An invariance property of quadratic forms in random vectors with a selection distribution, with application to sample variogram and covariogram estimators

ARELLANO-VALLE, Reinaldo B., GENTON, Marc

ARELLANO-VALLE, Reinaldo B., GENTON, Marc. An invariance property of quadratic forms in random vectors with a selection distribution, with application to sample variogram and covariogram estimators. *Annals of the Institute of Statistical Mathematics*, 2010, vol. 62, no. 2, p. 363-381

DOI : 10.1007/s10463-008-0175-3

Ann Inst Stat Math (2010) 62:363–381
DOI 10.1007/s10463-008-0175-3

An invariance property of quadratic forms in random vectors with a selection distribution, with application to sample variogram and covariogram estimators

Reinaldo B. Arellano-Valle · Marc G. Genton

Received: 10 May 2007 / Revised: 25 January 2008 / Published online: 22 May 2008

Abstract We study conditions under which an invariance property holds for the class of selection distributions. First, we consider selection distributions arising from two uncorrelated random vectors. In that setting, the invariance holds for the so-called $\mathcal{C}$-class and for elliptical distributions. Second, we describe the invariance property for selection distributions arising from two correlated random vectors. The particular case of the distribution of quadratic forms and its invariance, under various selection distributions, is investigated in more details. We describe the application of our invariance results to sample variogram and covariogram estimators used in spatial statistics and provide a small simulation study for illustration. We end with a discussion about other applications, for example such as linear models and indices of temporal/spatial dependence.

Keywords Kurtosis · Multivariate · Non-normal · Selection mechanism · Skewness · Spatial statistics · Time series

R. B. Arellano-Valle
Departamento de Estadística, Facultad de Matemática,
Pontificia Universidad Católica de Chile, Santiago 22, Chile
e-mail: reivalle@mat.puc.cl

M. G. Genton (✉)
Department of Econometrics, University of Geneva,
Bd du Pont-d'Arve 40, 1211 Geneva 4, Switzerland
e-mail: Marc.Genton@metri.unige.ch

M. G. Genton
Department of Statistics, Texas A&M University,
College Station, TX 77843-3143, USA
e-mail: genton@stat.tamu.edu

1 Introduction

In time series analysis, the autocovariance function is an important characteristic of a stationary stochastic process and is used for modeling the data and for forecasting. Similarly, in spatial analysis, the variogram and the covariogram (a spatial analogue of the autocovariance function) of a stationary random field are crucial for spatial interpolation of data. The sample moment estimators of the variogram and covariogram are quadratic forms in the vector of data. Genton et al. (2001) have proved an invariance property of these estimators. Specifically, for the class of skew-normal random vectors introduced by Azzalini and Dalla Valle (1996), the joint distribution of the sample variogram (respectively, sample covariogram) at various spatial lags does not depend on the skewness. That is, this joint distribution is the same as for normal vectors. This invariance property has been extended to a wider class of skew-symmetric random vectors by Wang et al. (2004a). This is an important robustness property of these estimators with respect to skewness in the data for those classes of distributions. The main goal of this paper is to identify and characterize multivariate distributions for which this invariance property holds.

Multivariate distributions that arise from certain selection mechanisms play an important role in our investigation. Let $\mathbf{U} \in \mathbb{R}^q$ and $\mathbf{V} \in \mathbb{R}^p$ be two random vectors, and denote by C a measurable subset of $\mathbb{R}^q$. Arellano-Valle et al. (2006) have defined a selection distribution as the conditional distribution of $\mathbf{V}$ given $\mathbf{U} \in C$, that is, as the distribution of $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$. If $\mathbf{V}$ has a probability density function (pdf), $f_{\mathbf{V}}$ say, then $\mathbf{X}$ has a pdf $f_{\mathbf{X}}$ given by

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{V}}(\mathbf{x}) \frac{P(\mathbf{U} \in C|\mathbf{V} = \mathbf{x})}{P(\mathbf{U} \in C)}, \tag{1}$$

which is determined by the marginal distribution of $\mathbf{V}$, the conditional distribution of $\mathbf{U}$ given $\mathbf{V}$ and the form of the subset C. Moreover, the cumulative distribution function (cdf) $F_{\mathbf{X}}$ of $\mathbf{X}$ can be computed easily from the joint distribution of $(\mathbf{U}^T, \mathbf{V}^T)^T$ as

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{V} \leq \mathbf{x}|\mathbf{U} \in C) = \frac{P(\mathbf{U} \in C, \mathbf{V} \leq \mathbf{x})}{P(\mathbf{U} \in C)}. \tag{2}$$

The most common specification of C is $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \mathbf{0}\}$, where the inequality between vectors is meant component-wise. In such case, the distribution corresponding to (1) is called *fundamental skew* (FUS). It is called *fundamental skew-symmetric* (FUSS), *fundamental skew-elliptical* (FUSE), and *fundamental skew-normal* (FUSN) when the parent pdf $f_{\mathbf{V}}$ is symmetric, elliptical, and normal, respectively; see Arellano-Valle and Genton (2005). If the distribution of $\mathbf{V}$ and the conditional distribution of $\mathbf{U}|\mathbf{V}$ are both elliptical or both normal, then the resulting FUS selection distribution is called *unified skew-elliptical* (SUE) and *unified skew-normal* (SUN), respectively, by Arellano-Valle and Azzalini (2006); see also Genton (2004) for various particular distributions with pdf of the form (1). Table 1 summarizes the aforementioned selection distributions.

Table 1 Some selection distributions

Distribution of: **X**	**V**	**U|V**
FUSS	Symmetric	Arbitrary
FUSE	Elliptical	Arbitrary
FUSN	Normal	Arbitrary
SUE	Elliptical	Elliptical
SUN	Normal	Normal

Although most distributions can be represented by (1), our main interest here lies in situations where $f_{\mathbf{V}}$ is symmetric, i.e., $f_{\mathbf{V}}(-\mathbf{x}) = f_{\mathbf{V}}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^p$, for which the pdf (1) is generally skewed, unless the conditional probabilities satisfy $P(\mathbf{U} \in C|\mathbf{V} = \mathbf{x}) = P(\mathbf{U} \in C|\mathbf{V} = -\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^p$. We note, however, that if $P(\mathbf{U} \in C|\mathbf{V} = \mathbf{x}) = P(\mathbf{U} \in C)$ for all $\mathbf{x} \in \mathbb{R}^p$, then $f_{\mathbf{X}} = f_{\mathbf{V}}$, i.e., $\mathbf{X} \overset{d}{=} \mathbf{V}$, and so $\psi(\mathbf{X}) \overset{d}{=} \psi(\mathbf{V})$ for all Borel functions ψ. This invariance property holds trivially when $\mathbf{U}$ and $\mathbf{V}$ are independent random vectors, but it can hold as well under weaker conditions when $\mathbf{U}$ and $\mathbf{V}$ are uncorrelated random vectors. Moreover, this property can also hold without the condition $P(\mathbf{U} \in C|\mathbf{V} = \mathbf{x}) = P(\mathbf{U} \in C)$ for some types of functions ψ. For example, it is a well known result that any even function ψ, i.e., $\psi(-\mathbf{x}) = \psi(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^p$, satisfies such an invariance property for the FUSS subclass of distributions corresponding to $q = 1$ and with pdf of the form

$$f_{\mathbf{X}}(\mathbf{x}) = 2 f_{\mathbf{V}}(\mathbf{x}) Q_U(\mathbf{x}),$$

where $f_{\mathbf{V}}$ is a symmetric pdf satisfying $f_{\mathbf{V}}(-\mathbf{x}) = f_{\mathbf{V}}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^p$ and Q_U is a skewing function satisfying $Q_U(\mathbf{x}) \geq 0$ and $Q_U(-\mathbf{x}) = 1 - Q_U(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^p$. Specifically, we have $\psi(\mathbf{X}) \overset{d}{=} \psi(\mathbf{V})$ in this case. The proof of this result can be found in Wang et al. (2004a) and in Azzalini and Capitanio (2003) for the equivalent parameterization $Q_U(\mathbf{x}) = F_U(w(\mathbf{x}))$, where F_U is the cdf of U and $w(-\mathbf{x}) = -w(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^p$. Similar results have been derived for particular cases by Azzalini (1985), Azzalini and Capitanio (1999), Wang et al. (2004b), and Genton and Loperfido (2005). However, we do not know the conditions under which this invariance property holds for the more general case with $q > 1$. In this paper, we explore the invariance property $\psi(\mathbf{X}) \overset{d}{=} \psi(\mathbf{V})$ for quadratic forms ψ when $P(\mathbf{U} \in C|\mathbf{V} = \mathbf{x}) \neq P(\mathbf{U} \in C)$.

The structure of the article is the following. In Sect. 2, we describe some basic properties of selection distributions in order to derive our main results. In Sect. 3, we present two examples with uncorrelated random vectors $\mathbf{U}$ and $\mathbf{V}$ for which the invariance property holds. Those are based on the so-called $\mathcal{C}$-class and on elliptical distributions. Examples with correlated random vectors $\mathbf{U}$ and $\mathbf{V}$ are described in Sect. 4. The particular case of the distribution of quadratic forms and its invariance, under various selection distributions, is investigated in Sect. 5. Application of our invariance results to sample variogram and covariogram estimators is reported in Sect. 6 along with a small simulation study for illustration. We end the paper with a discussion in Sect. 7.

2 Some basic properties of selection distributions

Many properties of a selection random vector $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$ can be studied directly from its definition, see Arellano-Valle et al. (2006) for details. Here, we describe some particular properties of direct relevance to our paper.

(P1) If the conditional distribution of $(\mathbf{V}|\mathbf{U} = \mathbf{u})$ has a moment generating function (mgf), $M_{\mathbf{V}|\mathbf{U}=\mathbf{u}}(\mathbf{t})$ say, then the mgf of $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$ can be computed by

$$M_{\mathbf{X}}(\mathbf{t}) = \frac{\int_{\mathbf{u}\in C} M_{\mathbf{V}|\mathbf{U}=\mathbf{u}}(\mathbf{t})\mathrm{d}F_{\mathbf{U}}(\mathbf{u})}{P(\mathbf{U} \in C)}, \tag{3}$$

where $F_{\mathbf{U}}$ denotes the cdf of $\mathbf{U}$.

(P2) We have $\psi(\mathbf{X}) \stackrel{d}{=} (\psi(\mathbf{V})|\mathbf{U} \in C)$ for any Borel function ψ. Moreover, since $(\psi(\mathbf{V})|\mathbf{U} \in C)$ is determined by the transformation $(\mathbf{U}, \mathbf{V}) \to (\mathbf{U}, \psi(\mathbf{V}))$, then when $\psi(\mathbf{V})$ has a pdf $f_{\psi(\mathbf{V})}$, we have that $\psi(\mathbf{X})$ has also a pdf of the form

$$f_{\psi(\mathbf{X})}(\mathbf{y}) = f_{\psi(\mathbf{V})}(\mathbf{y})\frac{P(\mathbf{U} \in C|\psi(\mathbf{V}) = \mathbf{y})}{P(\mathbf{U} \in C)}, \tag{4}$$

and its mgf can be computed by

$$M_{\psi(\mathbf{X})}(\mathbf{t}) = \frac{\int_{\mathbf{u}\in C} M_{\psi(\mathbf{V})|\mathbf{U}=\mathbf{u}}(\mathbf{t})\mathrm{d}F_{\mathbf{U}}(\mathbf{u})}{P(\mathbf{U} \in C)}. \tag{5}$$

As a direct consequence of (P2), we have:

(P2a) $\psi(\mathbf{X}) \stackrel{d}{=} \psi(\mathbf{V})$ if and only if $P(\mathbf{U} \in C|\psi(\mathbf{V}) = \mathbf{y}) = P(\mathbf{U} \in C)$ for all $\mathbf{y}$. It is trivial that this condition holds when $\psi(\mathbf{V})$ and $\mathbf{U}$ are independent, but it can hold also when $\psi(\mathbf{V})$ and $\mathbf{U}$ are uncorrelated.

(P2b) If $\psi(\mathbf{X}) = B\mathbf{X} + \mathbf{b}$ is a non-singular linear function, then (4) and (5) reduce to

$$f_{B\mathbf{X}+\mathbf{b}}(\mathbf{y}) = f_{B\mathbf{V}+\mathbf{b}}(\mathbf{y})\frac{P(\mathbf{U} \in C|B\mathbf{V} + \mathbf{b} = \mathbf{y})}{P(\mathbf{U} \in C)}, \tag{6}$$

and

$$M_{B\mathbf{X}+\mathbf{b}}(\mathbf{t}) = \exp\{\mathbf{t}^T\mathbf{b}\}\frac{\int_{\mathbf{u}\in C} M_{\mathbf{V}|\mathbf{U}=\mathbf{u}}(B^T\mathbf{t})\mathrm{d}F_{\mathbf{U}}(\mathbf{u})}{P(\mathbf{U} \in C)}, \tag{7}$$

respectively, for any (rectangular) matrix B and vector $\mathbf{b}$ fixed. Note also from (6) (or (7)) that any selection random vector $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$ is closed under linear transformations and, in particular, under marginalizations, when the original vector $\mathbf{V}$ has such a property.

(P3) If $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} > \mathbf{0})$ is a FUS random vector, then $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|D\mathbf{U} > \mathbf{0})$ for any diagonal matrix $D > 0$, i.e., $P(D\mathbf{U} > \mathbf{0}) = P(\mathbf{U} > \mathbf{0})$ and $P(D\mathbf{U} > \mathbf{0}|\mathbf{V} = \mathbf{x}) = P(\mathbf{U} > \mathbf{0}|\mathbf{V} = \mathbf{x})$. Thus, the FUS distribution is unaffected by rescaling of the constraints $\mathbf{U} > \mathbf{0}$.

3 Two examples with uncorrelated U and V

We give next two examples where $P(\mathbf{U} \in C|\mathbf{V} = \mathbf{x}) = P(\mathbf{U} \in C)$ holds for all $\mathbf{x} \in \mathbb{R}^p$, and therefore $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C) \stackrel{d}{=} \mathbf{V}$, without the condition that $\mathbf{U}$ and $\mathbf{V}$ are independent.

3.1 $\mathcal{C}$-class

Suppose that the random vectors $\mathbf{U} = (U_1, \ldots, U_q)^T$ and $\mathbf{V} = (V_1, \ldots, V_p)^T$ satisfy the following conditions:

(i) $\mathbf{U}$ is such that $P(\mathbf{U} = \mathbf{0}) = 0$ and $\mathrm{sgn}(\mathbf{U}) = (S_1, \ldots, S_q)^T$, with $S_i = -1$, if $U_i < 0$, and $S_i = 1$, if $U_i > 0$, and $|\mathbf{U}| = (|U_1|, \ldots, |U_q|)^T$ are independent; and

(ii) $\mathbf{V}$ is independent of $\mathrm{sgn}(\mathbf{U})$ given $|\mathbf{U}|$.

Note that condition (i) is equivalent to

$$\mathbf{U} \stackrel{d}{=} D(\mathbf{S})\mathbf{T} = D(\mathbf{T})\mathbf{S}, \tag{8}$$

where $\mathbf{S}$, $\mathbf{T} \in \mathbb{R}^q$ are independent random vectors with $\mathbf{S} \stackrel{d}{=} \mathrm{sgn}(\mathbf{U})$ and $\mathbf{T} \stackrel{d}{=} |\mathbf{U}|$, and for any vector $\mathbf{w} = (w_1, \ldots, w_q)^T$, $D(\mathbf{w})$ is the $q \times q$ diagonal matrix with diagonal entries $w_1, \ldots, w_q$. Moreover, if $\mathbf{S}$ has a uniform distribution on $\{-1, 1\}^q$, say $\mathbf{S} \sim \mathcal{U}_q$, then $P(\mathbf{S} = \mathbf{1}_q) = 2^{-q}$, where $\mathbf{1}_q$ is the vector of q ones, and so $\mathbf{U}$ has a symmetric (around zero) distribution. In such case, we say that $\mathbf{U}$ is a $\mathcal{C}$-random vector, see Arellano-Valle et al. (2002) and Arellano-Valle and del Pino (2004) for more details. Note that spherical random vectors and also all those with independent symmetric (around zero) components are $\mathcal{C}$-random vectors.

Under conditions (i) and (ii) we have $P(\mathbf{U} > \mathbf{0}|\mathbf{V} = \mathbf{x}) = P(\mathbf{U} > \mathbf{0}) = P(\mathbf{S} = \mathbf{1}_q)$, for all $\mathbf{x} \in \mathbb{R}^p$, implying that the FUS random vector $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} > \mathbf{0})$ has the same distribution as $\mathbf{V}$. Indeed, note first by (8) that

$$P(\mathbf{U} > \mathbf{0}) = E[P(D(\mathbf{T})\mathbf{S} > \mathbf{0}|\mathbf{T})] = E[P(\mathbf{S} = \mathbf{1}_q|\mathbf{T})] = P(\mathbf{S} = \mathbf{1}_q).$$

Now, considering again (8), we have by condition (ii) that

$$P(\mathbf{U} > \mathbf{0}|\mathbf{V} = \mathbf{x}) = E[P(D(\mathbf{T})\mathbf{S} > \mathbf{0}|\mathbf{V} = \mathbf{x}, \mathbf{T})] = E[P(\mathbf{S} = \mathbf{1}_q|\mathbf{T})] = P(\mathbf{S}=\mathbf{1}_q),$$

for all $\mathbf{x} \in \mathbb{R}^p$.

Note finally that condition (ii) holds when the full random vector $(\mathbf{U}^T, \mathbf{V}^T)^T$ satisfies condition (i), which holds in particular when $(\mathbf{U}^T, \mathbf{V}^T)^T$ is a $\mathcal{C}$-random vector.

3.2 Elliptical distributions

Suppose that $(\mathbf{U}^T, \mathbf{V}^T)^T$ has an elliptical distribution, say $EC_{q+p}(\boldsymbol{\mu}, \Sigma, h^{(q+p)})$, with $\boldsymbol{\mu} = (\mathbf{0}^T, \boldsymbol{\mu}_{\mathbf{V}}^T)^T$ and $\Sigma = \mathrm{diag}\{\Sigma_{\mathbf{U}}, \Sigma_{\mathbf{V}}\}$, where $h^{(q+p)}$ is a density generator function (see Fang et al. 1990), i.e., with joint pdf of the form

$$f_{\mathbf{U},\mathbf{V}}(\mathbf{x},\mathbf{y}) = (|\Sigma_{\mathbf{U}}||\Sigma_{\mathbf{V}}|)^{-1/2} h^{(q+p)}(\mathbf{x}^T \Sigma_{\mathbf{U}}^{-1}\mathbf{x} + (\mathbf{y}-\boldsymbol{\mu}_{\mathbf{V}})^T \Sigma_{\mathbf{V}}^{-1}(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{V}})),\quad \mathbf{x}\in\mathbb{R}^q,\ \mathbf{y}\in\mathbb{R}^p.$$

Then, $P(\mathbf{U} > \mathbf{0}|\mathbf{V} = \mathbf{y}) = P(\mathbf{U} > \mathbf{0}) = \Phi_q(\mathbf{0}; \mathbf{0}, \Sigma_{\mathbf{U}})$ (see Fang et al. 1990, p. 53), for any $\mathbf{y} \in \mathbb{R}^p$ and density generator $h^{(q+p)}$, where $\Phi_p(\cdot; \boldsymbol{\mu}, \Sigma)$ denotes the cdf of the $N_p(\boldsymbol{\mu}, \Sigma)$ distribution, and $\mathbf{X} \overset{d}{=} (\mathbf{V}|\mathbf{U} > \mathbf{0}) \overset{d}{=} \mathbf{V} \sim EC_p(\boldsymbol{\mu}_{\mathbf{V}}, \Sigma_{\mathbf{V}}, h^{(p)})$. In fact, since $\mathbf{U} \sim EC_q(\mathbf{0}, \Sigma_{\mathbf{U}}, h^{(q)})$ and $(\mathbf{U}|\mathbf{V} = \mathbf{y}) \sim EC_q(\mathbf{0}, \Sigma_{\mathbf{U}}, h^{(q)}_{q(\mathbf{y})})$, where $q(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu}_{\mathbf{V}})^T \Sigma_{\mathbf{V}}^{-1}(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{V}})$ and $h^{(q)}_u(v) = h^{(q+p)}(u+v)/h^{(p)}(u)$, we have

$$P(\mathbf{U} > \mathbf{0}|\mathbf{V} = \mathbf{y}) = P(\mathbf{U} > \mathbf{0}) = \Phi_q(\mathbf{0}; \mathbf{0}, \Sigma_{\mathbf{U}}),$$

for any $\mathbf{y} \in \mathbb{R}^p$ and density generator $h^{(q+p)}$; i.e., in (1) we have

$$f_{\mathbf{X}}(\mathbf{y}) = f_{\mathbf{V}}(\mathbf{y}) = |\Sigma_{\mathbf{V}}|^{-1/2} h^{(p)}((\mathbf{y}-\boldsymbol{\mu}_{\mathbf{V}})^T \Sigma_{\mathbf{V}}^{-1}(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{V}})),$$

for all $\mathbf{y} \in \mathbb{R}^p$, so that $\mathbf{X} \overset{d}{=} (\mathbf{V}|\mathbf{U} > \mathbf{0}) \overset{d}{=} \mathbf{V} \sim EC_p(\boldsymbol{\mu}_{\mathbf{V}}, \Sigma_{\mathbf{V}}, h^{(p)})$.

4 Correlated U and V

In the construction of the selection pdf (1) we need to assume that **U** and **V** are associated (correlated) random vectors, in order to alter the original pdf $f_{\mathbf{V}}$. There are many ways to introduce a relation between **U** and **V**, for example such as:

(a) Assuming that $\mathbf{V} = g(\mathbf{U}, \mathbf{V}_0)$ for some Borel function g and for a (symmetric) random vector $\mathbf{V}_0$ which is independent of (or uncorrelated with) **U**. Thus, linear functions like $\mathbf{V} = A\mathbf{U} + B\mathbf{V}_0$ can be explored. For instance, by independence between **U** and $\mathbf{V}_0$,

$$\mathbf{X} \overset{d}{=} (A\mathbf{U} + B\mathbf{V}_0|\mathbf{U} \in C) \overset{d}{=} A\mathbf{U}(C) + B\mathbf{V}_0,$$

where $\mathbf{U}(C) \overset{d}{=} (\mathbf{U}|\mathbf{U} \in C)$ and is independent of $\mathbf{V}_0$.

(b) Assuming that $\mathbf{U} = h(\mathbf{U}_0, \mathbf{V})$ for some Borel function h and for a (symmetric) random vector $\mathbf{U}_0$ which is independent of (or uncorrelated with) **V**. Under this situation, functions of the form $\mathbf{U} = w(\mathbf{V}) - \mathbf{U}_0$, with $w(-\mathbf{v}) = -w(\mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^p$ can be studied by noting that $P(\mathbf{U} \in C|\mathbf{V} = \mathbf{x}) = F_{\mathbf{U}_0|\mathbf{V}=\mathbf{x}}(w(\mathbf{x}))$, which reduces to $F_{\mathbf{U}_0}(w(\mathbf{x}))$ when $\mathbf{U}_0$ is independent of **V** given $w(\mathbf{V})$.

(c) Assuming that $\mathbf{V} = A\mathbf{W}$ and $\mathbf{U} = B\mathbf{W}$, where $\mathbf{W} \in \mathbb{R}^n$ is a (symmetric) random vector with dimension $n \geq p,\ q$.

In (a) and (b) the independence restriction (between $\mathbf{U}_0$ and **U** or $\mathbf{V}_0$ and **V**, respectively) can be replaced by a weaker condition similar to the assumption (ii) given in Sect. 3.1. Also, the symmetric condition given in parenthesis above is required only to obtain a pdf of FUSS type.

5 Distribution of quadratic forms under selection distributions

In many situations, for example such as sample variogram and covariogram estimators, we are interested in the distribution of a quadratic form

$$\psi(\mathbf{X}-\mathbf{a}) = (\mathbf{X}-\mathbf{a})^T A(\mathbf{X}-\mathbf{a}), \tag{9}$$

for any symmetric matrix A and vector $\mathbf{a}$, where the random vector $\mathbf{X}$ has a selection distribution, i.e., $\mathbf{X} \overset{d}{=} (\mathbf{V}|\mathbf{U} \in C)$. In this section, we derive the moment generating function of $\psi(\mathbf{X}-\mathbf{a})$ and we explore the conditions under which the invariance property $\psi(\mathbf{X}-\mathbf{a}) \overset{d}{=} \psi(\mathbf{V}-\mathbf{a})$ holds.

5.1 Selection normal distributions

Suppose that $\mathbf{X} \overset{d}{=} (\mathbf{V}|\mathbf{U} \in C)$, where

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim N_{q+p}\left(\boldsymbol{\xi}_* = \begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\xi} \end{pmatrix}, \Omega_* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \Omega \end{pmatrix}\right), \tag{10}$$

with Ω_* a non-singular covariance matrix. From (1), the pdf of the selection normal distribution of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \phi_p(\mathbf{x}; \boldsymbol{\xi}, \Omega) \frac{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma} + \Delta^T\Omega^{-1}(\mathbf{x}-\boldsymbol{\xi}), \Gamma - \Delta^T\Omega^{-1}\Delta)d\mathbf{u}}{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, \Gamma)d\mathbf{u}}. \tag{11}$$

Here, the notation $\mathbf{X} \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}_*, \Omega_*, C)$ or more explicitly $\mathbf{X} \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \Omega, \Gamma, \Delta, C)$ will be used to say that a random vector $\mathbf{X}$ has the selection normal pdf (11). Considering the latter notation, the following basic properties follow directly from the genesis of $\mathbf{X}$:

- (N1) If $\Delta = O$, the zero matrix, then the $SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \Omega, \Gamma, \Delta, C)$ distribution reduces to the $N_p(\boldsymbol{\xi}, \Omega)$ one, whatever the selection set C and the value of $(\boldsymbol{\gamma}, \Gamma)$.
- (N2) $SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \Omega, \Gamma, \Delta, C) = SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \mathbf{0}, \Omega, \Gamma, \Delta, C-\boldsymbol{\gamma})$.
- (N3) If $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \boldsymbol{\gamma}\}$, then $C - \boldsymbol{\gamma} = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \mathbf{0}\}$, and by (P3) and (N2) we have $SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \Omega, \Gamma, \Delta, C) = SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \mathbf{0}, \Omega, D\Gamma D, \Delta D, C-\boldsymbol{\gamma})$ for any diagonal matrix $D > 0$.
- (N4) If $\mathbf{X} \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \Omega, \Gamma, \Delta, C)$, then for any $r \times p$ matrix B and vector $\mathbf{b} \in \mathbb{R}^p$, we have $B\mathbf{X}+\mathbf{b} \sim SLCT\text{-}N_{r,q}(B\boldsymbol{\xi}+\mathbf{b}, \boldsymbol{\gamma}, B\Omega B^T, \Gamma, B\Delta, C)$, which by (N1) reduces to the $N_r(B\boldsymbol{\xi}+\mathbf{b}, B\Omega B^T)$ distribution when $B\Delta = O$.

When $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \mathbf{0}\}$, (11) reduces to the SUN pdf of Arellano-Valle and Azzalini (2006) given by

$$f_{\mathbf{X}}(\mathbf{x}) = \phi_p(\mathbf{x}; \boldsymbol{\xi}, \Omega) \frac{\Phi_q(\boldsymbol{\gamma} + \Delta^T\Omega^{-1}(\mathbf{x}-\boldsymbol{\xi}); \Gamma - \Delta^T\Omega^{-1}\Delta)}{\Phi_q(\boldsymbol{\gamma}; \mathbf{0}, \Gamma)}.$$

Under this SUN distribution, Arellano-Valle and Azzalini (2006) showed that for any $p \times p$ symmetric matrix A, the quadratic form $\psi(\mathbf{X}-\boldsymbol{\xi}) = (\mathbf{X}-\boldsymbol{\xi})^T A(\mathbf{X}-\boldsymbol{\xi})$ has mgf given by

$$M_{\psi(\mathbf{X}-\boldsymbol{\xi})}(t) = |I_p - 2tA\Omega|^{-1/2} \frac{\Phi_q(\boldsymbol{\gamma}; \mathbf{0}, \Gamma + 2t\Delta^T(I_p - 2tA\Omega)^{-1}A\Delta)}{\Phi_q(\boldsymbol{\gamma}; \mathbf{0}, \Gamma)},$$

from which it follows that $\psi(\mathbf{X}-\boldsymbol{\xi}) \stackrel{d}{=} \psi(\mathbf{V}-\boldsymbol{\xi})$, where $\mathbf{V} \sim N_p(\boldsymbol{\xi}, \Omega)$, i.e.,

$$M_{\psi(\mathbf{X}-\boldsymbol{\xi})}(t) = M_{\psi(\mathbf{V}-\boldsymbol{\xi})}(t) = |I_p - 2tA\Omega|^{-1/2}, \quad \forall t : I_p - 2tA\Omega > 0,$$

under each of the following situations:

1. $A\Delta = O$. Moreover, if A has rank r and $A\Omega A = A$, then $|I_p - 2tA\Omega|^{-1/2} = (1-2t)^{-r/2}$, i.e., $\psi(\mathbf{X}-\boldsymbol{\xi}) \sim \chi^2_r$.
2. $\boldsymbol{\gamma} = \mathbf{0}$, $A = \Omega^{-1}$ and Γ and $\Delta^T\Omega^{-1}\Delta$ are diagonal matrices. In this case, $\psi(\mathbf{X}-\boldsymbol{\xi}) \sim \chi^2_p$.
3. $\boldsymbol{\gamma} = \mathbf{0}$, with $q = 1$.

Remark 1 The invariance result under the condition $A\Delta = O$ is actually a direct consequence of (N4), since we can write $A = B^T B$, for some $r \times p$ matrix B of rank r, and so $A\Delta = O \Leftrightarrow B\Delta = O$. However, a more formal proof follows by noting that $\psi(\mathbf{X}-\mathbf{a}) = \|B(\mathbf{X}-\mathbf{a})\|^2$, where by definition (see also (P2b)) $B\mathbf{X} \stackrel{d}{=} (B\mathbf{V}|\mathbf{U} \in C) \stackrel{d}{=} B\mathbf{V}$, since $Cov(B\mathbf{V}, \mathbf{U}) = B\Delta = O$, i.e., $B\mathbf{V}$ and $\mathbf{U}$ are independent. Hence $\|B(\mathbf{X}-\mathbf{a})\|^2 \stackrel{d}{=} \|B(\mathbf{V}-\mathbf{a})\|^2$, where $\mathbf{V} \sim N_p(\boldsymbol{\xi}, \Omega)$.

We consider next the more general selection normal distribution (11) to study the distribution of the quadratic form (9). For this we need some preliminary results on the mgf of quadratic forms in normal random vectors (see also Khatri 1980, Eq. 3.4).

Lemma 1 *Let* $\psi(\mathbf{Y}-\mathbf{a}) = (\mathbf{Y}-\mathbf{a})^T A(\mathbf{Y}-\mathbf{a})$, *where* $\mathbf{Y} \sim N_d(\boldsymbol{\mu}, \Sigma)$. *Then*

$$M_{\psi(\mathbf{Y}-\mathbf{a})}(t) = \frac{\exp\{t(\mathbf{a}-\boldsymbol{\mu})^T(I_p - 2tA\Sigma)^{-1}A(\mathbf{a}-\boldsymbol{\mu})\}}{|I_d - 2tA\Sigma|^{1/2}}.$$

Proof Since $M_{\psi(\mathbf{Y}-\mathbf{a})}(t) = \int_{\mathbb{R}^d} \exp\{t(\mathbf{y}-\mathbf{a})^T A(\mathbf{y}-\mathbf{a})\}\phi_d(\mathbf{y}; \boldsymbol{\mu}, \Sigma)d\mathbf{y}$, the proof follows by considering the identities:

$$(\mathbf{y}-\mathbf{a})^T A(\mathbf{y}-\mathbf{a}) = (\mathbf{y}-\boldsymbol{\mu})^T A(\mathbf{y}-\boldsymbol{\mu}) + 2(\mathbf{a}-\boldsymbol{\mu})^T A(\mathbf{y}-\boldsymbol{\mu}) + (\mathbf{a}-\boldsymbol{\mu})^T A(\mathbf{a}-\boldsymbol{\mu})$$

and

$$\begin{aligned} t(\mathbf{y}-\mathbf{a})^T A(\mathbf{y}-\mathbf{a}) &- \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}) \\ &= -\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu} - 2t(\Sigma^{-1}-2tA)^{-1}A(\mathbf{a}-\boldsymbol{\mu})\right)^T \\ &\quad \times (\Sigma^{-1}-2tA)(\mathbf{y}-\boldsymbol{\mu} - 2t(\Sigma^{-1}-2tA)^{-1}A(\mathbf{a}-\boldsymbol{\mu})) \\ &\quad + t(\mathbf{a}-\boldsymbol{\mu})^T[A + 2tA(\Sigma^{-1}-2tA)^{-1}A](\mathbf{a}-\boldsymbol{\mu}), \end{aligned}$$

and so

$$\begin{aligned}
&\exp\{t(\mathbf{y}-\mathbf{a})^T A(\mathbf{y}-\mathbf{a})\}\phi_d(\mathbf{y};\boldsymbol{\mu},\Sigma)\\
&\quad=\frac{\exp\{t(\mathbf{a}-\boldsymbol{\mu})^T[A+2tA(\Sigma^{-1}-2tA)^{-1}A](\mathbf{a}-\boldsymbol{\mu})\}}{|I_d-2tA\Sigma|^{1/2}}\\
&\qquad\times\phi_d(\mathbf{y};\boldsymbol{\mu}+2t(\Sigma^{-1}-2tA)^{-1}A(\mathbf{a}-\boldsymbol{\mu}),(\Sigma^{-1}-2tA)^{-1}),
\end{aligned}$$

from where the proof follows by noting that $A+2tA(\Sigma^{-1}-2tA)^{-1}A=(I_p-2tA\Sigma)^{-1}A$. □

Proposition 1 *Let* $\psi(\mathbf{X}-\mathbf{a})=(\mathbf{X}-\mathbf{a})^T A(\mathbf{X}-\mathbf{a})$, *where* $\mathbf{X}\sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}_*,\Omega_*,C)$. *Then,*

$$\begin{aligned}
&M_{\psi(\mathbf{X}-\mathbf{a})}(t)\\
&\quad=\frac{\exp\{t(\mathbf{a}-\boldsymbol{\xi})^T(I_p-2tA\Omega)^{-1}A(\mathbf{a}-\boldsymbol{\xi})\}}{|I_p-2tA\Omega|^{1/2}}\\
&\quad\times\frac{\int_{\mathbf{u}\in C}\phi_q(\mathbf{u};\boldsymbol{\gamma}+2t\Delta^T(I_p-2tA\Omega)^{-1}A(\mathbf{a}-\boldsymbol{\xi}),\Gamma+2t\Delta^T(I_p-2tA\Omega)^{-1}A\Delta)d\mathbf{u}}{\int_{\mathbf{u}\in C}\phi_q(\mathbf{u};\boldsymbol{\gamma},\Gamma)d\mathbf{u}}.
\end{aligned}$$

Proof Note first by (10) that $(\mathbf{V}|\mathbf{U}=\mathbf{u})\sim N_p(\boldsymbol{\xi}+\Delta\Gamma^{-1}(\mathbf{u}-\boldsymbol{\gamma}),\Omega-\Delta\Gamma^{-1}\Delta^T)$ and $\mathbf{U}\sim N_q(\boldsymbol{\gamma},\Gamma)$. Thus, by applying (5) in (P2) we have

$$M_{\psi(\mathbf{X}-\mathbf{a})}(t)=\frac{\int_{\mathbf{u}\in C}M_{\psi(\mathbf{V}-\mathbf{a})|\mathbf{U}=\mathbf{u}}(t)\phi_q(\mathbf{u};\boldsymbol{\gamma},\Gamma)d\mathbf{u}}{\int_{\mathbf{u}\in C}\phi_q(\mathbf{u};\boldsymbol{\gamma},\Gamma)d\mathbf{u}}, \tag{12}$$

where by Lemma 1, with $\boldsymbol{\mu}=\boldsymbol{\xi}+\Delta\Gamma^{-1}(\mathbf{u}-\boldsymbol{\gamma})$ and $\Sigma=\Omega-\Delta\Gamma^{-1}\Delta^T$,

$$M_{\psi(\mathbf{V}-\mathbf{a})|\mathbf{U}=\mathbf{u}}(t)=\frac{\exp\{t(B\mathbf{y}+\mathbf{b})^T G(B\mathbf{y}+\mathbf{b})\}}{|I_p-2tA(\Omega-\Delta\Gamma^{-1}\Delta^T)|^{1/2}},$$

where $\mathbf{y}=\mathbf{u}-\boldsymbol{\gamma}$, $\mathbf{b}=\boldsymbol{\xi}-\mathbf{a}$, $B=\Delta\Gamma^{-1}$ and $G=A+2tA[(\Omega-\Delta\Gamma^{-1}\Delta^T)^{-1}-2tA]^{-1}A$, which can be re-written as $G=[I_p-2tA(\Omega-\Delta\Gamma^{-1}\Delta^T)]^{-1}A$. Considering now the identities

$$(B\mathbf{y}+\mathbf{b})^T G(B\mathbf{y}+\mathbf{b})=\mathbf{y}^T B^T GB\mathbf{y}+2\mathbf{y}^T B^T G\mathbf{b}+\mathbf{b}^T G\mathbf{b}$$

and

$$\begin{aligned}
t(B\mathbf{y}+\mathbf{b})^T G(B\mathbf{y}+\mathbf{b})-\frac{1}{2}\mathbf{y}^T\Gamma^{-1}\mathbf{y}&=-\frac{1}{2}(\mathbf{y}-2tHB^TG\mathbf{b})^T H^{-1}(\mathbf{y}-2tHB^TG\mathbf{b})\\
&\quad+t\mathbf{b}^T(G+2tGBHB^TG)\mathbf{b},
\end{aligned}$$

where $H = (\Gamma^{-1} - 2tB^TGB)^{-1}$, we have in (12) that

$$\begin{aligned} M_{\psi(\mathbf{V}-\mathbf{a})|\mathbf{U}=\mathbf{u}}(t)\phi_q(\mathbf{u};\boldsymbol{\gamma},\Gamma) &= \frac{\exp\{t(B\mathbf{y}+\mathbf{b})^T G(B\mathbf{y}+\mathbf{b})\}}{|I_p - 2tA(\Omega - \Delta\Gamma^{-1}\Delta^T)|^{1/2}}\phi_q(\mathbf{y};\mathbf{0},\Gamma) \\ &= \frac{\exp\{t\mathbf{b}^T[G + 2tGBHB^TG]\mathbf{b}\}}{|I_p - 2tA(\Omega - \Delta\Gamma^{-1}\Delta^T)|^{1/2}|I_q - 2tB^TGB\Gamma|^{1/2}} \\ &\quad \times \phi_q(\mathbf{y}; 2tHB^TG\mathbf{b}, H). \end{aligned}$$

From the facts that $B\Gamma B^T = \Delta\Gamma^{-1}\Delta^T$ and $G = [I_p - 2tA(\Omega - \Delta\Gamma^{-1}\Delta^T)]^{-1}A$, it follows that

$$|I_q - 2tB^TGB\Gamma| = \frac{|I_p - 2tA\Omega|}{|I_p - 2tA(\Omega - \Delta\Gamma^{-1}\Delta^T)|}.$$

Thus, since $\mathbf{y} = \mathbf{u} - \boldsymbol{\gamma}$, we have for the numerator in (12) that

$$\begin{aligned} \int_{\mathbf{u}\in C} M_{\psi(\mathbf{V}-\mathbf{a})|\mathbf{U}=\mathbf{u}}(t)\phi_q(\mathbf{u};\boldsymbol{\gamma},\Gamma)\mathrm{d}\mathbf{u} &= \frac{\exp\{t\mathbf{b}^T[G + 2tGBHB^TG]\mathbf{b}\}}{|I_p - 2tA\Omega|^{1/2}} \\ &\quad \times \int_{\mathbf{u}\in C} \phi_q(\mathbf{u};\boldsymbol{\gamma} + 2tHB^TG\mathbf{b}, H)\mathrm{d}\mathbf{u}, \end{aligned}$$

which concludes the proof, by noting that $H = \Gamma + 2t\Delta^T(I_p - 2tA\Omega)^{-1}A\Delta$, and so $HB^TG = \Delta^T(I_p - 2tA\Omega)^{-1}A = \Delta^TA(I_p - 2t\Omega A)^{-1}$ and $G + 2tGBHB^TG = (I_p - 2tA\Omega)^{-1}A$. □

Proposition 2 *Let* $\psi(\mathbf{X} - \mathbf{a}) = (\mathbf{X} - \mathbf{a})^T A(\mathbf{X} - \mathbf{a})$, *where* $\mathbf{X} \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}_*, \Omega_*, C)$, *and let* $\mathbf{V} \sim N_p(\boldsymbol{\xi}, \Omega)$. *Then,* $\psi(\mathbf{X} - \mathbf{a}) \stackrel{d}{=} \psi(\mathbf{V} - \mathbf{a})$ *under each of the following conditions:*

(i) $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \boldsymbol{\gamma}\}$, $\mathbf{a} = \boldsymbol{\xi}$, $A = \Omega^{-1}$, *and* Γ *and* $\Delta^T\Omega^{-1}\Delta$ *are diagonal matrices.*
(ii) $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \boldsymbol{\gamma}\}$, $A(\mathbf{a} - \boldsymbol{\xi}) = \mathbf{0}$ *and* $q = 1$.
(iii) $A\Delta = O$.

Proof By Proposition 1 and Lemma 1, with $\mathbf{Y} = \mathbf{V}$, $\boldsymbol{\mu} = \boldsymbol{\xi}$ and $\Sigma = \Omega$, we have that $M_{\psi(\mathbf{X}-\mathbf{a})}(t) = M_{\psi(\mathbf{V}-\mathbf{a})}(t)$ for all t where they are defined if and only if the following identity holds:

$$\frac{\int_{\mathbf{u}\in C}\phi_q(\mathbf{u};\boldsymbol{\gamma} + 2t\Delta^T(I_p - 2tA\Omega)^{-1}A(\mathbf{a} - \boldsymbol{\xi}), \Gamma + 2t\Delta^T(I_p - 2tA\Omega)^{-1}A\Delta)\mathrm{d}\mathbf{u}}{\int_{\mathbf{u}\in C}\phi_q(\mathbf{u};\boldsymbol{\gamma},\Gamma)\mathrm{d}\mathbf{u}} = 1.$$

From the latter, the proofs of (i)–(iii) follow easily, considering, in particular, that condition (iii) implies $\Delta^T(I_p - 2tA\Omega)^{-1}A = [(I_p - 2t\Omega A)^{-1}A\Delta]^T = O$. □

Remark 2 In (i)–(ii), we note by the condition $\boldsymbol{\xi} - \mathbf{a} = \mathbf{0}$ or $A(\boldsymbol{\xi} - \mathbf{a}) = \mathbf{0}$ that $\psi(\mathbf{X} - \mathbf{a}) = \psi(\mathbf{X} - \boldsymbol{\xi})$. Moreover, the conditions on C and on Γ are equivalent to assuming (without loss of generality) that $\boldsymbol{\gamma} = \mathbf{0}$ and $\Gamma = I_q$, respectively (see (N3)). In the case (iii), we note that the condition $A\Delta = O$ has sense when $p > 1$, which holds for any matrix A of the form $A = P - P\Delta(\Delta^T P\Delta)^{-1}\Delta^T P$, where P is a $p \times p$ symmetric matrix. Moreover, when the matrix P has rank r and $P\Omega P = P$, we have that A has rank r and $A\Omega A = A$.

Corollary 1 *Let* $\psi(\mathbf{X}-\mathbf{a}) = (\mathbf{X}-\mathbf{a})^T A(\mathbf{X}-\mathbf{a})$, *where* $\mathbf{X} \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}_*, \Omega_*, C)$, $\mathbf{a} \in \mathbb{R}^p$ *is a fixed column vector and A is a symmetric $p \times p$ fixed matrix of rank r. If* $\mathbf{a}$ *and A are such that* $A(\mathbf{a} - \boldsymbol{\xi}) = \mathbf{0}$ *and* $A\Omega A = A$, *then* $\psi(\mathbf{X} - \mathbf{a}) = \psi(\mathbf{X} - \boldsymbol{\xi}) \sim \chi^2_r$ *under each of the conditions (i)–(iii) of Proposition 2.*

Proof Since $A(\boldsymbol{\xi} - \mathbf{a}) = \mathbf{0}$, it is clear that $\psi(\mathbf{X} - \mathbf{a}) = \psi(\mathbf{X} - \boldsymbol{\xi})$. Now, note by applying the condition $A(\boldsymbol{\xi} - \mathbf{a}) = \mathbf{0}$ in Proposition 1 that

$$M_{\psi(\mathbf{X}-\mathbf{a})}(t) = |I_p - 2tA\Omega|^{-1/2}\ \frac{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, \Gamma + 2t\Delta^T(I_p - 2tA\Omega)^{-1}A\Delta)\mathrm{d}\mathbf{u}}{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, \Gamma)\mathrm{d}\mathbf{u}},$$

where, from the conditions that A has rank r and $A\Omega A = A$, $|I_p - 2tA\Omega|^{-1/2} = (1 - 2t)^{-r/2}$. Thus, the proof follows by noting that under each of the conditions (i)–(iii) we have

$$\frac{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, \Gamma + 2t\Delta^T(I_p - 2tA\Omega)^{-1}A\Delta)\mathrm{d}\mathbf{u}}{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, \Gamma)\mathrm{d}\mathbf{u}} = 1,$$

and so $M_{\psi(\mathbf{X}-\mathbf{a})}(t) = (1 - 2t)^{-r/2}$. □

Remark 3 Since $(\mathbf{V}|\mathbf{U}) \sim N_p(\boldsymbol{\xi} + \Delta\Gamma^{-1}(\mathbf{U} - \boldsymbol{\gamma}), \Omega - \Delta\Gamma^{-1}\Delta^T)$ by (10), we have that the condition $A\Delta = O$ implies that $\psi(\mathbf{V}-\mathbf{a}) = (\mathbf{V}-\mathbf{a})^T A(\mathbf{V}-\mathbf{a})$ is independent of $\mathbf{U}$, hence $\psi(\mathbf{X} - \mathbf{a}) \stackrel{d}{=} (\psi(\mathbf{V} - \mathbf{a})|\mathbf{U} \in C) \stackrel{d}{=} \psi(\mathbf{V} - \mathbf{a})$ by (P2), see also Remark 1.

5.2 Selection elliptical distributions

We consider now a selection elliptical random vector $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$, where

$$\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \sim EC_{q+p}\left(\boldsymbol{\xi}_* = \begin{pmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\xi}\end{pmatrix},\quad \Omega_* = \begin{pmatrix}\Gamma & \Delta^T\\ \Delta & \Omega\end{pmatrix}, h^{(q+p)}\right), \tag{13}$$

with Ω_* a positive definite dispersion matrix. By (1), it has a selection pdf given by

$$f_{\mathbf{X}}(\mathbf{x}) = f_p(\mathbf{x}; \boldsymbol{\xi}, \Omega, h^{(p)}) \times \frac{\int_{\mathbf{u}\in C} f_q(\mathbf{u}; \boldsymbol{\gamma} + \Delta^T\Omega^{-1}(\mathbf{x} - \boldsymbol{\xi}), \Gamma - \Delta^T\Omega^{-1}\Delta, h^{(q)}_{w(\mathbf{x})})\mathrm{d}\mathbf{u}}{\int_{\mathbf{u}\in C} f_q(\mathbf{u}; \boldsymbol{\gamma}, \Gamma, h^{(q)})\mathrm{d}\mathbf{u}}, \tag{14}$$

where $f_k(\mathbf{y}; \boldsymbol{\mu}, \Sigma, h^{(k)}) = |\Sigma|^{-1/2} h^{(k)}[(\mathbf{y}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})]$, $w(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\xi})^T \Omega^{-1}(\mathbf{x}-\boldsymbol{\xi})$ and $h_w^{(p)}(u) = h^{(q+p)}(u+w)/h^{(q)}(w)$. Here, the notation $\mathbf{X} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}_*, \Omega_*, h^{(q+p)}, C)$ or, more explicitly, $\mathbf{X} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \Omega, \Gamma, \Delta, h^{(q+p)}, C)$, will be used to say that a random vector $\mathbf{X}$ has the selection elliptical pdf (14). From the latter notation, the basic properties (N1)–(N4) of the selection normal distribution can also be extended easily to this case. When $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \mathbf{0}\}$, (14) reduces to the SUE pdf introduced by Arellano-Valle and Azzalini (2006) given by

$$f_{\mathbf{X}}(\mathbf{x}) = f_p(\mathbf{x}; \boldsymbol{\xi}, \Omega, h^{(p)}) \frac{F_q\left(\boldsymbol{\gamma} + \Delta^T \Omega^{-1}(\mathbf{x}-\boldsymbol{\xi}); \mathbf{0}, \Gamma - \Delta^T \Omega^{-1}\Delta, h^{(q)}_{w(\mathbf{x})}\right)}{F_q(\boldsymbol{\gamma}; \mathbf{0}, \Gamma, h^{(q)})}. \tag{15}$$

Under (15) we have $\psi(\mathbf{X}-\boldsymbol{\xi}) = (\mathbf{X}-\boldsymbol{\xi})^T \Omega^{-1}(\mathbf{X}-\boldsymbol{\xi})$ has the same distribution as $\psi(\mathbf{V}-\boldsymbol{\xi}) = (\mathbf{V}-\boldsymbol{\xi})^T \Omega^{-1}(\mathbf{V}-\boldsymbol{\xi})$, where $\mathbf{V} \sim EC_p(\boldsymbol{\xi}, \Omega, h^{(p)})$, when $\boldsymbol{\gamma} = \mathbf{0}$ and $q = 1$. In fact, under such conditions, (15) is in the generalized skew-elliptical class of pdf's introduced by Genton and Loperfido (2005), with skewing function $F_1(\boldsymbol{\delta}^T \Omega^{-1}(\mathbf{x}-\boldsymbol{\xi}); 0, 1-\boldsymbol{\delta}^T\Omega^{-1}\boldsymbol{\delta}, h^{(1)}_{w(\mathbf{x})})$, for which the invariance property holds for any even function of $\Omega^{-1/2}(\mathbf{X}-\boldsymbol{\xi})$. It is not clear if this fact can be extended to the full SUE class when $q > 1$, and thus to the selection elliptical distributions. However, under the latter class we can give the following result.

Proposition 3 *Let $\psi(\mathbf{X}-\mathbf{a}) = (\mathbf{X}-\mathbf{a})^T A(\mathbf{X}-\mathbf{a})$, where $\mathbf{X} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}_*, \Omega_*, h^{(q+p)}, C)$, $\mathbf{a} \in \mathbb{R}^p$ is a fixed column vector and A is a symmetric $p \times p$ fixed matrix of rank r. If $A\Delta = O$, and $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \boldsymbol{\gamma}\}$, then $\psi(\mathbf{X}-\mathbf{a}) \stackrel{d}{=} \psi(\mathbf{V}-\mathbf{a}) = (\mathbf{V}-\mathbf{a})^T A(\mathbf{V}-\mathbf{a})$, where $\mathbf{V} \sim EC_p(\boldsymbol{\xi}, \Omega, h^{(p)})$.*

Proof Note first that the condition on C implies that $C - \boldsymbol{\gamma} = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \mathbf{0}\}$, so that without loss of generality we can assume that $\boldsymbol{\gamma} = \mathbf{0}$. Now, let B be a $r \times p$ matrix of rank r such that $B^T B = A$. Without loss of generality assume that $\mathbf{a} = \boldsymbol{\xi} = \mathbf{0}$. Thus we need to show, under the above conditions, that $\psi(\mathbf{X}) = \|B\mathbf{X}\|^2$ and $\psi(\mathbf{V}) = \|B\mathbf{V}\|^2$ are equally distributed. By (P2), $B\mathbf{X} \stackrel{d}{=} (B\mathbf{V}|\mathbf{U} > \mathbf{0})$, where from (13)

$$\begin{pmatrix} \mathbf{U} \\ B\mathbf{V} \end{pmatrix} \sim EC_{q+r}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Gamma & (B\Delta)^T \\ B\Delta & B\Omega B^T \end{pmatrix}, h^{(q+r)}\right). \tag{16}$$

Thus, to prove the result, we note that the condition $A\Delta = O$ is equivalent to $B\Delta = O$, i.e., in (16), $\mathbf{U}$ and $B\mathbf{V}$ are uncorrelated elliptical random vectors. Hence, $B\mathbf{X} \stackrel{d}{=} B\mathbf{V}$ (see Sect. 3.2) and $\psi(\mathbf{X}) \stackrel{d}{=} \psi(\mathbf{V})$ by (P2). □

5.3 Scale mixtures of selection normal distributions

An important subclass of selection elliptical distributions is obtained when the joint distribution of $\mathbf{U}$ and $\mathbf{V}$ is such that

$$\left[\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}\Big| W = w\right] \sim N_{q+p}\left(\boldsymbol{\xi}_* = \begin{pmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\xi}\end{pmatrix},\quad w\Omega_* = \begin{pmatrix} w\Gamma & w\Delta^T\\ w\Delta & w\Omega\end{pmatrix}\right), \tag{17}$$

for some non-negative random variable W with cdf G. Some properties and examples of the resulting selection elliptical subfamily are considered in Arellano-Valle et al. (2006). For instance, we note that any selection random vector $\mathbf{X} \overset{d}{=} (\mathbf{V}|\mathbf{U} \in C)$ that follows from (17) can be represented as $(\mathbf{X}|W = w) \sim SLCT\text{-}N_{p,q}(\boldsymbol{\xi}_*, w\Omega_*, C)$, where $W \sim G$. In particular, for the pdf of $\mathbf{X}$ we have $f_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty f_{\mathbf{X}|W=w}(\mathbf{x})\mathrm{d}G(w)$, where $f_{\mathbf{X}|W=w}(\mathbf{x})$ is the selection normal pdf (11) with Ω_* replaced by $w\Omega_*$, i.e.,

$$f_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty \phi_p(\mathbf{x}; \boldsymbol{\xi}, w\Omega)\frac{\int_{\mathbf{u}\in C}\Phi_q(\mathbf{u}; \boldsymbol{\gamma} + \Delta^T\Omega^{-1}(\mathbf{x}-\boldsymbol{\xi}), w\{\Gamma - \Delta^T\Omega^{-1}\Delta\})\mathrm{d}\mathbf{u}}{\int_{\mathbf{u}\in C}\Phi_q(\mathbf{u}; \boldsymbol{\gamma}, w\Gamma)\mathrm{d}\mathbf{u}}\mathrm{d}G(w).$$

Similarly, the mgf of $\psi(\mathbf{X}-\mathbf{a}) = (\mathbf{X}-\mathbf{a})^T A(\mathbf{X}-\mathbf{a})$ can be computed as

$$M_{\psi(\mathbf{X}-\mathbf{a})}(t) = \int_0^\infty M_{\psi(\mathbf{X}-\mathbf{a})|W=w}(t)\mathrm{d}G(w),$$

where, by Proposition 1, with (Γ, Δ, Ω) replaced by $(w\Gamma, w\Delta, w\Omega)$, we obtain for the conditional mgf $M_{\psi(\mathbf{X}-\mathbf{a})|W=w}(t)$ that

$$\begin{aligned} M_{\psi(\mathbf{X}-\mathbf{a})|W=w}(t) &= M_{\psi(\mathbf{V}-\mathbf{a})|W=w}(t)\\ &\times\frac{\int_{\mathbf{u}\in C}\phi_q(\mathbf{u}; \boldsymbol{\gamma}+2tw\Delta^T(I_p-2twA\Omega)^{-1}A(\mathbf{a}-\boldsymbol{\xi}), w\Gamma+2tw^2\Delta^T(I_p-2twA\Omega)^{-1}A\Delta)\mathrm{d}\mathbf{u}}{\int_{\mathbf{u}\in C}\phi_q(\mathbf{u}; \boldsymbol{\gamma}, w\Gamma)\mathrm{d}\mathbf{u}}, \end{aligned} \tag{18}$$

and, by Lemma 1,

$$M_{\psi(\mathbf{V}-\mathbf{a})|W=w}(t) = \frac{\exp\{t(\mathbf{a}-\boldsymbol{\xi})^T(I_p - 2twA\Omega)^{-1}A(\mathbf{a}-\boldsymbol{\xi})\}}{|I_p - 2twA\Omega|^{1/2}}$$

since $(\mathbf{V}|W = w) \sim N_p(\boldsymbol{\xi}, w\Omega)$.

Thus, we can extend the results given in Proposition 2 for the selection normal distribution to the more general subclass of selection elliptical distributions defined above.

Proposition 4 *Let $\psi(\mathbf{X}-\mathbf{a}) = (\mathbf{X}-\mathbf{a})^T A(\mathbf{X}-\mathbf{a})$, where $(\mathbf{X}|W = w) \sim SLCT\text{-}N_{p,q}(\mathbf{0}, w\Omega_*, C)$ for some non-negative random variable $W \sim G$, fixed column vector $\mathbf{a} \in \mathbb{R}^p$, and fixed symmetric $p\times p$ matrix A of rank r. Then, $\psi(\mathbf{X}-\mathbf{a}) \overset{d}{=} \psi(\mathbf{V}-\mathbf{a})$, where $(\mathbf{V}|W = w) \sim N_p(\boldsymbol{\xi}, w\Omega)$, under each of the conditions (i)–(iii) of Proposition 2.*

Proof Under each of the conditions (i)–(iii) given in Proposition 2 we have in (18) that $M_{\psi(\mathbf{X}-\mathbf{a})|W=w}(t) = M_{\psi(\mathbf{V}-\mathbf{a})|W=w}(t)$ for all t and each $w > 0$. Thus, $\psi(\mathbf{X}-\mathbf{a}) \stackrel{d}{=} \psi(\mathbf{V}-\mathbf{a})$, where $\mathbf{V}$ is such that $(\mathbf{V}|W = w) \sim N_p(\boldsymbol{\xi}, w\Omega)$. □

From Proposition 4 , we can also extend the result in Corollary 1 as follows.

Corollary 2 *Let $\psi(\mathbf{X}-\mathbf{a}) = (\mathbf{X}-\mathbf{a})^T A(\mathbf{X}-\mathbf{a})$, where $(\mathbf{X}|W = w) \sim$ SLCT-$N_{p,q}(\boldsymbol{\xi}_*, w\Omega_*, C)$, for some non-negative random variable $W \sim G$, fixed column vector $\mathbf{a} \in \mathbb{R}^p$, and fixed symmetric $p \times p$ matrix A of rank r. If $\mathbf{a}$ and A are such that $A(\mathbf{a}-\boldsymbol{\xi}) = \mathbf{0}$ and $A\Omega A = A$, then, under each of the conditions (i)–(iii) of Proposition 2, we have $\psi(\mathbf{X}-\mathbf{a}) = \psi(\mathbf{X}-\boldsymbol{\xi}) \stackrel{d}{=} WS$, where $S \sim \chi^2_r$ and is independent of W.*

Proof Note first by applying the condition $A(\boldsymbol{\xi}-\mathbf{a}) = \mathbf{0}$ in (18) that

$$M_{\psi(\mathbf{X}-\mathbf{a})|W=w}(t) = |I_p - 2twA\Omega|^{-1/2} \frac{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, w\Gamma + 2tw^2\Delta^T(I_p - 2twA\Omega)^{-1}A\Delta)d\mathbf{u}}{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, w\Gamma)d\mathbf{u}},$$

where from the conditions that A has rank r and $A\Omega A = A$, $|I_p - 2twA\Omega|^{-1/2} = (1-2wt)^{-r/2}$. Note now that for each of the conditions (i)–(iii) of Proposition 2 we have

$$\frac{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, \Gamma + 2t\Delta^T(I_p - 2tA\Omega)^{-1}A\Delta)d\mathbf{u}}{\int_{\mathbf{u}\in C} \phi_q(\mathbf{u}; \boldsymbol{\gamma}, \Gamma)d\mathbf{u}} = 1,$$

and so $M_{\psi(\mathbf{X}-\mathbf{a})|W=w}(t) = (1-2wt)^{-r/2}$, i.e., $(\psi(\mathbf{X}-\mathbf{a})|W = w) \stackrel{d}{=} wS$, where $S \sim \chi^2_r$ does not dependent on w. Hence $\psi(\mathbf{X}-\mathbf{a}) \stackrel{d}{=} WS$, where W and S are independent. Finally, we note that the condition $A(\mathbf{a}-\boldsymbol{\xi}) = \mathbf{0}$ implies also that $\psi(\mathbf{X}-\mathbf{a}) = \psi(\mathbf{X}-\boldsymbol{\xi})$. □

Additional properties of this subclass of selection elliptical models can be explored for particular choices of C. For example, when $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \boldsymbol{\gamma}\}$, we have the stochastic representation (19) presented in Lemma 2 below. In such case, the resulting selection distributions belong to the SUE family considered by Arellano-Valle and Azzalini (2006). The special case of the multivariate skew-t distribution that follows when G is the inverse Gamma $IG(\nu/2, \nu/2)$ distribution is studied in some details by these authors.

Lemma 2 *Let $\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} \in C)$, where $[(\mathbf{U}^T, \mathbf{V}^T)^T|W = w] \sim N_{q+p}(\boldsymbol{\xi}_*, w\Omega_*)$ for some non-negative random variable W with cdf G. If $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \boldsymbol{\gamma}\}$, then*

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\xi} + \sqrt{W}\mathbf{X}_N, \tag{19}$$

where $\mathbf{X}_N \sim SUN_{p,q}(\mathbf{0}, \Omega_*)$ and is independent of W.

Proof By (17), $(\mathbf{U}^T, \mathbf{V}^T)^T \stackrel{d}{=} (\boldsymbol{\gamma}^T + \sqrt{W}\mathbf{U}_N^T, \boldsymbol{\xi}^T + \sqrt{W}\mathbf{V}_N^T)^T$, where $(\mathbf{U}_N^T, \mathbf{V}_N^T)^T \sim N_{q+p}(\mathbf{0}, \Omega_*)$ and is independent of W. Thus, since $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \boldsymbol{\gamma}\}$ we have by (P2) and the independence between W and $(\mathbf{U}_N^T, \mathbf{V}_N^T)^T$ that

$$\mathbf{X} \stackrel{d}{=} (\mathbf{V}|\mathbf{U} > \boldsymbol{\gamma}) \stackrel{d}{=} \left(\boldsymbol{\xi} + \sqrt{W}\mathbf{V}_N|\boldsymbol{\gamma} + \sqrt{W}\mathbf{U}_N > \boldsymbol{\gamma}\right)$$
$$\stackrel{d}{=} \boldsymbol{\xi} + \sqrt{W}(\mathbf{V}_N|\mathbf{U}_N > \mathbf{0}) \stackrel{d}{=} \boldsymbol{\xi} + \sqrt{W}\mathbf{X}_N,$$

where $\mathbf{X}_N \stackrel{d}{=} (\mathbf{V}_N|\mathbf{U}_N > \mathbf{0}) \sim SUN_{p,q}(\mathbf{0}, \Omega_*)$ and is independent of W. □

For the subfamily of selection distributions defined by (19), the results in Proposition 4 and Corollary 2 are more direct, since

$$\psi(\mathbf{X} - \boldsymbol{\xi}) \stackrel{d}{=} W\psi(\mathbf{X}_N),$$

where $\psi(\mathbf{X}_N) = \mathbf{X}_N^T A\mathbf{X}_N$ and is independent of W. Moreover, since $\mathbf{X}_N \sim SUN_{p,q}(\mathbf{0}, \Omega_*)$ we have $\psi(\mathbf{X}_N) \sim \chi_r^2$, where r is the rank of A, under each of the following conditions:

(a) $A = \Omega^{-1}$, and Γ and $\Delta^T\Omega^{-1}\Delta$ are diagonal matrices.
(b) $A\Delta = O$ and $A\Omega A = A$.

6 Application to sample variogram and covariogram estimators

We study the application of the previous invariance results to sample variogram and covariogram estimators in spatial statistics. Kim et al. (2004) have shown that skew-normal processes are useful for spatial prediction of rainfall. Allard and Naveau (2007) have proposed a spatial skew-normal random field based on the observation that skewness is often present in geostatistical data. The construction of their field is made via a selection mechanism as described in this paper. This motivates the small simulation study reported below.

Let $\{X(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$, $d \geq 1$, be a second-order stationary spatial process in a region D. The estimator of the variogram of this process, based on a sample $\mathbf{X} = (X(\mathbf{s}_1), \ldots, X(\mathbf{s}_p))^T$, is a quadratic form given by

$$\psi_1(\mathbf{X}; \mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (X(\mathbf{s}_i) - X(\mathbf{s}_j))^2 = \mathbf{X}^T A(\mathbf{h})\mathbf{X}, \tag{20}$$

where $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h} \in D\}$ and $A(\mathbf{h})$ is a spatial design matrix of the data at lag $\mathbf{h}$; see Gorsich et al. (2002), and Hillier and Martellosio (2006). Similarly, the sample covariogram estimator is a quadratic form given by

$$\psi_2(\mathbf{X}; \mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (X(\mathbf{s}_i) - \bar{X})(X(\mathbf{s}_j) - \bar{X}) = \mathbf{X}^T A(\mathbf{h})\mathbf{X}, \tag{21}$$

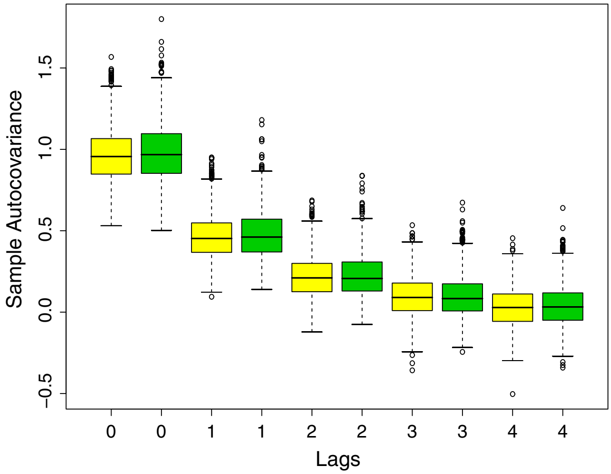

Fig. 1 Boxplots of sample autocovariance estimates at various lags under the normal (*yellow light shading*) and skew-normal (*green dark shading*) distributions

where $\bar{X} = \frac{1}{p}\sum_{i=1}^{p} X(\mathbf{s}_i)$ and $A(\mathbf{h})$ is a spatial design matrix of the data at lag $\mathbf{h}$. When $d = 1$, the expression (21) reduces to the sample autocovariance estimator of a second-order time series process given by the quadratic form

$$\psi_3(\mathbf{X}; h) = \frac{1}{p-h}\sum_{i=1}^{p-h}(X(i+h)-\bar{X})(X(i)-\bar{X}) = \mathbf{X}^T A(h)\mathbf{X}, \qquad (22)$$

where $A(h) = MD(h)M$, $D(0) = I_p$, $D(h) = (1/[2(p-h)])(P(h) + P(h)^T)$, $P(h) = \sum_{i=h}^{p-1} \mathbf{e}_i\mathbf{e}_i^T$, $\mathbf{e}_i = (0,\ldots,0,1_i,0,\ldots,0)^T \in \mathbb{R}^p$, and $M = I_p - J_p$ is the centering matrix, $J_p = (1/p)\mathbf{1}_p\mathbf{1}_p^T$, $M^2 = M$, $\mathrm{tr}\{M\} = p-1$.

In the definition of the above statistics, we are assuming that $\mathbf{a} = \boldsymbol{\xi}$, and $\boldsymbol{\xi} = \xi\mathbf{1}_p$ due to stationarity. Because the matrix $A(\mathbf{h})$ satisfies the condition $A(\mathbf{h})\mathbf{1}_p = \mathbf{0}$, we have $\psi(\mathbf{X}-\boldsymbol{\xi}) = \psi(\mathbf{X})$. By assumption $A(\mathbf{h})(\mathbf{a}-\boldsymbol{\xi}) = \mathbf{0}$. Thus, according to Proposition 3, for the SUE class of distributions, a condition to obtain invariance for any of the above statistics is that $A(\mathbf{h})\Delta = O$. Since, by assumption, $A(\mathbf{h})\mathbf{1}_p = \mathbf{0}$, the condition $A(\mathbf{h})\Delta = O$ holds for example when $\Delta = \delta\mathbf{1}_p\mathbf{1}_q^T$ or, more generally, when $\Delta = \mathbf{1}_p\boldsymbol{\delta}^T$, where $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_q)^T$. However, according to Proposition 4, $A(\mathbf{h})\Delta = O$ yields invariance for the scale mixtures of the selection normal model whatever the selection subset C.

Next, we illustrate the previous results for $d = 1$, i.e., in the context of time series. We simulate 1,000 samples of size $p = 100$ from a normal and a skew-normal (Azzalini and Dalla Valle 1996) distribution, i.e., $\mathbf{X} \sim N_p(\mathbf{0}, \Sigma)$ and $\mathbf{X} \sim SN_p(\mathbf{0}, \Sigma, \boldsymbol{\alpha})$, where Σ is the correlation matrix of an AR(1) process with correlation 0.5 and $\boldsymbol{\alpha} = (1 - \boldsymbol{\delta}^T\Sigma^{-1}\boldsymbol{\delta})^{-1}\Sigma^{-1}\boldsymbol{\delta}$ with $\boldsymbol{\delta} = 0.15\,\mathbf{1}_p$. On each sample, we estimate the autocovariance with the estimator $\psi_3(\mathbf{X}; h)$ for $h = 0, 1, 2, 3, 4$. Figure 1 depicts boxplots of the estimates at those lags under the normal (yellow light shading) and skew-normal (green dark shading) distributions. It can be appreciated that the distribution of the sample autocovariance estimator does not change in presence of skewness for that class of distributions.

7 Discussion

We have studied conditions under which an invariance property holds for the class of selection distributions. In particular, we have focused on the distribution of quadratic forms and described the implication of the invariance on sample variogram and covariogram estimators in spatial statistics. We have illustrated the latter situation by means of a small simulation study.

The invariance property studied in this paper has also implications in other important settings. For example, consider a linear model (multiple regression) $\mathbf{X} = \boldsymbol{\xi} + \mathbf{e}$, with $\boldsymbol{\xi} = Z\boldsymbol{\beta}$, where Z is a $p \times k$ known matrix of rank k, $\boldsymbol{\beta} \in \mathbb{R}^k$ is a vector of unknown parameters and $\mathbf{e} \in \mathbb{R}^p$ is an error random vector. For the classical normal linear model, we have $\mathbf{e} \sim N_p(\mathbf{0}, \sigma^2 I_p)$. Let $\mathbf{z}_1, \ldots, \mathbf{z}_k$ be the columns of Z an suppose that $\mathbf{z}_1 = \mathbf{1}_p$, i.e., $\boldsymbol{\xi} = Z\boldsymbol{\beta} = \beta_1\mathbf{1}_p + \beta_2\mathbf{z}_2 + \cdots + \beta_k\mathbf{z}_k$. Let also $\hat{\boldsymbol{\xi}} = P\mathbf{X}$ and $\hat{\mathbf{e}} = \mathbf{X} - \hat{\boldsymbol{\xi}} = (I_p - P)\mathbf{X}$, where $P = Z(Z^TZ)^{-1}Z^T$, be the classical linear predictor and residual vectors, respectively, and consider the well known variance decomposition:

$$\underbrace{\sum_{i=1}^{p}(X_i - \bar{X})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{p}(\hat{\xi}_i - \bar{X})^2}_{\text{SSR}} + \underbrace{\sum_{i=1}^{p}\hat{e}_i^2}_{\text{SSE}}$$

$$\Updownarrow$$

$$\|\mathbf{X} - \bar{X}\mathbf{1}_p\|^2 = \|\hat{\boldsymbol{\xi}} - \bar{X}\mathbf{1}_p\|^2 + \|\hat{\mathbf{e}}\|^2$$

$$\Updownarrow$$

$$\underbrace{\mathbf{X}^T(I_p - J_p)\mathbf{X}}_{\psi_4(\mathbf{X})} = \underbrace{\mathbf{X}^T(P - J_p)\mathbf{X}}_{\psi_5(\mathbf{X})} + \underbrace{\mathbf{X}^T(I_p - P)\mathbf{X}}_{\psi_6(\mathbf{X})},$$

where SST, SSR, and SSE represent the total, regression, and error sums of squares, respectively. Thus, the three statistics ψ_4, ψ_5 and ψ_6 are quadratic forms for which $\mathbf{a} = \boldsymbol{\xi}$ and the associated matrix A satisfies the condition $A\mathbf{1}_p = \mathbf{0}$, since $PZ = Z$ and so $P\mathbf{1}_p = \mathbf{1}_p$. Hence, by Proposition 4, if we consider the class of selection linear models obtained as scale mixtures of the normal linear model defined above, with skewness matrix of the form $\Delta = \mathbf{1}_p\boldsymbol{\delta}^T$, for some $\boldsymbol{\delta} \in \mathbb{R}^q$, then the distribution of $\psi_i(\mathbf{X})$, $i = 4, 5, 6$, is neither affected by the skewness parameter nor by the selection subset C. Within that class of selection linear models, we have also that the Fisher statistic is robust (in distribution). In fact, since this statistic is given by

$$T(\mathbf{X}) = \left(\frac{p-k}{k-1}\right)\left(\frac{\mathbf{X}^T(P - J_p)\mathbf{X}}{\mathbf{X}^T(I_p - P)\mathbf{X}}\right) = \left(\frac{p-k}{k-1}\right)\left(\frac{\psi_5(\mathbf{X})}{\psi_6(\mathbf{X})}\right),$$

we have that it satisfies the invariance by the common rescaling condition $T(a\mathbf{x}) = T(\mathbf{x})$, for all $a > 0$, and so $T(\mathbf{X}) \overset{d}{=} T(\mathbf{V}) \overset{d}{=} T(\mathbf{V}_0)$, for all $\mathbf{V} \sim EC_p(Z\boldsymbol{\beta}, \sigma^2 I_p, h^{(p)})$ random vectors, where $\mathbf{V}_0 \sim N_p(Z\boldsymbol{\beta}, \sigma^2 I_p)$, see, e.g., Fang et al. (1990). In particular, under the null hypothesis $H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$, we have $T(\mathbf{X}) \sim F_{k-1,p-k}$. The extension of this result to the more general Fisher statistic given by

$$T(\mathbf{X}) = \left(\frac{p-k}{k-k_0}\right)\left(\frac{\mathbf{X}^T(P-P_0)\mathbf{X}}{\mathbf{X}^T(I_p-P)\mathbf{X}}\right),$$

is straightforward. Here, P_0 is a $p \times p$ projection matrix, with $\mathrm{tr}(P_0) = k_0$ $(1 \le k_0 < k)$ and $P_0 P = P_0$. In fact, if we consider the condition that $P_0\Delta = \Delta$, then we have that $P\Delta = \Delta$, and so that $(P - P_0)\Delta = O$ and $(I_p - P)\Delta = O$. In particular, if the skewness matrix has the form $\Delta = \mathbf{1}_p\boldsymbol{\delta}^T$, then the above invariance condition reduces to $P_0\mathbf{1}_p = \mathbf{1}_p$. In such situations, the distribution of $T(\mathbf{X})$ will be invariant within the full class of selection elliptical distributions when the selection subset $C = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{u} > \boldsymbol{\gamma}\}$ (see Proposition 3), and within the normal scale mixtures subclass whatever the selection subset C. In such settings, more general null hypotheses of the form $H_0 : B\boldsymbol{\beta} = \mathbf{0}$, where B is a $(k - k_0) \times k$ matrix of rank $k - k_0$, can be tested by using the fact that, under H_0, $T(\mathbf{X}) \sim F_{k-k_0,p-k}$. The aforementioned invariances follow from the fact that in such a class of models all the components of the error term have a common random scale factor (see, e.g., Breusch et al. 1997).

Another well known statistic that has the same invariance property under the conditions considered above is the Durbin–Watson (DW) statistic. It is used in time series for testing the null hypothesis that there is not serial correlation in the error terms $e_1, \ldots, e_p$ and is given by

$$\mathrm{DW} = \frac{\sum_{i=2}^{p}(\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^{p}\hat{e}_i^2} = \frac{\mathbf{X}^T(I_p-P)D^T D(I_p-P)\mathbf{X}}{\mathbf{X}^T(I_p-P)\mathbf{X}},$$

where D is the $(p-1)\times p$ difference matrix such that $D\mathbf{x} = (x_2-x_1, \ldots, x_p-x_{p-1})^T$. For this, the assumption that the e_is follow an AR(1) process is incorporated in the linear model considered above. In the spatial setting, the analogues to the Durbin–Watson statistic are Moran's I and Geary's c. Both of them are defined by means of quadratic forms and therefore will possess the invariance property described in this paper.

Acknowledgments The authors would like to thank the Editor, the Associate Editor, and a referee for comments that improved this manuscript. Several remarks by Adelchi Azzalini are also greatfully acknowledged. This research was partially supported by a grant FONDECYT 1040865/7070137-Chile, by a Swiss National Foundation grant, and by NSF grants DMS-0504896 and CMG ATM-0620624.

References

Allard, D., Naveau, P. (2007). A new spatial skew-normal random field model. *Communications in Statistics Theory and Methods*, *36*, 1821–1834.

Arellano-Valle, R. B., Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, *33*, 561–574.

Arellano-Valle, R. B., Branco, M. D., Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *The Canadian Journal of Statistics*, *34*, 581–601.

Arellano-Valle, R. B., del Pino, G. (2004). From symmetric to asymmetric distributions: a unified approach. In M. G. Genton, (Ed.), *Skew-elliptical distributions and their applications*: *a journey beyond normality* (pp. 113–130). Boca Raton, FL: Chapman & Hall/CRC.

Arellano-Valle, R. B., del Pino, G., San Martín, E. (2002). Definition and probabilistic properties of skew-distributions. *Statistics and Probability Letters*, *58*, 111–121.

Arellano-Valle, R. B., Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, *96*, 93–116.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171–178.

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society Series B*, *61*, 579–602.

Azzalini, A., Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t* distribution. *Journal of the Royal Statistical Society Series B*, *65*, 367–389.

Azzalini, A., Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715–726.

Breusch, T. S., Robertson, J. C., Welsh, A. H. (1997). The emperor's new clothes: a critique of the multivariate *t* regression model. *Statistica Neerlandica*, *51*, 269–286.

Fang, K.-T., Kotz, S., Ng, K.-W. (1990). Symmetric multivariate and related distributions. Monographs on statistics and applied probability, vol 36. London: Chapman and Hall.

Genton, M. G. (Ed.) (2004). *Skew-elliptical distributions and their applications*: *a journey beyond normality*. Boca Raton, FL: Chapman & Hall/CRC.

Genton, M. G., He, L., Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statistics and Probability Letters*, *51*, 319–325.

Genton, M. G., Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, *57*, 389–401.

Gorsich, D. J., Genton, M. G., Strang, G. (2002). Eigenstructures of spatial design matrices. *Journal of Multivariate Analysis*, *80*, 138–165.

Hillier, G., Martellosio, F. (2006). Spatial design matrices and associated quadratic forms: structure and properties. *Journal of Multivariate Analysis*, *97*, 1–18.

Khatri, C. G. (1980). Quadratic forms in normal variables. In P. R. Krishnaiah, (Ed.), *Handbook of statistics* (Vol. 1, pp. 443–469). Amsterdam: North-Holland.

Kim, H.-M., Ha, E., Mallick, B. K. (2004). Spatial prediction of rainfall using skew-normal processes. In M. G. Genton, (Ed.), *Skew-elliptical distributions and their applications*: *a journey beyond normality* (pp. 279–289). Boca Raton, FL: Chapman & Hall/CRC.

Wang, J., Boyer, J., Genton, M. G. (2004a). A skew-symmetric representation of multivariate distributions. *Statistica Sinica*, *14*, 1259–1270.

Wang, J., Boyer, J., Genton, M. G. (2004b). A note on an equivalence between chi-square and generalized skew-normal distributions. *Statistics and Probability Letters*, *66*, 395–398.