On the calculations of tail-based risk functionals under the truncated multivariate normal distributions

Chong Yu [1], Jianxi Su [1], and Wei Wei*[2]

[1]Department of Statistics, Purdue University, West Lafayette, IN, USA

[2]Program of Actuarial Science and Risk Management, University of Illinois Urbana-Champaign, Champaign, IL, USA

Abstract

Multivariate normal distributions have long been a popular choice among practitioners for parametric modeling. However, their direct application to positive or lower-bounded data commonly encountered in insurance and risk management, is inherently limited due to their unbounded support. To address this limitation, we contribute to the growing literature that advocates for the truncated multivariate normal distributions as a natural solution. Specifically, we establish their tractability in risk functional calculations. We focus on two widely adopted tail-based risk functionals, namely the quantile conditional allocation and tail conditional allocation. We derive computationally tractable formulas for calculating these risk functionals under the assumption that losses follow a truncated multivariate normal distribution. An efficient numerical scheme is designed to facilitate the practical implementation of these formulas.

Keywords and phrases: Euler allocation, conditional tail expectation, quantile capital allocation, conditional moment generating functions, differential equations

JEL Classification: C34, C46

*Corresponding author; postal address: 605 E Spingfield Ave., Champaign, IL 61801, USA; email: weiw@illinois.edu

1 Introduction

Suppose that $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$, $d \in \mathbb{N}$, represents a portfolio of dependent insurance losses. Using multivariate distributions to analyze the joint behavior of $\boldsymbol{X}$ is an important task actuaries oftentimes need to deal with. In this context, a large group of practitioners advocate for the use of multivariate normal distributions. The widespread preference for multivariate normal distributions among actuaries arguably stems from their structural transparency in terms of using a simple covariance matrix parameter to capture the heterogeneous pair-wise correlations inherent in $\boldsymbol{X}$, as well as their analytical tractability for computing various risk quantities.

Due to the inherent nature of insurance problems including the non-negativity of losses and/or the presence of deductibles in insurance policies, actuarial modeling often involves data that are positive or bounded from below. In contrast, multivariate normal distributions have unbounded support. Hence, they may not be immediately suitable for modeling insurance loss data. To address this issue, one approach is to apply a (shifted) logarithm transform to the lower bounded data, converting them into an unbounded form. The transformed data can then be modeled using multivariate normal distributions. Essentially, this approach is equivalent to fitting the multivariate (shifted) log-normal distributions (Fang et al., 1990) to the original data. This transformation, however, significantly reduces the mathematical tractability of the original multivariate normal distributions, making the transformed model less appealing for practical applications, particularly when efficient calculations of risk functionals are required. For instance, even when the components of $\boldsymbol{X}$ follow log-normal distributions with mutual independence, advanced mathematical tools such as generalized gamma convolution, Padé approximation, numerical inversion of Laplace transform, are already required to handle its aggregate distribution and risk functional calculations (Furman et al., 2020). The complexity is further compounded when dependencies among the components of $\boldsymbol{X}$ are introduced.

Another way to address the aforementioned problem when dealing with lower bounded data is by truncating the multivariate normal distributions. The study of truncated multivariate normal distributions has attracted considerable attention from the statistical research community (see Cohen, 1991, for a comprehensive treatment). Recent theoretical investigations suggest that several key tractable

properties of the multivariate normal distributions can largely be preserved in their truncated counterparts. This includes properties such as moment calculation (Arismendi, 2013), dependence properties (Horrace, 2005), independence properties (Levine et al., 2020), and maximum likelihood estimation properties (Levine et al., 2023). Consequently, the truncated multivariate normal distributions have found fruitful applications in numerous domains, including but not limited to multivariate regression (Amemiya, 1974), econometrics (Hong and Shum, 2003), and educational studies (Cohen, 1991). Somewhat surprisingly, to the best of our knowledge, the truncated multivariate normal distributions have not yet been widely adopted as modeling tools by actuaries. Our paper carries a significant effort to illuminate the their suitability for actuarial applications, with a particular focus placed on risk functional calculations. By doing so, we hope to draw more attention from the actuarial community to this compelling yet underrated class of models.

In this paper, we focus on the calculations of *quantile condition allocation* (QCA) and *tail conditional allocation* (TCA) under the truncated multivariate normal distributions. QCA and TCA are two tail-based risk functionals that play a pivotal role in quantitative risk management within the current regulatory framework. However, computing QCA via simulation is not straightforward, as the estimation process requires careful tuning of the invovled kernel parameters to ensure the validity and efficiency of the associated estimate (Gribkova et al., 2023). While TCA can be often computed via simulation, our paper contributes to deriving a deterministic and efficient approach for its evaluation, which eliminates the need for conducting repeated trials to ensure the result's accuracy, as required in the simulation method.

It is noteworthy that the QCA and TCA of truncated multivariate normal distributions can be, respectively, computed as the component-wise conditional expectation of the original untruncated multivariate normal vector $\boldsymbol{Y}$, given that $\boldsymbol{Y}$ falls into regions:

$$\mathcal{S}_{\text{QCA}} = \{\boldsymbol{y} \in \mathbb{R}^d : y_1 > c_1, \ldots, y_d > c_d,\, y_1 + \cdots + y_d = s\}$$

and

$$\mathcal{S}_{\mathrm{TCA}} = \{\boldsymbol{y} \in \mathbb{R}^d : y_1 > c_1, \ldots, y_d > c_d,\, y_1 + \cdots + y_d > s\}$$

for some constants $c_1, \ldots, c_d, s \in \mathbb{R}$. These regions correspond to tail scenarios in which the risks/losses of all marginal units exceed certain (default) thresholds, and the total risk/loss is either equal to (in the case of QCA) or greater than (in the case of TCA) another (warning) level. Although earlier works have studied the conditional expectations of multivariate normal distributions given various tail conditions (Landsman et al., 2016, 2018; Landsman and Valdez, 2003; Ogasawara, 2021; Zuo et al., 2024), to the best of our knowledge, none has addressed the conditional expectation on the set $\mathcal{S}_{\mathrm{QCA}}$ or $\mathcal{S}_{\mathrm{TCA}}$ mentioned above. Assuming that the distribution of $\boldsymbol{Y}$ belongs to a boarder class of exponential dispersion models (Jørgensen, 1987), Shushi and Yao (2020) studied the conditional expectation of $\boldsymbol{Y}$ given that $\boldsymbol{Y}$ falls into a general set Ω. However, in order to pursue theoretical generality, the results derived in Shushi and Yao (2020) are presented as integrals over the region Ω. Despite their extensive theoretical investigation, the numerical implementation of these integration expressions remains largely unexplored, especially if Ω is non-rectangular, such as $\mathcal{S}_{\mathrm{QCA}}$ and $\mathcal{S}_{\mathrm{TCA}}$.

It is important for us to clarify the intention of this paper. We do not aim to argue that the (truncated) multivariate normal distributions always outperform other multivariate models in fitting insurance data, such as those constructed using copulas (e.g., Cossette et al., 2013; Frees and Valdez, 1998; Su and Hua, 2017), stochastic factors (e.g., Landsman and Shushi, 2022; Su and Furman, 2017), and mixtures (e.g., Bladt, 2023; Lee and Lin, 2012; Sarabia et al., 2018). Instead, by establishing their tractability in risk functional calculations, our message is that if a risk analyst strongly favors a multivariate normal framework for handling data with lower-bounded support, then the truncated multivariate normal distributions should be given serious consideration.

The rest of this paper is organized as follows. In Section 2, we provide formal definitions for the QCA and TCA risk functionals, as well as the truncated multivariate normal distributions. Under this distributional assumption, we derive computationally tractable expressions for QCA and TCA, which consist of weighted sums of normal distribution probabilities over simplex regions. In Section 3, we

treat these simplex-area probabilities as solutions of differential equations and introduce a numerical approach for their efficient evaluation. Section 4 contains numerical examples to illustrate the precision and efficiency of the proposed methodology. Section 5 concludes this paper.

2 The truncated multivariate normal distributions and the tail-based risk functionals

To define the truncated multivariate normal distributions, let us begin by recalling the component-wise partial ordering on d-dimensional Euclidean space, $\mathbb{R}^d$: For column vectors $\boldsymbol{u} = (u_1, \ldots, u_d)^\top$ and $\boldsymbol{v} = (v_1, \ldots, v_d)^\top$ in $\mathbb{R}^d$, $\boldsymbol{u} \geq \boldsymbol{v}$ means $u_i \geq v_i$ for all $i \in \mathcal{D} := \{1, \ldots, d\}$. In the context of risk management, it is conventional for loss models to have infinite upper bounds, in order to adequately capture tail risks. For this reason, the truncated multivariate normal distributions we consider in this paper only involve lower truncation.

A random vector (RV) $\boldsymbol{X} \in \mathbb{R}^d$ is said to follow a truncated multivariate normal distribution, with truncation point $\boldsymbol{c} \in \mathbb{R}^d$, if its probability density function is given by

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{c}) = C^{-1} \frac{1}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right), \qquad \boldsymbol{x} > \boldsymbol{c}, \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}$ is a $d \times d$ positive definite matrix, representing the location and dispersion parameters, respectively. Here, the normalizing constant is computed by

$$C := \overline{\Phi}_d(\boldsymbol{c}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) := \int_{\boldsymbol{x} > \boldsymbol{c}} \phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathrm{d}\boldsymbol{x},$$

where $\overline{\Phi}_d$ and ϕ_d denote the joint survival function and probability density function (PDF) of the d-dimensional multivariate normal distribution. Succinctly, we write $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{c})$.

Remark 1. *Let $\boldsymbol{c} \downarrow -\boldsymbol{\infty}$, then $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{c})$ reduces to the usual untruncated multivariate normal distribution, which is denoted by $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In this case, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix of $\boldsymbol{X}$, respectively. However, when $\boldsymbol{c} > -\boldsymbol{\infty}$, this distributional interpretation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ no longer holds. For calculations of the mean and covariance matrix of $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{c})$, we refer readers to*

Lemma 4.1 of Levine et al. (2023).

Next, we proceed to defining the two tail-based risk functionals we aim to study in this paper. Recall that the well-known Value-at-Risk (VaR) risk measure for a loss RV $X \in \mathbb{R}$ is defined as

$$\mathrm{VaR}_q(X) := \inf\{x \in \mathbb{R} : F_X(x) \geq q\}, \quad q \in [0,1),$$

where F_X denotes the cumulative distribution function (CDF) of X. Let $S_X := X_1 + \cdots + X_d$ be the aggregate risk RV of $\boldsymbol{X} = (X_1, \ldots, X_d)$, and set $s_q := \mathrm{VaR}_q(S_X)$. The QCA risk functional is formulated as

$$\mathrm{QCA}_q(X_i, S_X) := \mathbb{E}(X_i \,|\, S_X = s_q), \qquad i \in \mathcal{D} := \{1, \ldots, d\},\ q \in [0,1). \tag{2}$$

The TCA risk functional is given by

$$\mathrm{TCA}_q(X_i, S_X) := \mathbb{E}(X_i \,|\, S_X > s_q), \qquad i \in \mathcal{D},\ q \in [0,1). \tag{3}$$

In actuarial science, risk functionals (2) and (3) have been extensively studied in the context of economic capital allocation. These functionals can be derived using the celebrated notion of Euler allocation (Denault, 2001; Kalkbrener, 2005), which is widely recognized as the only capital allocation rule compatible with the return on risk-adjusted capital (McNeil et al., 2015). Specifically, if the total capital for the aggregate risk S is measured by the VaR and the tail conditional expectation (TCE):

$$\mathrm{TCE}_q(S_X) := \mathrm{TCA}_q(S_X, S_X) = \mathbb{E}(S_X \,|\, S_X > s_q),$$

then the corresponding Euler allocation rules are given by QCA (2) and TCA (3), respectively. In addition to capital allocation, risk functionals (2) and (3) have also found relevant applications in risk sharing (e.g., Denuit, 2019; Feng, 2023), sensitivity analysis (e.g., Asimit et al., 2019), and systemic risk management (e.g., Dhaene et al., 2022; Liu and Yang, 2021).

For the original untruncated multivariate normal distributions, QCA (2) can be readily computed

using the multivariate normal conditional distribution property (see Section 2.5 in Anderson, 2003), while the calculation of TCA (3) was studied by Landsman and Valdez (2003). The remainder of this current section is devoted to deriving computationally tractable expressions for calculating QCA and TCA under the truncated multivariate normal distributions.

Remark 2. *The truncated multivariate normal distributions are location invariant. Specifically, if* $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{c})$ *and* $\boldsymbol{X}^* \sim N_d(\boldsymbol{\mu} - \boldsymbol{c}, \boldsymbol{\Sigma}, \boldsymbol{0})$*, then it holds that*

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{X}^* + \boldsymbol{c},$$

where "$\stackrel{d}{=}$*" signifies equality in distribution. In light of this location invariance property, for any* $i \in \mathcal{D}$ *and* $q \in [0, 1)$*, we have*

$$\mathrm{QCA}_q(X_i, S_X) = c_i + \mathrm{QCA}_q(X_i^*, S_{X^*}) \tag{4}$$

and

$$\mathrm{TCA}_q(X_i, S_X) = c_i + \mathrm{TCA}_q(X_i^*, S_{X^*}), \tag{5}$$

where S_X *and* S_{X^*} *denote the aggregations of the elements in* $\boldsymbol{X}$ *and* $\boldsymbol{X}^*$*, respectively. For this reason, from here on, we will confine ourselves to the case of truncated multivariate normal distributions with* $\boldsymbol{c} = \boldsymbol{0}$*, whcih simplifies the presentation without any loss of generality for our results.*

We note that computing the QCA of truncated multivariate normal distributions is equivalent to computing TCA through the notion of left-tail expectation (LTE):

$$\mathrm{LTE}(X_i, S_X; s) = \mathbb{E}\big(X_i \,\big|\, S_X < s\big), \qquad i \in \mathcal{D},\ s \in \mathbb{R}.$$

The following lemma is crucial in establishing this desirable equivalence. While the result has been heuristically mentioned in the literature (Altmann et al., 2014), we were not able to identify a formal proof. In this paper, we provide a formal proof, and the involved techniques significantly inspire the

derivations of the other results presented in this section.

Before all else, we need to define a few necessary matrix notations. Let us first introduce a vector insertion operator: $\mathrm{insert}(\mathbf{A}, \boldsymbol{a}; k)$, which indicates that a row vector $\boldsymbol{a}$ of proper dimension is inserted immediately after the k-th row of $\mathbf{A}$, $k \in \mathbb{N}$. We denote by $\mathbf{I}_k$ a $k \times k$ identify matrix. Moreover, we let

$$\mathbf{B}_d := \mathrm{insert}(\mathbf{I}_{d-1}, -\mathbf{1}; d-1) \in \mathbb{R}^{d \times (d-1)}, \tag{6}$$

and

$$\boldsymbol{i}_d = (\underbrace{0, \ldots, 0}_{d-1}, 1)^\top. \tag{7}$$

Lemma 1. *Suppose that* $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0})$ *and* $\boldsymbol{X}^* \sim N_{d-1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \mathbf{0})$*, where*

$$\boldsymbol{\mu}^* = -(\mathbf{B}_d^\top \boldsymbol{\Sigma}^{-1} \mathbf{B}_d)^{-1} \mathbf{B}_d^\top \boldsymbol{\Sigma}^{-1} \big(s\, \boldsymbol{i}_d - \boldsymbol{\mu}\big) \qquad \textit{and} \qquad \boldsymbol{\Sigma}^* = (\mathbf{B}_d^\top \boldsymbol{\Sigma}^{-1} \mathbf{B}_d)^{-1}. \tag{8}$$

Then, it holds that

$$(X_i | S_X = s) \overset{d}{=} (X_i^* | S_{X^*} < s), \quad i = 1, \ldots, d-1, \textit{ and } s > 0.$$

Proof. See Appendix B. □

Lemma 1 tells that the condition $S_X = s$ reduces one degree of freedom among the random elements in $\boldsymbol{X}$. Therefore, after changing the condition to an inequality, its counterpart $\boldsymbol{X}^*$ has only $(d-1)$ dimension. Importantly, the change from equality to inequality in the condition creates a natural connection between the studies of QCA and TCA risk functionals, allowing us to develop a unified set of mathematical treatments to address their computations at one stroke. The succeeding assertion follows immediately from Lemma 1, thus its proof is omitted.

Lemma 2. *Suppose that* $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0})$. *For* $q \in [0,1)$, *the QCA of* $\boldsymbol{X}$ *can be computed via*

$$\mathrm{QCA}_q(X_i, S_X) = \mathrm{LTE}(X_i^*, S_{X^*}; s_q), \qquad i = 1, \ldots, d-1,$$

where $\boldsymbol{X}^* \sim N_{d-1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ *with* $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ *is defined according to* (8).

Moreover, when the truncation point $\boldsymbol{c}$ *of the truncated multivariate normal distribution is not zero, then the QCA can be computed via the location invariance property noted in Remark 2.*

Remark 3. *Note that the QCA formula reported in Lemma 2 applies only to the first* $(d-1)$ *elements of* $\boldsymbol{X}$. *To compute the QCA for* X_d, *one can either rearrange the position of the elements in* $\boldsymbol{X}$ *or use the fact that* $\sum_{i=1}^{d} \mathrm{QCA}_q(X_i, S_X) = s_q$, *and thus* $\mathrm{QCA}_q(X_d, S_X) = s_q - \sum_{i=1}^{d-1} \mathrm{QCA}_q(X_i, S_X)$.

For any continuous $\boldsymbol{X}$, the TCA can be expressed in terms of LTE via

$$\begin{aligned} \mathrm{TCA}_q(X_i, S_X) &= \frac{\mathbb{E}\big(X_i \, \mathbb{1}(S_X > s_q)\big)}{1-q} \\ &= \frac{1}{1-q}\big(\mathbb{E}(X_i) - q \times \mathrm{LTE}(X, S; s_q)\big), \quad i \in \mathcal{D}, \ q \in [0,1). \end{aligned} \tag{9}$$

If $\boldsymbol{X}$ follows the truncated multivariate normal distribution, then we can use Lemma 4.1 of Levine et al. (2023) to calculate its marginal mean, $\mathbb{E}(X_i)$, involved in (9). In light of Lemma 2 and Equation (9), computing the QCA and TCA of truncated multivariate normal distributions essentially boils down to the evaluation of LTE, which is further studied in the succeeding assertion.

Some additional notions are needed beforehand. Define a simplex set:

$$\mathcal{S}_d(s) := \big\{\boldsymbol{x} \in \mathbb{R}^d : x_1 > 0, \ldots, x_d > 0, \, x_1 + \cdots + x_d < s\big\}, \tag{10}$$

and denote the probability of the original untruncated multivariate normal distribution over this set by

$$\Psi(\mathcal{S}_d(s); \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\mathcal{S}_d(s)} \phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \mathrm{d}\boldsymbol{x}.$$

Moreover, for $k = 1, \ldots, d$, let

$$\boldsymbol{\mu}_k = (\mathbf{D}_k^\top \boldsymbol{\Sigma}^{-1} \mathbf{D}_k)^{-1} \mathbf{D}_k^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \qquad \text{and} \qquad \boldsymbol{\Sigma}_k = (\mathbf{D}_k^\top \boldsymbol{\Sigma}^{-1} \mathbf{D}_k)^{-1}, \tag{11}$$

where $\mathbf{D}_k := \text{insert}(\mathbf{I}_{d-1}, \mathbf{0}; k-1) \in \mathbb{R}^{d \times (d-1)}$, $k \in \mathcal{D}$. Additionally, we set

$$\alpha_k := \frac{(2\pi)^{(d-1)/2} \det(\boldsymbol{\Sigma}_k)^{1/2}}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp\Big(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}_k^\top \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k\Big), \qquad k = 1, \ldots, d, \tag{12}$$

and

$$\beta(s) := \frac{(2\pi)^{(d-1)/2} \det(\boldsymbol{\Sigma}^*)^{1/2}}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp\Big(\big(s\, \boldsymbol{i}_d - \boldsymbol{\mu}\big)^\top \boldsymbol{\Sigma}^{-1} \big(s\, \boldsymbol{i}_d - \boldsymbol{\mu}\big) - \boldsymbol{\mu}^{*\top} \boldsymbol{\Sigma}^{*-1} \boldsymbol{\mu}^*\Big), \qquad s \in \mathbb{R}, \tag{13}$$

where $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ is given in (8). Note that the expressions for α_k and β involve only standard matrix manipulations, hence they are straightforward to evaluate. We are now ready to spell out the main result of this current section.

Theorem 3. *Suppose that $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0})$. Fix $s > 0$, the LTE of $\boldsymbol{X}$ can be computed via*

$$\text{LTE}(X_i, S_X\,;\, s) = \mu_i + \frac{1}{\Psi(\mathcal{S}_d(s); \boldsymbol{\mu}, \boldsymbol{\Sigma})} \sum_{k=1}^{d+1} a_{i,k} \times \Psi(\mathcal{S}_{d-1}(s); \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \qquad i \in \mathcal{D},$$

where $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ for $k = 1, \ldots, d$, are defined according to (11) and $(\boldsymbol{\mu}_{d+1}, \boldsymbol{\Sigma}_{d+1}) = (\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ which is given in (8). Further,

$$a_{i,k} = \sigma_{i,k} \times \alpha_k, \;\; k = 1, \ldots, d, \qquad \text{and} \qquad a_{i,d+1} = \Big(-\sum_{j=1}^{d} \sigma_{i,j}\Big) \times \beta(s),$$

where $\sigma_{i,k}$ is the (i,k)-th element of $\boldsymbol{\Sigma}$, $\alpha_k(\cdot)$ and $\beta(\cdot)$ are given as per (12) and (13), respectively.

Proof. See Appendix B. □

As shown in Theorem 3, evaluating the LTE relies on the probabilities of multivariate normal distribution truncated over simplex areas. These probabilities cannot be computed in closed form. One way to estimate these probabilities is via rejection sampling, where samples from the untruncated

normal distribution are generated, and only those satisfying the simplex constraints are accepted. The probability is then estimated as the ratio of the number of accepted samples to the number of total samples. However, the simulation approach can become excessively computationally demanding if the acceptance rate is low, i.e., when the simplex area is small, which is indeed the case encountered in the context of quantitative risk management where the quantile levels in QCA and TCA are set close to one.

More advanced sampling methods, such as subset simulation, the Holmes-Diaconis-Ross algorithm, and the linear expectation sampling scheme, can help mitigate the low acceptance rate issue of naive sampling to some extent (Gessner et al., 2020). However, these methods come at the expense of utilizing much more complex sequential simulation techniques to handle difficult-to-sample regions. Meanwhile, the randomness inherent in the simulation estimate can not be fully removed. The added complexities involved in the mathematical mechanisms of these advanced sampling methods complicate the process of determining the sample size needed to achieve the desired robustness in the results.

Compared to simulation-based methods, numerical approaches which can provide deterministic results with guaranteed accuracy, are often preferred by risk analysts in practice. For computing the probability of a multivariate normal distribution truncated over a simplex area, the only non-simulation deterministic method we are aware of is presented by Adams (2022). His method involves decomposing the simplex into several regions, transforming these regions into rectangles through a change of variables, and then calculating the normal probabilities over hyperrectangles. For a d-dimensional normal distribution, the number of decomposed regions is 2^d, making the implementation cumbersome and tedious. Moreover, inaccuracies in computing multivariate normal probabilities over hyperrectangles can accumulate, leading to a significant reduction in accuracy even for only moderately high dimensions, such as $d = 5$ (see numerical comparisons in Adams, 2022).

The numerical method we propose for evaluating the simplex-area probabilities involved in Theorem 3 is motivated by the holonomic gradient method studied in Koyama (2015). However, the approach we take to establish the differential equations underlying the calculations of these probabilities differs significantly. Specifically, unlike the abstract algebraic approach taken in Koyama (2015), our proof is based on standard algebraic analysis, which is more familiar to researchers in applied

mathematics and quantitative risk management. From an application standpoint, a key advantage of adopting the proof we develop in this paper is that it only relies on an assumption which is very convenient to verify in practice, rather than relying on a more complicated notion of general position considered in Koyama (2015).

3 Computing the probability over a simplex area

At the outset, we remark that the simplex-area probabilities based on different normal distributions as specified in Theorem 3, can be computed in an unified manner with the aids of the notion of general simplex:

$$\begin{aligned}\mathcal{S}_d(\boldsymbol{\Theta}, \boldsymbol{\theta}) &= \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{\Theta}\boldsymbol{x} + \boldsymbol{\theta} > \mathbf{0}\} \\ &= \Big\{\boldsymbol{x} \in \mathbb{R}^d : \sum_{j=1}^{d} \theta_{i,j}\, x_j + \theta_i > 0, \text{ for all } i = 1, \cdots, d+1\Big\},\end{aligned}$$

where $\theta_{i,j} \in \mathbb{R}$ denotes the (i,j)-th entry of $\boldsymbol{\Theta}$ which is a $(d+1) \times d$ matrix, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{d+1}) \in \mathbb{R}^{d+1}$. The simplex $\mathcal{S}_d(s)$ given in Equation (10) can be expressed in terms of a general simplex via $\mathcal{S}_d(s) = \mathcal{S}_d(\mathbf{B}_{d+1}, s\,\boldsymbol{i}_{d+1})$, where $s > 0$, $\mathbf{B}_{d+1}$ and $\boldsymbol{i}_{d+1}$ are given in Equations (6) and (7), respectively. Now, consider a multivariate normal RV $\boldsymbol{Y} = (Y_1, \ldots, Y_d) \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\boldsymbol{Z} = (Z_1, \ldots, Z_d) \sim N_d(\mathbf{0}, \mathbf{I}_d)$. It holds that

$$\begin{aligned}\Psi(\mathcal{S}_d(s); \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Psi(\mathcal{S}_d(\mathbf{B}_{d+1}, s\,\boldsymbol{i}_{d+1}); \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \mathbb{P}(\mathbf{B}_{d+1}\boldsymbol{Y} + s\,\boldsymbol{i}_{d+1} > \mathbf{0}) \\ &= \mathbb{P}\big(\mathbf{B}_{d+1}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z} + \boldsymbol{\mu}) + s\,\boldsymbol{i}_{d+1} > \mathbf{0}\big) \\ &= \mathbb{P}\big(\mathbf{B}_{d+1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z} + (\mathbf{B}_{d+1}\boldsymbol{\mu} + s\,\boldsymbol{i}_{d+1}) > \mathbf{0}\big) \\ &= \mathbb{P}(\widetilde{\boldsymbol{\Theta}}\boldsymbol{Z} + \widetilde{\boldsymbol{\theta}}(s) > \mathbf{0}) \\ &= \Psi(\mathcal{S}_d(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s)); \mathbf{0}, \mathbf{I}_d), \end{aligned} \tag{14}$$

where

$$\widetilde{\boldsymbol{\Theta}} = \mathbf{B}_{d+1}\boldsymbol{\Sigma}^{\frac{1}{2}} \qquad \text{and} \qquad \widetilde{\boldsymbol{\theta}}(s) = \mathbf{B}_{d+1}\boldsymbol{\mu} + s\,\boldsymbol{i}_{d+1}. \tag{15}$$

Thereby, calculating the simplex-region probabilities of a non-standard normal distribution with any given set of location and dispersion parameters, can be consolidated into computing the simplex-region probability of the standard normal distribution.

3.1 The system of partial differential equations

To calculate the simplex-region probability of the standard normal distribution given in Equation (14), we propose treating it as the solution to a system of differential equations and applying the Euler method to evaluate the solution. To facilitate the presentation, for a given set of constraint parameters $\boldsymbol{\Theta} \in \mathbb{R}^{(d+1)\times d}$ and $\boldsymbol{\theta} \in \mathbb{R}^{d+1}$, let us denote the standard normal simplex-region probability by $g(\boldsymbol{\Theta}, \boldsymbol{\theta}) := \Psi\big(\mathcal{S}_d(\boldsymbol{\Theta}, \boldsymbol{\theta}); \mathbf{0}, \mathbf{I}_d\big)$. Denote by $\mathcal{I}$ the set $\{1, \ldots, d+1\}$. For a subset $\mathcal{H} \subseteq \mathcal{I}$, let us write

$$g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) = \frac{\partial^{|\mathcal{H}|}}{\prod_{i\in\mathcal{H}} \partial\theta_i}\, g(\boldsymbol{\Theta}, \boldsymbol{\theta}),$$

where $|\mathcal{H}|$ denotes the cardinality of $\mathcal{H}$. Suppose that $\boldsymbol{\Theta}$ has full rank, we set $\boldsymbol{\Lambda} := \boldsymbol{\Theta}^\top \boldsymbol{\Theta}$ and denote its inverse by $\boldsymbol{\Lambda}^{-1}$. The following assertion outlines the differential equations governing g, which form a fundamental component of our proposed numerical algorithm.

Theorem 4. *Suppose that $\boldsymbol{\Theta}$ is full rank. Fix any $\mathcal{H} \subset \mathcal{I} = \{1, \ldots, d+1\}$, then for any $m \in \mathcal{I}$, it holds that*

$$\frac{\partial}{\partial\theta_m}\, g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) = \begin{cases} g^{(\mathcal{H}\cup\{m\})}(\boldsymbol{\Theta}, \boldsymbol{\theta}), & \text{if} \quad m \notin \mathcal{H}; \\ -\sum\limits_{j\in\mathcal{H}} [\boldsymbol{\Lambda}^{-1}]_{j,m}\, \theta_j\, g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) - \sum\limits_{i\in\mathcal{H}^c} \Big(\sum\limits_{j\in\mathcal{H}} \lambda_{i,j}\ [\boldsymbol{\Lambda}^{-1}]_{j,m} \Big) g^{(\mathcal{H}\cup\{i\})}(\boldsymbol{\Theta}, \boldsymbol{\theta}), & \text{if} \quad m \in \mathcal{H}, \end{cases} \tag{16}$$

where $\mathcal{H}^c = \mathcal{I} \setminus \mathcal{H}$, and $\lambda_{i,j}$ and $[\boldsymbol{\Lambda}^{-1}]_{i,j}$ denote the (i,j)-th elements of the matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}^{-1}$, respectively. Moreover if $\mathcal{H} = \mathcal{I}$, then $\partial/\partial\theta_m\, g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) = 0$ for any $m \in \mathcal{I}$.

Proof. See Appendix B. □

Remark 4. *Recall that the simplex-region probabilities underpinning the calculations of QCA and TCA share a unified form* (14)*, wherein the constraint parameters* $\widetilde{\boldsymbol{\Theta}}$ *and* $\widetilde{\boldsymbol{\theta}}(s)$ *are given in* (15)*. By construction,* $\widetilde{\boldsymbol{\Theta}}$ *has rank* d*, and thus it is full rank. In other words,* $\widetilde{\boldsymbol{\Theta}}$ *satisfies the assumption of Theorem 4, and thus the simplex-region probability of form* (14) *satisfies the system of differential equations as specified by* (16)*.*

3.2 Proposed Euler's algorithm

Based on the differential equations established in Theorem 4, we apply the Euler method to numerically evaluate $g(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s))$ for some given $s > 0$. Let us recall the definition of $\widetilde{\boldsymbol{\theta}}(s)$ given in Equation (15). We note that only the $(d+1)$-th element of the constraint parameter $\widetilde{\boldsymbol{\theta}}(s)$ is dependent of the variable s, whereas the first d elements are constant. Moreover, the dependence relationship is linear with a slope of one. Given an easy-to-evaluate initial value $g(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_0))$ for some reference point $s_0 \in \mathbb{R}$, in principle, we only need $g^{(d+1)}(\cdot) = \partial/\partial\theta_{d+1}\, g(\cdot)$ to approximate $g(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s))$ for other values of $s \neq s_0$. Specifically, consider a uniform grid of points over the interval $[s_0, s]$, and choose the step size to be $\Delta s = (s - s_0)/m$ for some $m \in \mathbb{N}$. The k-th grid point is defined via $s_k = s_0 + k \times \Delta s$, where $k = 1, \ldots, m$. Assuming that, at the k-th step of the Euler algorithm, the value of $g\big(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_{k-1})\big)$ has been determined, then the value of $g\big(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_k)\big)$ can be determined via

$$g\big(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_k)\big) \approx g\big(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_{k-1})\big) + \Delta s \left.\frac{\partial}{\partial\theta_{d+1}} g\big(\widetilde{\boldsymbol{\Theta}}, \boldsymbol{\theta}\big)\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}(s_{k-1})}. \tag{17}$$

That said, the system of differential equations in Theorem 4 is highly dependent. Computing the partial derivative of g with respect to θ_{d+1} necessitates evaluating the partial derivatives with respect to all other components of $\boldsymbol{\theta}$. In other words, the Euler approximation (17) can not be implemented in isolation. Instead, we must simultaneously evaluate the entire row vector of derivatives $\boldsymbol{g}^{(\cdot)} = (g^{(\mathcal{H})} : \forall \mathcal{H} \subseteq \mathcal{I})$, which consists of 2^{d+1} components.

At a glance, simultaneously evaluating such a large vector of functions may seem excessively onerous. However, a more careful inspection of the differential equations in (16) reveals a handful

of significant simplifications. First, fix $m = d + 1$ in Theorem 4, and let $\partial / \partial\theta_{d+1}\, \boldsymbol{g}^{(\cdot)}$ denote the component-wise partial derivative of $\boldsymbol{g}^{(\cdot)}$ with respect to θ_{d+1}. Then, for any $\boldsymbol{\Theta} \in \mathbb{R}^{(d+1)\times d}$ and $\boldsymbol{\theta} \in \mathbb{R}^{d+1}$, we can treat (16) as a system of ordinary differential equations (ODE's) for $\boldsymbol{g}^{(\cdot)}$ and express it compactly in a matrix form as

$$\frac{\partial}{\partial\theta_{d+1}}\, \boldsymbol{g}^{(\cdot)}(\boldsymbol{\Theta}, \boldsymbol{\theta}) = \mathbf{M}(\boldsymbol{\Theta}, \boldsymbol{\theta}) \;\; \boldsymbol{g}^{(\cdot)}(\boldsymbol{\Theta}, \boldsymbol{\theta}), \tag{18}$$

where $\mathbf{M}(\boldsymbol{\Theta}, \boldsymbol{\theta})$ is a $2^{d+1} \times 2^{d+1}$ matrix collecting the coefficients in the differential equations from Theorem 4. Consequently, Euler's method can be naturally applied to solve the ODE system (18). Second, calculating the coefficients of the differential equations in (16), or equivalently, the entries of $\mathbf{M}$, only involves basic algebraic manipulations that can be performed fastly. Third, all the coefficients involved in $\mathbf{M}$ are independent of θ_{d+1}, except for the diagonal entries. Thereby, most entries in $\mathbf{M}(\boldsymbol{\Theta}, \boldsymbol{\theta})$ only need to be computed once throughout the entire algorithm.

To illustrate the set-up of $\mathbf{M}$, a two-dimensional example is presented below. For higher dimensional applications, we have developed an efficient algorithm for populating the relevant entries of $\mathbf{M}$, which is presented in Appendix A.

Example 1. *Suppose that $d = 2$, then the function $g(\boldsymbol{\Theta}, \boldsymbol{\theta})$ represents the probability that a bivariate standard normal RV falls within the simplex region defined by $\boldsymbol{\Theta} + \boldsymbol{\theta} \geq \mathbf{0}$, where $\boldsymbol{\Theta} \in \mathbb{R}^{3\times 2}$ and $\boldsymbol{\theta} \in \mathbb{R}^3$. We arrange the partial derivatives in $\boldsymbol{g}^{(\cdot)}$ as follows:*

$$\boldsymbol{g}^{(\cdot)} = (g,\, g^{(1)},\, g^{(2)},\, g^{(3)},\, g^{(1,2)},\, g^{(1,3)},\, g^{(2,3)},\, g^{(1,2,3)}).$$

It is worth noting that the order of the partial derivative elements in $\boldsymbol{g}^{(\cdot)}$ does not affect the validity of ODE's in (18), as long as the coefficients in $\mathbf{M}$ are populated accordingly.

According to Theorem 4, the ODE system (18) holds with $\mathbf{M}(\boldsymbol{\Theta}, \boldsymbol{\theta})$ *given by*

$$\begin{pmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & -[\boldsymbol{\Lambda}^{-1}]_{3,3}\,\theta_3 & 0 & -[\boldsymbol{\Lambda}^{-1}]_{3,3}\,\lambda_{1,3} & -[\boldsymbol{\Lambda}^{-1}]_{3,3}\,\lambda_{2,3} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & -\sum_{j\in\{1,3\}}[\boldsymbol{\Lambda}^{-1}]_{j,3}\,\theta_j & 0 & -\sum_{j\in\{1,3\}}[\boldsymbol{\Lambda}^{-1}]_{j,3}\,\lambda_{2,j} \\
0 & 0 & 0 & 0 & 0 & 0 & -\sum_{j\in\{2,3\}}[\boldsymbol{\Lambda}^{-1}]_{j,3}\,\theta_j & -\sum_{j\in\{2,3\}}[\boldsymbol{\Lambda}^{-1}]_{j,3}\,\lambda_{1,j} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.$$

Remark 5. *At first sight, the coefficient matrix,* $\mathbf{M}$*, appears computationally onerous due to its large dimension,* 2^{d+1}*. However,* $\mathbf{M}$ *is highly sparse. In actual coding, only its non-zero entries and their positions need to be stored, and those can be efficiently determined using the expressions provided in Appendix A. This sparsity feature enables the efficient implementation of the associated Euler method, without requiring excessively large computer memory to store the full matrix.*

Collectively, at the k-th step of the Euler method for computing $\boldsymbol{g}^{(\cdot)}$, we proceed with

$$\begin{aligned}
\boldsymbol{g}^{(\cdot)}\big(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_k)\big) &\approx \boldsymbol{g}^{(\cdot)}\big(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_{k-1})\big) + \Delta s \;\; \frac{\partial}{\partial \theta_{d+1}} \boldsymbol{g}^{(\cdot)}\big(\widetilde{\boldsymbol{\Theta}}, \boldsymbol{\theta}\big)\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}(s_{k-1})} \\
&= \Big(\mathbf{I}_{2^{d+1}} + \Delta s\; \mathbf{M}\big(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_{k-1})\big)\Big) \;\; \boldsymbol{g}^{(\cdot)}\big(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_{k-1})\big). \qquad (19)
\end{aligned}$$

At the final step of the algorithm, the component of $\boldsymbol{g}^{(\cdot)}$ corresponding to g provides an numerical approximation of the simplex-region probability of interest, $g(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s))$. The convergence of the proposed algorithm is guaranteed, as verified in the following assertion.

Lemma 5. *The Euler algorithm in (19) is convergent.*

Proof. See Appendix B. □

To implement the Euler method, what remains to be specified is an appropriate reference point, denoted by s_0, at which $g^{(\mathcal{H})}(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_0))$ can be computed conveniently for every $\mathcal{H} \subseteq \mathcal{I}$. This is

addressed in the following assertion.

Proposition 6. *Consider the reference point $s_0 = 0$. For any $\mathcal{H} \subseteq \mathcal{I}$, the following holds:*

$$g^{(\mathcal{H})}(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s_0)) = \begin{cases} \det(\widetilde{\boldsymbol{\Theta}}_{\mathcal{H}\times\mathcal{H}})^{-1} (2\pi)^{-d/2} \exp(-\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/2), & \text{if } |\mathcal{H}| = d; \\ 0, & \text{if } |\mathcal{H}| \neq d. \end{cases} \tag{20}$$

Proof. See Appendix B. □

Finally, we remark that the proposed Euler algorithm can be also used to evaluate the VaR involved in the calculation of QCA and TCA. Specifically, recall that for $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0})$, the CDF of the aggregate RV S_X can be expressed as

$$\mathbb{P}(S_X \leq s) = \frac{g(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s))}{\overline{\Phi}_d(\mathbf{0}; \boldsymbol{\mu}, \boldsymbol{\Sigma})}, \tag{21}$$

where $\widetilde{\boldsymbol{\Theta}}$ and $\widetilde{\boldsymbol{\theta}}(\cdot)$ are given in (15), and the denominator can be computed using standard functions available in most computing platforms for evaluating the CDF of the multivariate normal distribution. Given a step size Δs and starting from the initial reference point s_0, Euler formula (19) is iteratively applied until the CDF (21) first exceeds q, requiring say, k steps. The VaR of the aggregated variable S then can be approximated as $\mathrm{VaR}_q(S) \approx s_0 + \Delta s \times k$.

4 Numerical illustration

In this section, we aim at conducting a comparative analysis against the simulation method in order to illustrate the advantages of adopting the proposed numerical approach. Note that directly using simulation to estimate the QCA is more challenging than to estimate the TCA, as QCA represents the conditional expectation at a specific point, whereas TCA is defined over an interval. Therefore, this section focuses on the more demanding problem of computing QCA, which in turn sheds light on how the proposed method effectively addresses challenges that the commonly used simulation approach may struggle to handle adequately. Throughout this numerical study, we consider a confidence level

of $q = 99\%$, which is stipulated by the recent best practice in the insurance industry (Leiser et al., 2023).

For the simulation method, we use the standard rejection sampling method to generate samples from the truncated multivariate normal distribution. To estimate the QCA, we apply the empirical estimator considered in Gribkova et al. (2023). Specifically, given m simulated data $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,d})$, and let $s_i = \sum_{j=1}^{d} x_{i,j}$, $i = 1, \ldots, m$. Let us use

$$x_{i,(k)} = \sum_{j=1}^{m} x_{i,j}\, \mathbb{1}(s_i = s_{[k]}),$$

to denote the i-th marginal component of the concomitant induced by $s_{[k]}$, which is the k-th order statistic of $s_1, \ldots, s_m$. The empirical QCA considered in Gribkova et al. (2023), is given by

$$\widehat{\mathrm{QCA}}_{q,m}(X_i, S_X) = \frac{1}{\delta_m} \sum_{j=1}^{m} x_{i,(j)}\, \mathbb{1}\left(q - \Delta_{1,m} < \frac{j}{m} < q + \Delta_{2,m}\right), \tag{22}$$

where

$$\delta_m = \sum_{j=1}^{m} \mathbb{1}\left(q - \Delta_{1,m} < \frac{j}{m} < q + \Delta_{2,m}\right).$$

For the empirical QCA (22) to be consistent, the bandwidth parameters must be chosen such that $\max(\Delta_{1,m},\, \Delta_{2,m}) \to 0$ as $m \to \infty$, and $\liminf_{n\to\infty} \sqrt{m}\,(\Delta_{1,m} + \Delta_{2,m}) > 0$. Inspired by the numerical study conducted in Gribkova et al. (2023), we let $\Delta_{1,m} = \Delta_{2,m} = m^{-1/2}$, which guarantees the consistency of $\widehat{\mathrm{QCA}}_{q,m}$, and thus the simulation method can yield an asymptotically accurate estimate of the true QCA value.

For the sake of presentation convenience while without impacting insights we can obtain, throughout this section, we intentionally work with set-ups in which the marginal distributions are identical and the inherent dependence structure is symmetric. Such a setting ensures that the QCA values along each margins are the same, so we only need to report the results on $\mathrm{QCA}_q(X_1, S_X)$.

Let us begin with the baseline case in which the truncated multivariate standard normal distribution in dimension $d = 3$, with $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}_3$, is considered. To implement the proposed Euler

method for computing the QCA, selecting an appropriate step size, Δs, is crucial to strike a balance between accuracy and computational efficiency. Note that the Euler method is a first-order numerical scheme for solving an ODE system, with a global error that decays at a rate of $\mathcal{O}(h)$. Given this linear convergence, we set $h = 10^{-4}$ in an effort to achieve accuracy up to three decimal places. In our unreported analysis, we perturbed the chosen Δs, and the resulting changes in the QCA calculations, with differences limited to the fourth decimal place.

We apply the simulation method using varying simulation sizes $m \in \{1, 3, 10, 30\} \times 10^4$. To illustrate the variability inherent in the simulation and how it changes in response to different sample sizes, the same estimation process will be repeated 100 times for each simulate size. The box plot of the QCA estimate is presented in Figure 1,with the deterministic QCA value, computed using the proposed Euler method, highlighted by the dashed line. As can be seen, increasing the sample size leads to a reduction in the width of the confidence intervals, which coincides with the consistency property of the empirical QCA (22). Note that the mean of the empirical QCA estimates may vary significantly across different sample sizes. This is because, the empirical QCA (22) estimate the condition expectation at a point by the conditional expectation over an interval surrounding that point. As a result, the empirical QCA is inherently biased, and the bias can be particularly pronounced when the sample size is small (e.g., the case of $m = 10^4$). However, as the sample size increases, the length of the estimation interval, controlled by the bandwidth parameters $\Delta_{1,m}$ and $\Delta_{2,m}$, decreases, so does the bias. Consequently, both the spread of the estimates and their means converge to the QCA value computed using the proposed Euler method, which suggests that the proposed numerical approach provides an accurate calculation of the true QCA value.

Remark 6. *Under the current choice of distributional parameters and in the scenarios considered later, the QCA values range from one to two, which may not coincide with what we typically observe in practice where losses are in the thousands or millions. This discrepancy does not impact the applicability of the observations derived from this numerical study. Specifically, note that the truncated multivariate normal distribution is scale-invariant. When it is fitted to data, the data may have already been scaled for numerical stability. Otherwise, the fitted model can always be rescaled through a linear transformation, so that the parameter magnitudes are comparable to those considered here. After the*

QCA is computed, it can be rescaled to match the original magnitude of the loss variables.

It is also noteworthy that, as discussed in Gribkova et al. (2023), the empirical QCA does not attain the typical $\sqrt{n}$ convergence rate of empirical estimators for other quantities, such as the mean or TCA. The coverage pattern observed in Figure 1 further suggests a relatively slow convergence rate of the empirical QCA (22). This underscores another advantage of the proposed numerical approach, as achieving a desirable level of accuracy and stability through the simulation method may require an excessively large number of samples.

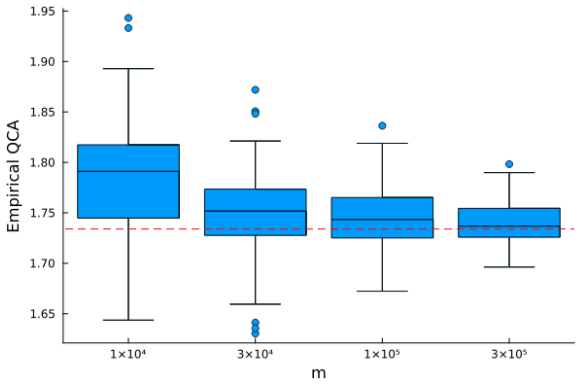

Figure 1: The box plot of the emperical QCA estimates for varying sample sizes, with the deterministic QCA value computed using the proposed Euler method, overlaid as a dashed line. The interquartile range (IQR) represents the middle 50% of the estimates, while the whiskers extend to the values within 1.5 times the IQR. Outliers beyond this range are displayed as individual points.

The sensitivity of the simulation approach and the proposed numerical method in response to different choices of the distributional parameters is further explored. Specifically, we deviate from the baseline case and consider sensitivity scenarios with a non-zero location parameter and a non-identify dispersion matrix. In the scenario of non-zero location parameter, we keep the dispersion parameter as the identify matrix but consider two different directions of shifting of the location parameter vector: $\boldsymbol{\mu}_1 = (-0.5, -0.5, -0.5)^\top$ and $\boldsymbol{\mu}_2 = (0.5, 0.5, 0.5)^\top$. For the case of non-identify dispersion matrix, we consider

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & -0.25 & -0.25 \\ -0.25 & 1 & -0.25 \\ -0.25 & -0.25 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 1 \end{pmatrix},$$

which describe a negative dependence structure and a positive dependence structure in $\boldsymbol{X}$, respectively.

We compare the performance of the simulation method and the proposed Euler method for calculating the QCA, across the two aforementioned sensitivity scenarios. The variability of the simulation method is quantified using the coefficient of variation (CV), defined as the ratio of the standard deviation to the mean. Since the QCA computed using the proposed Euler method is deterministic, its accuracy is assessed by comparing it to the 25th and 75th percentiles, as well as the mean, of the QCA estimates obtained from the simulation method. The computation time is measured by implementing both methods on a standard computer with an 2.3 GHz CPU and 16GB RAM. For the simulation method, the reported computation time corresponds only to the time required to complete a single trial of QCA estimation using a simulation of size m.

Scenario		**Simulation**				**Euler**
		$\boldsymbol{m = 10^4}$	$\boldsymbol{m = 3 \times 10^4}$	$\boldsymbol{m = 10^5}$	$\boldsymbol{m = 3 \times 10^5}$	
$\boldsymbol{\mu}_1$	25th percentile	1.4795	1.4545	1.4524	1.4603	1.4723
	75th percentile	1.5533	1.5099	1.4927	1.4846	
	Mean	1.5202	1.4851	1.4721	1.4724	
	CV	3.48%	2.89%	2.27%	1.38%	
	Time	0.22s	0.68s	2.27s	6.90s	0.25s
Baseline	25th percentile	1.7447	1.7276	1.7250	1.7258	1.7398
	75th percentile	1.8173	1.7734	1.7652	1.7544	
	Mean	1.7834	1.7476	1.7452	1.7392	
	CV	3.36%	2.71%	1.86%	1.23%	
	Time	0.06s	0.18s	0.61s	1.88s	0.29s
$\boldsymbol{\mu}_2$	25th percentile	2.0720	2.0375	2.0534	2.0495	2.0662
	75th percentile	2.1407	2.1045	2.0859	2.0842	
	Mean	2.1093	2.0697	2.0694	2.0669	
	CV	2.52%	2.28%	1.47%	1.09%	
	Time	0.02s	0.07s	0.24s	0.78s	0.34s

Table 1: Comparison of the simulation method and the proposed Euler method in terms of accuracy, stability, and computational efficiency, with respect to varying location parameter vectors. The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean of the empirical QCA estimates, is used as a measure of stability.

Let us first focus on the sensitivity scenario with changes in the location parameter vector. The comparison results are summarized in Table 1. First, we observe that across all choices of the location parameter, the 25th and 75th percentiles, as well as the mean of the QCA estimates, converge to the QCA values produced by the proposed Euler method. This again confirms the accuracy of our Euler-based numerical approach. Second, for a fixed simulation size m, as the elements of the location parameter $\boldsymbol{\mu}$ increase, the QCA estimates obtained via the simulation method become more volatile, as indicated by larger CV values. In contrast, the performance of the proposed Euler method remains stable and satisfactory across different values of $\boldsymbol{\mu}$. Third, when $\boldsymbol{\mu}$ is larger, the VaR of the aggregate RV S_X increases, leading to a higher value of QCA along X_1. Consequently, the computation time required for the Euler method increases, because more iteration steps are needed to arrive at the terminate point, i.e., s_q. However, the increase in computation time is rather modest. On the other hand, the simulation method becomes significantly more computationally demanding for smaller location parameter $\boldsymbol{\mu}$. This is because a negative $\boldsymbol{\mu}$ results in a lower truncation probability, $\overline{\Phi}_d(\mathbf{0}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, which increases the rejection rate in the rejection sampling process. Compared to the Euler method, the computation time of the simulation method is much more sensitive to changes in the location parameter. In particular, to achieve a comparable accuracy to the Euler method, the simulation method requires a sample size larger than $m = 3 \times 10^5$, as illustrated by Table 1. Even in this simulation size, the Euler method can be about 27 times faster than the simulation method when the location parameter vector is chosen to be $\boldsymbol{\mu}_1$.

Table 2 compares the sensitivities of the two methods under the scenario of varying dispersion parameter matrices. Naturally, the negative dependence described by $\boldsymbol{\Sigma}_1$ reduces the variability of the aggregate RV S_X, leading to lower values of the VaR of S_X and the QCA along X_1. In this case, computing QCA using the Euler method requires fewer steps and, consequently, less time. However, the impact on the computation speed of the simulation method is the opposite for the same reason as in the varying location parameter scenario. Specifically, the negative dependence inherent in $\boldsymbol{\Sigma}_1$ decreases the truncation probability, $\overline{\Phi}_d(\mathbf{0}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, making the simulation process less efficient. Across all scenarios, as demonstrated by the convergence of the 25th and 75th percentiles and the mean of the empirical QCA estimates, the proposed Euler method is able to capture the true QCA value with

Scenario		Simulation				Euler
		$m = 10^4$	$m = 3 \times 10^4$	$m = 10^5$	$m = 3 \times 10^5$	
$\boldsymbol{\Sigma}_1$	25th percentile	1.2761	1.2787	1.2796	1.2731	1.2887
	75th percentile	1.3672	1.3200	1.3136	1.2971	
	mean	1.3178	1.2996	1.2958	1.2878	
	CV	4.38%	2.71%	1.98%	1.81%	
	Time	0.17s	0.52s	1.73s	5.18s	0.22s
Baseline	25th percentile	1.7447	1.7276	1.7250	1.7258	1.7398
	75th percentile	1.8173	1.7734	1.7652	1.7544	
	mean	1.7834	1.7476	1.7452	1.7392	
	CV	3.36%	2.71%	1.86%	1.23%	
	Time	0.06s	0.18s	0.61s	1.88s	0.29s
$\boldsymbol{\Sigma}_2$	25th percentile	2.0640	2.0261	2.0377	2.0415	2.0516
	75th percentile	2.1310	2.0884	2.0736	2.0666	
	mean	2.0968	2.0606	2.0548	2.0534	
	CV	2.56%	1.93%	1.41%	0.90%	
	Time	0.06s	0.18s	0.63s	1.92s	0.35s

Table 2: Comparison of the simulation method and the proposed Euler method in terms of accuracy, stability, and computational efficiency, with respect to varying dispersion parameters.

high accuracy.

5 Conclusions

This paper studies the computation of QCA and TCA risk functionals under the assumption that losses follow a truncated multivariate normal distribution. We first establish the equivalence between the QCA and TCA calculations through the notion of LTE. We then show that LTE can be expressed as a linear combination of multivariate normal probabilities over simplex regions. By formulating these probabilities as solutions to systems of differential equations, we propose solving them via Euler's method. Numerical experiments demonstrate that the proposed method is stable, accurate, and computationally efficient. Through verifying the tractability for risk functional calculations, the practical message we aim to convey is that when a multivariate normal-type model is preferred for analyzing

positive data, the truncated normal distributions warrant a serious consideration.

References

Adams, M. P. (2022). Integral, mean and covariance of the simplex-truncated multivariate normal distribution. *PLOS ONE*, 17(7):e0272014.

Altmann, Y., McLaughlin, S., and Dobigeon, N. (2014). Sampling from a multivariate Gaussian distribution truncated on a simplex: A review. In *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, pages 113–116.

Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica*, 42(6):999–1012.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd Edition*. Wiley, New York.

Arismendi, J. C. (2013). Multivariate truncated moments. *Journal of Multivariate Analysis*, 117:41–75.

Asimit, V., Peng, L., Wang, R., and Yu, A. (2019). An efficient approach to quantile capital allocation and sensitivity analysis. *Mathematical Finance*, 29(4):1131–1156.

Bladt, M. (2023). A tractable class of multivariate phase-type distributions for loss modeling. *North American Actuarial Journal*, pages 1–21.

Cohen, A. C. (1991). *Truncated and Censored Samples: Theory and Applications*. CRC Press, Boca Raton.

Cossette, H., Côté, M.-P., Marceau, E., and Moutanabbir, K. (2013). Multivariate distribution defined with Farlie–Gumbel–Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics*, 52(3):560–572.

Denault, M. (2001). Coherent allocation of risk capital. *Journal of Risk*, 4:1–34.

Denuit, M. (2019). Size-biased transform and conditional mean risk sharing, with application to P2P insurance and tontines. *ASTIN Bulletin*, 49(3):591–617.

Dhaene, J., Laeven, R. J., and Zhang, Y. (2022). Systemic risk: Conditional distortion risk measures. *Insurance: Mathematics and Economics*, 102:126–145.

Fang, K., Kotz, S., and Ng, K. (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London.

Feng, R. (2023). *Decentralized Insurance: Technical Foundation of Business Models*. Springer, Cham.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25.

Furman, E., Hackmann, D., and Kuznetsov, A. (2020). On log-normal convolutions: An analytical–numerical method with applications to economic capital determination. *Insurance: Mathematics and Economics*, 90:120–134.

Gessner, A., Kanjilal, O., and Hennig, P. (2020). Integrals over Gaussians under linear domain constraints. In *International conference on artificial intelligence and statistics*, pages 2764–2774.

Gribkova, N., Su, J., and Zitikis, R. (2023). Estimating the VaR-induced Euler allocation rule. *ASTIN Bulletin*, 53(3):619–635.

Hong, H. and Shum, M. (2003). Econometric models of asymmetric ascending auctions. *Journal of Econometrics*, 112(2):327–358.

Horrace, W. C. (2005). Some results on the multivariate truncated normal distribution. *Journal of multivariate analysis*, 94(1):209–221.

Iserles, A. (2009). *A First Course in the Numerical Analysis of Differential Equations*. Cambridge University Press, Cambridge.

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49(2):127–145.

Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, 15(3):425–437.

Koyama, T. (2015). Holonomic gradient method for the probability content of a simplex region with a multivariate normal distribution. *arXiv:1512.06564*.

Landsman, Z., Makov, U., and Shushi, T. (2016). Multivariate tail conditional expectation for elliptical distributions. *Insurance: Mathematics and Economics*, 70:216–223.

Landsman, Z., Makov, U., and Shushi, T. (2018). A multivariate tail covariance measure for elliptical distributions. *Insurance: Mathematics and Economics*, 81:27–35.

Landsman, Z. and Shushi, T. (2022). Modelling random vectors of dependent risks with different elliptical components. *Annals of Actuarial Science*, 16(1):6–24.

Landsman, Z. M. and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4):55–71.

Lee, S. C. and Lin, X. S. (2012). Modeling dependent risks with multivariate Erlang mixtures. *ASTIN Bulletin*, 42(1):153–180.

Leiser, B., Bender, J., and Kaul, B. (2023). Regulatory capital adequacy for life insurance companies. *Technical report: The Society of Actuaries*, pages 1–43.

Levine, M., Richards, D., and Su, J. (2020). Independence properties of the truncated multivariate elliptical distributions. *Statistics and Probability Letters*, 161:108729.

Levine, M., Richards, D., and Su, J. (2023). Non-steepness and maximum likelihood estimation properties of the truncated multivariate normal distributions. *arXiv:2303.10287*.

Liu, J. and Yang, Y. (2021). Asymptotics for systemic risk with dependent heavy-tailed losses. *ASTIN Bulletin: The Journal of the IAA*, 51(2):571–605.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools, Revised Edition*. Princeton University Press, Princeton.

Ogasawara, H. (2021). A non-recursive formula for various moments of the multivariate normal distribution with sectional truncation. *Journal of Multivariate Analysis*, 183:104729.

Sarabia, J. M., Gómez-Déniz, E., Prieto, F., and Jordá, V. (2018). Aggregation of dependent risks in mixtures of exponential distributions and extensions. *ASTIN Bulletin*, 48(3):1079–1107.

Shushi, T. and Yao, J. (2020). Multivariate risk measures based on conditional expectation and systemic risk for exponential dispersion models. *Insurance: Mathematics and Economics*, 93:178–186.

Su, J. and Furman, E. (2017). Multiple risk factor dependence structures: Distributional properties. *Insurance: Mathematics and Economics*, 76:56–68.

Su, J. and Hua, L. (2017). A general approach to full-range tail dependence copulas. *Insurance: Mathematics and Economics*, 77:49–64.

Zuo, B., Yin, C., and Yao, J. (2024). Multivariate range Value-at-Risk and covariance risk measures for elliptical and log-elliptical distributions. *Communications in Statistics-Theory and Methods*, in press.

Appendix A Algorithm for constructing the coefficient matrix

In order to implement the Euler algorithm given in (19),it is necessary to populate the coefficients involved in $\mathbf{M}(\boldsymbol{\Theta}, \boldsymbol{\theta})$, for a given $\boldsymbol{\Theta} \in \mathbb{R}^{(d+1)\times d}$ and $\boldsymbol{\theta} \in \mathbb{R}^{d+1}$. We start by setting up the partial derivative elements involved in $\boldsymbol{g}^{(\cdot)}$, whose order would not impact the outcome of the Euler algorithm, as long as the coefficients are positioned accordingly in $\mathbf{M}$. Recall that $\boldsymbol{g}^{(\cdot)}$ contains all the partial derivatives of g with respect to θ_i for all $i \in \mathcal{H} \subseteq \mathcal{I} = \{1, \ldots, d+1\}$. Thereby, there are 2^{d+1} elements in $\boldsymbol{g}^{(\cdot)}$. For the sake of programming convenience, we use binary strings of length $d+1$ to encode integers in the range between 0 and $(2^{d+1}-1)$. This step can be done with the aids of library functions in various computing platforms, e.g., the "bin()" function in `Julia` and `Python`. For $i \in \{0, \ldots, (2^{d+1}-1)\}$, let us denote the binary string of i by $\boldsymbol{b}^{[i]} = (b_1^{[i]}, \ldots, b_{d+1}^{[i]})$, where $b_j^{[i]} \in \{0, 1\}$

for $j = 1, \ldots, d+1$. Moreover, we let the bits of the binary string indicate which elements of $\boldsymbol{\theta}$ are involved in the partial derivative set. Formally, let $\mathcal{H}_i = \{j \in \mathcal{I} : b_j^{[i-1]} = 1\}$, $i = 1, \ldots, 2^{d+1}$, denote the i-th derivative set, which is associated with the binary representation of integer $(i-1)$. The partial derivative elements in $\boldsymbol{g}^{(\cdot)}$ are then indexed accordingly, such that $\boldsymbol{g}^{(\cdot)} = \big(g^{(\mathcal{H}_i)}, i = 1, \ldots, 2^{d+1}\big)$.

Next we proceed to setting up $\mathbf{M}$ row by row. Some additional notations are needed herein. For a $(d+1)$-dimensional binary string $\boldsymbol{b}$, let us define

$$\text{int}(\boldsymbol{b}) = \sum_{i=1}^{d+1} b_i \, 2^{d+1-i},$$

which coverts $\boldsymbol{b}$ into its corresponding decimal integer. Moreover, let $\mathcal{H}_i^c = \mathcal{I}/\mathcal{H}_i = \{k \in \mathcal{I} : b_k^{[i-1]} = 0\}$, which gathers all positions of zero bits in the binary string $\boldsymbol{b}^{[i-1]}$, $i \in \{1, \ldots, 2^{d+1}\}$. For a given $k \in \mathcal{H}_i^c$, define $\boldsymbol{b}^{[i;k]} \in \{0,1\}^{d+1}$, whose elements are given by

$$b_j^{[i;k]} = \begin{cases} 1, & \text{if } j = k; \\ b_j^{[i]}, & \text{otherwise.} \end{cases}$$

Compared to $\boldsymbol{b}^{[i]}$, the string $\boldsymbol{b}^{[i;k]}$ flips the k-th bit of $\boldsymbol{b}^{[i]}$ from zero to one.

Recall that the i-th row of $\mathbf{M}$ collects the coefficients needed to specify the relationship:

$$\frac{\partial}{\partial \theta_{d+1}} g^{(\mathcal{H}_i)}(\boldsymbol{\Theta}, \boldsymbol{\theta}) = \mathbf{M}(\boldsymbol{\Theta}, \boldsymbol{\theta}) \; \boldsymbol{g}^{(\cdot)}(\boldsymbol{\Theta}, \boldsymbol{\theta}), \qquad i = 1, \ldots, 2^{d+1}.$$

For a given $i \in \{1, \ldots, 2^{d+1}\}$, suppose that $(d+1) \notin \mathcal{H}_i$, or equivalently $b_{d+1}^{[i-1]} = 0$. Then, according to the differential equations specified in (16), all entries in the i-th row of $\mathbf{M}$ are zero, except for the entry in column $j = \text{int}(\boldsymbol{b}^{[i-1;d+1]}) + 1 = \text{int}(\boldsymbol{b}^{[i-1]}) + 2 = i + 1$. The entry at this position is equal to one.

In contrast, suppose $(d+1) \in \mathcal{H}_i$, or equivalently $b_{d+1}^{[i-1]} = 1$. In this case, the non-zero columns in the i-th row of $\mathbf{M}$ include $j = i$ and $j = \text{int}(\boldsymbol{b}^{[i-1;k]}) + 1 = \text{int}(\boldsymbol{b}^{[i-1]}) + 2^{d+1-k} + 1 = i + 2^{d+1-k}$, for all $k \in \mathcal{H}_i^c$. At $j = i$, the coefficient is calculated as $[\boldsymbol{\Lambda}^{-1}]_{\{d+1\} \times \mathcal{H}_i} \, \boldsymbol{\theta}_{\mathcal{H}_i}$. At $j = \text{int}(\boldsymbol{b}^{[i-1;k]}) = i + 2^{d+1-k}$, $k \in \boldsymbol{\xi}_i$, the coefficient is given by $\boldsymbol{\Lambda}_{\{k\} \times \mathcal{H}_i} \, \boldsymbol{\Lambda}^{-1}_{\mathcal{H}_i \times \{d+1\}}$.

Collectively, the entries in $\mathbf{M}$ are populated in the following manner. For $i \in \{1, \ldots, 2^{d+1}\}$ with $(d+1) \notin \mathcal{H}_i$, the entries in the i-th row of $\mathbf{M}$ are given by

$$m_{i,j} = \begin{cases} 1, & \text{if } j = i+1; \\ 0, & \text{otherwise.} \end{cases}$$

When $(d+1) \in \mathcal{H}_i$, the entries are given by

$$m_{i,j} = \begin{cases} [\mathbf{\Lambda}^{-1}]_{\{d+1\} \times \mathcal{H}_i}\ \boldsymbol{\theta}_{\mathcal{H}_i}, & \text{if } j = i; \\ \mathbf{\Lambda}_{\{k\} \times \mathcal{H}_i}\ \mathbf{\Lambda}^{-1}_{\mathcal{H}_i \times \{d+1\}}, & \text{if } j = i + 2^{d+1-k} \quad \text{for each } k \in \mathcal{H}_i^c; \\ 0, & \text{otherwise.} \end{cases}$$

Appendix B Technical proofs

Proof of Lemma 1. We will only prove the desired equality for X_1 as the same proof applies to the other elements. For $s > 0$, it holds that

$$\begin{aligned} F_{S_X}(s) &= C^{-1} \int\limits_{\substack{x_i > 0,\ i=1,\ldots,d \\ x_1 + \cdots + x_d < s}} \phi_d(x_1, \ldots, x_d; \boldsymbol{\mu}, \mathbf{\Sigma})\, \mathrm{d}x_d \cdots \mathrm{d}x_1 \\ &= C^{-1} \int_0^s \int_0^{s-x_1} \cdots \int_0^{s-x_1-\cdots-x_{d-1}} \phi_d(x_1, \ldots, x_d; \boldsymbol{\mu}, \mathbf{\Sigma})\, \mathrm{d}x_d \cdots \mathrm{d}x_1. \end{aligned} \tag{23}$$

Now, for $x_i > 0$, $i \in \mathcal{D}$, and $s > 0$, let us define

$$h_d(x_1, \ldots, x_{d-1}; s) = C^{-1} \times \int_0^{s - \sum_{j=1}^{d-1} x_j} \phi_d(x_1, \ldots, x_d; \boldsymbol{\mu}, \mathbf{\Sigma})\, \mathrm{d}x_d,$$

and

$$h_k(x_1, \ldots, x_{k-1}; s) = \int_0^{s - \sum_{j=1}^{k-1} x_j} h_{k+1}(x_1, \ldots, x_k; s)\, \mathrm{d}x_k, \qquad k \in \{2, \cdots, d-1\},$$

and

$$h_1(s) = \int_0^s h_2(x; s)\,\mathrm{d}x_1\,.$$

Then the CDF of S_X in (23) can be also expressed implicitly as

$$F_{S_X}(s) = h_1(s), \qquad s > 0.$$

By the Leibniz integral rule, for $k = 2, \ldots, d-1$, we have

$$\begin{aligned}\frac{\partial h_k(s)}{\partial s} &= h_{k+1}\Big(x_1, \ldots, x_{k-1}, s - \sum_{j=1}^{k-1} x_j; s\Big) + \int_0^{s-\sum_{j=1}^{k-1} x_j} \frac{\partial}{\partial s} h_{k+1}(x_1, \ldots, x_k; s)\,\mathrm{d}x_k \\ &= \int_0^{s-\sum_{j=1}^{k-1} x_j} \frac{\partial}{\partial s} h_{k+1}(x_1, \ldots, x_k; s)\,\mathrm{d}x_k.\end{aligned}$$

Thereby, the PDF of S_X can be computed via

$$\begin{aligned}f_{S_X}(s) &= \frac{\partial}{\partial x} F_{S_X}(x) \\ &= \int_0^s \frac{\partial}{\partial s} h_2(x_1; s)\,\mathrm{d}x_1 \\ &= \int_0^s \int_0^{s-x_1} \frac{\partial}{\partial s} h_3(x_1, x_2; s)\,\mathrm{d}x_2\,\mathrm{d}x_1 \\ &\;\;\vdots \\ &= \int_0^s \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-2} x_j} \frac{\partial}{\partial s}\, h_d\Big(x_1, \ldots, x_{d-1}; s\Big)\,\mathrm{d}x_{d-1} \cdots \mathrm{d}x_1 \\ &= C^{-1} \int_0^s \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-2} x_j} \phi_d\Big(x_1, \ldots, x_{d-1}, s - \sum_{j=1}^{d-1} x_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}\Big)\,\mathrm{d}x_{d-1} \cdots \mathrm{d}x_1.\end{aligned}$$

After some standard algebraic manipulations, we can obtain

$$\phi_d\Big(x_1, \ldots, x_{d-1}, s - \sum_{j=1}^{d-1} x_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}\Big) = \phi_{d-1}\big(x_1, \ldots, x_{d-1}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*\big) \times \beta(s), \tag{24}$$

where $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ is defined according to (8), and

$$\beta(s) = \frac{(2\pi)^{(d-1)/2}\det(\boldsymbol{\Sigma}^*)^{1/2}}{(2\pi)^{d/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\Big(\big(s\,\boldsymbol{i}_d - \boldsymbol{\mu}\big)^\top \boldsymbol{\Sigma}^{-1}\big(s\,\boldsymbol{i}_d - \boldsymbol{\mu}\big) - \boldsymbol{\mu}^{*\top}\,\boldsymbol{\Sigma}^{*-1}\,\boldsymbol{\mu}^*\Big).$$

Collectively, we readily have

$$\begin{aligned} f_{S_X}(s) &= \beta(s) \times C^{-1} \int_0^s \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-2} x_j} \phi_{d-1}\big(x_1, \ldots, x_{d-1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\big)\, \mathrm{d}x_{d-1} \cdots \mathrm{d}x_1 \\ &= \beta(s) \times C^{-1} \times \overline{\Phi}_{d-1}(\mathbf{0}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \times \mathbb{P}\big(S_{X^*} < s\big), \end{aligned}$$

where $S_{X^*} = \sum_{i=1}^{d-1} X_i^*$ is the aggregation of elements in $\boldsymbol{X}^* = (X_1^*, \ldots, X_{d-1}^*) \sim N_{d-1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \mathbf{0})$.

Next, we consider the joint distribution of (X_1, S_X). For $(x, s) \in \mathbb{R}_+^2$, it holds that

$$F_{X_1, S_X}(x, s) = C^{-1} \int_0^x \int_0^{s-x_1} \cdots \int_0^{s-x_1-\cdots-x_{d-1}} \phi_d(x_1, \ldots, x_d; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \mathrm{d}x_d \cdots \mathrm{d}x_1,$$

which has the same form as the CDF of S_X as per (23), except that the upper bound of the outermost integration is replaced by x. Applying the same arguments used to derive the PDF of S_X previously, we can get

$$\begin{aligned} \frac{\partial}{\partial s} F_{X_1, S_X}(x, s) &= \frac{\partial}{\partial s} C^{-1} \int_0^x \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-1} x_j} \phi_d(x_1, \ldots, x_d; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \mathrm{d}x_d \cdots \mathrm{d}x_1 \\ &= C^{-1} \int_0^x \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-2} x_j} \phi_d\Big(x_1, \ldots, x_{d-1}, s - \sum_{j=1}^{d-1} x_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}\Big)\, \mathrm{d}x_{d-1} \cdots \mathrm{d}x_1 \\ &= \beta(s) \times C^{-1} \int_0^x \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-2} x_j} \phi_{d-1}\big(x_1, \ldots, x_{d-1}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\big)\, \mathrm{d}x_{d-1} \cdots \mathrm{d}x_1 \\ &= \beta(s) \times C^{-1} \times \overline{\Phi}_{d-1}(\mathbf{0}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \times \mathbb{P}\big(X_1^* \le x, S_{X^*} < s\big). \end{aligned}$$

All in all, for any $(x, s) \in \mathbb{R}_+^2$, we have readily obtained

$$\begin{aligned} F_{X_1 | S_X = s}(x) &= \frac{\partial F_{X_1, S_X}(x, s)/\partial s}{f_{S_X}(s)} \\ &= \frac{\beta(s) \times C^{-1} \times \overline{\Phi}_{d-1}(\mathbf{0}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \times \mathbb{P}\big(X_1^* \le x, S_{X^*} < s\big)}{\beta(s) \times C^{-1} \times \overline{\Phi}_{d-1}(\mathbf{0}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \times \mathbb{P}\big(S_{X^*} < s\big)} \end{aligned}$$

$$= \mathbb{P}(X_1^* \leq x \,|\, S_{X^*} < s),$$

which establishes the desired result. The proof is now finished. □

Proof of Theorem 3. To simplify the presentation, we will demonstrate the result for X_1 only. The result for the other elements of $\boldsymbol{X}$ is proved in the same manner. Suppose that $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0})$, let $\mu_+ = \sum_{i=1}^d \mu_i$ and $\pi(s) = \overline{\Phi}_d(\mathbf{0}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \mathbb{P}(S_X < s)$ for any $s > 0$. For $\boldsymbol{x}, \boldsymbol{w} \in \mathbb{R}^d$, the following string of relationships holds:

$$\begin{aligned}
\mathrm{LTE}(X_1, S_X; s) &= \frac{1}{\mathbb{P}(S_X < s)} \mathbb{E}\big(X_1 \, \mathbb{1}(S_X \leq s)\big) \\
&= \frac{\overline{\Phi}_d(\mathbf{0}; \boldsymbol{\mu}, \boldsymbol{\Sigma})^{-1}}{\mathbb{P}(S_X < s)} \int\limits_{\substack{x_i > 0, \ i=1,\ldots,d \\ x_1 + \cdots + x_d < s}} x_1 \, \phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \mathrm{d}\boldsymbol{x} \\
&= \mu_1 + \frac{1}{\pi(s)} \int\limits_{\substack{x_i > 0, \ i=1,\ldots,d \\ x_1 + \cdots + x_d < s}} (x_1 - \mu_1) \, \phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \mathrm{d}\boldsymbol{x} \\
&= \mu_1 + \frac{1}{\pi(s)} \int\limits_{\substack{w_i > -\mu_i, \ i=1,\ldots,d \\ w_1 + \cdots + w_d < s - \mu_+}} w_1 \, \phi_d(\boldsymbol{w}; \mathbf{0}, \boldsymbol{\Sigma}) \, \mathrm{d}\boldsymbol{w} \\
&= \mu_1 + \mathbb{E}(W_1), \qquad (25)
\end{aligned}$$

where W_1 is the first element of $\boldsymbol{W} = (W_1, \ldots, W_d)^\top$, which follows a multivariate normal distribution truncated over the area:

$$\mathcal{W} = \Big\{ \boldsymbol{w} \in \mathbb{R}^d : \boldsymbol{w} > -\boldsymbol{\mu}, \ \sum_{j=1}^d w_j < s - \mu_+ \Big\}.$$

The joint PDF of $\boldsymbol{W}$ can be written as

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{\pi(s)} \, \phi_d(\boldsymbol{w}, \mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{w} \in \mathcal{W}.$$

To compute the expectation of W_1, we resort to the moment generating function of $\boldsymbol{W}$. For $\boldsymbol{t} = (t_1, \ldots, t_d)^\top \in \mathbb{R}^d$, we have

$$
\begin{aligned}
M_{\boldsymbol{W}}(\boldsymbol{t}) &:= \mathbb{E}\big(\exp(\boldsymbol{t}^\top \boldsymbol{W})\big) \\
&= \frac{1}{\pi(s)} \, \frac{1}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \int_{\boldsymbol{w} \in \mathcal{W}} \exp\left(\boldsymbol{t}'\boldsymbol{w} - \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{w}\right) \mathrm{d}\boldsymbol{w} \\
&= \frac{1}{\pi(s)} \, \frac{1}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \int_{\boldsymbol{w} \in \mathcal{W}} \exp\left(\frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t} - \frac{1}{2}(\boldsymbol{w} - \boldsymbol{\Sigma}\boldsymbol{t})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{w} - \boldsymbol{\Sigma}\boldsymbol{t})\right) \mathrm{d}\boldsymbol{w} \\
&= \frac{1}{\pi(s)} \, \exp\left(\frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right) \int_{\boldsymbol{w} \in \mathcal{W}} \phi_d(\boldsymbol{w}; \boldsymbol{\Sigma}\boldsymbol{t}, \boldsymbol{\Sigma}) \mathrm{d}\boldsymbol{w} \\
&= \frac{1}{\pi(s)} \times m_1(\boldsymbol{t}) \times m_2(\boldsymbol{t}),
\end{aligned}
$$

where

$$
m_1(\boldsymbol{t}) := \exp\left(\frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right) \qquad \text{and} \qquad m_2(\boldsymbol{t}) := \int_{\boldsymbol{w} \in \mathcal{W}} \phi_d(\boldsymbol{w}; \boldsymbol{\Sigma}\boldsymbol{t}, \boldsymbol{\Sigma}) \mathrm{d}\boldsymbol{w}. \tag{26}
$$

Thereby,

$$
\begin{aligned}
\mathbb{E}(W_1) &= \frac{\partial}{\partial t_1} M_{\boldsymbol{W}}(\boldsymbol{t}) \Big|_{\boldsymbol{t}=\boldsymbol{0}} \\
&= \frac{1}{\pi(s)} \times \left(m_2(\boldsymbol{t}) \times \frac{\partial}{\partial t_1} m_1(\boldsymbol{t}) \Big|_{\boldsymbol{t}=\boldsymbol{0}} + m_1(\boldsymbol{t}) \times \frac{\partial}{\partial t_1} m_2(\boldsymbol{t}) \Big|_{\boldsymbol{t}=\boldsymbol{0}}\right).
\end{aligned} \tag{27}
$$

First, we have

$$
\frac{\partial m_1(\boldsymbol{t})}{\partial t_1} \Big|_{\boldsymbol{t}=\boldsymbol{0}} = \exp\left(\frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}\right) \times \sum_{j=1}^{d} \sigma_{1j} \, t_j \Big|_{\boldsymbol{t}=\boldsymbol{0}} = 0. \tag{28}
$$

Next, we consider the derivative of m_2, which is more demanding to calculate. Our first step is to transform the $\boldsymbol{t}$ variable from the integrand to the integration boundaries. To do so, we apply a change of variables $x_i = w_i - \gamma_i(\boldsymbol{t}) + \mu_i$, where $\gamma_i(\boldsymbol{t})$ represents the i-th element of vector $\boldsymbol{\gamma}(\boldsymbol{t}) = \boldsymbol{\Sigma}\,\boldsymbol{t}$, $i \in \mathcal{D}$.

Let $\gamma_+(\boldsymbol{t}) = \sum_{i=1}^{d} \gamma_i(\boldsymbol{t})$, we can rewrite the m_2 function in (26) as follows:

$$\begin{aligned} m_2(\boldsymbol{t}) &= \int_{\boldsymbol{w}\in\mathcal{W}} \phi_d(\boldsymbol{w}; \boldsymbol{\Sigma}\boldsymbol{t}, \boldsymbol{\Sigma})\, \mathrm{d}\boldsymbol{w} \\ &= \int\limits_{\substack{x_i > -\gamma_i(\boldsymbol{t}),\ i=1,\ldots,d \\ x_1+\cdots+x_d < s-\gamma_+(\boldsymbol{t})}} \phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \mathrm{d}\boldsymbol{x} \\ &= \int_{-\gamma_1(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})} \int_{-\gamma_2(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})-x_1} \cdots \int_{-\gamma_d(\boldsymbol{t})}^{s_q-\gamma_+(\boldsymbol{t})-\sum_{j=1}^{d-1} x_j} \phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \mathrm{d}\boldsymbol{x}. \end{aligned}$$

The technique we use to tackle the derivative of m_2 is similar to the one employed in the proof of Lemma 1. To this end, let us define

$$\zeta_d(x_1, \ldots, x_{d-1}; \boldsymbol{t}) = \int_{-\gamma_d(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})-\sum_{j=1}^{d-1} x_j} \phi_d(x_1, \ldots, x_d; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \mathrm{d}x_d,$$

and

$$\zeta_k(x_1, \ldots, x_{k-1}; \boldsymbol{t}) = \int_{-\gamma_k(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})-\sum_{j=1}^{k-1} x_j} \zeta_{k+1}(x_1, \ldots, x_k; \boldsymbol{t})\, \mathrm{d}x_k, \quad k \in \{2, \ldots, d-1\}, \tag{29}$$

and

$$\zeta_1(\boldsymbol{t}) = \int_{-\gamma_1(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})} \zeta_2(x_1; \boldsymbol{t})\, \mathrm{d}x_1.$$

Then, we can rewrite the m_2 function implicitly as

$$m_2(\boldsymbol{t}) = \zeta_1(\boldsymbol{t}).$$

By applying the Leibniz integral rule, for $k = 2, \ldots, d-1$, we get

$$\begin{aligned} &\frac{\partial}{\partial t_1} \zeta_k(x_1, \ldots, x_{k-1}; \boldsymbol{t}) \\ &\quad = -\gamma_1(\mathbf{1}) \times \zeta_{k+1}\Big(x_1, \ldots, x_{k-1}, s - \gamma_+(\boldsymbol{t}) - \sum_{j=1}^{k-1} x_j; \boldsymbol{t}\Big) \end{aligned}$$

$$+ \sigma_{k,1} \times \zeta_{k+1}\big(x_1, \ldots, x_{k-1}, -\gamma_k(\boldsymbol{t}); \boldsymbol{t}\big) + \int_{-\gamma_k}^{s-\gamma_+(\boldsymbol{t})-\sum_{j=1}^{k-1} x_j} \frac{\partial}{\partial t_1} \zeta_{k+1}(x_1, \ldots, x_k; \boldsymbol{t})\, \mathrm{d}x_k.$$

Therefore, we can write

$$\begin{aligned}
\frac{\partial}{\partial t_1} m_2(\boldsymbol{t}) &= \frac{\partial}{\partial t_1} \zeta_1(\boldsymbol{t}) \\
&= -\gamma_1(\mathbf{1}) \times \zeta_2\big(s - \gamma_+(\boldsymbol{t}); \boldsymbol{t}\big) + \sigma_{1,1} \times \zeta_2\big(-\gamma_1(\boldsymbol{t}); \boldsymbol{t}\big) + \int_{-\gamma_1(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})} \frac{\partial}{\partial t_1} \zeta_2(x_1; \boldsymbol{t})\, \mathrm{d}x_1 \\
&\quad \vdots \\
&= I_1(\boldsymbol{t}) + I_2(\boldsymbol{t}), \qquad (30)
\end{aligned}$$

where

$$\begin{aligned}
I_1(\boldsymbol{t}) := -\gamma_1(\mathbf{1}) \times \Bigg[& \zeta_2\big(s - \gamma_+(\boldsymbol{t}); \boldsymbol{t}\big) \\
& + \sum_{k=2}^{d-1} \int_{-\gamma_1(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})} \cdots \int_{-\gamma_{k-1}(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})-\sum_{j=1}^{k-2} x_j} \zeta_{k+1}\Big(x_1, \ldots, x_{k-1}, s - \gamma_+(\boldsymbol{t}) - \sum_{j=1}^{k-1} x_j; \boldsymbol{t}\Big)\, \mathrm{d}x_{k-1} \cdots \mathrm{d}x_1 \\
& + \int_{-\gamma_1(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})} \cdots \int_{-\gamma_{d-1}(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})-\sum_{j=1}^{d-2} x_j} \phi_d\Big(x_1, \ldots, x_{d-1}, s - \gamma_+(\boldsymbol{t}) - \sum_{j=1}^{d-1} x_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}\Big)\, \mathrm{d}x_{d-1} \cdots \mathrm{d}x_1 \Bigg],
\end{aligned}$$

and

$$\begin{aligned}
I_2(\boldsymbol{t}) := & \sigma_{1,1} \times \zeta_2\big(-\gamma_1(\boldsymbol{t}); \boldsymbol{t}\big) \\
& + \sum_{k=2}^{d-1} \sigma_{k,1} \times \int_{-\gamma_1(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})} \cdots \int_{-\gamma_{k-1}(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})-\sum_{j=1}^{k-2} x_j} \zeta_{k+1}\big(x_1, \ldots, x_{k-1}, -\gamma_k(\boldsymbol{t}); \boldsymbol{t}\big)\, \mathrm{d}x_{k-1} \cdots \mathrm{d}x_1 \\
& + \sigma_{d,1} \times \int_{-\gamma_1(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})} \cdots \int_{-\gamma_{d-1}(\boldsymbol{t})}^{s-\gamma_+(\boldsymbol{t})-\sum_{j=1}^{d-2} x_j} \phi_d\big(x_1, \ldots, x_{d-1}, -\gamma_d(\boldsymbol{t}); \boldsymbol{\mu}, \boldsymbol{\Sigma}\big)\, \mathrm{d}x_{d-1} \cdots \mathrm{d}x_1. \qquad (31)
\end{aligned}$$

Note that when $\boldsymbol{t} = \mathbf{0}$, then $\boldsymbol{\gamma}(\mathbf{0}) = \mathbf{0}$. Moreover, by the definition of ζ function, we have

$$\zeta_2\big(s - \gamma_+(\mathbf{0}); \mathbf{0}\big) = 0,$$

and for $k = 2, \ldots, d-1$,

$$\zeta_{k+1}\Big(x_1, \ldots, x_{k-2}, s - \gamma_+(\mathbf{0}) - \sum_{j=1}^{k-1} x_j; \mathbf{0}\Big) = 0.$$

Consequently, we have

$$\begin{aligned} I_1(\mathbf{0}) &= -\gamma_1(\mathbf{1}) \times \int_0^s \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-1} x_j} \phi_d\Big(z_1, \ldots, z_{d-1}, s - \sum_{j=1}^{d-1} z_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}\Big)\, \mathrm{d}z_{d-1} \cdots \mathrm{d}z_1 \\ &= -\gamma_1(\mathbf{1}) \times \beta(s) \times \int_0^s \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-1} z_j} \phi_{d-1}\Big(z_1, \ldots, z_{d-1}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*\Big)\, \mathrm{d}z_{d-1} \cdots \mathrm{d}z_1 \\ &= -\gamma_1(\mathbf{1}) \times \beta(s) \times \Psi(\mathcal{S}_{d-1}(s); \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*), \end{aligned} \tag{32}$$

where the second equality holds due to (24), $\mathcal{S}_{d-1}(\cdot)$ is given in (10), and $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ is defined in accordance with (8).

Now, let us turn to studying I_2. By the definition of ζ function, it is not difficult to check

$$\zeta_2(0; \mathbf{0}) = \int_0^s \int_0^{s-x_2} \cdots \int_0^{s-\sum_{i=2}^{d-1} x_j} \phi_d\big(0, \ldots, x_{d-1}, x_d; \boldsymbol{\mu}, \boldsymbol{\Sigma}\big)\, \mathrm{d}x_d \cdots \mathrm{d}x_2. \tag{33}$$

For $k = 2, \ldots, d-1$, we have

$$\begin{aligned} &\int_0^s \cdots \int_0^{s-\sum_{j=1}^{k-2} x_j} \zeta_{k+1}\big(x_1, \ldots, x_{k-1}, 0; \mathbf{0}\big)\, \mathrm{d}x_{k-1} \cdots \mathrm{d}x_1 \\ &= \int_0^s \cdots \int_0^{s-\sum_{j=1}^{k-2} x_j} \int_0^{s-\sum_{j=1}^{k-1} x_j} \zeta_{k+2}\big(x_1, \ldots, x_{k-1}, 0, x_{k+1}; \mathbf{0}\big)\, \mathrm{d}x_{k+1}\, \mathrm{d}x_{k-1} \cdots \mathrm{d}x_1 \\ &\vdots \\ &= \int_0^s \cdots \int_0^{s-\sum_{j=1}^{k-2} x_j} \int_0^{s-\sum_{j=1}^{k-1} x_j} \int_0^{s-\sum_{j=1, j\neq k}^{k+1} x_j} \cdots \int_0^{s-\sum_{j=1, j\neq k}^{d-1} x_j} \\ &\qquad \phi_d\big(x_1, \ldots, x_{k-1}, 0, x_{k+1}, \ldots, x_d; \boldsymbol{\mu}, \boldsymbol{\Sigma}\big)\, \mathrm{d}x_d \cdots \mathrm{d}x_{k+1}\, \mathrm{d}x_{k-1} \cdots \mathrm{d}x_1. \end{aligned} \tag{34}$$

Plug (33) and (34) into (31), we can simplify the expression of I_2 when $\boldsymbol{t} = \mathbf{0}$ into

$$I_2(\mathbf{0}) = \sum_{k=1}^{d-1} \sigma_{k,1} \int_0^s \cdots \int_0^{s-\sum_{j=1}^{k-1} x_j} \int_0^{s-(\sum_{j=1,j\neq k}^{k+1} x_j)} \cdots \int_0^{s-\sum_{j=1,j\neq k}^{d-1} x_j} \phi_d\big(x_1, \ldots, x_{k-1}, 0, x_{k+1}, \ldots, x_d; \boldsymbol{\mu}, \boldsymbol{\Sigma}\big)\, \mathrm{d}x_d \cdots \mathrm{d}x_{k+1}\, \mathrm{d}x_{k-1} \cdots \mathrm{d}x_1$$
$$+ \sigma_{d,1} \times \int_0^s \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-2} x_j} \phi_d\big(x_1, \ldots, x_{d-1}, 0\,;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}\big)\, \mathrm{d}x_{d-1} \cdots \mathrm{d}x_1. \tag{35}$$

Moreover, note that for $k = 1, \ldots, d$, and $x_k \in \mathbb{R}$, it holds that

$$\phi_d\big(x_1, \ldots, x_{k-1}, 0, x_{k+1}, \ldots, x_d; \boldsymbol{\mu}, \boldsymbol{\Sigma}\big) = \phi_{d-1}\big(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_d; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\big) \times \alpha_k,$$

where α_k is defined according to (12), and $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is defined as per (11). As a result, we can further simplify (35) into

$$\begin{aligned} I_2(\mathbf{0}) &= \sum_{k=1}^{d-1} \sigma_{k,1} \times \alpha_k \times \int_0^s \cdots \int_0^{s-\sum_{j=1}^{k-1} x_j} \int_0^{s-(\sum_{j=1,j\neq k}^{k+1} x_j)} \cdots \int_0^{s-\sum_{j=1,j\neq k}^{d-1} x_j} \\ &\qquad \phi_{d-1}\big(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_d; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\big)\, \mathrm{d}x_d \cdots \mathrm{d}x_{k+1}\, \mathrm{d}x_{k-1} \cdots \mathrm{d}x_1 \\ &\quad + \sigma_{d,1} \times \alpha_d \times \int_0^s \int_0^{s-x_1} \cdots \int_0^{s-\sum_{j=1}^{d-2} x_j} \phi_{d-1}\big(x_1, \ldots, x_{d-1},\,;\, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\big)\, \mathrm{d}x_{d-1} \cdots \mathrm{d}x_1 \\ &= \sum_{k=1}^{d} \sigma_{k,1} \times \alpha_k \times \int_0^s \int_0^{s-z_1} \cdots \int_0^{s-\sum_{j=1}^{d-1} z_j} \phi_{d-1}\big(z_1, \ldots, z_{d-1}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\big)\, \mathrm{d}z_{d-1} \cdots \mathrm{d}z_1 \\ &= \sum_{k=1}^{d} \sigma_{k,1} \times \alpha_k \times \Psi(\mathcal{S}_{d-1}(s); \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \end{aligned} \tag{36}$$

Substitute (32) and (36) into (30), and then apply the relationships noted in (27) and (28), we obtain

$$\mathbb{E}(W_1) = \frac{1}{\pi(s)} \left[-\Big(\sum_{k=1}^{d} \sigma_{k,1}\Big) \times \beta(s) \times \Psi(\mathcal{S}_{d-1}(s); \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) + \sum_{k=1}^{d} \sigma_{k,1} \times \alpha_k \times \Psi(\mathcal{S}_{d-1}(s); \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right].$$

Finally, by combining the above formula with the expression for the LTE derived in (25), we readily obtain the desired result of the proposition. This completes the proof. □

Proof of Theorem 4. If $m \notin \mathcal{H}$, then the derivative expression holds simply by definition, so we are going to focus on proving the result when $m \in \mathcal{H}$.

Let us first consider the case of $\mathcal{H} \subset \mathcal{I}$. Note that the order of constraints does not affect the simplex region they describe, so without loss of generality we can assume $\mathcal{H} = \{1, \cdots, p\}$, where $1 \leq p \leq d$. For now, we assume that the parameter $\boldsymbol{\Theta}$ satisfies the following condition:

$$\theta_{i,j} = 0, \quad 1 \leq i \leq p \quad \text{and} \quad (p+1) \leq j \leq (d+1). \tag{37}$$

At the end of this proof, we will show that this condition can be relaxed without affecting the expression of the differential equation system.

Next, let us define a $d \times d$ matrix $\mathbf{U}$ such that

$$\mathbf{U} = \begin{pmatrix} \boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d-p} \end{pmatrix}, \tag{38}$$

where $\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}} = \boldsymbol{\Theta}\big(\{i_1 | i_2 \in \mathcal{H}\}, \{i_2 | i_2 \in \mathcal{H}\}\big)$ is a $p \times p$ matrix containing the i_1-th rows and i_2-th columns of $\boldsymbol{\Theta}$ for all $i_1, i_2 \in \mathcal{H}$. Consider a transformation of $\boldsymbol{y} = \mathbf{U}\boldsymbol{x}$. Correspondingly, we have

$$x_i = \begin{cases} \sum_{j=1}^{p} [\mathbf{U}^{-1}]_{i,j} \times y_j, & 1 \leq i \leq p; \\ y_i, & p+1 \leq i \leq d. \end{cases}$$

Herein, $\mathbf{U}^{-1}$ exists because we assume $\boldsymbol{\Theta}$ has full rank.

Note that $\det(\mathbf{U}) = \det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})$. We can write

$$\begin{aligned} g(\boldsymbol{\Theta}, \boldsymbol{\theta}) &= \int_{\mathbb{R}^d} (2\pi)^{-d/2} \exp\Big(-\frac{1}{2}\sum_{i=1}^{d} x_i^2\Big) \prod_{i=1}^{d+1} \mathbb{1}\Big(\sum_{j=1}^{d} \theta_{i,j}\, x_j + \theta_i > 0\Big)\, \mathrm{d}\boldsymbol{x} \\ &= \frac{(2\pi)^{-d/2}}{\det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})} \int_{\mathbb{R}^d} \exp\Big(-\frac{1}{2}\Big(\sum_{i=1}^{p}(\sum_{j=1}^{p}[\mathbf{U}^{-1}]_{i,j}\, y_j)^2 + \sum_{i=p+1}^{d} y_i^2\Big)\Big) \prod_{i=1}^{p} \mathbb{1}(y_i + \theta_i > 0) \\ &\qquad \prod_{i=p+1}^{d+1} \mathbb{1}\Big(\sum_{j=1}^{p} \theta_{i,j} \sum_{k=1}^{p} [\mathbf{U}^{-1}]_{j,k}\, y_k + \sum_{j=p+1}^{d} \theta_{i,j}\, y_j + \theta_i > 0\Big)\, \mathrm{d}\boldsymbol{y}. \end{aligned} \tag{39}$$

For $y_1, \ldots, y_p \in \mathbb{R}$, define

$$\eta_1(y_1, \ldots, y_p) = (2\pi)^{-p/2} \exp\Big(-\frac{1}{2} \sum_{i=1}^{p} (\sum_{j=1}^{p} [\mathbf{U}^{-1}]_{i,j}\, y_j)^2 \Big),$$

and

$$\eta_2(y_1, \ldots, y_p; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-(d-p)/2} \exp\Big(-\frac{1}{2} \sum_{i=p+1}^{d} y_i^2 \Big) \prod_{i=p+1}^{d+1} \mathbb{1}\Big(\sum_{j=1}^{p} \theta_{i,j} \sum_{k=1}^{p} [\mathbf{U}^{-1}]_{j,k}\, y_k + \sum_{j=p+1}^{d} \theta_{i,j}\, y_j + \theta_i > 0 \Big) \, \mathrm{d}y_{p+1} \cdots \mathrm{d}y_d, \quad (40)$$

where $\boldsymbol{\theta}_{\mathcal{A}}$ denotes a subvector that contain the i-th elements of $\boldsymbol{\theta}$ for all $i \in \mathcal{A} \subseteq \mathcal{I}$. Moreover, set $\eta(y_1, \ldots, y_p; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}) = \det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})^{-1}\, \eta_1(y_1, \ldots, y_p) \times \eta_2(y_1, \ldots, y_p; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}})$. Then we can simplify the expression of $g(\boldsymbol{\Theta}, \boldsymbol{\theta})$ in Equation (39) into

$$\int_{-\theta_p}^{\infty} \cdots \int_{-\theta_1}^{\infty} \eta(y_1, \ldots, y_p; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}})\, \mathrm{d}y_1 \cdots \mathrm{d}y_p.$$

Thereby, we have

$$g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) = \frac{\partial^p}{\partial\theta_1 \cdots \partial\theta_p} g(\boldsymbol{\Theta}, \boldsymbol{\theta}) = \eta(-\boldsymbol{\theta}_{\mathcal{H}}; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}). \quad (41)$$

Now we proceed to calculate

$$\frac{\partial}{\partial\theta_m} g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) = \frac{\partial}{\partial\theta_m}\, \eta(-\boldsymbol{\theta}_{\mathcal{H}}; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}), \qquad m \in \mathcal{H}.$$

Recall that $\eta(\cdot) = \det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})^{-1} \eta_1(\cdot) \times \eta_2(\cdot)$. Applying the product rule, on the one hand, we have, for $m \in \mathcal{H}$,

$$\begin{aligned} \det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})^{-1} \frac{\partial}{\partial\theta_m} \eta_1(-\boldsymbol{\theta}_{\mathcal{H}}) \times \eta_2(-\boldsymbol{\theta}_{\mathcal{H}}; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}) &= \eta(-\boldsymbol{\theta}_{\mathcal{H}}; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}) \times -\frac{\partial}{\partial\theta_m} \frac{1}{2} \sum_{i=1}^{p} (-\sum_{j=1}^{p} [\mathbf{U}^{-1}]_{i,j}\, \theta_j)^2 \\ &= -g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) \sum_{j=1}^{p} (\sum_{i=1}^{p} [\mathbf{U}^{-1}]_{i,j} \times [\mathbf{U}^{-1}]_{i,m} \times \theta_j) \end{aligned}$$

$$= -g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) \sum_{j=1}^{p} [\boldsymbol{\Lambda}^{-1}]_{j,m} \times \theta_j. \tag{42}$$

On the other hand, to compute the partial derivative of η_2, define

$$\eta_2^*(-\boldsymbol{\theta}_{\mathcal{H}}; \widetilde{\boldsymbol{\theta}}_{\mathcal{I}\setminus\mathcal{H}}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi_{d-p}(y_{p+1}, \ldots, y_d) \prod_{i=p+1}^{d+1} \mathbb{1}\Big(\sum_{j=p+1}^{d} \theta_{i,j}\, y_j + \theta_i^* > 0 \Big)\, \mathrm{d}y_{p+1} \cdots \mathrm{d}y_d,$$

where $\theta_i^* = -\sum_{j=1}^{p} \theta_{i,j} \sum_{k=1}^{p} [\mathbf{U}^{-1}]_{j,k}\, \theta_k + \theta_i$, $i \in \mathcal{I}\setminus\mathcal{H}$. It is elementary to check that $\eta_2(-\boldsymbol{\theta}_{\mathcal{H}}; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}) = \eta_2^*(-\boldsymbol{\theta}_{\mathcal{H}}; \widetilde{\boldsymbol{\theta}}_{\mathcal{I}\setminus\mathcal{H}})$. Moreover, for $i \in \mathcal{I} \setminus \mathcal{H}$, because $\mathrm{d}\theta_i^*/\mathrm{d}\theta_i = 1$, we have

$$\frac{\partial}{\partial \theta_i} \eta_2(-\boldsymbol{\theta}_{\mathcal{H}}; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}) = \frac{\partial}{\partial \theta_i^*} \eta_2^*(-\boldsymbol{\theta}_{\mathcal{H}}; \widetilde{\boldsymbol{\theta}}_{\mathcal{I}\setminus\mathcal{H}}).$$

Thereby, we get, for $i \in \mathcal{I} \setminus \mathcal{H}$,

$$\begin{aligned} g^{(\mathcal{H}\cup\{i\})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) &= \frac{\partial}{\partial \theta_i} g^{(\mathcal{H})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) \\ &= \frac{1}{\det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})} \times \eta_1(-\boldsymbol{\theta}_{\mathcal{H}}) \times \frac{\partial}{\partial \theta_i} \eta_2(-\boldsymbol{\theta}_{\mathcal{H}}; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}) \\ &= \frac{1}{\det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})} \times \eta_1(-\boldsymbol{\theta}_{\mathcal{H}}) \times \frac{\partial}{\partial \theta_i^*} \eta_2(-\boldsymbol{\theta}_{\mathcal{H}}; \widetilde{\boldsymbol{\theta}}_{\mathcal{I}\setminus\mathcal{H}}). \end{aligned}$$

Now we can conclude that, for $m \in \mathcal{H}$,

$$\begin{aligned} \det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})^{-1}\, \eta_1(-\boldsymbol{\theta}_{\mathcal{H}}) \times \frac{\partial}{\partial \theta_m} \eta_2(-\boldsymbol{\theta}_{\mathcal{H}}; \boldsymbol{\theta}_{\mathcal{I}\setminus\mathcal{H}}) &= \det(\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}})^{-1} \sum_{i=p+1}^{d+1} \eta_1(-\boldsymbol{\theta}_{\mathcal{H}}) \times \frac{\partial}{\partial \theta_i^*} \eta_2^*(\widetilde{\boldsymbol{\theta}}_{\mathcal{I}\setminus\mathcal{H}}) \times \frac{\mathrm{d}\theta_i^*}{\mathrm{d}\theta_m} \\ &= \sum_{i=d+1}^{d+1} g^{(\mathcal{H}\cup\{i\})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) \times \frac{\mathrm{d}\theta_i^*}{\mathrm{d}\theta_m} \\ &= -\sum_{i=p+1}^{d+1} g^{(\mathcal{H}\cup\{i\})}(\boldsymbol{\Theta}, \boldsymbol{\theta}) \sum_{j=1}^{p} \theta_{i,j}\, [\mathbf{U}^{-1}]_{j,m}. \end{aligned} \tag{43}$$

Further, because $\mathbf{U}_{\mathcal{H}\times\mathcal{H}}^{-1}\,\boldsymbol{\Theta}_{\mathcal{H}\times\mathcal{H}} = \mathbf{I}_p$, for $1 \le j_1 \ne j_2 \le p$, the following relationships hold:

$$\sum_{k=1}^{p} [\mathbf{U}^{-1}]_{j_1,k}\,\theta_{k,j_2} = \begin{cases} 1, & \text{if } j_1 = j_2; \\ 0, & \text{if } j_1 \ne j_2. \end{cases}$$

So we can further rewrite the expression in (43) into

$$\begin{aligned} & -\sum_{i=p+1}^{d+1} g^{(\mathcal{H}\cup\{i\})}(\boldsymbol{\Theta},\boldsymbol{\theta}) \sum_{j=1}^{p} \theta_{i,j}\,[\mathbf{U}^{-1}]_{j,m} \sum_{k=1}^{p} [\mathbf{U}^{-1}]_{j,k}\,\theta_{k,j} \\ = & -\sum_{i=p+1}^{d+1} g^{(\mathcal{H}\cup\{i\})}(\boldsymbol{\Theta},\boldsymbol{\theta}) \sum_{k=1}^{p} \Big(\sum_{j_1=1}^{p} [\mathbf{U}^{-1}]_{j_1,k}\,[\mathbf{U}^{-1}]_{j_1,m}\Big)\Big(\sum_{j_2=1}^{p} \theta_{i,j_2}\,\theta_{k,j_2}\Big) \\ = & -\sum_{i=p+1}^{d+1} g^{(\mathcal{H}\cup\{i\})}(\boldsymbol{\Theta},\boldsymbol{\theta}) \sum_{k=1}^{p} [\boldsymbol{\Lambda}^{-1}]_{k,m} \times \lambda_{i,k}. \end{aligned} \tag{44}$$

Combining the results in (42) and (44), we readily obtain that if $\boldsymbol{\Theta}$ satisfies the condition specified in (37), then it holds that

$$\frac{\partial}{\partial\theta_m} g^{(\mathcal{H})}(\boldsymbol{\Theta},\boldsymbol{\theta}) = -\sum_{j=1}^{p} [\boldsymbol{\Lambda}^{-1}]_{j,m}\,\theta_j\, g^{(\mathcal{H})}(\boldsymbol{\Theta},\boldsymbol{\theta}) - \sum_{j=1}^{p} [\boldsymbol{\Lambda}^{-1}]_{j,m} \sum_{i=p+1}^{d+1} \lambda_{i,j}\; g^{(\mathcal{H}\cup\{i\})}(\boldsymbol{\Theta},\boldsymbol{\theta}).$$

Recall the notation $\mathcal{D} = \{1,\ldots,d\}$. In the case of $\mathcal{H} = \mathcal{I}$, we have already known from (41) that

$$g^{(\mathcal{D})}(\boldsymbol{\Theta},\boldsymbol{\theta}) = \frac{(2\pi)^{-d/2}}{\det(\boldsymbol{\Theta}_{\mathcal{D}\times\mathcal{D}})} \exp\Big(-\frac{1}{2}\Big(\sum_{i=1}^{d}\Big(\sum_{j=1}^{d} [\boldsymbol{\Theta}_{\mathcal{D}\times\mathcal{D}}^{-1}]_{i,j}\,\theta_j\Big)^2\Big)\Big),$$

which is independent of θ_{d+1}. Thus, $g^{(\mathcal{H})}(\boldsymbol{\Theta},\boldsymbol{\theta})$ equals zero, and so does $\partial/\partial\theta_m\, g^{(\mathcal{H})}(\boldsymbol{\Theta},\boldsymbol{\theta})$ for any $m \in \mathcal{I}$.

Throughout the rest of this proof, we will demonstrate how to relax the constraint specified in (37). Firstly, we note that if $\boldsymbol{\Theta}$ does not satisfy the condition in (37), there always exists an orthogonal matrix, says Ω, such that $\widetilde{\boldsymbol{\Theta}} = \boldsymbol{\Theta}\boldsymbol{\Omega}^\top$ can satisfy (37). Further, it holds that, for $\boldsymbol{Z} \sim N_d(\mathbf{0}, \mathbf{I}_d)$,

$$g(\boldsymbol{\Theta},\boldsymbol{\theta}) = \mathbb{P}(\boldsymbol{\Theta}\,\boldsymbol{Z} + \boldsymbol{\theta} > 0) = \mathbb{P}(\boldsymbol{\Theta}\boldsymbol{\Omega}^\top\boldsymbol{\Omega}\boldsymbol{Z} + \boldsymbol{\theta} > 0) = \mathbb{P}(\widetilde{\boldsymbol{\Theta}}\boldsymbol{Z} + \boldsymbol{\theta} > 0) = g(\widetilde{\boldsymbol{\Theta}},\boldsymbol{\theta}),$$

where the second last equality holds because $\boldsymbol{\Omega Z} \overset{d}{=} \boldsymbol{Z}$. Moreover, the coefficients involved in the PDE system, $\boldsymbol{\Lambda}$, remain invariant under any orthogonal transformation of the underlying parameter matrix $\boldsymbol{\Theta}$. Namely, it always holds that $\boldsymbol{\Lambda} = \boldsymbol{\Theta}^\top \boldsymbol{\Theta} = \widetilde{\boldsymbol{\Theta}}^\top \widetilde{\boldsymbol{\Theta}}$. Collectively, this implies that even if $\boldsymbol{\Theta}$ does not satisfy (37), the system of PDE's governing g remains identical to that associated with $\widetilde{\boldsymbol{\Theta}}$, which satisfies the condition in (37).

The proof is now completed. □

Proof of Lemma 5. Consider a system of ODE's:

$$\boldsymbol{y}' = \boldsymbol{f}(s, \boldsymbol{y}), \qquad s \in [s_1, s_2], \qquad \boldsymbol{f} : [s_1, s_2] \times \mathbb{R}^d \to \mathbb{R}^d.$$

By Theorem 1 of Iserles (2009), the Euler method is convergent if $\boldsymbol{f}$ is Lipchitz continuous with respect to $\boldsymbol{y}$, and $\boldsymbol{f}$ and $\boldsymbol{y}$ is continuously differentiable.

Recall that the system of ODE's we aim to solve is given by (18). Treat $\boldsymbol{g}^{(\cdot)}$ as $\boldsymbol{y}$, the function $\boldsymbol{f}$ corresponding to the ODE system (18), becomes $\boldsymbol{f}(s, \boldsymbol{y}) = \mathbf{M}(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}(s))\, \boldsymbol{y}$, which satisfies both the Lipschitz continuity and continuous differentiability conditions. Moreover, since $g(\cdot)$ is the integral of an exponential function, it is continuously differentiable, so is $\boldsymbol{g}^{(\cdot)}$. Consequently, the associated Euler algorithm is convergent.

This completes the proof. □

Proof of Proposition 6. In this proof, let us shorthand $\widetilde{\boldsymbol{\theta}}(0)$ by $\widetilde{\boldsymbol{\theta}}^0 = (\widetilde{\theta}^0_1, \ldots, \widetilde{\theta}^0_{d+1})$. Let us first consider the case of $\mathcal{H} \subset \mathcal{I}$. Similar to the proof of Theorem 4, without losing any generality, let us assume $\mathcal{H} = \{1, \ldots, p\}$, $1 \le p \le d$, and $\widetilde{\boldsymbol{\Theta}}$ satisfies the condition in (37). From Equation (41) in the proof of Theorem 4, we can obtain

$$g^{(\mathcal{H})}(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}^0) = \frac{(2\pi)^{-d/2}}{\det(\widetilde{\boldsymbol{\Theta}}_{\mathcal{H}\times\mathcal{H}})} \exp\Big(-\frac{1}{2}\sum_{i=1}^{p}\big(\sum_{j=1}^{p}[\mathbf{U}^{-1}]_{i,j}\,\widetilde{\theta}^0_j\big)^2\Big) \times \int_{\mathbb{R}^{d-p}} \exp\Big(-\frac{1}{2}\sum_{i=p+1}^{d} y_i^2\Big)$$
$$\prod_{i=p+1}^{d+1} \mathbb{1}\Big(-\sum_{j=1}^{p}\widetilde{\theta}_{i,j}\sum_{k=1}^{p}[\mathbf{U}^{-1}]_{j,k}\,\widetilde{\theta}^0_k + \sum_{j=p+1}^{d}\widetilde{\theta}_{i,j}\,y_j + \widetilde{\theta}^0_i \ge 0\Big)\,\mathrm{d}y_{p+1}\cdots\mathrm{d}y_d. \tag{45}$$

Recall that the definition of $\widetilde{\boldsymbol{\theta}}(s)$ is given in Equation (14). We note that the simplex region $\mathcal{S}_d(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}^0) = \{\boldsymbol{x} : \mathbf{B}_{d+1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{x} + \mathbf{B}_{d+1}\boldsymbol{\mu} \geq \mathbf{0}\}$ contains only one point $\boldsymbol{x}^* = -\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$. After applying the linear transform $\boldsymbol{y} = \mathbf{U}\boldsymbol{x}$, where $\mathbf{U}$ is defined as in (38), the simplex region becomes $\{\boldsymbol{y} : \mathbf{B}_{d+1}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{U}^{-1}\boldsymbol{y} + \mathbf{B}_{d+1}\boldsymbol{\mu} \geq \mathbf{0}\}$, which also contains only one point $\boldsymbol{y}^* = \mathbf{U}\boldsymbol{x}^*$. The integral region of (45) corresponding to $\{\boldsymbol{y} : \mathbf{B}_{d+1}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{U}^{-1}\boldsymbol{y} + \mathbf{B}_{d+1}\boldsymbol{\mu} \geq \mathbf{0},\ \boldsymbol{y}_{\mathcal{H}} = \boldsymbol{y}^*_{\mathcal{H}}\}$, also contains one point only. Consequently, $g^{(\mathcal{H})}(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}^0)$ in Equation (45) equals zero unless $|\mathcal{H}| = d$. In this specific case, the integral component in (45) vanishes, and the expression reduces to

$$\frac{(2\pi)^{-d/2}}{\det(\widetilde{\boldsymbol{\Theta}}_{\mathcal{H}\times\mathcal{H}})} \exp(-\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/2). \tag{46}$$

When it comes to the case of $\mathcal{H} = \mathcal{I}$, we have

$$g^{(\mathcal{I})}(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}^0) = \frac{\partial}{\partial \widetilde{\theta}^0_{d+1}} g^{(\mathcal{D})}(\widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\theta}}^0) = 0,$$

because the expression in Equation (46) is independent of any $\widetilde{\theta}^0_i$, $i \in \mathcal{I}$.

This completes the proof. □