



Conjugacy properties of multivariate unified skew-elliptical distributions

Maicon J. Karling^{a,*}, Daniele Durante^b, Marc G. Genton^a

^a Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

^b Department of Decision Sciences and Institute for Data Science and Analytics, Bocconi University, Milan, Italy

ARTICLE INFO

AMS 2020 subject classifications:

primary 62F15
secondary 62H10

Keywords:

Bayesian statistics
Multivariate unified skew-elliptical distribution
Unified skew-normal distribution
Unified skew-*t* distribution

ABSTRACT

The family of multivariate unified skew-normal (SUN) distributions has been recently shown to possess fundamental conjugacy properties. When used as priors for the vector of coefficients in probit, tobit, and multinomial probit models, these distributions yield posteriors that still belong to the SUN family. Although this result has led to important advancements in Bayesian inference and computation, its applicability beyond likelihoods associated with fully-observed, discretized, or censored realizations from multivariate Gaussian models remains yet unexplored. This article covers such a gap by proving that the wider family of multivariate unified skew-elliptical (SUE) distributions, which extends SUNs to more general perturbations of elliptical densities, guarantees conjugacy for broader classes of models, beyond those relying on fully-observed, discretized or censored Gaussians. Such a result leverages the closure under linear combinations, conditioning and marginalization of SUE to prove that this family is conjugate to the likelihood induced by regression models for fully-observed, censored or dichotomized realizations from skew-elliptical distributions. This key advancement enlarges the set of models that enable conjugate Bayesian inference to general formulations arising from elliptical and skew-elliptical families, including the multivariate Student's *t* and skew-*t*, among others.

1. Introduction

Conjugacy is a fundamental property in Bayesian statistics. A prior $p(\beta)$ for the vector of parameters $\beta \in \mathcal{B} \subset \mathbb{R}^p$ is conjugate to the likelihood $p(y | \beta)$ of the observed data $y \in \mathcal{Y} \subset \mathbb{R}^n$, if the induced posterior $p(\beta | y)$ still belongs to the same class of distributions of the assumed prior. This important property implies that when $p(\beta)$ belongs to a known and tractable family, Bayesian inference under the induced posterior distribution can also leverage the tractability of such a family, thereby circumventing the challenges that arise in Bayesian computation and inference under intractable posterior distributions. Despite the relevance of this property, identifying known and tractable conjugate priors for the likelihoods induced by commonly-used statistical models is often challenging. Remarkably, until recently, conjugacy in regression settings was mainly established for univariate or multivariate normal responses y with Gaussian priors for the coefficients β , thus hindering potentials of conjugate Bayesian inference beyond this specific setting.

To address the aforementioned gap, Durante [1] has recently shown that also general probit models admit conjugate priors, with these priors belonging to the known family of unified skew-normal (SUN) distributions [2]. Such a class includes multivariate Gaussians as a special case, and extends these symmetric distributions through the perturbation of the corresponding density via a factor that coincides with the cumulative distribution function of a multivariate normal, thereby inducing skewness. Crucially,

* Corresponding author.

E-mail address: maicon.karling@kaust.edu.sa (M.J. Karling).

<https://doi.org/10.1016/j.jmva.2024.105357>

Received 15 February 2024; Received in revised form 4 August 2024; Accepted 5 August 2024

Available online 6 August 2024

0047-259X/© 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

SUNs (i) have a known normalizing constant and moment-generating function, (ii) admit a tractable stochastic representation, and (iii) preserve the closure under linear combinations, conditioning and marginalization of the original multivariate Gaussians [2–4]. These properties facilitate Bayesian inference under the induced SUN posterior and, consequently, have motivated rapid subsequent research to establish SUN conjugacy for broader classes of models beyond classical probit representations. Relevant advancements along these lines include dynamic multivariate probit [5], multinomial probit [6], probit Gaussian processes [7], tobit models [8] and, more generally, any representation inducing likelihoods proportional to the kernel of a SUN [8]. Such a latter result crucially includes also important skewed extensions of classical probit, multinomial probit, and tobit [e.g., 9–18], along with earlier conjugacy results for the parameters of skew-normal distributions [19–21]; see also Fasano et al. [22] and Onorati and Liseo [23] for additional results in binary regression settings, and Durante et al. [24] for an extension of the Bernstein–von Mises theorem which clarifies the crucial role played by skewed extensions of multivariate Gaussians in Bayesian approximations and asymptotic theory.

Although the above contributions substantially enlarge the class of routinely-implemented models that admit conjugate priors, all these formulations are based on fully-observed, discretized or censored Gaussian or skew-normal representations. As discussed above, such a class includes multivariate linear regression along with probit, multinomial probit, and tobit models, among others, thus covering a core subset of formulations that are often employed within statistics. Nonetheless, in several applications, there is still interest in extensions of these representations which replace the Gaussian or skew-normal assumption for the error terms with alternative distributions. Popular examples in applications are generalizations of linear regression, probit, multinomial probit, and tobit models that rely on Student's t or skew- t error terms to incorporate robustness [e.g., 25–35]. More generally, several important contributions [e.g., 36–40] have also focused on multivariate elliptical [41] and skew-elliptical [42–47] distributions, which include multivariate Gaussians, skew-normals, Student's t and skew- t as special cases, thereby providing a large class of practically-relevant models. Despite the relevance of such a family, there is still a lack of general, unified and tractable solutions for Bayesian inference within these settings. This is arguably due to the fact that, to date, no general conjugacy results have been established for generic models arising from fully-observed, censored, or dichotomized realizations from elliptical and skew-elliptical distributions.

Motivated by the above discussion, we cover this gap by proving that multivariate unified skew-elliptical (SUE) distributions [2, 46] are the conjugate priors to the aforementioned class of models. From a technical perspective, the derivation of this result is based on specifying a general joint SUE distribution for the parameters β and the noise vector ϵ of the response \mathbf{y} , and then leveraging the closure under linear combination, marginalization and conditioning of SUE to prove that both $p(\beta)$ and $p(\beta | \mathbf{y})$ are SUE, whenever $p(\mathbf{y} | \beta)$ is proportional to a suitable likelihood induced by a fully-observed, censored or dichotomized elliptical or skew-elliptical distribution. This focus on the joint distribution serves only as a technical strategy to identify, under a classical Bayesian setting, which SUE priors are conjugate to specific likelihoods, thereby yielding SUE posterior distributions $p(\beta | \mathbf{y}) \propto p(\beta)p(\mathbf{y} | \beta)$ via the standard application of the Bayes rule. These novel results are obtained within Section 3, leveraging both available and newly-derived SUE properties outlined in Section 2. As discussed in Sections 3.1–3.3 (see Examples 1–6), these advancements include, as a special case, the conjugacy properties derived in Anceschi et al. [8] for SUNs, while extending these properties to other models of potential practical interest, such as, for example, generalizations of linear regression, probit and tobit models to Student's t or skew- t error terms. For these formulations, we show that the corresponding conjugate priors are multivariate unified skew- t (SUT) [46,48]. Concluding remarks can be found in Section 4, where we also clarify that besides the practical consequences for some special cases of the general results in Section 3, the conjugacy properties we derive are of broader and independent interest in expanding the theoretical analysis of the SUE family.

2. General overview and properties of multivariate unified skew-elliptical (SUE) distributions

Sections 2.1–2.2 provide an overview of the SUE family along with its special cases, whereas Section 2.3 comprises both available closure properties and newly-derived ones that are required to prove the novel conjugacy results within Section 3. To ease the presentation, in the following, we adopt a different notation between random variables and the associated realizations only when the distinction between the two is not clear from the context.

2.1. Multivariate unified skew-elliptical distributions

Multivariate unified skew-elliptical (SUE) distributions [e.g., 2,46] arise from the perturbation of elliptical densities [e.g., 41], defined as

$$f_m(\bar{\mathbf{z}} - \boldsymbol{\xi}; \boldsymbol{\Omega}, g^{(m)}) = |\boldsymbol{\Omega}|^{-1/2} g^{(m)}[(\bar{\mathbf{z}} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1} (\bar{\mathbf{z}} - \boldsymbol{\xi})], \quad \bar{\mathbf{z}} \in \mathbb{R}^m,$$

where $\boldsymbol{\xi} \in \mathbb{R}^m$ is a location parameter, $\boldsymbol{\Omega} \in \mathbb{R}^{m \times m}$ denotes a symmetric positive-definite dispersion matrix, and $g^{(m)}(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ characterizes the so-called density generator. Recalling Fang et al. [41], different choices of such a density generator $g^{(m)}$ lead to a broad class of routinely-implemented elliptical densities, covering multivariate Gaussians, Student's t , Cauchy, logistic and Laplace, among others. As such, for a generic vector $\bar{\mathbf{z}}$ from an elliptical distribution it is customary to adopt the general notation $\bar{\mathbf{z}} \sim EC_m(\boldsymbol{\xi}, \boldsymbol{\Omega}, g^{(m)})$, with $\boldsymbol{\xi}$, $\boldsymbol{\Omega}$ and $g^{(m)}$ parameterizing such a distribution. Refer to Chapters 1–3 of Fang et al. [41] for an in-depth treatment of the multivariate elliptical family, including details on the definition and properties of the density generator $g^{(m)}$.

Due to its generality, the multivariate elliptical family has been subject of a substantial interest that has led to the development of broader classes of distributions introducing skewness in the above representation. An important and comprehensive example in this direction is provided by the SUE family [2,46]. Leveraging a parameterization that agrees with the unified skew-normal (SUN) sub-family introduced by Arellano-Valle and Azzalini [2], and with the general selection representation in Arellano-Valle et al. [49]

(see also equation 19 in Arellano-Valle and Genton [46] and Section 7.1.3 of Azzalini and Capitanio [3]) a random vector $\mathbf{z} \in \mathbb{R}^m$ has a multivariate unified skew-elliptical (SUE) distribution, i.e., $\mathbf{z} \sim \text{SUE}_{m,q}(\xi, \mathbf{\Omega}, \mathbf{A}, \tau, \bar{\Gamma}, g^{(m+q)})$, if its density $\rho(\mathbf{z})$ is defined as

$$\rho(\mathbf{z}) = f_m(\mathbf{z} - \xi; \mathbf{\Omega}, g^{(m)}) \frac{F_q[\tau + \mathbf{A}^\top \bar{\mathbf{\Omega}}^{-1} \boldsymbol{\omega}^{-1}(\mathbf{z} - \xi); \bar{\Gamma} - \mathbf{A}^\top \bar{\mathbf{\Omega}}^{-1} \mathbf{A}, g_{Q(\mathbf{z})}^{(q)}]}{F_q(\tau; \bar{\Gamma}, g^{(q)}), \quad \mathbf{z} \in \mathbb{R}^m, \tag{1}$$

where $f_m(\mathbf{z} - \xi; \mathbf{\Omega}, g^{(m)})$ corresponds to the previously-defined elliptical density – evaluated at $\mathbf{z} - \xi$ – with density generator $g^{(m)}$, location $\xi \in \mathbb{R}^m$, and positive-definite dispersion matrix $\mathbf{\Omega} \in \mathbb{R}^{m \times m}$ with associated scales and correlations in $\boldsymbol{\omega} = \text{diag}(\mathbf{\Omega})^{1/2} \in \mathbb{R}^{m \times m}$ and $\bar{\mathbf{\Omega}} = \boldsymbol{\omega}^{-1} \mathbf{\Omega} \boldsymbol{\omega}^{-1} \in \mathbb{R}^{m \times m}$, respectively. In addition, $\tau \in \mathbb{R}^q$ is a truncation parameter, $\mathbf{A} \in \mathbb{R}^{m \times q}$ denotes a shape matrix, whereas $\bar{\Gamma} \in \mathbb{R}^{q \times q}$ corresponds to a positive-definite dispersion matrix. Finally, $Q(\mathbf{z})$ is a quadratic form defined as $Q(\mathbf{z}) = (\mathbf{z} - \xi)^\top \boldsymbol{\omega}^{-1}(\mathbf{z} - \xi) \in \mathbb{R}^+$, whereas $g_{Q(\mathbf{z})}^{(q)}(u) = g^{(m+q)}[u + Q(\mathbf{z})]/g^{(m)}[Q(\mathbf{z})]$ denotes the elliptical conditional density generator.

In (1), the quantity responsible for inducing skewness is the q -dimensional centered elliptical cumulative distribution function $F_q(\cdot; \bar{\Gamma} - \mathbf{A}^\top \bar{\mathbf{\Omega}}^{-1} \mathbf{A}, g_{Q(\mathbf{z})}^{(q)})$ with density generator $g_{Q(\mathbf{z})}^{(q)}$, and dispersion matrix defined in the second argument. As clarified in Lemma 5, the density in (1) includes as a special case the one of classical elliptical distributions, which can be obtained by setting $\tau = \mathbf{0}$ and $\mathbf{A} = \mathbf{0}$. Such a result highlights that τ and \mathbf{A} play a crucial role in inducing skewness. Let us also emphasize that, in this article, the notation $\bar{\Gamma}$ is used to denote the Pearson-correlation matrix defined as $\bar{\Gamma} = \boldsymbol{\gamma}^{-1} \boldsymbol{\Gamma} \boldsymbol{\gamma}^{-1}$, where $\boldsymbol{\gamma} = \text{diag}(\boldsymbol{\Gamma})^{1/2}$.

To further clarify the SUE construction, it shall be emphasized that the density expressed in (1) can be directly obtained from the selection representation

$$\mathbf{z} \stackrel{d}{=} (\bar{\mathbf{z}} \mid \bar{\mathbf{z}}_0 > \mathbf{0}), \quad \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{z}}_0 \end{bmatrix} \sim \text{EC}_{m+q} \left(\begin{bmatrix} \xi \\ \tau \end{bmatrix}, \begin{bmatrix} \mathbf{\Omega} & \boldsymbol{\omega} \mathbf{A} \\ \mathbf{A}^\top \boldsymbol{\omega} & \bar{\Gamma} \end{bmatrix}, g^{(m+q)} \right), \tag{2}$$

where $\bar{\mathbf{z}}_0 > \mathbf{0}$ indicates the event “each component of $\bar{\mathbf{z}}_0$ is positive”. More specifically, under the above representation and leveraging the closure under marginalization of elliptical distributions, a direct application of the Bayes rule yields

$$\rho(\mathbf{z}) = \rho(\bar{\mathbf{z}} = \mathbf{z}) \frac{\mathbb{P}(-\bar{\mathbf{z}}_0 \leq \mathbf{0} \mid \bar{\mathbf{z}} = \mathbf{z})}{\mathbb{P}(-\bar{\mathbf{z}}_0 \leq \mathbf{0})} = f_m(\mathbf{z} - \xi; \mathbf{\Omega}, g^{(m)}) \frac{\mathbb{P}(-\bar{\mathbf{z}}_0 \leq \mathbf{0} \mid \bar{\mathbf{z}} = \mathbf{z})}{\mathbb{P}(-\bar{\mathbf{z}}_0 \leq \mathbf{0})}, \quad \mathbf{z} \in \mathbb{R}^m. \tag{3}$$

Letting $\boldsymbol{\omega} \mathbf{\Omega}^{-1} = \bar{\mathbf{\Omega}}^{-1} \boldsymbol{\omega}^{-1}$ and $\boldsymbol{\omega} \mathbf{\Omega}^{-1} \boldsymbol{\omega} = \bar{\mathbf{\Omega}}^{-1}$, and recalling again the closure properties of elliptical distributions [3,41], it follows that $-\bar{\mathbf{z}}_0 \sim \text{EC}_q(-\tau, \bar{\Gamma}, g^{(q)})$ and $(-\bar{\mathbf{z}}_0 \mid \bar{\mathbf{z}} = \mathbf{z}) \sim \text{EC}_q(-\tau - \mathbf{A}^\top \boldsymbol{\omega} \mathbf{\Omega}^{-1}(\mathbf{z} - \xi), \bar{\Gamma} - \mathbf{A}^\top \boldsymbol{\omega} \mathbf{\Omega}^{-1} \boldsymbol{\omega} \mathbf{A}, g_{Q(\mathbf{z})}^{(q)})$; refer also to Arellano-Valle et al. [49]. As a consequence, the two probabilities at the denominator and numerator of (3) coincide with the cumulative distribution functions, evaluated at τ and $\tau + \mathbf{A}^\top \bar{\mathbf{\Omega}}^{-1} \boldsymbol{\omega}^{-1}(\mathbf{z} - \xi)$, of the centered elliptical distributions $\text{EC}_q(\mathbf{0}, \bar{\Gamma}, g^{(q)})$ and $\text{EC}_q(\mathbf{0}, \bar{\Gamma} - \mathbf{A}^\top \bar{\mathbf{\Omega}}^{-1} \mathbf{A}, g_{Q(\mathbf{z})}^{(q)})$, respectively, thereby allowing to recover (1).

Besides providing additional insights on the quantities defining the joint density $\rho(\mathbf{z})$ in (1), the selection representation in (2)–(3) is also useful to derive the cumulative distribution function $\mathcal{P}(\mathbf{z})$ of the SUE distribution having density as in (1). More specifically, since the SUE random vector $\mathbf{z} \sim \text{SUE}_{m,q}(\xi, \mathbf{\Omega}, \mathbf{A}, \tau, \bar{\Gamma}, g^{(m+q)})$ admits the equivalent selection representation in (2), it follows that

$$\mathcal{P}(\mathbf{z}) = \mathbb{P}(\bar{\mathbf{z}} \leq \mathbf{z}, -\bar{\mathbf{z}}_0 \leq \mathbf{0}) / \mathbb{P}(-\bar{\mathbf{z}}_0 \leq \mathbf{0}) = \mathbb{P}(\bar{\mathbf{z}} - \xi \leq \mathbf{z} - \xi, -\bar{\mathbf{z}}_0 + \tau \leq \tau) / F_q(\tau; \bar{\Gamma}, g^{(q)}).$$

Therefore, leveraging again the closure under linear combinations of elliptical distributions [3,41], it is possible to derive the following closed-form expression for the cumulative distribution function

$$\mathcal{P}(\mathbf{z}) = \frac{F_{m+q} \left(\begin{bmatrix} \mathbf{z} - \xi \\ \tau \end{bmatrix}; \begin{bmatrix} \mathbf{\Omega} & -\boldsymbol{\omega} \mathbf{A} \\ -\mathbf{A}^\top \boldsymbol{\omega} & \bar{\Gamma} \end{bmatrix}, g^{(m+q)} \right)}{F_q(\tau; \bar{\Gamma}, g^{(q)})}. \tag{4}$$

Notice that the vector $\bar{\mathbf{z}}_0$ is often called the latent part of the distribution, whereas q is the latent dimension.

Before discussing important SUE examples, we shall also emphasize that, as a result of the closure under linear combinations of elliptical and SUE distributions [41,46], an alternative to representation (2) is

$$\mathbf{z} \stackrel{d}{=} \xi + \boldsymbol{\omega} \mathbf{z}^*, \quad \mathbf{z}^* \stackrel{d}{=} (\bar{\mathbf{z}} \mid \bar{\mathbf{z}}_0 + \tau > \mathbf{0}), \quad \begin{bmatrix} \bar{\mathbf{z}} \\ \bar{\mathbf{z}}_0 \end{bmatrix} \sim \text{EC}_{m+q} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{\Omega}} & \mathbf{A} \\ \mathbf{A}^\top & \bar{\Gamma} \end{bmatrix}, g^{(m+q)} \right). \tag{5}$$

Such a representation is the one adopted by [2] for the sub-family of SUN distributions and is particularly convenient for deriving the mean vector and covariance matrix of \mathbf{z} . More specifically, leveraging (5), the law of total expectation, and the previously-discussed closure properties of elliptical distributions, we have that

$$\mathbb{E}(\mathbf{z}) = \xi + \boldsymbol{\omega} \mathbb{E}(\bar{\mathbf{z}} \mid \bar{\mathbf{z}}_0 + \tau > \mathbf{0}) = \xi + \boldsymbol{\omega} \mathbf{A} \bar{\Gamma}^{-1} \mathbb{E}(\bar{\mathbf{z}}_0 \mid \bar{\mathbf{z}}_0 + \tau > \mathbf{0}). \tag{6}$$

Recalling Arellano-Valle and Genton [46], a related reasoning yields the following covariance matrix

$$\text{var}(\mathbf{z}) = \boldsymbol{\psi} \mathbf{\Omega} + \boldsymbol{\omega} \mathbf{A} \left[\bar{\Gamma}^{-1} \text{var}(\bar{\mathbf{z}}_0 \mid \bar{\mathbf{z}}_0 + \tau > \mathbf{0}) \bar{\Gamma}^{-1} - \boldsymbol{\psi} \bar{\Gamma}^{-1} \right] \mathbf{A}^\top \boldsymbol{\omega}, \tag{7}$$

where $\boldsymbol{\psi}$ is a scalar whose form can be obtained from the derivations in Arellano-Valle and Genton [46]. The above expressions clarify that moments of SUE random vectors can be directly obtained from those of multivariate truncated elliptical distributions [e.g., 50–54]. Moreover, (7) shows that to enforce a lack of correlation among the entries in \mathbf{z} , it is not sufficient to impose suitable diagonal or block-diagonal structures within $\mathbf{\Omega}$. Rather, these constraints should be combined with additional ones, e.g., on the shape matrix \mathbf{A} . Such a result is useful in Section 3 to derive examples of practically-meaningful priors and likelihoods inducing SUE posterior

distributions. To this end, Sections 3.1–3.3 state general conjugacy properties and then specialize these results to the two most popular examples of SUE distributions that are presented in detail in Section 2.2 below, namely, multivariate unified skew-normals (SUN) [2] and multivariate unified skew- t (SUT) [48]. The results for SUN clarify that the conjugacy properties derived by Durante [1], Fasano and Durante [6] and Anceschi et al. [8] can be obtained as a special case, and under a different proof technique, of the more general SUE framework introduced in the present article. Conversely, the conjugacy results stated for SUT are a novel contribution that extends to a broader class of models of potential practical interest the findings in Song and Xia [55] and Zhang et al. [40] on specific Student’s t linear regressions and multivariate probit formulations based on skew-elliptical link functions.

2.2. Relevant sub-classes of multivariate unified skew-elliptical distributions

The SUN and SUT families arise from skewed perturbations of multivariate Gaussians and Student’s t densities, respectively. As such, these formulations are arguably the most relevant and practically-impactful sub-classes in the SUE family. In addition, recalling Arellano-Valle and Azzalini [2], Arellano-Valle and Genton [46], and Wang et al. [48], both SUN and SUT admit additive stochastic representations which allow for i.i.d. sampling under posterior distributions belonging to these two sub-classes, thus facilitating Bayesian inference; see also Yin and Balakrishnan [56] for a recent extension of these stochastic representations to more general multivariate skew-elliptical distributions, beyond SUN and SUT. Sections 2.2.1–2.2.2 provide a concise overview of SUN and SUT sub-classes, respectively. A more extensive treatment can be found in, e.g., [2,48].

2.2.1. Multivariate unified skew-normal distributions

The SUN family has been introduced by Arellano-Valle and Azzalini [2] to provide a single class of distributions capable of unifying several extensions of the original multivariate skew-normal [57]. Relevant examples of representations that belong to such a wide class are extended multivariate skew-normals [58,59] and closed skew-normals [60,61], among others. Recalling Arellano-Valle and Azzalini [2], a random vector $\mathbf{z} \in \mathbb{R}^m$ has SUN distribution, i.e., $\mathbf{z} \sim \text{SUN}_{m,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma})$, if its density is

$$p(\mathbf{z}) = \phi_m(\mathbf{z} - \xi; \Omega) \frac{\Phi_q[\tau + \Delta^T \bar{\Omega}^{-1} \omega^{-1}(\mathbf{z} - \xi); \bar{\Gamma} - \Delta^T \bar{\Omega}^{-1} \Delta]}{\Phi_q(\tau; \bar{\Gamma})}, \quad \mathbf{z} \in \mathbb{R}^m, \tag{8}$$

where $\xi, \Omega, \tau, \Delta$ and $\bar{\Gamma}$ have the same role and interpretation of the corresponding quantities within the general SUE representation in (1), whereas $\phi_m(\mathbf{z} - \xi; \Omega), \Phi_q[\tau + \Delta^T \bar{\Omega}^{-1} \omega^{-1}(\mathbf{z} - \xi); \bar{\Gamma} - \Delta^T \bar{\Omega}^{-1} \Delta]$ and $\Phi_q(\tau; \bar{\Gamma})$ denote the density and cumulative distribution functions, evaluated at $\mathbf{z} - \xi, \tau + \Delta^T \bar{\Omega}^{-1} \omega^{-1}(\mathbf{z} - \xi)$ and τ , respectively, of the centered multivariate Gaussians with covariance matrices $\Omega \in \mathbb{R}^{m \times m}, \bar{\Gamma} - \Delta^T \bar{\Omega}^{-1} \Delta \in \mathbb{R}^{q \times q}$ and $\bar{\Gamma} \in \mathbb{R}^{q \times q}$. Notice that, when $\Delta = \mathbf{0}$, the above density coincides with that of the multivariate Gaussian $N_m(\xi, \Omega)$, which can be therefore recovered as a special case, irrespectively of the value of τ and $\bar{\Gamma}$.

Although the above representation is originally derived in Arellano-Valle and Azzalini [2] under a selection representation similar to (5) applied to an underlying Gaussian, the density in (8) can also be derived directly from (1) under a suitable choice of the density generators. In particular, for $u \geq 0$, define $g^{(m)} = \phi^{(m)}(u) = (2\pi)^{-m/2} \exp(-u/2)$, $g^{(q)} = \phi^{(q)}(u) = (2\pi)^{-q/2} \exp(-u/2)$ and recall also that, in the particular Gaussian setting, the conditional generator $\phi_{Q(v)}^{(q)}$ coincides with the unconditional one $\phi^{(q)}$. Then, recalling related derivations in Arellano-Valle and Genton [46], and replacing these density generators in (1), directly yields expression (8), thus clarifying that SUNs are special cases of SUE distributions.

Recalling Section 1, the SUN family has been at the basis of recent advancements in conjugate Bayesian inference for routinely-implemented representations relying on fully-observed, discretized or partially-discretized Gaussian and multivariate skew-normal models [1,6,8]. As discussed in Section 3, these conjugacy properties can be obtained as a special case of those we derive for the SUE family, which, in turn, allow us to extend such results to larger classes, including the SUT one introduced in Section 2.2.2.

2.2.2. Multivariate unified skew- t distributions

The success of the SUN family has motivated several extensions aimed at deriving similar skewed representations for other sub-classes of the elliptical family. A noticeable and natural generalization is provided by the class of SUT distributions [e.g., 48] which can be obtained by replacing Gaussians with suitably-defined multivariate Student’s t in the original selection representation of SUNs. Recalling, e.g., Wang et al. [48], this yields a density for the SUT random vector $\mathbf{z} \sim \text{SUT}_{m,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$, defined as

$$p(\mathbf{z}) = t_m(\mathbf{z} - \xi; \Omega, \nu) \frac{T_q[\alpha_{\nu, Q(\mathbf{z})}^{-1/2} \{ \tau + \Delta^T \bar{\Omega}^{-1} \omega^{-1}(\mathbf{z} - \xi) \}; \bar{\Gamma} - \Delta^T \bar{\Omega}^{-1} \Delta, \nu + m]}{T_q(\tau; \bar{\Gamma}, \nu)}, \quad \mathbf{z} \in \mathbb{R}^m, \tag{9}$$

where $\alpha_{\nu, Q(\mathbf{z})} = [\nu + Q(\mathbf{z})]/(\nu + m)$, $Q(\mathbf{z}) = (\mathbf{z} - \xi)^T \Omega^{-1} (\mathbf{z} - \xi)$, and $\nu > 0$ are the degrees of freedom. The remaining parameters have, likewise, similar interpretations to those in (1). Analogously, $t_m(\mathbf{z} - \xi; \Omega, \nu), T_q[\alpha_{\nu, Q(\mathbf{z})}^{-1/2} \{ \tau + \Delta^T \bar{\Omega}^{-1} \omega^{-1}(\mathbf{z} - \xi) \}; \bar{\Gamma} - \Delta^T \bar{\Omega}^{-1} \Delta, \nu + m]$ and $T_q(\tau; \bar{\Gamma}, \nu)$ denote the density and cumulative distribution functions, evaluated at the corresponding first argument, respectively, of the centered multivariate Student’s t distributions with scale matrices $\Omega \in \mathbb{R}^{m \times m}, \bar{\Gamma} - \Delta^T \bar{\Omega}^{-1} \Delta \in \mathbb{R}^{q \times q}$ and $\bar{\Gamma} \in \mathbb{R}^{q \times q}$, and degrees of freedom $\nu, \nu + m$ and ν , respectively. Notice that, for $q = 1$, one retrieves the multivariate extended skew- t in [62]. Moreover, when $\tau = \mathbf{0}$ and $\Delta = \mathbf{0}$, the numerator and denominator in (9) coincide, and, therefore, the density reduces to that of a multivariate Student’s t distribution $\mathcal{T}_m(\xi, \Omega, \nu)$ with location ξ , scale Ω , and degrees of freedom ν . This implies that the multivariate Student’s t is obtained as a special case of SUT.

As for the SUN, also the SUT density within (9) can be derived from (1) under a suitable choice for the density generators. In particular, let $g^{(m)} = t_v^{(m)}(u) = c(\nu, m)[1 + u/\nu]^{-(\nu+m)/2}$, and $g^{(q)} = t_v^{(q)}(u) = c(\nu, q)[1 + u/\nu]^{-(\nu+q)/2}$, where the generic quantity $c(a, b)$ is

defined as $c(a, b) = \Gamma((a+b)/2)/[\Gamma(a/2)\Gamma(b/2)]$, while $\Gamma(\cdot)$ denotes the usual gamma function. Then, recalling, for example, Arellano-Valle and Genton [46], by replacing the generic density generators in (1) with those defined above and with the induced conditional density generator, yields, as a result of straightforward calculations, the SUT density in (9). Interestingly, as clarified in, e.g., Wang et al. [48], when $\nu \rightarrow \infty$, this density reduces to (8), thereby establishing a direct connection between SUT and SUN distributions. This suggests that the conjugacy properties derived in Durante [1], Fasano and Durante [6] and Anceschi et al. [8] for SUN might extend to SUT, and, more generally, to SUE distributions. Two promising results in this direction have been derived in Song and Xia [55] and in Zhang et al. [40], but only with a focus on Student's t linear regression and on specific multivariate binary data settings. Leveraging the SUE properties in Section 2.3, we prove in Section 3 that SUE conjugacy holds for a substantially larger class of models which further embraces formulations of potential interest in practice.

2.3. Properties of unified skew-elliptical distributions

Lemmas 1–5 below state several central properties of SUE distributions that are at the core of the novel conjugacy results derived in Section 3. More specifically, Lemmas 1–2 establish closure under linear combinations, marginalization and conditioning of SUE distributions. All these properties have appeared also in Arellano-Valle and Genton [46], but under a different parameterization. Conversely, Lemmas 3–5 state novel results that are useful to study SUE conjugacy under broad classes of models and to derive special cases of potential interest in practical contexts.

Lemma 1. Let $\mathbf{z} \sim \text{SUE}_{m,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, g^{(m+q)})$. In addition, denote with $\mathbf{A} \in \mathbb{R}^{r \times m}$ a matrix with rank $r \leq m$, and let $\mathbf{b} \in \mathbb{R}^r$ be a vector of constants. Then

$$\mathbf{Az} + \mathbf{b} \sim \text{SUE}_{r,q}(\mathbf{A}\xi + \mathbf{b}, \mathbf{A}\Omega\mathbf{A}^\top, \Delta_{\mathbf{A}}, \tau, \bar{\Gamma}, g^{(r+q)}),$$

where $\Delta_{\mathbf{A}} = \omega_{\mathbf{A}}^{-1}\mathbf{A}\omega_{\mathbf{A}}\Delta$ with $\omega_{\mathbf{A}} = \text{diag}(\mathbf{A}\Omega\mathbf{A}^\top)^{1/2}$. Moreover, let $\mathbf{z}_C \in \mathbb{R}^{|\mathcal{C}|}$ be a generic sub-vector comprising the entries in \mathbf{z} with indexes in $C \subset \{1, \dots, m\}$, and denote with $\xi_C \in \mathbb{R}^{|\mathcal{C}|}$, $\Omega_{CC} \in \mathbb{R}^{|\mathcal{C}| \times |\mathcal{C}|}$ and $\Delta_C \in \mathbb{R}^{|\mathcal{C}| \times q}$ the associated location sub-vector, dispersion sub-matrix and shape sub-matrix, respectively. Then

$$\mathbf{z}_C \sim \text{SUE}_{|\mathcal{C}|,q}(\xi_C, \Omega_{CC}, \Delta_C, \tau, \bar{\Gamma}, g^{(|\mathcal{C}|+q)}),$$

for any $C \subset \{1, \dots, m\}$.

Proof. The proof adapts related derivations in Arellano-Valle and Genton [46] to the parameterization considered in the present article. More specifically, by applying to (2) the linearity properties of elliptical distributions, it follows that

$$\begin{bmatrix} \mathbf{Az} \\ -\bar{\mathbf{z}}_0 \end{bmatrix} \sim \text{EC}_{r+q} \left(\begin{bmatrix} \mathbf{A}\xi \\ -\tau \end{bmatrix}, \begin{bmatrix} \mathbf{A}\Omega\mathbf{A}^\top & -\omega_{\mathbf{A}}\Delta_{\mathbf{A}} \\ -\Delta_{\mathbf{A}}^\top\omega_{\mathbf{A}} & \bar{\Gamma} \end{bmatrix}, g^{(r+q)} \right),$$

where $\omega_{\mathbf{A}} = \text{diag}(\mathbf{A}\Omega\mathbf{A}^\top)^{1/2}$ and $\Delta_{\mathbf{A}} = \omega_{\mathbf{A}}^{-1}\mathbf{A}\omega_{\mathbf{A}}\Delta$. Now, as a direct consequence of the selection representation in (2), we have that $\mathbf{Az} + \mathbf{b}$ is distributed as $(\mathbf{Az} + \mathbf{b} \mid \bar{\mathbf{z}}_0 > \mathbf{0})$. Therefore, leveraging the results and discussions in Section 2.1, the cumulative distribution function of $\mathbf{Az} + \mathbf{b}$ can be expressed as

$$\mathbb{P}(\mathbf{Az} + \mathbf{b} \leq \mathbf{z}^*) = \frac{\mathbb{P}(\mathbf{Az} \leq \mathbf{z}^* - \mathbf{b}, -\bar{\mathbf{z}}_0 \leq \mathbf{0})}{\mathbb{P}(-\bar{\mathbf{z}}_0 \leq \mathbf{0})} = \frac{F_{r+q} \left(\begin{bmatrix} \mathbf{z}^* - \mathbf{A}\xi - \mathbf{b} \\ \tau \end{bmatrix}; \begin{bmatrix} \mathbf{A}\Omega\mathbf{A}^\top & -\omega_{\mathbf{A}}\Delta_{\mathbf{A}} \\ -\Delta_{\mathbf{A}}^\top\omega_{\mathbf{A}} & \bar{\Gamma} \end{bmatrix}, g^{(r+q)} \right)}{F_q(\tau; \bar{\Gamma}, g^{(q)})},$$

for any $\mathbf{z}^* \in \mathbb{R}^r$, thus obtaining the cumulative distribution function of the $\text{SUE}_{r,q}(\mathbf{A}\xi + \mathbf{b}, \mathbf{A}\Omega\mathbf{A}^\top, \Delta_{\mathbf{A}}, \tau, \bar{\Gamma}, g^{(r+q)})$. The result for the marginals follows as a direct consequence by letting $\mathbf{b} = \mathbf{0}$ and setting \mathbf{A} equal to a suitably-defined binary selection matrix such that $\mathbf{Az} = \mathbf{z}_C$. □

Lemma 1 ensures that linear combinations and marginals of multivariate SUE distributions are still within the same class, and the associated parameters can be derived in closed form via tractable analytical calculations. Lemma 2 below clarifies that a related result holds for the conditional distributions.

Lemma 2. Let $\mathbf{z} = (\mathbf{z}_1^\top, \mathbf{z}_2^\top)^\top \sim \text{SUE}_{m,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, g^{(m+q)})$ with $\mathbf{z}_1 \in \mathbb{R}^{m_1}$, $\mathbf{z}_2 \in \mathbb{R}^{m_2}$, and parameters partitioned as

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \quad \omega = \begin{bmatrix} \omega_1 & \mathbf{0} \\ \mathbf{0} & \omega_2 \end{bmatrix}, \quad \bar{\Omega} = \begin{bmatrix} \bar{\Omega}_{11} & \bar{\Omega}_{12} \\ \bar{\Omega}_{21} & \bar{\Omega}_{22} \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix}, \tag{10}$$

where $m_1 + m_2 = m$, $\bar{\Omega}_{21} = \omega_2^{-1}\Omega_{21}\omega_1^{-1}$ and $\bar{\Omega}_{12} = \omega_1^{-1}\Omega_{12}\omega_2^{-1}$. Then, for $i, j \in \{1, 2\}$ and $j \neq i$, we have

$$(\mathbf{z}_i \mid \mathbf{z}_j) \sim \text{SUE}_{m_i,q}(\xi_{i|j}, \Omega_{i|j}, \Delta_{i|j}, \tau_{i|j}, \bar{\Gamma}_{i|j}, g_{Q_j(\mathbf{z}_j)}^{(m_i+q)}), \quad \mathbf{z}_j \in \mathbb{R}^{m_j}, \tag{11}$$

with parameters defined as

$$\begin{aligned} \xi_{i|j} &= \xi_i + \Omega_{ij}\Omega_{jj}^{-1}(\mathbf{z}_j - \xi_j), & \Omega_{i|j} &= \Omega_{ii} - \Omega_{ij}\Omega_{jj}^{-1}\Omega_{ji}, & \omega_{i|j} &= \text{diag}(\Omega_{i|j})^{1/2}, & \gamma_{i|j} &= \text{diag}(\bar{\Gamma} - \Delta_j^\top\bar{\Omega}_{jj}^{-1}\Delta_j)^{1/2}, \\ \Delta_{i|j} &= \omega_{i|j}^{-1}(\omega_i\Delta_i - \Omega_{ij}\Omega_{jj}^{-1}\omega_j\Delta_j)\gamma_{i|j}^{-1}, & \tau_{i|j} &= \gamma_{i|j}^{-1}[\tau + \Delta_j^\top\bar{\Omega}_{jj}^{-1}\omega_j^{-1}(\mathbf{z}_j - \xi_j)], & \bar{\Gamma}_{i|j} &= \gamma_{i|j}^{-1}(\bar{\Gamma} - \Delta_j^\top\bar{\Omega}_{jj}^{-1}\Delta_j)\gamma_{i|j}^{-1}, \end{aligned} \tag{12}$$

and conditional density generator $g_{Q_j(\mathbf{z}_j)}^{(m_i+q)}(u) = g^{(m+q)}[Q_j(\mathbf{z}_j) + u]/g^{(m_j)}[Q_j(\mathbf{z}_j)]$, with $Q_j(\mathbf{z}_j) = (\mathbf{z}_j - \xi_j)^\top\Omega_{jj}^{-1}(\mathbf{z}_j - \xi_j)$.

Proof. To prove Lemma 2, let us leverage again (2). To this end, consider the following elliptical distribution

$$\begin{bmatrix} \bar{\mathbf{z}}_1 \\ \bar{\mathbf{z}}_2 \\ \bar{\mathbf{z}}_0 \end{bmatrix} \sim \text{EC}_{m_1+m_2+q} \left(\begin{bmatrix} \xi_1 \\ \xi_2 \\ \tau \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} & \omega_1 \mathbf{A}_1 \\ \Omega_{21} & \Omega_{22} & \omega_2 \mathbf{A}_2 \\ \mathbf{A}_1^\top \omega_1 & \mathbf{A}_2^\top \omega_2 & \bar{\Gamma} \end{bmatrix}, g^{(m_1+m_2+q)} \right). \tag{13}$$

Then, by the closure under linear combinations and conditioning of elliptical distributions [e.g., 41], we have that

$$([\bar{\mathbf{z}}_i^\top, (-\gamma_{ij}^{-1} \bar{\mathbf{z}}_0)^\top]^\top | \bar{\mathbf{z}}_j) \sim \text{EC}_{m_j+q} \left(\begin{bmatrix} \xi_{ij} \\ -\tau_{ij} \end{bmatrix}, \begin{bmatrix} \Omega_{ij} & -\omega_{ij} \mathbf{A}_{ij} \\ -\mathbf{A}_{ij}^\top \omega_{ij} & \bar{\Gamma}_{ij} \end{bmatrix}, g_{Q_j(\bar{\mathbf{z}}_j)}^{(m_j+q)} \right), \quad Q_j(\bar{\mathbf{z}}_j) = (\bar{\mathbf{z}}_j - \xi_j)^\top \Omega_{jj}^{-1} (\bar{\mathbf{z}}_j - \xi_j),$$

which also implies $(-\gamma_{ij}^{-1} \bar{\mathbf{z}}_0 | \bar{\mathbf{z}}_j) \sim \text{EC}_q(-\tau_{ij}, \bar{\Gamma}_{ij}, g_{Q_j(\bar{\mathbf{z}}_j)}^{(q)})$, as a direct consequence of the closure under marginalization.

By combining these results with the selection representation in (2), and noticing that the event $-\bar{\mathbf{z}}_0 \leq \mathbf{0}$ is equivalent to $-\gamma_{ij}^{-1} \bar{\mathbf{z}}_0 \leq \mathbf{0}$ (since γ_{ij}^{-1} is diagonal with non-negative entries), it follows that

$$\mathcal{P}(\mathbf{z}_i | \mathbf{z}_j) = \frac{\mathbb{P}(\bar{\mathbf{z}}_i \leq \mathbf{z}_i, -\gamma_{ij}^{-1} \bar{\mathbf{z}}_0 \leq \mathbf{0} | \bar{\mathbf{z}}_j = \mathbf{z}_j)}{\mathbb{P}(-\gamma_{ij}^{-1} \bar{\mathbf{z}}_0 \leq \mathbf{0} | \bar{\mathbf{z}}_j = \mathbf{z}_j)} = \frac{F_{m_i+q} \left(\begin{bmatrix} \mathbf{z}_i - \xi_{ij} \\ \tau_{ij} \end{bmatrix}; \begin{bmatrix} \Omega_{ij} & -\omega_{ij} \mathbf{A}_{ij} \\ -\mathbf{A}_{ij}^\top \omega_{ij} & \bar{\Gamma}_{ij} \end{bmatrix}, g_{Q_j(\mathbf{z}_j)}^{(m_i+q)} \right)}{F_q(\tau_{ij}; \bar{\Gamma}_{ij}, g_{Q_j(\mathbf{z}_j)}^{(q)})}, \tag{14}$$

which coincides with the cumulative distribution function of the SUE in (11) having parameters as in (12); see (4) for the expression of the cumulative distribution function of a generic SUE $_{m,q}(\xi, \Omega, \mathbf{A}, \tau, \bar{\Gamma}, g^{(m+q)})$. Notice that the first equality in (14) follows from the fact that $p(\mathbf{z}_i | \mathbf{z}_j) = p(\mathbf{z}_i, \mathbf{z}_j) / p(\mathbf{z}_j)$, where $(\mathbf{z}_i^\top, \mathbf{z}_j^\top)^\top = \mathbf{z}$ is distributed as a SUE and, by Lemma 1, the same holds for \mathbf{z}_j . Hence, from the selection representation in (2), it follows that

$$p(\mathbf{z}_i | \mathbf{z}_j) = \frac{p(\bar{\mathbf{z}}_i = \mathbf{z}_i, \bar{\mathbf{z}}_j = \mathbf{z}_j)}{p(\bar{\mathbf{z}}_j = \mathbf{z}_j)} \frac{\mathbb{P}(\bar{\mathbf{z}}_0 \geq \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i, \bar{\mathbf{z}}_j = \mathbf{z}_j)}{\mathbb{P}(\bar{\mathbf{z}}_0 \geq \mathbf{0} | \bar{\mathbf{z}}_j = \mathbf{z}_j)} = p(\bar{\mathbf{z}}_i = \mathbf{z}_i | \bar{\mathbf{z}}_j = \mathbf{z}_j) \frac{\mathbb{P}(\bar{\mathbf{z}}_0 \geq \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i, \bar{\mathbf{z}}_j = \mathbf{z}_j)}{\mathbb{P}(\bar{\mathbf{z}}_0 \geq \mathbf{0} | \bar{\mathbf{z}}_j = \mathbf{z}_j)}, \tag{15}$$

and, hence, $\mathcal{P}(\mathbf{z}_i | \mathbf{z}_j) = \mathbb{P}(\bar{\mathbf{z}}_i \leq \mathbf{z}_i, -\gamma_{ij}^{-1} \bar{\mathbf{z}}_0 \leq \mathbf{0} | \bar{\mathbf{z}}_j = \mathbf{z}_j) / \mathbb{P}(-\gamma_{ij}^{-1} \bar{\mathbf{z}}_0 \leq \mathbf{0} | \bar{\mathbf{z}}_j = \mathbf{z}_j)$. \square

Lemma 2 guarantees that when $\mathbf{z} = (\mathbf{z}_i^\top, \mathbf{z}_j^\top)^\top$ is distributed as a SUE, then also the conditional distribution for a generic sub-vector \mathbf{z}_i belongs to the same family. Such a result conditions on a given realization \mathbf{z}_j for the remaining entries in \mathbf{z} . In this respect, Lemma 3 states a novel finding which clarifies that the properties in Lemma 2 can be preserved also when conditioning on a truncation event. Similar results can be found in [4,48] with a focus on SUN and SUT sub-classes. Lemma 3 states this property for the SUE family.

Lemma 3. Let $\mathbf{z} = (\mathbf{z}_1^\top, \mathbf{z}_2^\top)^\top \sim \text{SUE}_{m,q}(\xi, \Omega, \mathbf{A}, \tau, \bar{\Gamma}, g^{(m+q)})$ with parameters partitioned as in (10). Then

$$(\mathbf{z}_i | \mathbf{z}_j > \mathbf{0}) \sim \text{SUE}_{m_i, m_j+q}(\xi_i, \Omega_{ii}, \bar{\mathbf{A}}_{ij}, \bar{\tau}_{ij}, \bar{\Gamma}_{ij}, g^{(m+q)}), \tag{16}$$

where the quantities $\bar{\mathbf{A}}_{ij}$, $\bar{\tau}_{ij}$, and $\bar{\Gamma}_{ij}$ are defined as

$$\bar{\mathbf{A}}_{ij} = [\bar{\Omega}_{ij} \quad \mathbf{A}_i], \quad \bar{\tau}_{ij} = \begin{bmatrix} \omega_j^{-1} \xi_j \\ \tau \end{bmatrix}, \quad \bar{\Gamma}_{ij} = \begin{bmatrix} \bar{\Omega}_{jj} & \mathbf{A}_j \\ \mathbf{A}_j^\top & \bar{\Gamma} \end{bmatrix}, \tag{17}$$

for every $i, j \in \{1, 2\}$, with $j \neq i$ and $\bar{\Omega}_{ij} = \omega_i^{-1} \Omega_{ij} \omega_j^{-1}$.

Proof. Consider the selection representation in (2) based on the underlying elliptical distribution in (13). Leveraging derivations and arguments similar to those considered in the proof of Lemma 2, we have that

$$\mathbb{P}(\mathbf{z}_j > \mathbf{0} | \mathbf{z}_i) = \frac{\mathbb{P}(\bar{\mathbf{z}}_j > \mathbf{0}, \bar{\mathbf{z}}_0 > \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i)}{\mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i)}, \quad \mathbb{P}(\mathbf{z}_j > \mathbf{0}) = \frac{\mathbb{P}(\bar{\mathbf{z}}_j > \mathbf{0}, \bar{\mathbf{z}}_0 > \mathbf{0})}{\mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0})} = \frac{\mathbb{P}(\omega_j^{-1} \bar{\mathbf{z}}_j > \mathbf{0}, \bar{\mathbf{z}}_0 > \mathbf{0})}{\mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0})}. \tag{18}$$

Moreover, recall that by representation (2) and the closure properties of SUE, the marginal density for \mathbf{z}_i is defined as $p(\mathbf{z}_i) = p(\bar{\mathbf{z}}_i = \mathbf{z}_i) \mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i) / \mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0})$. Combining such an expression with those in (18) leads to

$$\begin{aligned} p(\mathbf{z}_i | \mathbf{z}_j > \mathbf{0}) &= p(\mathbf{z}_i) \frac{\mathbb{P}(\mathbf{z}_j > \mathbf{0} | \mathbf{z}_i)}{\mathbb{P}(\mathbf{z}_j > \mathbf{0})} = p(\bar{\mathbf{z}}_i = \mathbf{z}_i) \frac{\mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i)}{\mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0})} \frac{\mathbb{P}(\bar{\mathbf{z}}_j > \mathbf{0}, \bar{\mathbf{z}}_0 > \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i)}{\mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i)} \frac{\mathbb{P}(\bar{\mathbf{z}}_0 > \mathbf{0})}{\mathbb{P}(\omega_j^{-1} \bar{\mathbf{z}}_j > \mathbf{0}, \bar{\mathbf{z}}_0 > \mathbf{0})} \\ &= p(\bar{\mathbf{z}}_i = \mathbf{z}_i) \frac{\mathbb{P}(\bar{\mathbf{z}}_j > \mathbf{0}, \bar{\mathbf{z}}_0 > \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i)}{\mathbb{P}(\omega_j^{-1} \bar{\mathbf{z}}_j > \mathbf{0}, \bar{\mathbf{z}}_0 > \mathbf{0})} = p(\bar{\mathbf{z}}_i = \mathbf{z}_i) \frac{\mathbb{P}(-\omega_j^{-1} \bar{\mathbf{z}}_j \leq \mathbf{0}, -\bar{\mathbf{z}}_0 \leq \mathbf{0} | \bar{\mathbf{z}}_i = \mathbf{z}_i)}{\mathbb{P}(-\omega_j^{-1} \bar{\mathbf{z}}_j \leq \mathbf{0}, -\bar{\mathbf{z}}_0 \leq \mathbf{0})}. \end{aligned} \tag{19}$$

Leveraging the closure under linear combinations, marginalization and conditioning of elliptical distributions [e.g., 41], we have

$$\begin{aligned} [(-\omega_j^{-1} \bar{\mathbf{z}}_j)^\top, -\bar{\mathbf{z}}_0^\top]^\top &\sim \text{EC}_{m_j+q}(-\bar{\tau}_{ij}, \bar{\Gamma}_{ij}, g^{(m_j+q)}), \\ [(-\omega_j^{-1} \bar{\mathbf{z}}_j)^\top, -\bar{\mathbf{z}}_0^\top]^\top | \bar{\mathbf{z}}_i = \mathbf{z}_i &\sim \text{EC}_{m_j+q}(-\bar{\tau}_{ij}, \bar{\mathbf{A}}_{ij}^\top \bar{\Omega}_{ii}^{-1} \omega_i^{-1} (\mathbf{z}_i - \xi_i), \bar{\Gamma}_{ij} - \bar{\mathbf{A}}_{ij}^\top \bar{\Omega}_{ii}^{-1} \bar{\mathbf{A}}_{ij}, g_{Q_i(\mathbf{z}_i)}^{(m_j+q)}), \end{aligned}$$

where $Q_i(\mathbf{z}_i) = (\mathbf{z}_i - \xi_i)^\top \Omega_{ii}^{-1} (\mathbf{z}_i - \xi_i)$.

Hence, combining the above results with expression (19), and recalling that $\rho(\bar{\mathbf{z}}_i = \mathbf{z}_i)$ coincides with the density of the elliptical distribution $EC_{m_i}(\xi_i, \Omega_{ii}, g^{(m_i)})$, we obtain

$$\rho(\mathbf{z}_i | \mathbf{z}_j > \mathbf{0}) = f_{m_i}(\mathbf{z}_i - \xi_i; \Omega_{ii}, g^{(m_i)}) \frac{F_{m_j+q}(\bar{\tau}_{ij} + \bar{\Delta}_{ij}^\top \bar{\Omega}_{ii}^{-1} \omega_i^{-1}(\mathbf{z}_i - \xi_i); \bar{\Gamma}_{ij} - \bar{\Delta}_{ij}^\top \bar{\Omega}_{ii}^{-1} \bar{\Delta}_{ij}, g_{Q_i(\mathbf{z}_i)}^{(m_j+q)})}{F_{m_j+q}(\bar{\tau}_{ij}; \bar{\Gamma}_{ij}, g^{(m_j+q)})},$$

which coincides with the density of the SUE in Lemma 3. \square

Remark 1. Under a similar argument and derivations, it easily follows that also $(\mathbf{z}_i | \mathbf{z}_j < \mathbf{0})$ is a SUE distribution.

Lemma 4 below is useful for converting a SUE distribution parameterized by a matrix Γ that is not in the form of a Pearson correlation matrix to a standard SUE meeting such a constraint.

Lemma 4. Let Γ be a positive-definite matrix, then $SUE_{m,q}(\xi, \Omega, \Delta, \tau, \Gamma, g^{(m+q)}) \stackrel{d}{=} SUE_{m,q}(\xi, \Omega, \Delta\gamma^{-1}, \gamma^{-1}\tau, \bar{\Gamma}, g^{(m+q)})$, where $\bar{\Gamma}$ is a Pearson correlation matrix defined as $\bar{\Gamma} = \gamma^{-1}\Gamma\gamma^{-1}$, with $\gamma = \text{diag}(\Gamma)^{1/2}$.

Proof. Let $\mathbf{z} \sim SUE_{m,q}(\xi, \Omega, \Delta, \tau, \Gamma, g^{(m+q)})$. Then, according to the selection representation in (2), $\mathbf{z} = \bar{\mathbf{z}} | \bar{\mathbf{z}}_0 > \mathbf{0}$, and its density function is given by (3). Notice that, since γ is a diagonal matrix with non-negative entries, then the numerator and denominator in (3) can be alternatively re-written as $\mathbb{P}(-\bar{\mathbf{z}}_0 \leq \mathbf{0} | \bar{\mathbf{z}} = \mathbf{z}) = \mathbb{P}(-\gamma^{-1}\bar{\mathbf{z}}_0 \leq \mathbf{0} | \bar{\mathbf{z}} = \mathbf{z})$ and $\mathbb{P}(-\bar{\mathbf{z}}_0 \leq \mathbf{0}) = \mathbb{P}(-\gamma^{-1}\bar{\mathbf{z}}_0 \leq \mathbf{0})$. Therefore, the proof follows directly from (2)–(3) and by the closure under linear combinations and conditioning of elliptical distributions. \square

Lemma 5 concludes this section by presenting particular cases of SUE distributions obtained under specific constraints on the associated parameters. These results are useful for detecting redundant latent dimensions and identifying interesting examples of constrained representations yielding specific models of interest under the conjugacy results derived in Section 3.

Lemma 5. Let $\mathbf{z} \sim SUE_{m,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, g^{(m+q)})$ with parameters Δ, τ , and $\bar{\Gamma}$ partitioned as

$$\Delta = [\Delta_1 \quad \Delta_2], \quad \tau = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix}, \quad \bar{\Gamma} = \begin{bmatrix} \bar{\Gamma}_{11} & \bar{\Gamma}_{12} \\ \bar{\Gamma}_{21} & \bar{\Gamma}_{22} \end{bmatrix},$$

where $\Delta_i \in \mathbb{R}^{m \times q_i}$, $\tau_i \in \mathbb{R}^{q_i}$, and $\bar{\Gamma}_{ij} \in \mathbb{R}^{q_i \times q_j}$, for every $i, j \in \{1, 2\}$, such that $q_1 + q_2 = q$. Then, (i) if $\Delta = \mathbf{0}$ and $\tau = \mathbf{0}$, it follows that $\mathbf{z} \sim EC_m(\xi, \Omega, g^{(m)})$. Additionally, (ii) if $\Delta_i = \mathbf{0}$, $\tau_i = \mathbf{0}$, $\bar{\Gamma}_{ij} = \mathbf{0}$, for i and j fixed, with $j \neq i$, then $\mathbf{z} \sim SUE_{m,q_j}(\xi, \Omega, \Delta_j, \tau_j, \bar{\Gamma}_{jj}, g^{(m+q_j)})$. Finally, (iii) $F_{m+q}(\{[(\mathbf{z} - \xi)^\top, \mathbf{0}^\top]^\top; \text{diag}(\Omega, \bar{\Gamma}), g^{(m+q)}\}) = F_m(\mathbf{z} - \xi; \Omega, g^{(m)}) \cdot F_q(\mathbf{0}; \bar{\Gamma}, g^{(q)})$, where $\text{diag}(\Omega, \bar{\Gamma})$ denotes a block-diagonal matrix with blocks Ω and $\bar{\Gamma}$, respectively.

Proof. To prove (i) note that, due to the invariance of orthant probabilities under centered elliptical distributions [e.g., 41], we have $F_q(\mathbf{0}; \bar{\Gamma}, g_{Q(\mathbf{z})}^{(q)}) = F_q(\mathbf{0}; \bar{\Gamma}, g^{(q)})$. Hence, including the constraints $\Delta = \mathbf{0}$ and $\tau = \mathbf{0}$ in (1), yields

$$\rho(\mathbf{z}) = f_m(\mathbf{z} - \xi; \Omega, g^{(m)}) \frac{F_q(\mathbf{0}; \bar{\Gamma}, g_{Q(\mathbf{z})}^{(q)})}{F_q(\mathbf{0}; \bar{\Gamma}, g^{(q)})} = f_m(\mathbf{z} - \xi; \Omega, g^{(m)}),$$

which coincides with the density of the elliptical distribution $EC_m(\xi, \Omega, g^{(m)})$. This result allows us to prove also (iii). In particular, since $\mathbf{z} \sim EC_m(\xi, \Omega, g^{(m)})$ when both $\Delta = \mathbf{0}$ and $\tau = \mathbf{0}$, then $\mathcal{P}(\mathbf{z}) = F_m(\mathbf{z} - \xi; \Omega, g^{(m)})$. Conversely, by including the constraints $\Delta = \mathbf{0}$ and $\tau = \mathbf{0}$ within the general expression for the SUE cumulative distribution function in (4) yields $\mathcal{P}(\mathbf{z}) = F_{m+q}(\{[(\mathbf{z} - \xi)^\top, \mathbf{0}^\top]^\top; \text{diag}(\Omega, \bar{\Gamma}), g^{(m+q)}\}) / F_q(\mathbf{0}; \bar{\Gamma}, g^{(q)})$. Therefore $\mathcal{P}(\mathbf{z}) = F_m(\mathbf{z} - \xi; \Omega, g^{(m)}) = F_{m+q}(\{[(\mathbf{z} - \xi)^\top, \mathbf{0}^\top]^\top; \text{diag}(\Omega, \bar{\Gamma}), g^{(m+q)}\}) / F_q(\mathbf{0}; \bar{\Gamma}, g^{(q)})$ which implies the result stated in point (iii).

Finally, to prove (ii), assume for the sake of simplicity that $\Delta_2 = \mathbf{0}$, $\tau_2 = \mathbf{0}$, and $\bar{\Gamma}_{21} = \bar{\Gamma}_{12}^\top = \mathbf{0}$, i.e., $i = 2$ and $j = 1$ (the proof for $i = 1$ and $j = 2$ is analogous). Then, applying (iii) to both the numerator and the denominator of the expression for $\mathcal{P}(\mathbf{z})$ in (4), evaluated under the constrained parameters, leads to

$$\mathcal{P}(\mathbf{z}) = \frac{F_{m+q_1+q_2} \left(\begin{bmatrix} \mathbf{z} - \xi \\ \tau_1 \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} \Omega & -\omega\Delta_1 & \mathbf{0} \\ -\Delta_1^\top\omega & \bar{\Gamma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \bar{\Gamma}_{22} \end{bmatrix}, g^{(m+q_1+q_2)} \right)}{F_{q_1+q_2} \left(\begin{bmatrix} \tau_1 \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} \bar{\Gamma}_{11} & \mathbf{0} \\ \mathbf{0} & \bar{\Gamma}_{22} \end{bmatrix}, g^{(q_1+q_2)} \right)} = \frac{F_{m+q_1} \left(\begin{bmatrix} \mathbf{z} - \xi \\ \tau_1 \end{bmatrix}; \begin{bmatrix} \Omega & -\omega\Delta_1 \\ -\Delta_1^\top\omega & \bar{\Gamma}_{11} \end{bmatrix}, g^{(m+q_1)} \right)}{F_{q_1}(\tau_1; \bar{\Gamma}_{11}, g^{(q_1)})},$$

where the last equality follows from the fact that $F_{q_2}(\mathbf{0}; \bar{\Gamma}_{22}, g^{(q_2)})$ at the numerator and the denominator simplifies. To conclude the proof, it suffices to notice that the above cumulative distribution function is the one of a $SUE_{m,q_1}(\xi, \Omega, \Delta_1, \tau_1, \bar{\Gamma}_{11}, g^{(m+q_1)})$. \square

Remark 2. In the particular case of a SUN distribution, it can be shown that the restrictions $\tau = \mathbf{0}$ or $\tau_i = \mathbf{0}$ are not necessary in Lemma 5 since there are results analogous to (i)–(iii) that follow directly by the specific properties of Gaussian cumulative distribution functions. In particular, adapting the above proof to the SUN sub-family, it can be easily shown, for example, that if $\Delta_i = \mathbf{0}$ and $\bar{\Gamma}_{ij} = \bar{\Gamma}_{ji}^\top = \mathbf{0}$, $j \neq i$, then $\mathbf{z} \sim SUN_{m,q_j}(\xi, \Omega, \Delta_j, \tau_j, \bar{\Gamma}_{jj})$, even if $\tau_i \neq \mathbf{0}$.

Leveraging Lemmas 1–5, Section 3 derives the novel conjugacy properties of SUE distributions.

3. Conjugacy properties of multivariate unified skew-elliptical (SUE) distributions

Sections 3.1–3.3 present the new results on the SUE conjugacy properties under a broad class of regression models for fully-observed, censored or dichotomized realizations from elliptical or skew-elliptical variables. As anticipated in Section 1, the technical derivation of these results is based on specifying a general joint SUE distribution for the parameters in β and the noise vector ϵ underlying the response y . This allows to leverage the closure properties in Section 2.3 to obtain closed-form SUE priors $\rho(\beta)$ and meaningful likelihoods $\rho(y | \beta)$ whose combination, under the standard Bayes rule, yields posterior distributions $\rho(\beta | y) \propto \rho(\beta)\rho(y | \beta)$ that still belong to the SUE class.

The above technical focus on the joint distribution $\rho(\beta, y)$ for β and y is motivated by the fact that the results in Sections 3.1–3.3 clarify that not all the models arising from elliptical or skew-elliptical noise vectors admit conjugate SUE priors. For this property to hold generally within the SUE family, it is necessary to consider a form of dependence between β and ϵ due to the specific properties of the density generator. Notice that such a dependence is often weak. In particular, it allows to account for meaningful priors and models having β and ϵ uncorrelated, while reducing to full independence under the density generators of the multivariate Gaussians and unified skew-normals. Nonetheless, such a weak dependence combined with a technical focus on $\rho(\beta, y)$ allows for a more comprehensive investigation of SUE conjugacy properties that would not be as immediate to prove theoretically under a direct specification of the prior $\rho(\beta)$ and the likelihood $\rho(y | \beta)$.

As a simple example that clarifies the above arguments, consider a univariate setting with Cauchy(0, 1) prior for β and Cauchy(β , 1) likelihood for $(y | \beta)$. By application of the Bayes rule, we obtain $\rho(\beta | y) \propto \rho(\beta)\rho(y | \beta)$ where $\rho(\beta)\rho(y | \beta) = 1/[\pi^2(1+\beta^2)(1+(y-\beta)^2)]$ is not proportional to the kernel of a Cauchy density. Clearly, in this illustrative example and in general situations where conjugacy lacks, Bayesian inference can still proceed via routinely-implemented MCMC methods or deterministic approximations of the target posterior. Nonetheless, as clarified in Sections 3.1–3.3, SUE conjugacy can be still achieved under certain likelihoods induced by elliptical or skew-elliptical error terms (including instances of potential practical interest), thereby facilitating posterior inference.

3.1. Conjugacy properties of SUE distributions in multivariate linear models

Let us first study the SUE conjugacy properties under general multivariate linear models of the form

$$y = X\beta + \epsilon, \tag{20}$$

where $y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n \times p}$ corresponds to a known design matrix, $\beta \in \mathbb{R}^p$ is a vector of unknown parameters, often referred to as the regression coefficients, and $\epsilon \in \mathbb{R}^n$ is the error vector. Current results in Bayesian inference under the above model have established conjugacy of Gaussian or SUN priors for β when combined with Gaussian or SUN noise vectors ϵ [8]. Although these advancements cover a broad range of models, in practice, it is of interest to consider alternative representations in the wider elliptical or skew-elliptical family, which account for heavier tails and ensure increased robustness. However, conjugacy remains unexplored in these larger classes, undermining advancements in tractable Bayesian inference. Proposition 1 below covers such a gap.

Proposition 1. Assume that $(\beta^T, \epsilon^T)^T \sim \text{SUE}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, g^{(p+n+q)})$ with parameters partitioned as

$$\xi = \begin{bmatrix} \xi_\beta \\ \xi_\epsilon \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_\beta & \Omega_{\beta\epsilon} \\ \Omega_{\epsilon\beta} & \Omega_\epsilon \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_\beta \\ \Delta_\epsilon \end{bmatrix}. \tag{21}$$

Then, when y is defined as in (20), it follows that $(\beta^T, y^T)^T \sim \text{SUE}_{p+n,q}(\xi^\dagger, \Omega^\dagger, \Delta^\dagger, \tau, \bar{\Gamma}, g^{(p+n+q)})$, with

$$\xi^\dagger = \begin{bmatrix} \xi_\beta \\ X\xi_\beta + \xi_\epsilon \end{bmatrix} =: \begin{bmatrix} \xi_\beta \\ \xi_y \end{bmatrix}, \quad \Delta^\dagger = \begin{bmatrix} \Delta_\beta \\ \omega_y^{-1}(X\omega_\beta\Delta_\beta + \omega_\epsilon\Delta_\epsilon) \end{bmatrix} =: \begin{bmatrix} \Delta_\beta \\ \Delta_y \end{bmatrix}, \tag{22}$$

$$\Omega^\dagger = \begin{bmatrix} \Omega_\beta & \Omega_\beta X^T + \Omega_{\beta\epsilon} \\ X\Omega_\beta + \Omega_{\epsilon\beta} & X\Omega_\beta X^T + \Omega_{\epsilon\beta} X^T + X\Omega_{\beta\epsilon} + \Omega_\epsilon \end{bmatrix} =: \begin{bmatrix} \Omega_\beta & \Omega_{\beta y} \\ \Omega_{y\beta} & \Omega_y \end{bmatrix},$$

where $\omega_\epsilon = \text{diag}(\Omega_\epsilon)^{1/2}$, $\omega_\beta = \text{diag}(\Omega_\beta)^{1/2}$, and $\omega_y = \text{diag}(\Omega_y)^{1/2}$. Moreover

- (a) Prior distribution: $\beta \sim \text{SUE}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau, \bar{\Gamma}, g^{(p+q)})$.
- (b) Likelihood: $(y | \beta) \sim \text{SUE}_{n,q}(\xi_{y|\beta}, \Omega_{y|\beta}, \Delta_{y|\beta}, \tau_{y|\beta}, \bar{\Gamma}_{y|\beta}, g_{Q_\beta(\beta)}^{(n+q)})$, with parameters

$$\xi_{y|\beta} = \xi_y + \Omega_{y\beta}\Omega_\beta^{-1}(\beta - \xi_\beta), \quad \Omega_{y|\beta} = \Omega_y - \Omega_{y\beta}\Omega_\beta^{-1}\Omega_{\beta y}, \quad \Delta_{y|\beta} = \omega_{y|\beta}^{-1}(\omega_y\Delta_y - \Omega_{y\beta}\Omega_\beta^{-1}\omega_\beta\Delta_\beta)\gamma_{y|\beta}^{-1},$$

$$\tau_{y|\beta} = \gamma_{y|\beta}^{-1}[\tau + \Delta_\beta^T\bar{\Omega}_\beta^{-1}\omega_\beta^{-1}(\beta - \xi_\beta)], \quad \bar{\Gamma}_{y|\beta} = \gamma_{y|\beta}^{-1}(\bar{\Gamma} - \Delta_\beta^T\bar{\Omega}_\beta^{-1}\Delta_\beta)\gamma_{y|\beta}^{-1}, \quad Q_\beta(\beta) = (\beta - \xi_\beta)^T\Omega_\beta^{-1}(\beta - \xi_\beta),$$

where $\omega_{y|\beta} = \text{diag}(\Omega_{y|\beta})^{1/2}$ and $\gamma_{y|\beta} = \text{diag}(\bar{\Gamma} - \Delta_\beta^T\bar{\Omega}_\beta^{-1}\Delta_\beta)^{1/2}$.

- (c) Posterior distribution: $(\beta | y) \sim \text{SUE}_{p,q}(\xi_{\beta|y}, \Omega_{\beta|y}, \Delta_{\beta|y}, \tau_{\beta|y}, \bar{\Gamma}_{\beta|y}, g_{Q_y(y)}^{(p+q)})$, with parameters

$$\xi_{\beta|y} = \xi_\beta + \Omega_{\beta y}\Omega_y^{-1}(y - \xi_y), \quad \Omega_{\beta|y} = \Omega_\beta - \Omega_{\beta y}\Omega_y^{-1}\Omega_{y\beta}, \quad \Delta_{\beta|y} = \omega_{\beta|y}^{-1}(\omega_\beta\Delta_\beta - \Omega_{\beta y}\Omega_y^{-1}\omega_y\Delta_y)\gamma_{\beta|y}^{-1},$$

$$\tau_{\beta|y} = \gamma_{\beta|y}^{-1}[\tau + \Delta_y^T\bar{\Omega}_y^{-1}\omega_y^{-1}(y - \xi_y)], \quad \bar{\Gamma}_{\beta|y} = \gamma_{\beta|y}^{-1}(\bar{\Gamma} - \Delta_y^T\bar{\Omega}_y^{-1}\Delta_y)\gamma_{\beta|y}^{-1}, \quad Q_y(y) = (y - \xi_y)^T\Omega_y^{-1}(y - \xi_y),$$

where $\omega_{\beta|y} = \text{diag}(\Omega_{\beta|y})^{1/2}$ and $\gamma_{\beta|y} = \text{diag}(\bar{\Gamma} - \Delta_y^T\bar{\Omega}_y^{-1}\Delta_y)^{1/2}$.

Remark 3. Before proving Proposition 1, it shall be emphasized that the above results, along with those provided in Propositions 2 and 3, are purposely stated in a highly-general form in order to derive a comprehensive conjugacy theory for SUE distributions that is of broader and independent interest in expanding the theoretical analysis of such a family. As clarified in Examples 1–6, priors and likelihoods of potential interest in practice are only a subset of the general results in Propositions 1–3. More specifically, setting $\Omega_{\epsilon\beta} = \Omega_{\beta\epsilon}^\top = \mathbf{0}$, $\tau = \mathbf{0}$, $\xi_\epsilon = \mathbf{0}$, and either $\Delta_\beta = \mathbf{0}$ or $\Delta_\epsilon = \mathbf{0}$, would be sufficient to recover most of the priors and likelihoods of direct interest in applications.

Proof. To prove (22) in Proposition 1, first notice that $(\beta^\top, \mathbf{y}^\top)^\top = \mathbf{A}(\beta^\top, \epsilon^\top)^\top$, where \mathbf{A} is a known matrix of dimension $(p+n) \times (p+n)$ with blocks $\mathbf{A}_{11} = \mathbf{I}_p$, $\mathbf{A}_{12} = \mathbf{0}$, $\mathbf{A}_{21} = \mathbf{X}$ and $\mathbf{A}_{22} = \mathbf{I}_n$. Combining such a representation with the closure under linear combination properties of SUE distributions presented in Lemma 1, we have that $(\beta^\top, \mathbf{y}^\top)^\top \sim \text{SUE}_{p+n,q}(\mathbf{A}\xi, \mathbf{A}\Omega\mathbf{A}^\top, \mathbf{A}\Delta, \tau, \bar{\Gamma}, g_{Q_{\beta}(\beta)}^{(p+n+q)})$, where $\mathbf{A}\xi = \xi^\dagger$, $\mathbf{A}\Omega\mathbf{A}^\top = \Omega^\dagger$ and $\mathbf{A}\Delta = \Delta^\dagger$. As a result, the prior for β in Proposition 1 follows directly from the closure under marginalization of the SUE family outlined in Lemma 1. Similarly, the likelihood $(\mathbf{y} | \beta)$ and posterior $(\beta | \mathbf{y})$ in Proposition 1 can be readily derived by applying the closure under conditioning properties in Lemma 2 to the joint SUE distribution for $(\beta^\top, \mathbf{y}^\top)^\top$ presented in Proposition 1, with parameters ξ^\dagger , Ω^\dagger and Δ^\dagger partitioned as in (22). \square

As anticipated within Section 3, the results in Proposition 1 clarify that the joint distribution for β and ϵ requires some form of dependence to guarantee conjugacy. In this respect, notice that even when $\xi = \mathbf{0}$, $\Omega_{\epsilon\beta} = \Omega_{\beta\epsilon}^\top = \mathbf{0}$ and $\Delta = \mathbf{0}$, by the closure under conditioning properties of the unified skew-elliptical family, it follows that $(\epsilon | \beta) \sim \text{SUE}_{n,q}(\mathbf{0}, \Omega_\epsilon, \mathbf{0}, \tau, \bar{\Gamma}, g_{Q_{\beta}(\beta)}^{(n+q)})$, which clarifies that a weak form of dependence persists in the conditional density generator. Nonetheless, such a form of higher-level dependence still allows to include within the results in Proposition 1 interesting models with uncorrelated β and ϵ vectors. Recalling the expression for the SUE covariance matrix in (7), a sufficient condition to retrieve these uncorrelated representations is to assume either $\Delta_\beta = \mathbf{0}$ or $\Delta_\epsilon = \mathbf{0}$, and set $\Omega_{\epsilon\beta} = \Omega_{\beta\epsilon}^\top = \mathbf{0}$. When both Δ_β and Δ_ϵ are $\mathbf{0}$, and also $\tau = \mathbf{0}$, by point (i) in Lemma 5, $(\beta^\top, \epsilon^\top)^\top$ reduces to an elliptical distribution. As such, conjugacy under this latter class can be obtained as a special case of Proposition 1.

This discussion clarifies that full independence between β and ϵ cannot be generally enforced if the objective is to obtain broad conjugacy results as in Proposition 1 that hold for the whole SUE family. Nonetheless, in the specific setting of Gaussian density generators, which leads to the sub-class of SUN distributions, such a full independence can be enforced without undermining conjugacy. As discussed in Section 2.2.1, under this specific choice, the conditional density generator coincides with the unconditional one, thus allowing to enforce independence between β and ϵ while preserving conjugacy. This is clear from the results in Anceschi et al. [8], that establish SUN conjugacy via a classical Bayes rule perspective, without requiring to specify a joint distribution for β and ϵ or, alternatively, β and \mathbf{y} . As illustrated in Example 1 below, these conjugacy results can be obtained as a particular case of those in Proposition 1.

Example 1 (SUN Conjugacy). The supplementary materials of Anceschi et al. [8] present an example based on a classical linear regression with skew-normal errors, i.e., $(y_i | \beta) \sim \text{SN}(x_i^\top \beta, \sigma^2, \alpha)$, independently for every unit $i \in \{1, \dots, n\}$, and, consistent with our notation, $\text{SUN}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau_\beta, \bar{\Gamma}_\beta)$ prior for β . This model yields a likelihood $p(\mathbf{y} | \beta) \propto \phi_n(\mathbf{y} - \mathbf{X}\beta; \sigma^2 \mathbf{I}_n) \Phi_n(\alpha \mathbf{y} - \alpha \mathbf{X}\beta; \sigma^2 \mathbf{I}_n)$ that is proportional to a $\text{SUN}_{n,n}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n, \alpha \sigma \mathbf{I}_n, \mathbf{0}, (1 + \alpha^2)\sigma^2 \mathbf{I}_n)$ density. Leveraging Lemma 4, such a SUN is equivalent to $(\mathbf{y} | \beta) \sim \text{SUN}_{n,n}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n, [\alpha/(1 + \alpha^2)^{1/2}] \mathbf{I}_n, \mathbf{0}, \mathbf{I}_n)$. Before showing that such a Bayesian formulation is a special case of the broader family of models and priors in Proposition 1, it shall be emphasized that this construction also comprises classical multivariate Gaussian priors for β , when $\Delta_\beta = \mathbf{0}$, and Gaussian linear regression for \mathbf{y} if $\alpha = 0$. Replacing $\sigma^2 \mathbf{I}_n$ with a full covariance matrix also leads to general multivariate versions of such models. This yields an important class of routinely-implemented formulations.

In order to rephrase the above Bayesian formulation within those covered by Proposition 1, let us consider the case $(\beta^\top, \epsilon^\top)^\top \sim \text{SUN}_{p+n,q+n}(\xi, \Omega, \Delta, \tau, \bar{\Gamma})$ with parameters partitioned as

$$\xi = \begin{bmatrix} \xi_\beta \\ \mathbf{0} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_\beta & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_n \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_\beta & \mathbf{0} \\ \mathbf{0} & \bar{\alpha} \mathbf{I}_n \end{bmatrix}, \quad \tau = \begin{bmatrix} \tau_\beta \\ \mathbf{0} \end{bmatrix}, \quad \bar{\Gamma} = \begin{bmatrix} \bar{\Gamma}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix},$$

where $\bar{\alpha} = \alpha/(1 + \alpha^2)^{1/2}$. Adapting Proposition 1 to this setting, yields $(\beta^\top, \mathbf{y}^\top)^\top \sim \text{SUN}_{p+n,q+n}(\xi^\dagger, \Omega^\dagger, \Delta^\dagger, \tau, \bar{\Gamma})$, with

$$\xi^\dagger = \begin{bmatrix} \xi_\beta \\ \mathbf{X}\xi_\beta \end{bmatrix}, \quad \Omega^\dagger = \begin{bmatrix} \Omega_\beta & \Omega_\beta \mathbf{X}^\top \\ \mathbf{X}\Omega_\beta & \mathbf{X}\Omega_\beta \mathbf{X}^\top + \sigma^2 \mathbf{I}_n \end{bmatrix}, \quad \Delta^\dagger = \begin{bmatrix} \Delta_\beta & \mathbf{0} \\ \omega_y^{-1} \mathbf{X}\omega_\beta \Delta_\beta & \omega_y^{-1} \sigma \bar{\alpha} \mathbf{I}_n \end{bmatrix}, \quad \tau = \begin{bmatrix} \tau_\beta \\ \mathbf{0} \end{bmatrix}, \quad \bar{\Gamma} = \begin{bmatrix} \bar{\Gamma}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix}.$$

As a direct consequence of the closure under linear combinations of SUE, and hence SUN, distributions, together with point (ii) in Lemma 5 (see also Remark 2), the above formulation implies

$$\beta \sim \text{SUN}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau_\beta, \bar{\Gamma}_\beta), \quad \epsilon \sim \text{SUN}_{n,n}(\mathbf{0}, \sigma^2 \mathbf{I}_n, \bar{\alpha} \mathbf{I}_n, \mathbf{0}, \mathbf{I}_n),$$

where the marginal distribution for β coincides with the prior considered in Anceschi et al. [8], whereas ϵ corresponds to a noise vector comprising n independent skew-normals. Similarly, by applying to the above parameters the expressions for those of $(\mathbf{y} | \beta)$ in Proposition 1, and recalling Remark 2, yields after standard calculations the following likelihood

$$(\mathbf{y} | \beta) \sim \text{SUN}_{n,n}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n, [\alpha/(1 + \alpha^2)^{1/2}] \mathbf{I}_n, \mathbf{0}, \mathbf{I}_n).$$

Such a likelihood is proportional to the one considered in Anceschi et al. [8] for the general skew-normal regression setting, which includes the Gaussian as a special case and can be readily extended to more general multivariate models. As such, Proposition 1 also covers SUN conjugacy properties in commonly-implemented linear models.

Example 2 (SUT Conjugacy). As stated in Corollary 1, by specializing Proposition 1 to the SUT sub-family in Section 2.2.2, it is possible to derive new conjugacy results not yet explored in the literature, along with examples of potential interest in applications.

Corollary 1. Consider model (20), and assume $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUT}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$, with ξ, Ω, Δ partitioned as in (21). Then, the prior for β is $\beta \sim \text{SUT}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau, \bar{\Gamma}, \nu)$, whereas the likelihood and posterior are $(y | \beta) \sim \text{SUT}_{n,q}(\xi_{y|\beta}, \alpha_\beta \Omega_{y|\beta}, \Delta_{y|\beta}, \alpha_\beta^{-1/2} \tau_{y|\beta}, \bar{\Gamma}_{y|\beta}, \nu + p)$ and $(\beta | y) \sim \text{SUT}_{p,q}(\xi_{\beta|y}, \alpha_y \Omega_{\beta|y}, \Delta_{\beta|y}, \alpha_y^{-1/2} \tau_{\beta|y}, \bar{\Gamma}_{\beta|y}, \nu + n)$, respectively. In these expressions, $\alpha_\beta = [\nu + Q_\beta(\beta)]/(\nu + p)$ and $\alpha_y = [\nu + Q_y(y)]/(\nu + n)$. The remaining parameters, along with $Q_\beta(\beta)$ and $Q_y(y)$, are defined as in Proposition 1.

Proof. The proof follows directly by replacing the generic density generators in Proposition 1 with those of the Student’s t distribution presented in Section 2.2.2. Alternatively, it is possible to prove the statement leveraging the specific properties of unified skew- t distributions in Proposition 11 of Wang et al. [48]. \square

Corollary 1 states a general conjugacy result which includes those of the SUN as a limiting case, provided that SUT distributions converge to SUNs when $\nu \rightarrow \infty$ [e.g., 48]. In addition, suitable constraints on the parameters of the joint SUT distribution for $(\beta^\top, \varepsilon^\top)^\top$ in Corollary 1, yield priors and likelihoods of potential practical interest. In particular, consider $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUT}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$, with

$$\xi = \begin{bmatrix} \xi_\beta \\ \mathbf{0} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_\beta & \mathbf{0} \\ \mathbf{0} & \Omega_\varepsilon \end{bmatrix}, \quad \Delta = \begin{bmatrix} \mathbf{0} \\ \Delta_\varepsilon \end{bmatrix},$$

and $\tau = \mathbf{0}$. Then, by the closure under marginalization of SUE distributions combined with Lemma 5 and Corollary 1, we have that β and ε are uncorrelated and have marginals

$$\beta \sim \mathcal{F}_p(\xi_\beta, \Omega_\beta, \nu), \quad \varepsilon \sim \text{SUT}_{n,q}(\mathbf{0}, \Omega_\varepsilon, \Delta_\varepsilon, \mathbf{0}, \bar{\Gamma}, \nu),$$

which yield a Bayesian multivariate linear regression model $y = X\beta + \varepsilon$ with Student’s t prior on β and unified skew- t residuals ε , uncorrelated with β . In the above expression, $\mathcal{F}_p(\xi_\beta, \Omega_\beta, \nu)$ denotes the p -variate Student’s t distribution with location ξ_β , scale Ω_β and degrees of freedom ν . By Corollary 1, this implies the SUT likelihood

$$(y | \beta) \sim \text{SUT}_{n,q}(X\beta, \alpha_\beta \Omega_\varepsilon, \Delta_\varepsilon, \mathbf{0}, \bar{\Gamma}, \nu + p).$$

Including the additional constrain $\Delta_\varepsilon = \mathbf{0}$ within this formulation, and recalling again Lemma 5, it is possible to obtain the Bayesian Student’s t regression with prior and likelihood given by

$$\beta \sim \mathcal{F}_p(\xi_\beta, \Omega_\beta, \nu), \quad (y | \beta) \sim \mathcal{F}_n(X\beta, \alpha_\beta \Omega_\varepsilon, \nu + p),$$

which yields, by Corollary 1, a p -variate Student’s t posterior for β . This result provides an important finding which clarifies that, in specific contexts of potential practical interest, Student’s t – Student’s t conjugacy can be attained, thereby expanding some earlier findings in Song and Xia [55] on a simpler formulation. As is clear from the expression of the likelihood, for this property to hold it is necessary to incorporate the classical location dependence on β via $X\beta$, together with a weak form of additional dependence induced by the scaling term $\alpha_\beta = [\nu + (\beta - \xi_\beta)^\top \Omega_\beta^{-1} (\beta - \xi_\beta)]/(\nu + p)$. Recalling Zhang et al. [40], under weakly informative Student’s t priors employed in practice for β , such an effect tends to be small, and most of the dependence between β and y is through the classical linear predictor $X\beta$. In addition, notice that, when $\beta \sim \mathcal{F}_p(\xi_\beta, \Omega_\beta, \nu)$, then $(\beta - \xi_\beta)^\top \Omega_\beta^{-1} (\beta - \xi_\beta)/p$ has F distribution with degrees of freedom p and ν , which implies that for moderate p and ν , the term $(\beta - \xi_\beta)^\top \Omega_\beta^{-1} (\beta - \xi_\beta)/p$, and hence α_β , shrink to 1.

Section 3.2 shows that the conjugacy properties of SUE distributions derived in Proposition 1 extend even beyond multivariate linear models for continuous response vectors, to cover, in particular, also generalizations of multivariate probit and multinomial probit under elliptical or skew-elliptical link functions.

3.2. Conjugacy properties of SUE distributions in multivariate binary models

When the focus is on Bayesian modeling of multivariate binary data $y \in \{0, 1\}^n$, a natural strategy, which extends classical probit, multivariate probit and multinomial probit formulations [27,63], is to adapt the class of models studied in Section 3.1 to such a setting, by assuming

$$y = [1(\bar{y}_1 > 0), \dots, 1(\bar{y}_n > 0)]^\top, \quad \bar{y} = X\beta + \varepsilon, \tag{23}$$

where $1(\cdot)$ denotes the indicator function, $\bar{y} = (\bar{y}_1, \dots, \bar{y}_n)^\top \in \mathbb{R}^n$ and $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUE}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, g^{(p+n+q)})$. Proposition 2 clarifies from a general perspective that SUE conjugacy can be established also in these contexts. This unifies and extends the contributions by Durante [1], Fasano and Durante [6], Anceschi et al. [8] and Zhang et al. [40] on particular SUE sub-classes; i.e., SUNs and specific skew-elliptical distributions in the SUE family; see also Remark 3.

Proposition 2. Consider the binary random vector $y \in \{0, 1\}^n$, defined as in (23) and let $D_y = \text{diag}(2y_1 - 1, \dots, 2y_n - 1)$. Moreover, assume again $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUE}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, g^{(p+n+q)})$ with parameters partitioned as in (21). Then, $(\beta^\top, \bar{y}^\top)^\top$ is a SUE with dimensions $(p + n, q)$ and parameters defined as in (22). In addition

(a) Prior distribution: $\beta \sim \text{SUE}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau, \bar{\Gamma}, g^{(p+q)})$.

(b) Likelihood: $(\mathbf{y} \mid \beta)$ is a multivariate Bernoulli with probability $\Pi_{\mathbf{y}|\beta}$ for the generic configuration \mathbf{y} defined as

$$\Pi_{\mathbf{y}|\beta} = \mathbb{P}(\mathbf{y} \mid \beta) = \frac{F_{n+q} \left(\begin{bmatrix} \xi_{\mathbf{y}|\beta} \\ \tau_{\mathbf{y}|\beta} \end{bmatrix}; \begin{bmatrix} \Omega_{\mathbf{y}|\beta} & \omega_{\mathbf{y}|\beta} \mathbf{A}_{\mathbf{y}|\beta} \\ \Delta_{\mathbf{y}|\beta}^\top \omega_{\mathbf{y}|\beta} & \bar{\Gamma}_{\mathbf{y}|\beta} \end{bmatrix}, g_{Q_\beta(\beta)}^{(n+q)} \right)}{F_q(\tau_{\mathbf{y}|\beta}; \bar{\Gamma}_{\mathbf{y}|\beta}, g_{Q_\beta(\beta)}^{(q)})},$$

for all $\mathbf{y} \in \{0, 1\}^n$, and parameters available in closed form according to the following equations

$$\begin{aligned} \xi_{\mathbf{y}|\beta} &= \mathbf{D}_{\mathbf{y}} [\xi_{\mathbf{y}} + \Omega_{\mathbf{y}\beta} \Omega_\beta^{-1} (\beta - \xi_\beta)], & \Omega_{\mathbf{y}|\beta} &= \mathbf{D}_{\mathbf{y}} (\Omega_{\mathbf{y}} - \Omega_{\mathbf{y}\beta} \Omega_\beta^{-1} \Omega_{\beta\mathbf{y}}) \mathbf{D}_{\mathbf{y}}, & \mathbf{A}_{\mathbf{y}|\beta} &= \omega_{\mathbf{y}|\beta}^{-1} \mathbf{D}_{\mathbf{y}} (\omega_{\mathbf{y}} \mathbf{A}_{\mathbf{y}} - \Omega_{\mathbf{y}\beta} \Omega_\beta^{-1} \omega_\beta \mathbf{A}_\beta) \gamma_{\mathbf{y}|\beta}^{-1}, \\ \tau_{\mathbf{y}|\beta} &= \gamma_{\mathbf{y}|\beta}^{-1} [\tau + \Delta_\beta^\top \bar{\Omega}_\beta^{-1} \omega_\beta^{-1} (\beta - \xi_\beta)], & \bar{\Gamma}_{\mathbf{y}|\beta} &= \gamma_{\mathbf{y}|\beta}^{-1} (\bar{\Gamma} - \Delta_\beta^\top \bar{\Omega}_\beta^{-1} \Delta_\beta) \gamma_{\mathbf{y}|\beta}^{-1}, & Q_\beta(\beta) &= (\beta - \xi_\beta)^\top \Omega_\beta^{-1} (\beta - \xi_\beta), \end{aligned}$$

where $\omega_{\mathbf{y}|\beta} = \text{diag}(\Omega_{\mathbf{y}|\beta})^{1/2}$ and $\gamma_{\mathbf{y}|\beta} = \text{diag}(\bar{\Gamma} - \Delta_\beta^\top \bar{\Omega}_\beta^{-1} \Delta_\beta)^{1/2}$, while $\xi_{\mathbf{y}}$, $\Omega_{\mathbf{y}}$, $\omega_{\mathbf{y}}$, $\Omega_{\mathbf{y}\beta}$, $\Omega_{\beta\mathbf{y}}$, and $\mathbf{A}_{\mathbf{y}}$ are defined as in (22).

(c) Posterior distribution: $(\beta \mid \mathbf{y}) \sim \text{SUE}_{p,n+q}(\xi_{\beta|\mathbf{y}}, \Omega_{\beta|\mathbf{y}}, \mathbf{A}_{\beta|\mathbf{y}}, \tau_{\beta|\mathbf{y}}, \bar{\Gamma}_{\beta|\mathbf{y}}, g^{(p+n+q)})$, with parameters

$$\xi_{\beta|\mathbf{y}} = \xi_\beta, \quad \Omega_{\beta|\mathbf{y}} = \Omega_\beta, \quad \mathbf{A}_{\beta|\mathbf{y}} = [\bar{\Omega}_{\beta\mathbf{y}} \mathbf{D}_{\mathbf{y}} \quad \Delta_\beta], \quad \tau_{\beta|\mathbf{y}} = \begin{bmatrix} \omega_{\mathbf{y}}^{-1} \mathbf{D}_{\mathbf{y}} \xi_{\mathbf{y}} \\ \tau \end{bmatrix}, \quad \bar{\Gamma}_{\beta|\mathbf{y}} = \begin{bmatrix} \mathbf{D}_{\mathbf{y}} \bar{\Omega}_{\mathbf{y}} \mathbf{D}_{\mathbf{y}} & \mathbf{D}_{\mathbf{y}} \mathbf{A}_{\mathbf{y}} \\ \Delta_{\mathbf{y}}^\top \mathbf{D}_{\mathbf{y}} & \bar{\Gamma} \end{bmatrix},$$

where $\bar{\Omega}_{\beta\mathbf{y}} = \omega_\beta^{-1} \Omega_{\beta\mathbf{y}} \omega_{\mathbf{y}}^{-1}$ and $\bar{\Omega}_{\mathbf{y}} = \omega_{\mathbf{y}}^{-1} \Omega_{\mathbf{y}} \omega_{\mathbf{y}}^{-1}$, while $\xi_{\mathbf{y}}$, $\Omega_{\mathbf{y}}$, $\omega_{\mathbf{y}}$, $\Omega_{\mathbf{y}\beta}$, $\Omega_{\beta\mathbf{y}}$, and $\mathbf{A}_{\mathbf{y}}$ are defined as in (22).

Proof. To prove Proposition 2, first notice that under model (23), the probability of observing a given configuration \mathbf{y} coincides with that of the event $\mathbf{D}_{\mathbf{y}} \bar{\mathbf{y}} > \mathbf{0}$. Let $\mathbf{D}_{\mathbf{y}} \bar{\mathbf{y}} =: \bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}}$, then

$$\begin{bmatrix} \beta \\ \bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}} \end{bmatrix} = \mathbf{A}_{\mathbf{y}} \begin{bmatrix} \beta \\ \bar{\mathbf{y}} \end{bmatrix}, \quad \mathbf{A}_{\mathbf{y}} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\mathbf{y}} \end{bmatrix}, \tag{24}$$

where $(\beta^\top, \bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}}^\top)^\top \sim \text{SUE}_{p+n,q}(\xi^\dagger, \Omega^\dagger, \Delta^\dagger, \tau, \bar{\Gamma}, g^{(p+n+q)})$, with parameters as in (22).

Therefore, by Lemma 1, we have that $(\beta^\top, \bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}}^\top)^\top \sim \text{SUE}_{p+n,q}(\mathbf{A}_{\mathbf{y}} \xi^\dagger, \mathbf{A}_{\mathbf{y}} \Omega^\dagger \mathbf{A}_{\mathbf{y}}^\top, \Delta_{\mathbf{A}_{\mathbf{y}}}^\dagger, \tau, \bar{\Gamma}, g^{(p+n+q)})$, with

$$\mathbf{A}_{\mathbf{y}} \xi^\dagger = \begin{bmatrix} \xi_\beta \\ \mathbf{D}_{\mathbf{y}} \xi_{\mathbf{y}} \end{bmatrix}, \quad \mathbf{A}_{\mathbf{y}} \Omega^\dagger \mathbf{A}_{\mathbf{y}}^\top = \begin{bmatrix} \Omega_\beta & \Omega_{\beta\mathbf{y}} \mathbf{D}_{\mathbf{y}} \\ \mathbf{D}_{\mathbf{y}} \Omega_{\mathbf{y}\beta} & \mathbf{D}_{\mathbf{y}} \Omega_{\mathbf{y}} \mathbf{D}_{\mathbf{y}} \end{bmatrix}, \quad \Delta_{\mathbf{A}_{\mathbf{y}}}^\dagger = \begin{bmatrix} \Delta_\beta \\ \mathbf{D}_{\mathbf{y}} \mathbf{A}_{\mathbf{y}} \end{bmatrix}.$$

Under the above construction, the prior for β follows directly by the SUE closure under marginalization.

As for the likelihood of \mathbf{y} , recall that $\mathbb{P}(\mathbf{y} \mid \beta)$ is equal to $\mathbb{P}(\bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}} > \mathbf{0} \mid \beta)$. Moreover, by applying the results in Proposition 1 to the random vector $(\beta^\top, \bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}}^\top)^\top$, we have $(\bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}} \mid \beta) \sim \text{SUE}_{n,q}(\xi_{\mathbf{y}|\beta}, \Omega_{\mathbf{y}|\beta}, \mathbf{A}_{\mathbf{y}|\beta}, \tau_{\mathbf{y}|\beta}, \bar{\Gamma}_{\mathbf{y}|\beta}, g_{Q_\beta(\beta)}^{(n+q)})$, with parameters defined as in Proposition 2. Therefore, the likelihood $\mathbb{P}(\mathbf{y} \mid \beta)$ coincides with the cumulative distribution function, evaluated at $\mathbf{0}$, of the SUE random variable $(-\bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}} \mid \beta) \sim \text{SUE}_{n,q}(-\xi_{\mathbf{y}|\beta}, \Omega_{\mathbf{y}|\beta}, -\mathbf{A}_{\mathbf{y}|\beta}, \tau_{\mathbf{y}|\beta}, \bar{\Gamma}_{\mathbf{y}|\beta}, g_{Q_\beta(\beta)}^{(n+q)})$. As a consequence, by applying (4) to such a SUE yields

$$\mathbb{P}(\mathbf{y} \mid \beta) = \frac{F_{n+q} \left(\begin{bmatrix} \xi_{\mathbf{y}|\beta} \\ \tau_{\mathbf{y}|\beta} \end{bmatrix}; \begin{bmatrix} \Omega_{\mathbf{y}|\beta} & \omega_{\mathbf{y}|\beta} \mathbf{A}_{\mathbf{y}|\beta} \\ \Delta_{\mathbf{y}|\beta}^\top \omega_{\mathbf{y}|\beta} & \bar{\Gamma}_{\mathbf{y}|\beta} \end{bmatrix}, g_{Q_\beta(\beta)}^{(n+q)} \right)}{F_q(\tau_{\mathbf{y}|\beta}; \bar{\Gamma}_{\mathbf{y}|\beta}, g_{Q_\beta(\beta)}^{(q)})},$$

for all $\mathbf{y} \in \{0, 1\}^n$, as in Proposition 2. To conclude the proof, note that $(\beta \mid \mathbf{y})$ is distributed as $(\beta \mid \bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}} > \mathbf{0})$. As a result, the posterior distribution follows directly by applying Lemma 3 to the SUE random vector $(\beta^\top, \bar{\mathbf{y}}_{\mathbf{D}_{\mathbf{y}}}^\top)^\top$. \square

Proposition 2 clarifies that SUE distributions possess fundamental conjugacy properties also when combined with specific models for multivariate binary data. This result extends the one recently derived by Zhang et al. [40] under model (23) with a specific focus on a skew-elliptical joint distribution for $(\beta^\top, \epsilon^\top)^\top$ which enforces lack of correlation between β and ϵ while inducing an elliptical prior for β . Such a construction can be derived, under simple linear algebra operations, as a particular case of the general SUE assumption for $(\beta^\top, \epsilon^\top)^\top$ in Proposition 2, which crucially allows to recover more general Bayesian formulations, including priors beyond the symmetric elliptical family. This connection with the contribution by Zhang et al. [40] is helpful to showcase the practical impact of extending conjugacy to broader classes of models beyond classical multivariate and multinomial probit. Examples 3–4 further stress this aspect with a focus on SUN and SUT distributions.

Example 3 (SUN Conjugacy). A direct and natural strategy to adapt the model studied in Example 1 within the binary data context, is to consider $y_i = 1(\bar{y}_i > 0)$ with $(\bar{y}_i \mid \beta) \sim \text{SN}(x_i^\top \beta, \sigma^2, \alpha)$, independently for $i \in \{1, \dots, n\}$, and $\beta \sim \text{SUN}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau_\beta, \bar{\Gamma}_\beta)$. Such a model is studied in the supplementary materials of Anceschi et al. [8] as a broad extension of classical probit models to skewed link functions, further facilitating generalizations to multivariate and multinomial binary responses. Leveraging standard properties of multivariate Gaussian cumulative distribution functions, the resulting likelihood in Anceschi et al. [8] can be alternatively re-expressed as proportional to $\Phi_{2n}((\mathbf{D}_{\mathbf{y}} \mathbf{X} \beta)^\top, \mathbf{0}^\top; \Sigma)$, where Σ is a block matrix partitioned as $\Sigma_{11} = \sigma^2 \mathbf{I}_n$, $\Sigma_{21} = \Sigma_{12}^\top = \mathbf{D}_{\mathbf{y}} \bar{\alpha} \sigma \mathbf{I}_n$ and $\Sigma_{22} = \mathbf{I}_n$.

To recast the above Bayesian formulation within those studied in Proposition 2, consider the setting $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUN}_{p+n,q+n}(\xi, \Omega, \Delta, \tau, \bar{\Gamma})$ with parameters partitioned as in Example 1. This assumption, combined with model (23) and the proof of Proposition 2, implies $(\beta^\top, \bar{y}_{D_y}^\top)^\top \sim \text{SUN}_{p+n,q+n}(\mathbf{A}_y \xi^\dagger, \mathbf{A}_y \Omega^\dagger \mathbf{A}_y^\top, \Delta_{\mathbf{A}_y}^\dagger, \tau, \bar{\Gamma})$, with

$$\mathbf{A}_y \xi^\dagger = \begin{bmatrix} \xi_\beta \\ D_y \mathbf{X} \xi_\beta \end{bmatrix}, \quad \mathbf{A}_y \Omega^\dagger \mathbf{A}_y^\top = \begin{bmatrix} \Omega_\beta & \Omega_\beta \mathbf{X}^\top D_y \\ D_y \mathbf{X} \Omega_\beta & D_y (\mathbf{X} \Omega_\beta \mathbf{X}^\top + \sigma^2 \mathbf{I}_n) D_y \end{bmatrix}, \quad \Delta_{\mathbf{A}_y}^\dagger = \begin{bmatrix} \Delta_\beta & \mathbf{0} \\ \omega_y^{-1} D_y \mathbf{X} \omega_\beta \Delta_\beta & \omega_y^{-1} D_y \sigma \bar{\alpha} \mathbf{I}_n \end{bmatrix},$$

where $\bar{y}_{D_y} = D_y \bar{y}$, $\bar{\alpha} = \alpha / (1 + \alpha^2)^{1/2}$, and \mathbf{A}_y is defined in (24). The above representations, together with the closure properties of SUNs and point (ii) in Lemma 5, yield

$$\beta \sim \text{SUN}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau_\beta, \bar{\Gamma}_\beta), \quad \varepsilon \sim \text{SUN}_{n,n}(\mathbf{0}, \sigma^2 \mathbf{I}_n, \bar{\alpha} \mathbf{I}_n, \mathbf{0}, \mathbf{I}_n),$$

thereby recovering the SUN prior and skew-normal noise vector considered in Anceschi et al. [8]. Furthermore, by Lemma 2 and Remark 2, we have $(\bar{y}_{D_y} | \beta) \sim \text{SUN}_{n,n}(D_y \mathbf{X} \beta, \sigma^2 \mathbf{I}_n, D_y \bar{\alpha} \mathbf{I}_n, \mathbf{0}, \mathbf{I}_n)$ which implies

$$\mathbb{P}(y | \beta) \propto \Phi_{2n} \left(\begin{bmatrix} D_y \mathbf{X} \beta \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} \sigma^2 \mathbf{I}_n & D_y \bar{\alpha} \sigma \mathbf{I}_n \\ D_y \bar{\alpha} \sigma \mathbf{I}_n & \mathbf{I}_n \end{bmatrix} \right),$$

for all $y \in \{0, 1\}^n$, which coincides again with the likelihood in Anceschi et al. [8]. As discussed above, such a formulation includes several models of direct interest in practice. For instance, setting $\alpha = 0$ yields classical probit regression, whereas replacing $\sigma^2 \mathbf{I}_n$ with a full covariance matrix allows to recover multivariate probit and, for a suitable specification of \mathbf{X} , multinomial probit, under both skewed and non-skewed link functions.

As discussed in Example 4 below, these classes of models can be further extended to specific formulations relying on Student's t and skew- t link functions while preserving conjugacy. As such, the practical impact of Proposition 2 goes beyond the broad class of models studied in Anceschi et al. [8].

Example 4 (SUT Conjugacy). Corollary 2 specializes the conjugacy properties derived in Proposition 2 to the specific context of the SUT sub-family presented in Section 2.2.2.

Corollary 2. Consider model (23), with $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUT}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$, and parameters ξ, Ω, Δ partitioned as in (21). Then, the induced prior distribution is $\beta \sim \text{SUT}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau, \bar{\Gamma}, \nu)$, whereas the likelihood is equal to

$$\mathbb{P}(y | \beta) = \frac{T_{n+q} \left(\alpha_\beta^{-1/2} \begin{bmatrix} \xi_{y|\beta} \\ \tau_{y|\beta} \end{bmatrix}; \begin{bmatrix} \Omega_{y|\beta} & \omega_{y|\beta} \Delta_{y|\beta} \\ \Delta_{y|\beta}^\top \omega_{y|\beta} & \bar{\Gamma}_{y|\beta} \end{bmatrix}, \nu + p \right)}{T_q(\alpha_\beta^{-1/2} \tau_{y|\beta}; \bar{\Gamma}_{y|\beta}, \nu + p)},$$

for all $y \in \{0, 1\}^n$, with $\alpha_\beta = [\nu + (\beta - \xi_\beta)^\top \Omega_\beta^{-1} (\beta - \xi_\beta)] / (\nu + p)$, and $\xi_{y|\beta}, \tau_{y|\beta}, \Omega_{y|\beta}, \omega_{y|\beta}, \Delta_{y|\beta}, \bar{\Gamma}_{y|\beta}$ as in Proposition 2. The resulting posterior is $(\beta | y) \sim \text{SUT}_{p,n+q}(\xi_{\beta|y}, \Omega_{\beta|y}, \Delta_{\beta|y}, \tau_{\beta|y}, \bar{\Gamma}_{\beta|y}, \nu)$, with parameters as in Proposition 2.

Proof. To prove Corollary 2, it is sufficient to replace the generic density generators in Proposition 2 with those of the Student's t distribution provided in Section 2.2.2. \square

As for the continuous setting in Example 2, let us consider special cases of potential practical interest that arise from Corollary 2 under suitable constraints. In particular, define $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUT}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$, with

$$\xi = \begin{bmatrix} \xi_\beta \\ \mathbf{0} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_\beta & \mathbf{0} \\ \mathbf{0} & \Omega_\varepsilon \end{bmatrix}, \quad \Delta = \begin{bmatrix} \mathbf{0} \\ \Delta_\varepsilon \end{bmatrix},$$

and $\tau = \mathbf{0}$. Recalling Example 2, such a formulation implies

$$\beta \sim \mathcal{F}_p(\xi_\beta, \Omega_\beta, \nu), \quad \varepsilon \sim \text{SUT}_{n,q}(\mathbf{0}, \Omega_\varepsilon, \Delta_\varepsilon, \mathbf{0}, \bar{\Gamma}, \nu),$$

and hence, under (23), the resulting model for y coincides with a multivariate binary regression having unified skew- t link function, and Student's t prior for β uncorrelated with the underlying noise vector ε . As a direct consequence of Corollary 2, such a formulation yields the likelihood

$$\mathbb{P}(y | \beta) \propto T_{n+q} \left(\begin{bmatrix} \alpha_\beta^{-1/2} D_y \mathbf{X} \beta \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} D_y \Omega_\varepsilon D_y & \omega_\varepsilon D_y \Delta_\varepsilon \\ \Delta_\varepsilon^\top D_y \omega_\varepsilon & \bar{\Gamma} \end{bmatrix}, \nu + p \right),$$

for all $y \in \{0, 1\}^n$, which provides a natural extension to more general settings of classical binary regression with t link function; recall the discussion in Example 2 on the impact of $\alpha_\beta^{-1/2}$, relative to the standard linear dependence through $\mathbf{X}\beta$. By setting $\Delta_\varepsilon = \mathbf{0}$ within the above formulation, and recalling again Example 2, leads to

$$\beta \sim \mathcal{F}_p(\xi_\beta, \Omega_\beta, \nu), \quad \mathbb{P}(y | \beta) \propto T_n(\alpha_\beta^{-1/2} D_y \mathbf{X} \beta; D_y \Omega_\varepsilon D_y, \nu + p),$$

which yields a closed-form SUT posterior distribution, while further clarifying the direct link with statistical models implemented in practice.

Section 3.3 concludes our analysis by studying SUE conjugacy in regression models for random vectors comprising both fully-observed and dichotomized data. Such a class combines results in Sections 3.1–3.2 to explore a general set of formulations that extends classical tobit models in both multivariate and skew-elliptical contexts.

3.3. Conjugacy properties of SUE distributions in multivariate censored models

The classes of models studied in Sections 3.1–3.2 are designed for data that are either all continuous or all discretized. However, in practice, it is also possible to observe vectors comprising a combination of these two types of data. This is the case, for example, when a continuous variable is fully observed only if its value exceeds a certain threshold.

Such a form of censoring is common in several applications and is typically addressed via tobit models and related extensions [64, 65]. Although common implementations rely on Gaussian noise vectors, such a class can be naturally extended to the broader unified skew-elliptical family via the following formulation

$$y = [\bar{y}_1 1(\bar{y}_1 > 0), \dots, \bar{y}_n 1(\bar{y}_n > 0)]^T, \quad \bar{y} = X\beta + \varepsilon, \tag{25}$$

where $1(\cdot)$ is the indicator function, $\bar{y} = (\bar{y}_1, \dots, \bar{y}_n)^T \in \mathbb{R}^n$ and $(\beta^T, \varepsilon^T)^T \sim \text{SUE}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, g^{(p+n+q)})$.

Recent research on the above class of Bayesian models [8] has shown that when β and ε have independent SUN distributions, also the posterior $(\beta | y)$ is SUN. Proposition 3 below clarifies that similar conjugacy results can be obtained when the focus is on the whole SUE family; see also Remark 3.

Proposition 3. Let $y = (y_1^T, y_0^T)^T$ denote a generic realization from model (25), where $y_1 \in \mathbb{R}^{n_1}$ corresponds to the vector of fully-observed data and $y_0 = \mathbf{0}$ comprises the n_0 censored ones, with $n_1 + n_0 = n$. Moreover, assume again $(\beta^T, \varepsilon^T)^T \sim \text{SUE}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, g^{(p+n+q)})$ and consider the following partition of the parameters

$$\xi = \begin{bmatrix} \xi_\beta \\ \xi_\varepsilon \end{bmatrix} = \begin{bmatrix} \xi_\beta \\ \xi_{\varepsilon_1} \\ \xi_{\varepsilon_0} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_\beta & \Omega_{\beta\varepsilon} \\ \Omega_{\varepsilon\beta} & \Omega_\varepsilon \end{bmatrix} = \begin{bmatrix} \Omega_\beta & \Omega_{\beta\varepsilon_1} & \Omega_{\beta\varepsilon_0} \\ \Omega_{\varepsilon_1\beta} & \Omega_{\varepsilon_1} & \Omega_{\varepsilon_1\varepsilon_0} \\ \Omega_{\varepsilon_0\beta} & \Omega_{\varepsilon_0\varepsilon_1} & \Omega_{\varepsilon_0} \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_\beta \\ \Delta_\varepsilon \end{bmatrix} = \begin{bmatrix} \Delta_\beta \\ \Delta_{\varepsilon_1} \\ \Delta_{\varepsilon_0} \end{bmatrix}, \tag{26}$$

where ε_1 and ε_0 comprise the noise terms associated with the two vectors y_1 and y_0 in which the generic realization y is partitioned. Then, when \bar{y} is defined as in (25), we have $(\beta^T, \bar{y}^T)^T \sim \text{SUE}_{p+n_1+n_0,q}(\xi^\dagger, \Omega^\dagger, \Delta^\dagger, \tau, \bar{\Gamma}, g^{(p+n_1+n_0+q)})$ with

$$\begin{aligned} \xi^\dagger &= \begin{bmatrix} \xi_\beta \\ X_1 \xi_\beta + \xi_{\varepsilon_1} \\ X_0 \xi_\beta + \xi_{\varepsilon_0} \end{bmatrix} =: \begin{bmatrix} \xi_\beta \\ \xi_{y_1} \\ \xi_{y_0} \end{bmatrix} = \begin{bmatrix} \xi_{-y_0} \\ \xi_{y_0} \end{bmatrix}, & \Delta^\dagger &= \begin{bmatrix} \Delta_\beta \\ \omega_{y_1}^{-1}(X_1 \omega_\beta \Delta_\beta + \omega_{\varepsilon_1} \Delta_{\varepsilon_1}) \\ \omega_{y_0}^{-1}(X_0 \omega_\beta \Delta_\beta + \omega_{\varepsilon_0} \Delta_{\varepsilon_0}) \end{bmatrix} =: \begin{bmatrix} \Delta_\beta \\ \Delta_{y_1} \\ \Delta_{y_0} \end{bmatrix} = \begin{bmatrix} \Delta_{-y_0} \\ \Delta_{y_0} \end{bmatrix}, \\ \Omega^\dagger &= \begin{bmatrix} \Omega_\beta & \Omega_\beta X_1^T + \Omega_{\beta\varepsilon_1} & \Omega_\beta X_0^T + \Omega_{\beta\varepsilon_0} \\ X_1 \Omega_\beta + \Omega_{\varepsilon_1\beta} & X_1 \Omega_\beta X_1^T + \Omega_{\varepsilon_1\beta} X_1^T + X_1 \Omega_{\beta\varepsilon_1} + \Omega_{\varepsilon_1} & X_1 \Omega_\beta X_0^T + \Omega_{\varepsilon_1\beta} X_0^T + X_1 \Omega_{\beta\varepsilon_0} + \Omega_{\varepsilon_1\varepsilon_0} \\ X_0 \Omega_\beta + \Omega_{\varepsilon_0\beta} & X_0 \Omega_\beta X_1^T + \Omega_{\varepsilon_0\beta} X_1^T + X_0 \Omega_{\beta\varepsilon_1} + \Omega_{\varepsilon_0\varepsilon_1} & X_0 \Omega_\beta X_0^T + \Omega_{\varepsilon_0\beta} X_0^T + X_0 \Omega_{\beta\varepsilon_0} + \Omega_{\varepsilon_0} \end{bmatrix} \\ &=: \begin{bmatrix} \Omega_\beta & \Omega_{\beta y_1} & \Omega_{\beta y_0} \\ \Omega_{y_1\beta} & \Omega_{y_1} & \Omega_{y_1 y_0} \\ \Omega_{y_0\beta} & \Omega_{y_0 y_1} & \Omega_{y_0} \end{bmatrix} = \begin{bmatrix} \Omega_{-y_0} & \Omega_{y_0} \\ \Omega_{y_0} & \Omega_{y_0} \end{bmatrix}, & \omega_{\varepsilon_0} &= \text{diag}(\Omega_{\varepsilon_0})^{1/2}, \quad \omega_{\varepsilon_1} = \text{diag}(\Omega_{\varepsilon_1})^{1/2}, \quad \omega_\beta = \text{diag}(\Omega_\beta)^{1/2}, \\ & \omega_{y_0} &= \text{diag}(\Omega_{y_0})^{1/2}, \quad \omega_{y_1} = \text{diag}(\Omega_{y_1})^{1/2}, \end{aligned} \tag{27}$$

where $X_1 \in \mathbb{R}^{n_1 \times p}$ and $X_0 \in \mathbb{R}^{n_0 \times p}$ denote the two design matrices associated with the two sub-vectors \bar{y}_1 and \bar{y}_0 of $\bar{y} = (\bar{y}_1^T, \bar{y}_0^T)^T$ in (25), which in turn correspond to the partition $y = (y_1^T, y_0^T)^T$. Moreover

- (a) Prior distribution: $\beta \sim \text{SUE}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau, \bar{\Gamma}, g^{(p+q)})$.
- (b) Likelihood: Let $\eta_{-y_0} := (\beta^T, y_1^T)^T$, then $(y | \beta)$ is a multivariate random vector whose density is equal to

$$p(y | \beta) = p(y_1, y_0 | \beta) = p(\bar{y}_1 = y_1 | \beta) \cdot \mathbb{P}(\bar{y}_0 \leq \mathbf{0} | \bar{y}_1 = y_1, \beta),$$

where $p(\bar{y}_1 = y_1 | \beta)$ is the density of the $\text{SUE}_{n_1,q}(\xi_{y_1|\beta}, \Omega_{y_1|\beta}, \Delta_{y_1|\beta}, \tau_{y_1|\beta}, \bar{\Gamma}_{y_1|\beta}, g_{Q_\beta(\beta)}^{(n_1+q)})$, having parameters

$$\begin{aligned} \xi_{y_1|\beta} &= \xi_{y_1} + \Omega_{y_1\beta} \Omega_\beta^{-1} (\beta - \xi_\beta), & \Omega_{y_1|\beta} &= \Omega_{y_1} - \Omega_{y_1\beta} \Omega_\beta^{-1} \Omega_{\beta y_1}, & \Delta_{y_1|\beta} &= \omega_{y_1|\beta}^{-1} (\omega_{y_1} \Delta_{y_1} - \Omega_{y_1\beta} \Omega_\beta^{-1} \omega_\beta \Delta_\beta) \gamma_{y_1|\beta}^{-1}, \\ \tau_{y_1|\beta} &= \gamma_{y_1|\beta}^{-1} [\tau + \Delta_\beta^T \Omega_\beta^{-1} \omega_\beta^{-1} (\beta - \xi_\beta)], & \bar{\Gamma}_{y_1|\beta} &= \gamma_{y_1|\beta}^{-1} (\bar{\Gamma} - \Delta_\beta^T \Omega_\beta^{-1} \Delta_\beta) \gamma_{y_1|\beta}^{-1}, & Q_\beta(\beta) &= (\beta - \xi_\beta)^T \Omega_\beta^{-1} (\beta - \xi_\beta), \end{aligned}$$

with $\omega_{y_1|\beta} = \text{diag}(\Omega_{y_1|\beta})^{1/2}$ and $\gamma_{y_1|\beta} = \text{diag}(\bar{\Gamma} - \Delta_\beta^T \Omega_\beta^{-1} \Delta_\beta)^{1/2}$, whereas $\mathbb{P}(\bar{y}_0 \leq \mathbf{0} | \bar{y}_1 = y_1, \beta)$ corresponds to the cumulative distribution function, evaluated at $\mathbf{0}$, of the $\text{SUE}_{n_0,q}(\xi_{y_0|\beta}, \Omega_{y_0|\beta}, \Delta_{y_0|\beta}, \tau_{y_0|\beta}, \bar{\Gamma}_{y_0|\beta}, g_{Q_{\eta_{-y_0}}}^{(n_0+q)})$ with

$$\begin{aligned} \xi_{y_0|\beta} &= \xi_{y_0} + \Omega_{y_0\beta} \Omega_{y_0}^{-1} (\eta_{-y_0} - \xi_{-y_0}), & \Omega_{y_0|\beta} &= \Omega_{y_0} - \Omega_{y_0\beta} \Omega_{y_0}^{-1} \Omega_{y_0\beta}, & \Delta_{y_0|\beta} &= \omega_{y_0|\beta}^{-1} (\omega_{y_0} \Delta_{y_0} - \Omega_{y_0\beta} \Omega_{y_0}^{-1} \omega_{-y_0} \Delta_{-y_0}) \gamma_{y_0|\beta}^{-1}, \\ \tau_{y_0|\beta} &= \gamma_{y_0|\beta}^{-1} [\tau + \Delta_{-y_0}^T \Omega_{-y_0}^{-1} \omega_{-y_0}^{-1} (\eta_{-y_0} - \xi_{-y_0})], & \bar{\Gamma}_{y_0|\beta} &= \gamma_{y_0|\beta}^{-1} (\bar{\Gamma} - \Delta_{-y_0}^T \Omega_{-y_0}^{-1} \Delta_{-y_0}) \gamma_{y_0|\beta}^{-1}, \end{aligned}$$

where $Q_{\eta_{-y_0}}(\eta_{-y_0}) = (\eta_{-y_0} - \xi_{-y_0})^T \Omega_{-y_0}^{-1} (\eta_{-y_0} - \xi_{-y_0})$, $\omega_{y_0|\beta} = \text{diag}(\Omega_{y_0|\beta})^{1/2}$ and $\gamma_{y_0|\beta} = \text{diag}(\bar{\Gamma} - \Delta_{-y_0}^T \Omega_{-y_0}^{-1} \Delta_{-y_0})^{1/2}$.

(c) Posterior distribution: $(\beta | y) \sim \text{SUE}_{p,n_0+q}(\xi_{\beta|y}, \Omega_{\beta|y}, \Delta_{\beta|y}, \tau_{\beta|y}, \bar{\Gamma}_{\beta|y}, g_{Q_{y_1}(y_1)}^{(p+n_0+q)})$, with the following parameters

$$\begin{aligned} \xi_{\beta|y} &= \xi_{\beta} + \Omega_{\beta y_1} \Omega_{y_1}^{-1} (y_1 - \xi_{y_1}), & \Omega_{\beta|y} &= \Omega_{\beta} - \Omega_{\beta y_1} \Omega_{y_1}^{-1} \Omega_{y_1 \beta}, & \omega_{\beta|y} &= \text{diag}(\Omega_{\beta|y})^{1/2}, \\ \Delta_{\beta|y} &= \omega_{\beta|y}^{-1} \begin{bmatrix} -\Omega_{\beta y_0} + \Omega_{\beta y_1} \Omega_{y_1}^{-1} \Omega_{y_1 y_0} & \omega_{\beta} \Delta_{\beta} - \Omega_{\beta y_1} \Omega_{y_1}^{-1} \omega_{y_1} \Delta_{y_1} \end{bmatrix} \gamma_{\beta|y}^{-1}, & Q_{y_1}(y_1) &= (y_1 - \xi_{y_1})^{\top} \Omega_{y_1}^{-1} (y_1 - \xi_{y_1}), \\ \tau_{\beta|y} &= \gamma_{\beta|y}^{-1} [(-\xi_{y_0} - \Omega_{y_0 y_1} \Omega_{y_1}^{-1} (y_1 - \xi_{y_1}))^{\top} \quad (\tau + \Delta_{y_1}^{\top} \bar{\Omega}_{y_1}^{-1} \omega_{y_1}^{-1} (y_1 - \xi_{y_1}))^{\top}]^{\top}, \\ \bar{\Gamma}_{\beta|y} &= \gamma_{\beta|y}^{-1} \begin{bmatrix} \Omega_{y_0} - \Omega_{y_0 y_1} \Omega_{y_1}^{-1} \Omega_{y_1 y_0} & -\omega_{y_0} \Delta_{y_0} + \Omega_{y_0 y_1} \Omega_{y_1}^{-1} \omega_{y_1} \Delta_{y_1} \\ (-\omega_{y_0} \Delta_{y_0} + \Omega_{y_0 y_1} \Omega_{y_1}^{-1} \omega_{y_1} \Delta_{y_1})^{\top} & \bar{\Gamma} - \Delta_{y_1}^{\top} \bar{\Omega}_{y_1}^{-1} \Delta_{y_1} \end{bmatrix} \gamma_{\beta|y}^{-1}, \end{aligned}$$

where $\gamma_{\beta|y}$ is a block diagonal matrix with blocks $\text{diag}(\Omega_{y_0} - \Omega_{y_0 y_1} \Omega_{y_1}^{-1} \Omega_{y_1 y_0})^{1/2}$ and $\text{diag}(\bar{\Gamma} - \Delta_{y_1}^{\top} \bar{\Omega}_{y_1}^{-1} \Delta_{y_1})^{1/2}$.

Proof. The proof of (27) follows directly from the closure under linear combinations of SUE derived in Lemma 1, after noticing that $(\beta^{\top}, \bar{y}^{\top})^{\top} = (\beta^{\top}, \bar{y}_1^{\top}, \bar{y}_0^{\top})^{\top} = \mathbf{A}(\beta^{\top}, \varepsilon^{\top})^{\top} = \mathbf{A}(\beta^{\top}, \varepsilon_1^{\top}, \varepsilon_0^{\top})^{\top}$ with \mathbf{A} a block matrix having row blocks $\mathbf{A}_1 = [\mathbf{I}_p \ \mathbf{0} \ \mathbf{0}]$, $\mathbf{A}_2 = [\mathbf{X}_1 \ \mathbf{I}_{n_1} \ \mathbf{0}]$ and $\mathbf{A}_3 = [\mathbf{X}_0 \ \mathbf{0} \ \mathbf{I}_{n_0}]$. Under (27), the SUE prior distribution for β in Proposition 3 is a direct consequence of the closure under marginalization stated in Lemma 1 for the SUE class.

As for the likelihood, first notice that $p(y | \beta) = p(y_1, y_0 | \beta) = p(y_1 | \beta)p(y_0 | y_1, \beta)$. Under model (25), $p(y_1 | \beta)$ is equal to $p(\bar{y}_1 = y_1 | \beta)$, which in turn coincides with the density, evaluated at y_1 , of $(\bar{y}_1 | \beta)$. Therefore, by applying to the SUE random vector $(\beta^{\top}, \bar{y}^{\top})^{\top}$ – with parameters as in (27) – the closure under marginalization and conditioning presented in Lemmas 1–2, it directly follows that $(\bar{y}_1 | \beta)$ is a SUE having parameters as in Proposition 3. For what concerns the second term $p(y_0 | y_1, \beta)$, notice that, under model (25), $p(y_0 | y_1, \beta) = \mathbb{P}(\bar{y}_0 \leq \mathbf{0} | \bar{y}_1 = y_1, \beta)$, where $(\bar{y}_0 | \bar{y}_1 = y_1, \beta)$ is, again, a SUE whose parameters are defined in Proposition 3. Such a latter result follows directly from the closure under linear combinations and conditioning properties in Lemmas 1–2, applied to the SUE random vector $(\beta^{\top}, \bar{y}^{\top})^{\top} = (\beta^{\top}, \bar{y}_1^{\top}, \bar{y}_0^{\top})^{\top}$ partitioned as $(\eta_{-y_0}^{\top}, \bar{y}_0^{\top})^{\top}$.

To conclude the proof, notice that

$$\mathbb{P}(\beta \leq \mathbf{b} | y) = \mathbb{P}(\beta \leq \mathbf{b} | \bar{y}_1 = y_1, \bar{y}_0 \leq \mathbf{0}) = \mathbb{P}(\beta \leq \mathbf{b}, \bar{y}_0 \leq \mathbf{0} | \bar{y}_1 = y_1) / \mathbb{P}(\bar{y}_0 \leq \mathbf{0} | \bar{y}_1 = y_1).$$

By Lemma 2, the numerator in the above expression coincides with the cumulative distribution function, evaluated at $(\mathbf{b}^{\top}, \mathbf{0}^{\top})^{\top}$, of the random vector having SUE $_{p+n_0,q}(\xi_{nu}, \Omega_{nu}, \Delta_{nu}, \tau_{nu}, \bar{\Gamma}_{nu}, g_{Q_{y_1}(y_1)}^{(p+n_0+q)})$ distribution, with parameters

$$\begin{aligned} \xi_{nu} &= \begin{bmatrix} \xi_{\beta} + \Omega_{\beta y_1} \Omega_{y_1}^{-1} (y_1 - \xi_{y_1}) \\ \xi_{y_0} + \Omega_{y_0 y_1} \Omega_{y_1}^{-1} (y_1 - \xi_{y_1}) \end{bmatrix} =: \begin{bmatrix} \xi_{nu\beta} \\ \xi_{nu y_0} \end{bmatrix}, & \Omega_{nu} &= \begin{bmatrix} \Omega_{\beta} - \Omega_{\beta y_1} \Omega_{y_1}^{-1} \Omega_{y_1 \beta} & \Omega_{\beta y_0} - \Omega_{\beta y_1} \Omega_{y_1}^{-1} \Omega_{y_1 y_0} \\ \Omega_{y_0 \beta} - \Omega_{y_0 y_1} \Omega_{y_1}^{-1} \Omega_{y_1 \beta} & \Omega_{y_0} - \Omega_{y_0 y_1} \Omega_{y_1}^{-1} \Omega_{y_1 y_0} \end{bmatrix} =: \begin{bmatrix} \Omega_{nu\beta} & \Omega_{nu\beta y_0} \\ \Omega_{nu y_0 \beta} & \Omega_{nu y_0} \end{bmatrix}, \\ \Delta_{nu} &= \begin{bmatrix} \omega_{nu\beta}^{-1} (\omega_{\beta} \Delta_{\beta} - \Omega_{\beta y_1} \Omega_{y_1}^{-1} \omega_{y_1} \Delta_{y_1}) \gamma_{nu}^{-1} \\ \omega_{nu y_0}^{-1} (\omega_{y_0} \Delta_{y_0} - \Omega_{y_0 y_1} \Omega_{y_1}^{-1} \omega_{y_1} \Delta_{y_1}) \gamma_{nu}^{-1} \end{bmatrix} =: \begin{bmatrix} \Delta_{nu\beta} \\ \Delta_{nu y_0} \end{bmatrix}, & \tau_{nu} &= \gamma_{nu}^{-1} [\tau + \Delta_{y_1}^{\top} \bar{\Omega}_{y_1}^{-1} \omega_{y_1}^{-1} (y_1 - \xi_{y_1})], \\ & \bar{\Gamma}_{nu} &= \gamma_{nu}^{-1} (\bar{\Gamma} - \Delta_{y_1}^{\top} \bar{\Omega}_{y_1}^{-1} \Delta_{y_1}) \gamma_{nu}^{-1}, \end{aligned}$$

where $Q_{y_1}(y_1) = (y_1 - \xi_{y_1})^{\top} \Omega_{y_1}^{-1} (y_1 - \xi_{y_1})$, $\omega_{nu\beta} = \text{diag}(\Omega_{nu\beta})^{1/2}$, $\omega_{nu y_0} = \text{diag}(\Omega_{nu y_0})^{1/2}$ and $\gamma_{nu} = \text{diag}(\bar{\Gamma} - \Delta_{y_1}^{\top} \bar{\Omega}_{y_1}^{-1} \Delta_{y_1})^{1/2}$. Similarly, the denominator in the expression for $\mathbb{P}(\beta \leq \mathbf{b} | y)$ coincides with the cumulative distribution function, evaluated at $\mathbf{0}$, of a SUE random vector. By the closure under marginalization of SUE variables, the distribution of this vector can be directly derived from the one above to obtain a SUE $_{n_0,q}(\xi_{de}, \Omega_{de}, \Delta_{de}, \tau_{de}, \bar{\Gamma}_{de}, g_{Q_{y_1}(y_1)}^{(n_0+q)})$ with parameters

$$\xi_{de} = \xi_{nu y_0}, \quad \Omega_{de} = \Omega_{nu y_0}, \quad \Delta_{de} = \Delta_{nu y_0}, \quad \tau_{de} = \tau_{nu}, \quad \bar{\Gamma}_{de} = \bar{\Gamma}_{nu}, \quad Q_{y_1}(y_1) = (y_1 - \xi_{y_1})^{\top} \Omega_{y_1}^{-1} (y_1 - \xi_{y_1}).$$

Combining the above results and recalling the expression for the SUE cumulative distribution function in (4), we have

$$\begin{aligned} \mathbb{P}(\beta \leq \mathbf{b} | y) &= \frac{F_{p+n_0+q} \left(\begin{bmatrix} \mathbf{b} - \xi_{nu\beta} \\ -\xi_{nu y_0} \\ \tau_{nu} \end{bmatrix}; \begin{bmatrix} \Omega_{nu\beta} & \Omega_{nu\beta y_0} & -\omega_{nu\beta} \Delta_{nu\beta} \\ \Omega_{nu y_0 \beta} & \Omega_{nu y_0} & -\omega_{nu y_0} \Delta_{nu y_0} \\ -\Delta_{nu\beta}^{\top} \omega_{nu\beta} & -\Delta_{nu y_0}^{\top} \omega_{nu y_0} & \bar{\Gamma}_{nu} \end{bmatrix}; g_{Q_{y_1}(y_1)}^{(p+n_0+q)} \right) F_q \left(\tau_{nu}; \bar{\Gamma}_{nu}; g_{Q_{y_1}(y_1)}^{(q)} \right)}{F_q \left(\tau_{nu}; \bar{\Gamma}_{nu}; g_{Q_{y_1}(y_1)}^{(q)} \right) F_{n_0+q} \left(\begin{bmatrix} -\xi_{nu y_0} \\ \tau_{nu} \end{bmatrix}; \begin{bmatrix} \Omega_{nu y_0} & -\omega_{nu y_0} \Delta_{nu y_0} \\ -\Delta_{nu y_0}^{\top} \omega_{nu y_0} & \bar{\Gamma}_{nu} \end{bmatrix}; g_{Q_{y_1}(y_1)}^{(n_0+q)} \right)} \\ &= \frac{F_{p+n_0+q} \left(\begin{bmatrix} \mathbf{b} - \xi_{\beta|y} \\ \tau_{\beta|y} \end{bmatrix}; \begin{bmatrix} \Omega_{\beta|y} & -\omega_{\beta|y} \Delta_{\beta|y} \\ -\Delta_{\beta|y}^{\top} \omega_{\beta|y} & \bar{\Gamma}_{\beta|y} \end{bmatrix}; g_{Q_{y_1}(y_1)}^{(p+n_0+q)} \right)}{F_{n_0+q} \left(\tau_{\beta|y}; \bar{\Gamma}_{\beta|y}; g_{Q_{y_1}(y_1)}^{(n_0+q)} \right)}, \end{aligned}$$

which coincides with the cumulative distribution function of the SUE $_{p,n_0+q}(\xi_{\beta|y}, \Omega_{\beta|y}, \Delta_{\beta|y}, \tau_{\beta|y}, \bar{\Gamma}_{\beta|y}, g_{Q_{y_1}(y_1)}^{(p+n_0+q)})$ posterior for β whose parameters, after suitable standardizations based on Lemma 4, are defined as in Proposition 3. \square

Proposition 3 states a general result that establishes SUE conjugacy for a broad class of models whose likelihood factorizes as the product of multivariate elliptical densities and cumulative distribution functions. These likelihoods substantially extend classical tobit representations to multivariate and skewed contexts while covering a broader family of noise terms beyond the Gaussian ones.

As clarified in [Examples 5–6](#), albeit general, such a result allows to recover Bayesian formulations of potential interest in practice while ensuring conjugacy under these representations.

Example 5 (SUN Conjugacy). Classical tobit models consider $y_i = \bar{y}_i 1(\bar{y}_i > 0)$ with $(\bar{y}_i | \beta) \sim N(x_i^\top \beta, \sigma^2)$, independently for every $i \in \{1, \dots, n\}$. A natural extension that incorporates skewness within these representations replaces $N(x_i^\top \beta, \sigma^2)$ with $SN(x_i^\top \beta, \sigma^2, \alpha)$ [[16](#)]. Under this setting, which includes the classical and routinely-implemented tobit formulation when $\alpha = 0$, [Anceschi et al. \[8\]](#) have shown that SUN priors for β , i.e., $\beta \sim \text{SUN}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau_\beta, \bar{\Gamma}_\beta)$, yield posterior distributions within the same class. Such a result can be derived as a very special case of [Proposition 3](#) under a Gaussian density generator and suitable constraints on the parameters.

To clarify the above point, assume again $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUN}_{p+n,q+n}(\xi, \Omega, \Delta, \tau, \bar{\Gamma})$ with parameters partitioned as in [Example 1](#). In addition, consider the partitioning $(y_1^\top, y_0^\top)^\top$ defined in [Proposition 3](#) for a generic realization y from model [\(25\)](#). This construction, combined with the results in [\(27\)](#) and the proof of [Proposition 3](#), implies that $(\beta^\top, \bar{y}^\top)^\top \sim \text{SUN}_{p+n_1+n_0,q+n_1+n_0}(\xi^\dagger, \Omega^\dagger, \Delta^\dagger, \tau, \bar{\Gamma})$, with

$$\xi^\dagger = \begin{bmatrix} \xi_\beta \\ X_1 \xi_\beta \\ X_0 \xi_\beta \end{bmatrix}, \quad \Delta^\dagger = \begin{bmatrix} \Delta_\beta & \mathbf{0} & \mathbf{0} \\ \omega_{y_1}^{-1} X_1 \omega_\beta \Delta_\beta & \omega_{y_1}^{-1} \sigma \bar{\alpha} \mathbf{I}_{n_1} & \mathbf{0} \\ \omega_{y_0}^{-1} X_0 \omega_\beta \Delta_\beta & \mathbf{0} & \omega_{y_0}^{-1} \sigma \bar{\alpha} \mathbf{I}_{n_0} \end{bmatrix}, \quad \Omega^\dagger = \begin{bmatrix} \Omega_\beta & \Omega_\beta X_1^\top & \Omega_\beta X_0^\top \\ X_1 \Omega_\beta & X_1 \Omega_\beta X_1^\top + \sigma^2 \mathbf{I}_{n_1} & X_1 \Omega_\beta X_0^\top \\ X_0 \Omega_\beta & X_0 \Omega_\beta X_1^\top & X_0 \Omega_\beta X_0^\top + \sigma^2 \mathbf{I}_{n_0} \end{bmatrix},$$

where $\bar{\alpha} = \alpha/(1 + \alpha^2)^{1/2}$, $\omega_\beta = \text{diag}(\Omega_\beta)^{1/2}$, $\omega_{y_1} = \text{diag}(X_1 \Omega_\beta X_1^\top + \sigma^2 \mathbf{I}_{n_1})^{1/2}$ and $\omega_{y_0} = \text{diag}(X_0 \Omega_\beta X_0^\top + \sigma^2 \mathbf{I}_{n_0})^{1/2}$. These results, combined with the closure properties of SUNs and point (ii) in [Lemma 5](#), yield

$$\beta \sim \text{SUN}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau_\beta, \bar{\Gamma}_\beta), \quad \varepsilon \sim \text{SUN}_{n,n}(\mathbf{0}, \sigma^2 \mathbf{I}_n, \bar{\alpha} \mathbf{I}_n, \mathbf{0}, \mathbf{I}_n),$$

that coincide with the SUN prior and skew-normal noise vector for the extension of the tobit model analyzed in the supplementary materials of [Anceschi et al. \[8\]](#). In addition, leveraging [Lemma 2](#) and [Remark 2](#), we have $(\bar{y}_1 | \beta) \sim \text{SUN}_{n_1,n_1}(X_1 \beta, \sigma^2 \mathbf{I}_{n_1}, \bar{\alpha} \mathbf{I}_{n_1}, \mathbf{0}, \mathbf{I}_{n_1})$. Similarly, by the properties of the Gaussian density generators, it follows that under the above constraints for the SUN parameters, $(\bar{y}_0 \perp \bar{y}_1 | \beta)$. Therefore, $p(\bar{y}_0 | \bar{y}_1 = y_1, \beta) = p(\bar{y}_0 | \beta)$, and hence, by the same derivations that led to the SUN for $(\bar{y}_1 | \beta)$, we obtain $(\bar{y}_0 | \beta) \sim \text{SUN}_{n_0,n_0}(X_0 \beta, \sigma^2 \mathbf{I}_{n_0}, \bar{\alpha} \mathbf{I}_{n_0}, \mathbf{0}, \mathbf{I}_{n_0})$. Combining these results with [Proposition 3](#), [Lemma 2](#) and [Remark 2](#), and recalling the expression for the SUN density and cumulative distribution function (see e.g., [Arellano-Valle and Azzalini \[4\]](#)), leads to

$$p(y_1, y_0 | \beta) \propto \phi_{n_1}(y_1 - X_1 \beta; \sigma^2 \mathbf{I}_{n_1}) \Phi_{n_1+2n_0} \left(\begin{bmatrix} \alpha(y_1 - X_1 \beta) \\ -X_0 \beta \\ \mathbf{0} \end{bmatrix}; \begin{bmatrix} \sigma^2 \mathbf{I}_{n_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{n_0} & -\bar{\alpha} \sigma \mathbf{I}_{n_0} \\ \mathbf{0} & -\bar{\alpha} \sigma \mathbf{I}_{n_0} & \mathbf{I}_{n_0} \end{bmatrix} \right),$$

which coincides again with the likelihood in [Anceschi et al. \[8\]](#). As for the models explored within [Examples 1](#) and [3](#), the above representation also includes several formulations of direct interest in practice. In particular, setting $\alpha = 0$ yields classical tobit regression, whereas replacing $\sigma^2 \mathbf{I}_n$ with a full covariance matrix allows to recover multivariate extensions of tobit models, including those based on skewed link functions.

[Example 6](#) concludes our analysis by clarifying that similar, but yet-unexplored, conjugacy properties can be established also when the focus is on models for Student’s t or skew- t censored observations.

Example 6 (SUT Conjugacy). Conjugacy properties for generalizations of tobit models relying on Student’s t or skew- t censored observations are currently lacking. As stated in [Corollary 3](#), these properties can be derived as special cases of [Proposition 3](#) under Student’s t density generators.

Corollary 3. Consider model [\(25\)](#), with $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUT}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$, and parameters ξ, Ω, Δ partitioned as in [\(26\)](#). Then, the induced prior distribution is $\beta \sim \text{SUT}_{p,q}(\xi_\beta, \Omega_\beta, \Delta_\beta, \tau, \bar{\Gamma}, \nu)$, whereas the likelihood is equal to

$$p(y | \beta) = p(y_1, y_0 | \beta) = p(\bar{y}_1 = y_1 | \beta) \cdot \mathbb{P}(\bar{y}_0 \leq \mathbf{0} | \bar{y}_1 = y_1, \beta),$$

with $p(\bar{y}_1 = y_1 | \beta)$ the density of the $\text{SUT}_{n_1,q}(\xi_{y_1|\beta}, \alpha_\beta \Omega_{y_1|\beta}, \Delta_{y_1|\beta}, \alpha_\beta^{-1/2} \tau_{y_1|\beta}, \bar{\Gamma}_{y_1|\beta}, \nu + p)$ having parameters as in [Proposition 3](#), and $\mathbb{P}(\bar{y}_0 \leq \mathbf{0} | \bar{y}_1 = y_1, \beta)$ the cumulative distribution function, evaluated at $\mathbf{0}$, of a $\text{SUT}_{n_0,q}(\xi_{y_0|\cdot}, \alpha_{\eta_{-y_0}} \Omega_{y_0|\cdot}, \Delta_{y_0|\cdot}, \alpha_{\eta_{-y_0}}^{-1/2} \tau_{y_0|\cdot}, \bar{\Gamma}_{y_0|\cdot}, \nu + p + n_1)$ with parameters as in [Proposition 3](#). In these expressions $\alpha_\beta = [\nu + Q_\beta(\beta)]/(\nu + p)$ and $\alpha_{\eta_{-y_0}} = [\nu + Q_{\eta_{-y_0}}(\eta_{-y_0})]/(\nu + n_1 + p)$, with $Q_\beta(\beta)$ and $Q_{\eta_{-y_0}}(\eta_{-y_0})$ defined again in [Proposition 3](#). Finally, the posterior distribution for β is $(\beta | y) \sim \text{SUT}_{p,n_0+q}(\xi_{\beta|y}, \alpha_{y_1} \Omega_{\beta|y}, \Delta_{\beta|y}, \alpha_{y_1}^{-1/2} \tau_{\beta|y}, \bar{\Gamma}_{\beta|y}, \nu + n_1)$ with $\alpha_{y_1} = [\nu + Q_{y_1}(y_1)]/(\nu + n_1)$, and the remaining quantities defined as in [Proposition 3](#).

Proof. The proof of [Corollary 3](#) requires replacing the generic density generators in [Proposition 3](#) with those of the Student’s t , and then leveraging the properties of such generators described in [Section 2.2.2](#). \square

Let us conclude with some special cases of [Corollary 3](#) that yield priors and likelihoods of potential interest in practice. To this end, similarly to [Examples 2](#) and [4](#), consider $(\beta^\top, \varepsilon^\top)^\top \sim \text{SUT}_{p+n,q}(\xi, \Omega, \Delta, \tau, \bar{\Gamma}, \nu)$, with $n = n_1 + n_0$ and parameters partitioned as

$$\xi = \begin{bmatrix} \xi_\beta \\ \xi_\varepsilon \end{bmatrix} = \begin{bmatrix} \xi_\beta \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_\beta & \mathbf{0} \\ \mathbf{0} & \Omega_\varepsilon \end{bmatrix} = \begin{bmatrix} \Omega_\beta & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{\varepsilon_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Omega_{\varepsilon_0} \end{bmatrix}, \quad \Delta = \begin{bmatrix} \mathbf{0} \\ \Delta_\varepsilon \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \Delta_{\varepsilon_1} \\ \Delta_{\varepsilon_0} \end{bmatrix},$$

and $\tau = \mathbf{0}$.

Recalling [Examples 2](#) and [4](#), the above construction implies

$$\beta \sim \mathcal{F}_p(\xi_\beta, \Omega_\beta, \nu), \quad \varepsilon \sim \text{SUT}_{n,q}(\mathbf{0}, \Omega_\varepsilon, \mathbf{A}_\varepsilon, \mathbf{0}, \bar{\Gamma}, \nu),$$

and therefore, under [\(25\)](#), the model underlying a generic observation $\mathbf{y} = (\mathbf{y}_1^\top, \mathbf{y}_0^\top)^\top$ coincides with a multivariate extension of tobit regression having unified skew- t error terms, and Student's t prior for β uncorrelated with the noise vector ε . Applying [Corollary 3](#) to such a formulation yields the likelihood

$$\rho(\mathbf{y} \mid \beta) = \rho(\mathbf{y}_1, \mathbf{y}_0 \mid \beta) = \rho(\bar{\mathbf{y}}_1 = \mathbf{y}_1 \mid \beta) \cdot \mathbb{P}(\bar{\mathbf{y}}_0 \leq \mathbf{0} \mid \bar{\mathbf{y}}_1 = \mathbf{y}_1, \beta),$$

where $\rho(\bar{\mathbf{y}}_1 = \mathbf{y}_1 \mid \beta)$ coincides with the density function, computed at \mathbf{y}_1 , of the $\text{SUT}_{n_1,q}(\mathbf{X}_1\beta, \alpha_\beta\Omega_{\varepsilon_1}, \mathbf{A}_{\varepsilon_1}, \mathbf{0}, \bar{\Gamma}, \nu + p)$ random variable, whereas the quantity $\mathbb{P}(\bar{\mathbf{y}}_0 \leq \mathbf{0} \mid \bar{\mathbf{y}}_1 = \mathbf{y}_1, \beta)$ corresponds to the cumulative distribution function, evaluated at $\mathbf{0}$, of the $\text{SUT}_{n_0,q}(\mathbf{X}_0\beta, \alpha_{\eta_{-y_0}}\Omega_{\varepsilon_0}, \mathbf{A}_{\varepsilon_0}\gamma_0^{-1}, \alpha_{\eta_{-y_0}}^{-1/2}\gamma_0^{-1}\mathbf{A}_{\varepsilon_1}^\top\bar{\Omega}_{\varepsilon_1}^{-1}\omega_{\varepsilon_1}^{-1}(\mathbf{y}_1 - \mathbf{X}_1\beta), \gamma_0^{-1}(\bar{\Gamma} - \mathbf{A}_{\varepsilon_1}^\top\bar{\Omega}_{\varepsilon_1}^{-1}\mathbf{A}_{\varepsilon_1})\gamma_0^{-1}, \nu + p + n_1)$, with $\gamma_0 = \text{diag}(\bar{\Gamma} - \mathbf{A}_{\varepsilon_1}^\top\bar{\Omega}_{\varepsilon_1}^{-1}\mathbf{A}_{\varepsilon_1})^{1/2}$. This result clarifies that classical multivariate tobit representations admit extensions to suitable skew- t formulations while preserving conjugacy. Imposing additional constraints within such a formulation further highlights the practical potential of our contribution. For example, setting $\mathbf{A}_\varepsilon = \mathbf{0}$ in the above formulation, and recalling again [Examples 2](#) and [4](#), yields

$$\beta \sim \mathcal{F}_p(\xi_\beta, \Omega_\beta, \nu), \quad \rho(\mathbf{y} \mid \beta) = \rho(\bar{\mathbf{y}}_1 = \mathbf{y}_1 \mid \beta) \cdot \mathbb{P}(\bar{\mathbf{y}}_0 \leq \mathbf{0} \mid \bar{\mathbf{y}}_1 = \mathbf{y}_1, \beta),$$

where $\rho(\bar{\mathbf{y}}_1 = \mathbf{y}_1 \mid \beta)$ is the density of $\mathcal{F}_{n_1}(\mathbf{X}_1\beta, \alpha_\beta\Omega_{\varepsilon_1}, \nu + p)$ evaluated at \mathbf{y}_1 , while $\mathbb{P}(\bar{\mathbf{y}}_0 \leq \mathbf{0} \mid \bar{\mathbf{y}}_1 = \mathbf{y}_1, \beta)$ is a $\mathcal{F}_{n_0}(\mathbf{X}_0\beta, \alpha_{\eta_{-y_0}}\Omega_{\varepsilon_0}, \nu + p + n_1)$ cumulative distribution function computed at $\mathbf{0}$. As a consequence of [Corollary 3](#), the induced posterior distribution for β is still within the SUT family.

4. Conclusions

This article proves that SUE distributions have important conjugacy properties when combined with broad classes of likelihoods that generalize classical probit, tobit, multinomial probit, and linear models in several directions. These generalizations include multivariate formulations based on general elliptical noise terms and allow for asymmetric representations relying on unified skew-elliptical extensions. Our results leverage available and newly-derived closure properties of the SUE family to prove that priors within this class yield again SUE posterior distributions when combined with the likelihood of the models mentioned above, under the classical Bayes rule. Recalling [Propositions 1–3](#), these results are technically derived by starting from a joint SUE distribution for the parameters and the observed data. Such a proof technique is not meant to provide a different perspective on the standard specification of a prior and a likelihood in Bayesian statistics. Rather, it provides a convenient strategy that facilitates the derivation, within the SUE class, of meaningful priors and likelihoods yielding closed-form SUE posterior distributions.

More specifically, [Examples 1–6](#) clarify that our results include models of direct interest in practice, such as those based on multivariate Gaussian or Student's t formulations, along with the corresponding skewed extensions. In this respect, an interesting direction would be to specialize [Propositions 1–3](#) to other SUE sub-families, e.g., those based on Cauchy or logistic density generators. This goal can be accomplished by replacing the generic density generators in [Propositions 1–3](#), with those yielding the sub-family investigated. These extensions further motivate advancements in the study of other relevant SUE sub-families to derive results and properties similar to those characterizing SUN [e.g., [4](#)] and SUT [e.g., [48](#)] distributions. Particularly impactful, within our context, would be the derivation of additive stochastic representations as those obtained for SUNs and SUTs. Advancements along these lines would facilitate i.i.d. sampling under any SUE posterior, thus enlarging the class of models and priors that allow for tractable Bayesian inference. The recent additive stochastic representations derived by Yin and Balakrishnan [[56](#)] for general skew-elliptical distributions provide a promising advancement in this direction, which also suggests that related results could be derived even for the wider SUE family. Similarly, expanding the available strategies for the efficient evaluation of the moments of SUE distributions in, e.g., [\(6\)–\(7\)](#), would further facilitate Bayesian inference leveraging the conjugacy results derived in the present article. Current contributions [e.g., [50–54](#)] provide important results along these lines which motivate future research to showcase the computational advantages and the practical impact of these solutions when the focus is on Bayesian inference under the newly-derived SUE posterior distributions.

Finally, we shall emphasize that the SUE family can be itself rephrased as a particular case of selection elliptical distributions [[49](#)] arising from even more general conditioning mechanisms. As such, it would be interesting to expand the conjugacy properties derived in this article for SUE distributions to the broader selection elliptical family. Let us conclude by highlighting that not all the priors and likelihoods implied by the general results in [Propositions 1–3](#) have direct practical applicability. Nonetheless, from a theoretical perspective, also these instances are of interest in expanding the analysis of the probabilistic properties of the SUE family. Moreover, although conjugacy is a desirable property, it is important to emphasize that those priors and likelihoods in the SUE family that do not yield SUE posteriors can still allow for Bayesian inference leveraging, e.g., MCMC methods and deterministic approximations.

Acknowledgments

The authors are grateful to the Editor, Associate Editor and referees for the useful and constructive comments on the initial version of such a manuscript. Maicon J. Karling and Marc G. Genton acknowledge the support by the King Abdullah University of Science and Technology (KAUST) in Saudi Arabia.

References

- [1] D. Durante, Conjugate Bayes for probit regression via unified skew-normal distributions, *Biometrika* 106 (2019) 765–779.
- [2] R.B. Arellano-Valle, A. Azzalini, On the unification of families of skew-normal distributions, *Scand. J. Stat.* 33 (2006) 561–574.
- [3] A. Azzalini, A. Capitanio, *The Skew-Normal and Related Families*, Cambridge University Press, Cambridge, 2014.
- [4] R.B. Arellano-Valle, A. Azzalini, Some properties of the unified skew-normal distribution, *Statist. Papers* 63 (2022) 461–487.
- [5] A. Fasano, G. Rebaudo, D. Durante, S. Petrone, A closed-form filter for binary time series, *Stat. Comput.* 31 (2021) 47.
- [6] A. Fasano, D. Durante, A class of conjugate priors for multinomial probit models which includes the multivariate normal one, *J. Mach. Learn. Res.* 23 (2022) 1358–1383.
- [7] J. Cao, D. Durante, M.G. Genton, Scalable computation of predictive probabilities in probit models with Gaussian process priors, *J. Comput. Graph. Statist.* 31 (2022) 709–720.
- [8] N. Anceschi, A. Fasano, D. Durante, G. Zanella, Bayesian conjugacy in probit, tobit, multinomial probit and extensions: A review and new results, *J. Amer. Statist. Assoc.* 118 (2023) 1451–1469.
- [9] M.-H. Chen, D.K. Dey, Q.-M. Shao, A new skewed link model for dichotomous quantal response data, *J. Amer. Statist. Assoc.* 94 (1999) 1172–1186.
- [10] S.K. Sahu, D.K. Dey, M.D. Branco, A new class of multivariate skew distributions with applications to Bayesian regression models, *Canad. J. Statist.* 31 (2003) 129–150.
- [11] R.B. Arellano-Valle, S. Ozan, H. Bolfarine, V.H. Lachos, Skew normal measurement error models, *J. Multivariate Anal.* 96 (2005) 265–281.
- [12] H. Bolfarine, V.H. Lachos, Skew-probit measurement error models, *Stat. Methodol.* 4 (2007) 1–12.
- [13] V.H. Lachos, H. Bolfarine, R.B. Arellano-Valle, L.C. Montenegro, Likelihood-based inference for multivariate skew-normal regression models, *Comm. Statist. Theory Methods* 36 (2007) 1769–1786.
- [14] J.L. Bazán, H. Bolfarine, M.D. Branco, A framework for skew-probit links in binary regression, *Comm. Statist. Theory Methods* 39 (2010) 678–697.
- [15] V.H. Lachos, P. Ghosh, R.B. Arellano-Valle, Likelihood based inference for skew-normal independent linear mixed models, *Statist. Sinica* (2010) 303–322.
- [16] J.L. Hutton, E. Stanghellini, Modelling bounded health scores with censored skew-normal distributions, *Stat. Med.* 30 (2011) 368–376.
- [17] A. Benavoli, D. Azzimonti, D. Piga, Skew Gaussian processes for classification, *Mach. Learn.* 109 (2020) 1877–1902.
- [18] C.E. Galarza, L.A. Matos, V.H. Lachos, An EM algorithm for estimating the parameters of the multivariate skew-normal distribution with censored responses, *Metron* 80 (2022) 231–253.
- [19] R.B. Arellano-Valle, M.G. Genton, R.H. Loschi, Shape mixtures of multivariate skew-normal distributions, *J. Multivariate Anal.* 100 (2009) 91–101.
- [20] C.C. Vieira, R.H. Loschi, D. Duarte, Nonparametric mixtures based on skew-normal distributions: An application to density estimation, *Commun. Stat. - Theory Methods* 44 (2015) 1552–1570.
- [21] A. Canale, E.C. Kenne Pagui, B. Scarpa, Bayesian modeling of university first-year students' grades after placement test, *J. Appl. Stat.* 43 (2016) 3015–3029.
- [22] A. Fasano, D. Durante, G. Zanella, Scalable and accurate variational Bayes for high-dimensional binary regression models, *Biometrika* 109 (2022) 901–919.
- [23] P. Onorati, B. Liseo, An extension of the unified skew-normal family of distributions and application to Bayesian binary regression, 2022, arXiv:2209.03474.
- [24] D. Durante, F. Pozza, B. Szabo, Skewed Bernstein-von Mises theorem and skew-modal approximations, *Ann. Statist.* (2024) in press.
- [25] A. Zellner, Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms, *J. Amer. Statist. Assoc.* 71 (1976) 400–405.
- [26] K.L. Lange, R.J.A. Little, J.M.G. Taylor, Robust statistical modeling using the t distribution, *J. Amer. Statist. Assoc.* 84 (1989) 881–896.
- [27] J.H. Albert, S. Chib, Bayesian analysis of binary and polychotomous response data, *J. Amer. Statist. Assoc.* 88 (1993) 669–679.
- [28] R.B. Arellano-Valle, L.M. Castro, G. González-Farías, K.A. Muñoz-Gajardo, Student-t censored regression model: Properties and inference, *Stat. Methods Appl.* 21 (2012) 453–473.
- [29] Y.V. Marchenko, M.G. Genton, A Heckman selection-t model, *J. Amer. Statist. Assoc.* 107 (2012) 304–317.
- [30] G. Dagne, Y. Huang, Bayesian inference for a nonlinear mixed-effects tobit model with multivariate skew-t distributions: Application to AIDS studies, *Int. J. Biostat.* 8 (2012) 27.
- [31] A.M. Garay, H. Bolfarine, V.H. Lachos, C.R.B. Cabral, Bayesian analysis of censored linear regression models with scale mixtures of normal distributions, *J. Appl. Stat.* 42 (2015) 2694–2714.
- [32] L.A. Matos, L.M. Castro, C.R.B. Cabral, V.H. Lachos, Multivariate measurement error models based on Student-t distribution under censored responses, *Statistics* 52 (2018) 1395–1416.
- [33] V.H. Lachos, M.O. Prates, D.K. Dey, Heckman selection-t model: Parameter estimation via the EM-algorithm, *J. Multivariate Anal.* 184 (2021) 104737.
- [34] V.H. Lachos, J.L. Bazán, L.M. Castro, J. Park, The skew-t censored regression model: Parameter estimation via an EM-type algorithm, *Commun. Stat. Appl. Methods* 29 (2022) 333–351.
- [35] K.A.L. Valeriano, C.E. Galarza, L.A. Matos, V.H. Lachos, Likelihood-based inference for the multivariate skew-t regression with censored or missing responses, *J. Multivariate Anal.* 196 (2023) 105174.
- [36] A. Spanos, On modeling heteroskedasticity: The Student's t and elliptical linear regression models, *Econometric Theory* 10 (1994) 286–315.
- [37] M.D. Branco, D.K. Dey, Regression model under skew elliptical error distribution, *J. Math. Sci.* 1 (2002) 151–169.
- [38] M.Q. Islam, F. Yildirim, M. Yazici, Inference in multivariate linear regression models with elliptically distributed errors, *J. Appl. Stat.* 41 (2014) 1746–1766.
- [39] M. Barros, M. Galea, V. Leiva, M. Santos-Neto, Generalized tobit models: Diagnostics and application in econometrics, *J. Appl. Stat.* 45 (2018) 145–167.
- [40] Z. Zhang, R.B. Arellano-Valle, M.G. Genton, R. Huser, Tractable Bayes of skew-elliptical link models for correlated binary data, *Biometrics* 79 (2023) 1788–1800.
- [41] K.-T. Fang, S. Kotz, K.-W. Ng, *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London, 1990.
- [42] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 61 (1999) 579–602.
- [43] M.D. Branco, D.K. Dey, A general class of multivariate skew-elliptical distributions, *J. Multivariate Anal.* 79 (2001) 99–113.
- [44] A. Azzalini, A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 65 (2003) 367–389.
- [45] B.Q. Fang, The skew elliptical distributions and their quadratic forms, *J. Multivariate Anal.* 87 (2003) 298–314.
- [46] R.B. Arellano-Valle, M.G. Genton, Multivariate unified skew-elliptical distributions, *Chil. J. Stat.* 1 (2010) 17–33.
- [47] C. Adcock, A. Azzalini, A selective overview of skew-elliptical and related distributions and of their applications, *Symmetry* 12 (2020) 118.
- [48] K. Wang, M.J. Karling, R.B. Arellano-Valle, M.G. Genton, Multivariate unified skew-t distributions and their properties, *J. Multivariate Anal.* 203 (2024) 105322.
- [49] R.B. Arellano-Valle, M.D. Branco, M.G. Genton, A unified view on skewed distributions arising from selections, *Canad. J. Statist.* 34 (2006) 581–601.
- [50] J.C. Arismendi, S. Broda, Multivariate elliptical truncated moments, *J. Multivariate Anal.* 157 (2017) 29–44.
- [51] R.A. Morán-Vásquez, S.L.P. Ferrari, New results on truncated elliptical distributions, *Commun. Math. Stat.* 9 (2021) 299–313.
- [52] C.E. Galarza, L.A. Matos, L.M. Castro, V.H. Lachos, Moments of the doubly truncated selection elliptical distributions with emphasis on the unified multivariate skew-t distribution, *J. Multivariate Anal.* 189 (2022) 104944.
- [53] C.E. Galarza, L.A. Matos, D.K. Dey, V.H. Lachos, On moments of folded and doubly truncated multivariate extended skew-normal distributions, *J. Comput. Graph. Statist.* 31 (2022) 455–465.
- [54] K.A.L. Valeriano, C.E. Galarza, L.A. Matos, Moments and random number generation for the truncated elliptical family of distributions, *Stat. Comput.* 33 (2023) 32.

- [55] C. Song, S.-T. Xia, Bayesian linear regression with Student-t assumptions, 2016, arXiv:1604.04434.
- [56] C. Yin, N. Balakrishnan, Stochastic representations and probabilistic characteristics of multivariate skew-elliptical distributions, *J. Multivariate Anal.* 199 (2024) 105240.
- [57] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, *Biometrika* 83 (1996) 715–726.
- [58] B.C. Arnold, R.J. Beaver, Skewed multivariate models related to hidden truncation and/or selective reporting, *Test* 11 (2002) 7–54.
- [59] B.C. Arnold, R.J. Beaver, Hidden truncation models, *Sankhyā* 62 (2000) 23–35.
- [60] G. González-Farías, J.A. Domínguez-Molina, A.K. Gupta, Additive properties of skew normal random vectors, *J. Statist. Plann. Inference* 126 (2004) 521–534.
- [61] A.K. Gupta, G. González-Farías, J.A. Domínguez-Molina, A multivariate skew normal distribution, *J. Multivariate Anal.* 89 (2004) 181–190.
- [62] R.B. Arellano-Valle, M.G. Genton, Multivariate extended skew-t distributions and related families, *Metron* 68 (2010) 201–234.
- [63] S. Chib, E. Greenberg, Analysis of multivariate probit models, *Biometrika* 85 (1998) 347–361.
- [64] T. Amemiya, Tobit models: A survey, *J. Econometrics* 24 (1984) 3–61.
- [65] S. Chib, Bayes inference in the tobit censored regression model, *J. Econometrics* 51 (1992) 79–99.