sigma_i^2}{2}\right)t+\sigma_i\sqrt{t}W_i\right) := e^{X_i(t)},$$

for $i = 1, 2$, and $W = (W_1, W_2)$, a standard normal random vector. Following the notation in Section 4.6.1.1, we define for each of the two stocks

$$a_i(t) = \ln s_0(i) + \left(r-\frac{\sigma_i^2}{2}\right)t \quad \text{and} \quad b_i(t) = \sigma_i\sqrt{t},$$

the respective mean and standard deviation of the normal random variable denoting the continuously compounded return

$$X_i(t) = a_i(t) + b_i(t)W_i.$$

Since the random vector $X(t)$ is based on the bivariate standard vector W, for the moment this density function is our focus of interest. The Lebesgue density of W is written as

$$\phi_\rho(x) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(x-a(t))^\top \Sigma^{-1}(x-a(t))\right), \tag{4.32}$$

for $x = (x_1, x_2)^\top$, and $\top$ denotes the transpose operation. The covariance matrix Σ is given by

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

with $\rho \in [-1, 1]$ being the correlation of $X_1(t)$ and $X_2(t)$. Specifically,

$$\det\Sigma = 1-\rho^2 \quad \text{and} \quad \Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}.$$

For $x = (x_1, x_2)^\top$, we expand

$$x^\top \Sigma^{-1} x = \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2},$$

and therefore write the Lebesgue density in Equation (4.32) as a product of the density of the conditional random variable W_1 given W_2 and the univariate density denoting the random variable W_2:

$$\begin{aligned}\phi_\rho(x) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{x_1-\rho x_2}{\sqrt{1-\rho^2}}\right)^2\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_2^2\right).\end{aligned}$$

In the following we consider the spread option and consider the derivative of the quantile w.r.t. the parameter ρ. In this instance $N = 1$ and $J = 2$. Let

$$H(S(t)) = \max\{S_2(t) - S_1(t) - K, 0\},$$

for $K > 0$. For $x = (x_1, x_2)$ we set

$$h(x) = \max\left\{e^{x_2} - e^{x_1} - K, 0\right\},$$

then

$$H(S(t)) = h(X_1(t), X_2(t)).$$

The parameter ρ is only present in the conditional density function. As the parameters $a_i(t), b_i(t)$, for $i = 1,2$, are independent of ρ, the eventual final measure-valued derivative for $X(t)$ will just involve substituting $W_1|W_2$ for other distributions to attain $X_1^{\pm}(t)$.

Provided above, the conditional density function of W_1 given $W_2 = x_2$ is $\phi_{\rho x_2, \sqrt{1-\rho^2}}$. Evoking the MVD results for the normal distribution in Section 1.2.4.2 it follows that

$$\frac{\partial}{\partial \rho}\phi_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2) = x_2 \frac{\partial}{\partial \mu}\phi_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2) - \frac{\rho}{\sqrt{1-\rho^2}}\frac{\partial}{\partial \sigma}\phi_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2).$$

By conditioning w.r.t. W_2, we can attain the parameter derivative $\Phi_\rho(x_1|x_2)$ by integration. And by fixing W_2, $h(\cdot, x_2)$ becomes invertible and the derivative estimator in the form of Equation (4.4) can be written.

Instead, we show a method to represent the measure-valued derivatives so as to attain an estimator in the form of Equation (4.6), respectively (4.7). Rearranging the positive and negative parts, we arrive at

$$\begin{aligned}
&\frac{\partial}{\partial \rho}\phi_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2)\\
&= \frac{x_2}{\sqrt{2\pi}\sqrt{1-\rho^2}}\left(r^{+}_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2) - r^{-}_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2)\right)\\
&\quad - \frac{\rho}{1-\rho^2}\left(m_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2) - \phi_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2)\right)\\
&= (c_1(x_2)+c_2)\left(\frac{c_1(x_2)}{c_1(x_2)+c_2} r^{+}_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2) + \frac{c_2}{c_1(x_2)+c_2}\phi_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2)\right.\\
&\qquad\left. - \frac{c_1(x_2)}{c_1(x_2)+c_2} r^{-}_{!\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2) - \frac{c_2}{c_1(x_2)+c_2} m_{\rho x_2,\sqrt{1-\rho^2}}(x_1|x_2)\right),
\end{aligned}$$

in which

$$c_1(x_2) = \frac{x_2}{\sqrt{2\pi}\sqrt{1-\rho^2}} \quad \text{and} \quad c_2 = \frac{\rho}{1-\rho^2}.$$

The density functions $r^{+}_{\rho x_2,\sqrt{1-\rho^2}}$, $r^{-}_{\rho x_2,\sqrt{1-\rho^2}}$ are Rayleigh densities on opposite semi-infinite supports, $(-\infty, \rho x_2)$, and $[\rho x_2, \infty)$ respectively. As the mean is dependent on the random variable W_2, we denote the random variable associated with the density function by $R^{\pm}|W_2$, in which $R^{+}|(W_2 = x_2) \sim r^{+}_{\rho x_2,\sqrt{1-\rho^2}}$, and similarly for $R^{-}|(W_2 = x_2)$. In terms of an affine transformation of a standard Rayleigh random variable H, $R^{\pm}|W_2 = \rho W_2 \pm \sqrt{1-\rho^2}H$. Following the same reasoning, we denote the random variable corresponding to the Double-Maxwell density function with W_2-dependent mean by $M|W_2$, i.e., the random variable $M|(W_2 = x_2) \sim m_{\rho x_2, \sqrt{1-\rho^2}}$. From a standardized Double-Maxwell distributed random variable, the random variable $M|W_2$ is described by the same affine transformation. The final form of the positive component W_1^+ is

$$W_1^+ = \frac{c_1(W_2)}{c_1(W_2)+c_2}(R^+|W_2) + \frac{c_2}{c_1(W_2)+c_2}(W_1|W_2),$$

namely, a shifted Rayleigh random variable with normally distributed probability $c_1(W_2)/(c_1(W_2)+c_2)$ and the original conditioned normal random variable with probability $c_2/(c_1(W_2)+c_2)$. For the random variable W_1^-, this expression is similar

$$W_1^- = \frac{c_1(W_2)}{c_1(W_2)+c_2}(R^-|W_2) + \frac{c_2}{c_1(W_2)+c_2}(M|W_2),$$

namely, a shifted Rayleigh random variable with normally distributed probability $c_1(W_2)/(c_1(W_2)+c_2)$ or otherwise a conditional Double-Maxwell distributed random variable.

The corresponding measure-valued derivative random vectors are denoted as $X^{\pm}(t) = (X_1^{\pm}(t), X_2(t))^{\top}$, where $X_1^{\pm}(t)$ is obtained from

$$X_1^{\pm}(t) = a_1(t) + b_1(t) W_1^{\pm},$$

and the general estimator in Equation (4.6) becomes

$$-\,m(c_1(W_2)+c_2)\left(1\left\{h\left(X_1^+(t), X_2(t)\right) \le Z_{\lceil \alpha m\rceil:m}\right\} - 1\left\{h(\left(X_1^-(t), X_2(t)\right) \le Z_{\lceil \alpha m\rceil:m}\right\}\right).$$

The order statistic $Z_{\lceil \alpha m\rceil:m}$ is obtained from an i.i.d. sample $\mathbf{Z}$ consisting of m copies of the random variable $Z = h(X_1(t), X_2(t))$. The pre-factor gives the normalizing constant originating from the mixture interpretation of the weak derivative. The fact that Assumption **(C1)**, **(C2)** and **(C4)** hold follows readily from arguments provided in Section 4.6.2.1. For Assumption **(C3)** we cannot apply the product rule of weak differentiation to (W_1, W_2) as the components are not independent. However, since the mapping $1\{h(x_1,x_2) \le z\} \in \mathfrak{B}$, it follows from the Dominated Convergence Theorem that

$$\begin{aligned}
&\frac{\partial}{\partial\rho}\int_{\mathbb{R}^2} 1\{h(x_1,x_2)\le z\}\, \phi_{\rho x_2, \sqrt{1-\rho^2}}(x_1|x_2)\,\phi(x_2)\,dx_1\,dx_2 \\
&\quad = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} 1\{h(x_1,x_2)\le z\}\,\frac{\partial}{\partial\rho}\phi_{\rho x_2, \sqrt{1-\rho^2}}(x_1|x_2)\,dx_1\right)\phi(x_2)\,dx_2.
\end{aligned}$$

for any z, which implies that $h(X_1(t), X_2(t))$ is $\mathfrak{B}$-differentiable w.r.t. ρ. Assumption **(C3)** then follows from Lemma 4.4.

In Appendix B we provide related measure-valued derivatives w.r.t. a mean and standard deviation parameter of a marginal random variable for the bivariate normal distribution. The corresponding random generation methods are also provided. This Appendix also provides a variation in obtaining the measure-valued derivatives w.r.t. the correlation and these aforesaid parameters

4.6.3 VaR of a Stock Portfolio

We now consider a stock portfolio consisting of only long equity positions within a single index, i.e. as introduced in Section 4.6. This portfolio contains J stocks, with $w = (w_i : 1 \leq i \leq J)$, denoting the number of stocks purchased for each public listed company, i.e. $w_i > 0$. The price of each stock is simply given by $H_j(S(t)) = S_j(t)$. The uncertain value of the financial assets in the next period are given by $v(w, S(t))$. The value of this portfolio is then

$$v(w, S(t)) := \sum_{j=1}^{J} w_j S_j(t),\ t \geq 0.$$

We denote the distribution of the value of this portfolio by F_θ. We assume that θ is a distributional parameter, denoting a risk factor related to the stock prices at the end of the next period.

To provide structure to the portfolio model we consider the single-index model for the return of each stock, see [8]. This model first introduced in [95]. Specifically, let X_i, $i = 1, \ldots, J$, denote the return of stock j above the short-term interest rate r at a fixed time $t > 0$, given by

$$X_i - r = \alpha_i + \beta_i (X_M - r) + E_i. \tag{4.33}$$

This return is based on two components: the excess rate of the return of the stock index X_M above the short-term interest rate r given by $\beta_i(X_M - r)$ where β_i denotes a stock specific parameter expressing the extent of the incremental increase of the return of the stock due to an incremental increase in the index return, and the firm-specific return for each stock $\alpha_i + E_i$, defined as the sum of the mean return α_i and the random variable denoting the innovation E_i.

For the purposes of modelling prices, we convert the random variables in Equation (4.33) to stochastic processes, ensuring that the mean and variance match the single-index model. An early reference of this procedure is [79]. Particularly, we represent the stock return $X_i(t)$, $i = 1, \ldots, J$ as stochastic differential equations, with $X_j(0) = 0$ almost surely and $X_i(t)$, $t > 0$ being the sum of the return of a stock due to systematic market movements $X_i^s(t)$ and the stock return that is specific to the firm $X_i^f(t)$:

$$dX_i(t) = dX_i^s(t) + dX_i^f(t).$$

By construction, we define the process $X_i^s(t)$ as the sum of the return rt from the very liquid asset and the excess return from the index, where $X_M(t)$ represents the stochastic process of the index return:

$$dX_i^s(t) = r\,dt + \beta_i(dX_M(t) - r\,dt).$$

We define both the excess return component $X_M(t)$ and the firm-specific return components $X_i^f(t)$ as Brownian motions with respective Wiener processes $W_M(t)$ and $W_i(t)$. Then, the stochastic differential equations for $X_M^s(t)$ and $X_i^f(t)$ are written as

$$\begin{aligned} dX_i^s(t) &= \big(r - \beta_i(x_M - r)\big)\,dt + \beta_i\sigma_M\,dW_M(t) \\ dX_i^f(t) &= \alpha_i\,dt + \sigma_i\,dW_j(t), \end{aligned}$$

where x_M is the mean return for the stock index, σ_M is the volatility of the stock index, and σ_i is the firm-specific volatility of stock i separate form the stock index. Moreover, we denote by $\sigma_{i,M}$ the covariance between the firm-specific return of stock i and the index return, and with this notation, from [8], the parameter β_i is defined as

$$\beta_i = \frac{\sigma_{i,M}}{\sigma_M^2}. \tag{4.34}$$

All of the parameters are statistics of returns that are continuously compounded. For the previously unmentioned parameters, x_M is the mean return for the market index, σ_M is the volatility for the market index, and σ_i is the firm-specific volatility of stock j that is separated from the market index. For the firm-specific return component, the Wiener processes $(W_i(t) : 1 \le i \le J)$ are independent of $W_M(t)$ and are mutually independent of each other. The resulting price process S is a multivariate Geometric Brownian motion, obtained via the stochastic exponential, and for each of the J stocks, the price process is given by

$$d\hat{S}_i(t) = \hat{S}_i(t)dX_i(t).$$

The general solution of this equation is given in [60]. For our purposes, the marginal distribution of each price process is for $i = 1, \dots, J$:

$$\begin{aligned} S_i(t) = s_0(i)\exp&\left(\left((r + \beta_j(x_M - r)) - \frac{1}{2}\beta_i^2\sigma_M^2\right)t + |\beta_i|\sigma_M\sqrt{t}W_M\right) \\ &\cdot\exp\left(\left(\alpha_j - \frac{1}{2}\sigma_{\alpha_j}^2\right)t + \sigma_j\sqrt{t}W_j\right), \end{aligned} \tag{4.35}$$

where W_M and W_j are standard normal random variables. More beneficial to our calculations is the rewriting of the above the expression in the form of independent components

$$S_i(t) = s_0(i)\, e^{X_i^s(t)+X_i^f(t)},$$

similar to Section 4.6.1.

In the distributional expression, we replace β_i by the RHS of (4.34). For each stock and $t > 0$, $a_i^s := a_i^s(t)$ and $b_i^s := b_i^s(t)$ denotes the respective mean and standard deviation of the return component due to systematic index movements. Analogously, $a_i^f := a_i^f(t)$ and $b_i^f := b_i^f(t)$ are the respective mean and standard deviation of the firm-specific return, i.e.:

$$\begin{pmatrix} a_j^s \\ a_j^f \end{pmatrix} = \begin{pmatrix} \left(r + \frac{\sigma_{j,M}}{\sigma_M^2}(x_M - r) - \frac{1}{2}\frac{\sigma_{j,M}^2}{\sigma_M^2}\right)t \\ \left(\alpha_j - \frac{1}{2}\sigma_j^2\right)t \end{pmatrix},$$

$$\begin{pmatrix} b_j^s \\ b_j^f \end{pmatrix} = \begin{pmatrix} \frac{|\sigma_{j,M}|}{\sigma_M}\sqrt{t} \\ \sigma_j\sqrt{t} \end{pmatrix}.$$

For each stock, $i = 1,\ldots,J$, we define the distribution function for the systematic index component of the return by $F_{\theta,i}^s$. Similarly, the firm-specific return distribution is written as $F_{\theta,i}^f$. As before, θ denotes a distributional risk parameter. In this example, the return components for a stock are normally distributed with corresponding mean and standard deviation, i.e., $F_{\theta,i}^s(x) = \Phi_{a_i^s,b_i^s}(x)$ and $F_{\theta,i}^f(x) = \Phi_{a_i^f,b_i^f}(x)$.

The sensitivity $\partial_\theta q_{\alpha,t}(w,S(t))$ will allow us to analyse the uncertainty of the VaR estimator due to small changes in the parameters. We will analyse the VaR sensitivity with respect to the firm-specific return component $\theta = \alpha_1$, and the volatility of the index, σ_M. In this section, we will determine the sensitivity of the distribution of function F_θ w.r.t. these parameters. The sensitivity of the quantile of the portfolio can then be estimated by combining the distributional estimator with the inverse density estimator, see Section 4.4.3.

In the simulation, let $\mathbf{Z} = (Z_i : 1 \leq i \leq n)$ be an i.i.d. sample of $v(w,S(t))$. In turn, a random variable for $v(w.S(t))$ is composed from $J+1$ random variables: one innovation W_M due to the return from the index and the firm-specific innovations W_i from each of the J stocks. We assume that both the vector of initial prices for each stock $S(0) = (s_0(i) : 1 \leq i \leq J)$ and the number of shares purchased for each stock, w, is fixed.

For $i = 1,\ldots,J$, let $x_i^s := x_i^s(t)$, $x_i^f := x_i^f(t)$, denote respectively a realization for the continuously-compounded systematic index return component and the corresponding firm-specific return over the period $(0, t]$. Following Equation (4.35), a realization of the value of the portfolio is written as

$$v(w, S(t)) = \sum_{i=1}^{J} w_i s_0(i) e^{x_i^s} e^{x_i^t}$$

so that the loss incurred by the portfolio over $(0, t]$ is given by

$$V(w, S(t)) = -(v(w, S(t)) - v(w, S(0))) = \sum_{i=1}^{J} w_i s_0(i) \left(1 - e^{x_i^s} e^{x_i^f}\right).$$

Note that $V(w, S(t))$ takes values in $\left(-\infty, \sum_{i=1}^{J} w_i s_0(i)\right)$.

We proceed by first determining the VaR sensitivity w.r.t. α_1. For this sensitivity, we condition on all of the random variables except for $X_1^f(t)$. Let $\hat{x}_{\bar{1}} = (\hat{x}_2, \ldots, \hat{x}_J)$ where for $2 \le i \le J$, $\bar{x}_i = (x_i^s, x_i^f)$ and for $z \in \left(-\infty, \sum_{i=1}^{J} w_i s_0(i)\right)$ we define

$$V^{-1}(x_1^s, \hat{x}_{\bar{1}})(z) = \ln\left(1 - \frac{1}{w_1 s_0(1)}\left(z - \sum_{i=2}^{J} w_i s_0(i)\left(1 - e^{x_i^s} e^{x_i^f}\right)\right)\right) - x_1^s. \tag{4.36}$$

Then, by Lemma 4.7 it holds for this sensitivity that

$$\frac{\partial}{\partial\theta}\mathbb{P}(V(w, S(t)) \le z) = \frac{d}{d\theta} F_\theta(z) = -\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta} F_{\theta,1}^f\left(V^{-1}(X_1^s(t), \hat{X}_{\bar{1}}(t))(z)\right)\right],$$

for $z \in \left(-\infty, \sum_{i=1}^{J} w_i s_0(i)\right)$. Analogously, $\hat{X}_{\bar{1}}(t) = (\hat{X}_2(t), \ldots, \hat{X}_J(t))$ where for $2 \le i \le J$, $\hat{X}_i(t) = (X_i^s(t), X_i^f(t))$. Given differentiable functions a_i^p, $i = 1, \ldots, J$, $p = s, f$, then from Lemma 4.7, conditioned on the realizations $X_1^s(t) = x_1^s, \hat{X}_1, \hat{X}_2(t) = \hat{x}_2, \ldots, \hat{X}_J(t) = \hat{x}_J(t)$, we obtain

$$\begin{aligned}
\frac{\partial}{\partial\alpha_1} F_\theta(z) &= -\frac{\partial}{\partial\mu}\Phi_{a_1^f(\alpha_1), b_1^f(\alpha_1)}(V^{-1}(x_1^s, \hat{x}_{\bar{1}})(z))\frac{\partial a_1^f}{\partial\alpha_1} \\
&= \frac{t}{\sqrt{2\pi} b_1^f(\alpha_1)} \exp\left(-\frac{1}{2}\left(\frac{V^{-1}(x_1^s, \hat{x}_{\bar{1}})(z) - a_1^f(\alpha_1)}{b_1^f}\right)^2\right).
\end{aligned}$$

For $1 \le j \le l$, let $V_j^{-1}(z) = V^{-1}(X_1^s(t, j), \hat{X}_{\bar{1}}(t, j))(z)$, where $(X_1^s(t, j), \hat{X}_{\bar{1}}(t, j))$ are i.i.d. random variables. Hence, letting $z = Z_{\lceil\alpha m\rceil:m}$, we obtain

$$D^\alpha(m, k) = -\text{dens}_{m,k}^\alpha \times \frac{1}{\sqrt{2\pi} b_1^f(\alpha_1)} \frac{1}{l} \sum_{j=1}^{l} \exp\left(-\frac{1}{2}\left(\frac{V_j^{-1}(Z_{\lceil\alpha m\rceil:m}) - a_1^f(\alpha_1)}{b_1^f}\right)^2\right)$$

as our estimator. Note that $Z_{\lceil \alpha m \rceil : m}$ is by construction an element of $\left(-\infty, \sum_{i=1}^{J} w_i s_0(i)\right)$, which implies that $V^{-1}(X_1^s(t), \hat{X}_{\bar{1}}(t))(z)$ is almost surely well defined.

We now turn to the VaR sensitivity w.r.t. σ_M. Note that σ_M is present in both a_i^s, b_j^s, $i = 1, \dots J$. The method of computation follows from Lemma 4.7. Let $\rho = \{1, \dots, J\}$ be uniformly distributed and similar to (4.36) and let

$$V_\rho^{-1}(z) = V(\hat{x}_\rho^f, \hat{x}_{\bar{\rho}})(z) = \ln\left(1 - \frac{1}{w_\rho s_0(\rho)}\left(z - \sum_{\substack{i=1 \\ i \neq \rho}}^{J} w_i s_0(i)\left(1 - e^{x_j^s} e^{x_j^f}\right)\right)\right) - x_\rho^s,$$

for $z \in \left(-\infty, \sum_{i=1}^{J} w_i s_0(i)\right)$, where $\hat{x}_{\bar{\rho}} = (\bar{x}_1, \dots, \bar{x}_{\rho-1}, \bar{x}_{\rho+1}, \dots, \bar{x}_J)$. Together with the chain rule, the distribution function sensitivity is equal to

$$\begin{aligned}
\frac{\partial}{\partial \sigma_M} F_{\sigma_M}(z) &= -J\mathbb{E}_{\sigma_M}\left[\left(\frac{\partial}{\partial \mu}\Phi_{a_\rho^s(\sigma_M), b_\rho^s(\sigma_M)}(V_\rho^{-1}(z))\frac{\partial}{\partial \sigma_M} a_\rho^s(\sigma_M)\right.\right. \\
&\qquad \left.\left. + \frac{\partial}{\partial \sigma}\Phi_{a_\rho^s(\sigma_M), b_\rho^s(\sigma_M)}(V_\rho^{-1}(z))\frac{\partial}{\partial \sigma_M} b_\rho^s(\sigma_M)\right)\right] \\
&= -J\mathbb{E}_{\sigma_M}\left[\frac{1}{\sqrt{2\pi} b_\rho^s(\sigma_M)\sigma_M}\left(V_\rho^{-1}(z) - 2rt - a_\rho^s(\sigma_M)\right)\right. \\
&\qquad \left. \cdot \exp\left(-\frac{1}{2}\left(\frac{V_\rho^{-1}(z) - a_\rho^s(\sigma_M)}{b_\rho^s(\sigma_M)}\right)^2\right)\right].
\end{aligned}$$

For $1 \leq j \leq l$, let $\rho(j) \in \{1, \dots, J\}$ be l i.i.d. uniformly distributed random variables and, analogous to the earlier example, let $V_{\rho(j)}^{-1}(z) := V^{-1}(X_\rho^f(t,j), \hat{X}_{\rho(\bar{j})}(t,j))(z)$. The estimator $D^\alpha(m,k)$ is then given by

$$\begin{aligned}
D^\alpha(m,k) = -J\text{dens}_{m,k}^\alpha \times \frac{1}{l}\sum_{j=1}^{l} \frac{1}{\sqrt{2\pi} b_{\rho(j)}^s \sigma_M}&\left(V_{\rho(j)}^{-1}(Z_{\lceil \alpha m \rceil : m}) - 2rt - a_{\rho(j)}^s(\sigma_M)\right) \\
&\cdot \exp\left(-\frac{1}{2}\left(\frac{V_{\rho(j)}^{-1}(Z_{\lceil \alpha m \rceil : m}) - a_{\rho(j)}^s(\sigma_M)}{b_{\rho(j)}^s(\sigma_M)}\right)^2\right).
\end{aligned}$$

Alternatively, elaborating on the random variable interpretation of the distributional derivatives, the estimator put forward in Lemma 4.7 can be obtained.

Conclusion

In this chapter we provided expressions for quantile sensitivities that can be used for gradient estimation by means of Monte Carlo simulation. The examples illustrate the flexibility of our measure-valued differentiation and statistical spacing theory based approach. Specifically, multivariate problems and models containing the Variance-Gamma process can be dealt with as well. Future research will be on the extension of our approach to time-dependent queueing networks.