Skew-Normal Time Series Models With Nonlinear Heteroscedastic Predictors

Mohsen Pourahmadi
pourahm@math.niu.edu
Division of Statistics
Northern Illinois University

Abstract: Multivariate skew-normal (SN) distributions (Azzalini and Dalla Valle, 1996) have nonlinear heteroscedastic predictors but lack a **closure property** (sum of independent SN random variables is not SN) of normal distributions needed for introducing moving average and spectral representations of stationary SN processes. We study SN stationary processes and construction of SN ARMA models for various SN noises and show that their finite-dimensional distributions are skew-normal, seldom strictly stationary and their covariance functions differ from their normal ARMA counterparts in that they do not converge to zero for large lags. Using the Wold decomposition we characterize SN stationary processes as sum of independent Gaussian stationary processes and half-normal random variables. Their predictors are nonlinear and prediction variances are nonconstant, but explicit multi-index functions of the past variables. The piecewise linear appearances of their predictors are similar to those from threshold autoregressive models. The outline of a correlogram-based statistical model-fitting process for, and the role of an invariance property of SN ARMA models are reviewed.

Some key words: ARMA models; multi-index functions; nonnormal distributions; threshold autoregressive models; regression functions; shape parameters; strict and covariance stationarity; Wold decomposition.

1 Introduction

There is a well-developed theory of prediction and statistical model-fitting process for the class of linear Gaussian autoregressive and moving average (ARMA) models

$$X_t = \sum_{k=1}^{p} \beta_k X_{t-k} + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j}, \tag{1}$$

where $\{\varepsilon_t\}$ is a Gaussian white noise, i.e. a sequence of independent and identically distributed $N(0, \sigma^2)$ random variables, and $\beta_1, \cdots, \beta_q, \theta_0 = 1, \theta_1, \cdots, \theta_q$ are the AR and MA coefficients of the model (Brockwell and Davis, 1991; Box, Jenkins and Reinsel, 1994; Pourahmadi, 2001). Limitations of linearity and normality of ARMA models (Tong, 1990, Sec. 1.4.2), have generated considerable interest in nonnormal time series models with nonlinear and heteroscedastic predictors.

The first popular approach to nonnormal ARMA models which proceeds by retaining the linear ARMA models (1), but replacing the normal noise $\{\varepsilon_t\}$ by a nonnormal noise (Jacobs and Lewis, 1977; Lawrance and Lewis, 1985; Pourahmadi, 1988; McKenzie, 2003), usually leads to nonnormal marginals, but predictors that are linear with constant variance. The second approach based on the class of threshold autoregressive (TAR) models, introduced by H. Tong in 1977, abandons (1) and starts with conditionally linear AR models which lead to nonnormal marginals and nonlinear heteroscedastic predictors (Tong,1990).

In this paper we show that retaining (1) but replacing the noise $\{\varepsilon_t\}$ by certain SN distributions could lead to time series models having best features of the above two approaches. A recurrent theme in the first approach is that of finding a suitable distribution for the noise $\{\varepsilon_t\}$ for a given marginal (joint) distribution for $\{X_t\}$ and vice versa(Tong, 1990, Section 4.2.6). However, other than the normal (stable) case the distribution of the noise $\{\varepsilon_t\}$ is often quite different from that of $\{X_t\}$, though recently Tarami and Pourahmadi (2003) have shown that $\{X_t\}$ and $\{\varepsilon_t\}$ both can have Student-t distributions if the noise is allowed to be non-white (exchangeable). This departure from the standard white noise is becoming prevalent in the literature of nonlinear time series (Wu and Min, 2004). In particular, we show that ARMA models driven by correlated SN noises can have SN marginals with nonlinear and heteroscedastic predictors. Recall that Gaussian stationary processes can be constructed either via moving averages of Gaussian white noises or the Fourier transform of Gaussian processes with orthogonal increments (Pourahmadi, 2001, Chap. 5). Such constructions rely heavily on the following **closure property** of normal distributions: *linear combinations of independent normal random variables are normally distributed*. Unfortunately, this property does not hold for SN random variables. Thus, to emulate the success of Gaussian stationary processes we need to modify either the definition of the white noise or that of the multivariate SN (Azzalini and Dalla Valle, 1996; González-Farías et al. 2004). We modify a white noise by either taking correlated SN variables or non-identically distributed SN variables and show that in either case the solution of (1), i.e. the distribution of $\mathbf{X}_n = (X_1, \cdots, X_n)$ is SN. Also working within the family of closed skew-normal (CSN) distributions introduced by González-Farías et al. (2004) and using genuine CSN white noise in (1) we show that $\mathbf{X}_n$ remains in the CSN class, but at the additional cost of model and computational complexity at the estimation stage.

The outline of the paper is as follows: In Section 2, we present the simplest TAR model and review regression properties of multivariate SN and elliptically contoured distributions. In Section 3, we show that linear ARMA models (1) with exchangeable SN-distributed noise have SN stationary solutions if the shape vector of $\{\varepsilon_t\}$ is a scalar multiple of the identity vector and the roots of their AR characteristic polynomials are outside the unit circle. We use the stochastic representation of SN variables to obtain the spectral and Wold decompositions of a solution of (1). We also study aspects of ARMA models with CSN distributions. In Section 4, the basics of prediction problems and the correlation curves are discussed. Section 5 outlines the statistical model fitting process for the SN ARMA models and the appropriateness of correlogram-based model formulation and diagnostic. We conclude the paper in Section 6.

Just like the theory of nonlinear time series (Tong, 1990), our results on SN ARMA models are fragmentary at best and do not provide a coherent framework like the familiar one for Gaussian ARMA models (Brockwell and Davis, 1991). Though much work remains to be done, our partial and preliminary results in Section 3 do reveal that the success of any framework for SN ARMA models may hinge on having a suitable and flexible class of multivariate SN distributions.

2 Models with Nonlinear Heteroscedastic Predictors

Most useful features of real-life phenomena exhibiting nonlinear and nonconstant variation over time and space cannot be captured using Gaussian processes or linear models. In this section, the simple TAR models and multivariate SN distributions are introduced and their roles in providing prototypical models capturing the above features are reviewed.

2.1 Connection between the simplest TAR Model and SN Distribution

In general, it is difficult to write closed form formulae for the predictor and stationary distribution of a simple nonlinear AR model like $X_t = f(X_{t-1}) + \varepsilon_t$, where $f : R \to R$ is a known function. Fortunately, when f is the piecewise linear function,

$$f(x) = \begin{cases} \alpha x & \text{if } x \geq 0, \\ \\ \beta x & \text{if } x < 0, \end{cases} \tag{2}$$

one obtains a two-regime TAR model and necessary and sufficient conditions on the coefficients for stationarity and ergodicity of its solution are $\alpha < 1, \beta < 1$ and $\alpha\beta < 1$ (Petruccelli and Woolford, 1984), see Tong(1990) for a thorough review of results on the more general TAR models. Surprisingly, closed formulae for the stationary solutions of such simple TAR models have been found when the noise distributions are normal, Cauchy and Laplace (Andel et al. 1984; Andel and Barton, 1986; Loges, 2004). More specifically , when $\beta = -\alpha$, for the simplest TAR model

$$X_t = \alpha |X_{t-1}| + \varepsilon_t, \tag{3}$$

with $\{\varepsilon_t\}$ a $N(0,1)$-noise, Andel, Netuka and Svara (1984) have shown that its stationary distribution is of the form

$$2\phi\left(\frac{x}{\sigma_\alpha}\right)\Phi(\alpha x), \tag{4}$$

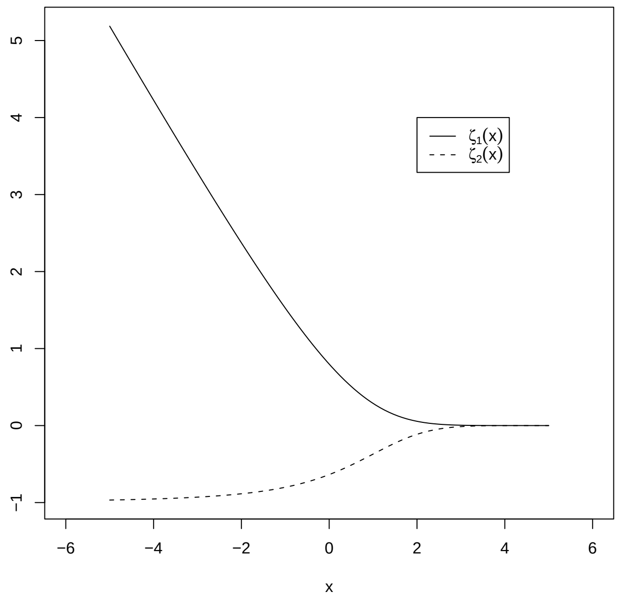

Figure 1: Plots of $\zeta_i(\cdot), i = 1, 2$.

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution functions of the standard normal random variable and $\sigma_\alpha = (1-\alpha^2)^{-1/2}$, see Tong (1990, Section 4.2) for more details. This encouraging early result may suggest deeper connections between densities of the form (4) and stationary distributions of simple threshold ARMA models like

$$X_t = \alpha|X_{t-1}| + \varepsilon_t + \theta|\varepsilon_{t-1}|. \tag{5}$$

A systematic study of densities of the form (4) were initiated by Azzalini (1985) who christened them skew-normal densities with skewness parameter α. Note that when $\alpha = 0$, then (4) reduces to the standard normal density function. A **stochastic representation** (Henze, 1986; Azzalini, 1986) of a standard SN random variable X as

$$X = \delta|Z| + (1-\delta^2)^{1/2}\varepsilon, \tag{6}$$

where Z and ε are independent $N(0,1)$ random variables and $|\delta| < 1$, plays a central role in this paper. It reveals the close connection among the SN, normal and half-normal random variables and the stationary distribution of (3) and (5) and their extensions.

2.2 Multivariate Skew-Normal Distributions

Multivariate skew-normal distributions defined by Azzalini and co-authors (Azzalini and Dalla Valle, 1996; Azzalini and Capitanio, 1999; Capitanio, Azzalini and Stanghellini, 2003) have density functions of the form

$$f(y) = \phi_k(y-\mu;\Sigma)\Phi\left(\alpha_0 + \boldsymbol{\alpha}' D^{-1}(y-\mu)\right)/\Phi(\tau), y\epsilon R^k, \tag{7}$$

with $\phi_k(y-\mu;\Sigma)$ the density of a k-dimensional $N_k(0,\Sigma), \mu$ a vector of location parameters, $\Sigma = (\sigma_{ij})$ a positive-definite matrix, $D = \text{ diag}(\sigma_{11},\cdots,\sigma_{kk})^{1/2}$ and $\boldsymbol{\alpha}\epsilon R^k$ is arbitrary and where τ is an additional parameter and α_0 as a function of (Σ,α,τ) is specified in (10) below. Note that $\tau = 0$ implies $\alpha_0 = 0$, see also Arnold and Beaver (2000). This class clearly extends the multivariate normal distributions through the vector parameter $\boldsymbol{\alpha}$ which regulates the shape of the distribution, and for $\boldsymbol{\alpha} = 0$ it reduces to the class of normal distributions.

A random vector Y with the density function (7) is said to have an *(extended) skew-normal* distribution with parameters (μ,Σ,α,τ), in symbol we write $Y \sim SN_k(\mu,\Sigma,\alpha,\tau)$. A stochastic representation of a random vector Y with distribution (7) is helpful in simulating its values and revealing the meaning of its new parameter τ. Following Capitanio et al. (2003), consider the $(k+1)$-dimensional normal random vector

$$W = (X_0, X_1,\cdots,X_k)' = \begin{pmatrix} X_0 \\ X \end{pmatrix} \sim N_{k+1}(0,R_W), \tag{8}$$

where

$$R_W = \begin{pmatrix} 1 & \delta' \\ \delta & R \end{pmatrix} \tag{9}$$

is a full-rank correlation matrix. Then, the probability density function of $Z = (X|X_0+\tau>0)$ is of the form $SN_k(0,R,\alpha,\tau)$ where

$$\alpha_0 = \tau(1-\delta' R^{-1}\delta)^{-1/2}, \alpha = (1-\delta' R^{-1}\delta)^{-1/2}R^{-1}\delta. \tag{10}$$

For later use, note that (10) can also be written as

$$\alpha_0 = \tau(1+\alpha' R\alpha)^{1/2}, \delta = (1+\alpha' R\alpha)^{-1/2}R\alpha. \tag{11}$$

For generality, one may introduce location and scale parameters for Z via the affine transformation $Y = \mu + DZ$, where $\mu\epsilon R^k$ and D is a $k\times k$ diagonal matrix with positive diagonal entries. Then, the density function of Y is of the form (7) with $\Sigma = DRD$ and cumulant generating function

$$K(t) = \mu' t + \frac{1}{2}t'\Sigma t + \zeta_0(\tau + \delta' Dt) - \zeta_0(\tau), t\epsilon R^k,$$

where $\zeta_0(u) = \log 2\Phi(u)$. Evaluating the first and second derivatives of $K(\cdot)$ at $t = 0$ gives

$$E(Y) = \mu + \zeta_1(\tau)D\delta,\ \text{Cov}\ (Y) = \Sigma + \zeta_2(\tau)D\delta\delta' D, \tag{12}$$

where δ is as in (11) and $\zeta_m(\cdot)$ is the mth derivative of $\zeta_0(\cdot)$. Since the mean and covariance of Y are related to its location μ, scale Σ, shape parameter α, from (12) it is evident that departures from μ and Σ are controlled by the vector $D\delta$ and the two univariate functions

$$\zeta_1(x) = \frac{\phi(x)}{\Phi(x)}, \zeta_2(x) = -\zeta_1(x)\left\{x + \zeta_1(x)\right\}, \tag{13}$$

where $\zeta_1(x)$ is always positive and from (13) it can be seen that $\zeta_2(x) < 0$ for $x > 0$. However, $\zeta_1(\cdot)$ is strictly decreasing (Gordon, 1941) and hence for all $x\epsilon R$, we have $\zeta_2(x) < 0$, so that the covariance matrix in (12) is always "smaller" than Σ, its multivariate normal counterpart. Plots of $\zeta_1(x)$ and $\zeta_2(x)$ in Figure 1 show their shapes and the degrees of departures from their normal regression counterparts. Of particular importance is the "hockey-stick" appearance of the plot of $\zeta_1(\cdot)$ or its near piecewise linearity with signs of the local slopes depending on the size of the threshold variable x which changes at about $x = 1$. This clearly suggests the appropriateness of distributions (7) and their progenies (Genton, 2004; Arellano-Valle and Azzalini, 2004) as models for situations where the conditional expectations and variances are known to be nonlinear and heteroscedastic functions.

2.3 Multi-Index Heteroscedastic Regressions

To obtain the regression and variance functions of the SN distribution in (7), we partition the random vector Y and its parameters conformally as

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \tag{14}$$

where Y_1 and Y_2 are of dimensions k_1, k_2 with $k_1 + k_2 = k$. We also partition R, the correlation matrix of Σ, similarly and set $R_{22.1} = R_{22} - R_{21}R_{11}^{-1}R_{12}$. Then, it is known (Capitanio et al., 2003) that

$$Y_1 \sim SN_{k_1}(\mu_1, \Sigma_{11}, \alpha_{1(2)}, \tau),\ \alpha_{1(2)} = \frac{\alpha_1 + R_{11}^{-1}R_{12}\alpha_2}{(1 + \alpha_2' R_{22.1}\alpha_2)^{1/2}}, \tag{15}$$

and

$$(Y_2|Y_1 = y_1) \sim SN_{k_2}(\mu_{2.1}, \Sigma_{22.1}, \alpha_{2.1}, \tau_{2.1}), \tag{16}$$

where

$$\alpha_{2.1} = D_{22.1}D_2^{-1}\alpha_2,\ \tau_{2.1} = \tau(1 + \alpha_{1(2)}' R_{11}\alpha_{1(2)})^{1/2} + \alpha_{1(2)}' D_1^{-1}(y_1 - \mu_1), \tag{17}$$

and D_1, D_2 and $D_{22.1}$ are diagonal matrices constructed from the square roots of the diagonal entries of Σ_{11}, Σ_{22} and $\Sigma_{22.1}$, respectively.

From (12) and (16), we obtain that

$$\begin{cases} E(Y_2|Y_1 = y_1) = \mu_{2.1} + \zeta_1(\tau_{2.1}) D_{22.1}\delta_{2.1}, \\ \text{Cov}(Y_2|Y_1 = y_1) = \Sigma_{22.1} + \zeta_2(\tau_{2.1}) D_{22.1}\delta_{2.1}\delta'_{2.1} D_{22.1}, \end{cases} \tag{18}$$

where $\delta_{2.1}$ is related to $\alpha_{2.1}$, the shape parameter in (16), via

$$\delta_{2.1} = (1 + \alpha'_{2.1} R_{22.1}\alpha_{2.1})^{-1/2} R_{22.1}\alpha_{2.1}. \tag{19}$$

Note that the regression function in (18) is a double-index function of the covariates in y_1 (Härdle, Hall and Ichimura, 1993; Xia, Tong, Li and Zhu, 2002), the first index $\mu_{2.1}$ is simply the regression function from a $N_k(\mu, \Sigma)$ and the second index $\tau_{2.2}$ is determined by $(\mu, \Sigma, \alpha_0\boldsymbol{\alpha})$ in (7); the conditional variance is a single-index function of $\tau_{2.1}$. Such multi-index models are becoming increasingly popular in statistics and econometrics as exploratory tools when looking for possible low-dimensional structures to approximate high-dimensional regression surfaces, see Xia et al. (2002), Härdle et al. (1993) and the references therein. The next example illustrates the formulae in (18).

Example 1. For a bivariate skew-normal distribution with $k = 2, k_1 = k_2 = 1$,

$$\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 1/2, \alpha = (1, 2)', \tau = 1,$$

the expressions in (15) - (20) reduce to

$$\alpha_{1(2)} = \frac{2}{\sqrt{5}},\ \alpha_{2.1} = \sqrt{3},\ \delta_{2.1} = \sqrt{3/2} \text{ and } \tau_{2.1} = (3 + 2y_1)/\sqrt{5}.$$

Thus, the regression and variance functions are given by

$$\begin{aligned} E(Y_2|Y_1 = y_1) &= \frac{1}{2}y_1 + \frac{3}{4}\zeta_1\left(\frac{5 + 2y_1}{\sqrt{5}}\right), \\ \text{Var}(Y_2|Y_1 = y_1) &= \frac{3}{4} + \frac{9}{16}\zeta_2\left(\frac{5 + 2y_1}{\sqrt{5}}\right). \end{aligned} \tag{20}$$

2.4 Elliptically Contoured Distributions

Comparing (18) and (20) with their bivariate normal counterparts it is evident that the linearity of regression and constancy of variance are violated mainly because of presence of $\zeta_1(\tau_{2.1})$ and

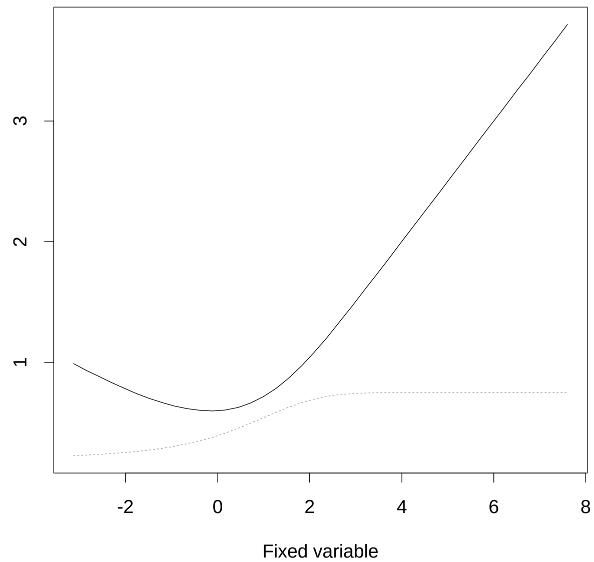

Figure 2: The regression function (solid) and the variance function (dotted) for Example 1.

$\zeta_2(\tau_{2.1})$. These which are nonlinear, multi-index functions (Härdle et al. 1993; Xia et al. 2002) of the fixed variable y_1, see Figure 2. The situation is somewhat different for the class of elliptically contoured distributions (Anderson, 2003, Sec. 2.7) where the regression is linear in the fixed variables y_1 just as in the multivariate normal case, but the conditional covariance matrix has the multiplicative form

$$m(y_1)\Sigma_{22.1}, \tag{21}$$

for a scalar function $m(\cdot)$. The function $m()$ is known (Cambanis et al., 2000) only when Y is a scale mixture of a normal vector, that is $Y = A^{1/2}G$ with A a positive random variable independent of $G \sim N_k(0, \Sigma)$, in this case $m(y_1) = E(A|Y_1 = y_1)$. The latter computed explicitly in the following two cases is a single-index function of the quadratic form $y_1' \sum_{11}^{-1} y_1$.

Example 2. (a) Scale mixture of normal with $A^{-1} \sim \chi^2_\nu, \nu > 0$. Then, Y has a multivariate Student-t distribution (Box et al., 1994, p. 286) and $m(\cdot)$ is given by

$$m(y_1) = \frac{\nu + y_1' \sum_{11}^{-1} y_1}{k + \nu - 2}, y_1 \epsilon R^{k_1}.$$

Thus, the conditional covariance matrix (21) is less than its normal counterpart $\Sigma_{22.1}$ only when the fixed variables y_1 are inside the ellipsoid $y_1'\Sigma_{11}^{-1}y_1 \leq k - 2$, for $k > 2$.

(b) Scale mixture of normal with $A \sim \Gamma(\nu, 1), \nu > 0$, a gamma distribution.

Recently, Fotopoulos (2004) has shown that

$$m(y_1) = (y_1' \Sigma_{11}^{-1} y_1)^{1/2} \frac{K_{\nu+1+k/2}\left((2y_1'\Sigma_{11}^{-1}y_1)^{1/2}\right)}{\sqrt{2}K_{\nu+k/2}\left((2y_1'\Sigma_{11}^{-1}y_1)^{1/2}\right)},$$

where $K_\nu(\cdot)$ is the modified Bessel function with $\nu \epsilon R$.In this case it is much harder to compare (21) with its normal counterpart.

The correlation curves for these cases are presented in Section 4.2.

3 SN Stationary Time Series

In this section, a general SN stationary time series is defined in analogy with the definition of Gaussian stationary processes. Two notions of SN noise are introduced and the probabilistic and statistical properties of their corresponding SN stationary ARMA models are studied and compared to their Gaussian counterparts. The bulk of the work is on the dependent SN noise with distribution (7), but for the sake of comparison we also discuss the role of the CSN noise in defining ARMA models.

3.1 General SN Stationary Process

A set of time-ordered random variables $X_t; t = 0, \pm 1, \pm 2, \cdots$, is said to be *SN stationary* if for every integer $n \geq 1$, and integers $t_1, \cdots, t_n$, the n-dimensional random vector $\mathbf{X}_n = (X_{t_1}, \cdots, X_{t_n})$ has an $SN_n(\mu_n, \Sigma_n, \boldsymbol{\alpha}_n, \tau)$ distribution as in (7), and for all integers t, k
$(S1)$ $E(X_t) \equiv \mu$, does not depend on t;
$(S2)$ Cov $(X_{t+k}, X_t) \equiv \gamma_k$, depends only on k but not t.
In this case, the function $\{\gamma_k; k = 0, 1, 2, \cdots\}$ is called the *autocovariance function* of $\{X_t\}$ and the $n \times n$ covariance matrix of $\mathbf{X}_n$, denoted by $\Gamma_n = (\gamma_{i-j})_{i,j=1}^n$, has a Toeplitz structure (with constant entries along diagonals). Using (12), one could express $E(\mathbf{X}_n)$ and Cov$(\mathbf{X}_n)$ in terms of the location μ_n, scale Σ_n, shape $\boldsymbol{\alpha}_n$ and τ. Intuitively, for $\{X_t\}$ to be SN and strictly stationary, at least all the shape parameters and variances of X_t's must be the same, this forces $\boldsymbol{\delta}_n$ and D_n in (12) to be of the form $\boldsymbol{\delta}_n = \delta 1_n$, $D_n = \gamma_0 I_n$, and hence the scale matrix $\sum_n$ must be a Toeplitz matrix.

Existence of general SN processes is guaranteed by noting that the family of finite-dimensional distributions $SN_n(\mu_n, \sum_n, \boldsymbol{\alpha}_n, \tau)$ in (7) or their moment generating functions satisfy the consistency condition of the Kolmogorov's theorem (Brockwell and Davis, 1991, p.11). However, construction of SN stationary processes via either their spectral or moving average representations requires having a suitable notion of white noise. In fact, starting with any i.i.d. SN random variables $\{\varepsilon_t\}$, one can construct general linear processes in the usual manner (Pourahmadi, 2001, Sec. 5.2). For instance, the one-sided $MA(\infty)$ process

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{22}$$

is well-defined (the series is convergent in the mean) if $\sum_{j=0}^{\infty} |\psi_j| < \infty$, and the process $\{X_t\}$ is strictly stationary with mean and covariance function given by

$$E(X_t) = \mu_\varepsilon \sum_{j=0}^{\infty} \psi_j, \ \text{Cov}\ (X_{t+k}, X_t) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_{j+k}\psi_j,$$

where μ_ε and σ_ε^2 are the mean and variance of ε_t. However, the distribution of $\mathbf{X}_n = (X_1, \cdots, X_n)$ is not SN unless $\{\varepsilon_t\}$ has the aforementioned closure property.

To study the distribution of $\mathbf{X}_n$, its covariance matrix and stationarity of the solution of ARMA models, we rely extensively on the fact that $\mathbf{X}_n$ is an affine transformation of the noise $\{\varepsilon_t\}$ and the closure under affine transformations of the SN distributions (Capitanio et al.,2003; González-Farías et al., 2004). More precisely, writing the n equations (1) for $t = 1, \cdots, n$, in terms of the vectors $\mathbf{X}_n, \boldsymbol{\varepsilon}_n = (\varepsilon_1, \cdots, \varepsilon_n)$ and the $(p+q)$-dimensional vector of initial values $\mathbf{e} = (X_{1-p} \cdots, X_0, \varepsilon_{1-q}, \cdots, \varepsilon_0)$ gives

$$L_\beta \mathbf{X}_n = L_\theta \boldsymbol{\varepsilon}_n + L_{p,q}\, \mathbf{e}\, , L_{p,q} = \begin{pmatrix} L_{p,\beta} & L_{q,\theta} \\ 0 & 0 \end{pmatrix}, \tag{23}$$

where for a vector $\beta = (\beta_1, \cdots, \beta_p), L_\beta$ is an $n \times n$ unit lower triangular matrix with 1's on the leading diagonal, β_1 on the first subdiagonal, β_2 on the second subdiagonal, and so on, with 0's on the ith subdiagonal for $i > p$, and

$$L_{p,\beta} = \begin{pmatrix} \beta_p & \beta_{p-1} & \cdots & \cdots & \beta_1 \\ 0 & \beta_p & \cdots & \cdots & \beta_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \beta_p \end{pmatrix},$$

L_θ and $L_{q,\theta}$ are defined similarly for a vector $\theta = (\theta_1, \cdots, \theta_q)$, see Box et al. (1994, p.292). Multiplying both sides of the first equation in (23) by L_β^{-1}, we obtain

$$\mathbf{X}_n = L_\beta^{-1} L_\theta \boldsymbol{\varepsilon}_n + L_\beta^{-1} L_{p,q}\, \mathbf{e} = A \begin{pmatrix} \boldsymbol{\varepsilon}_n \\ \mathbf{e} \end{pmatrix}. \tag{24}$$

Thus, identifying an appropriate distribution for $\begin{pmatrix} \boldsymbol{\varepsilon}_n \\ \mathbf{e} \end{pmatrix}$ holds the key to find the distribution of $\mathbf{X}_n$. This topic is studied in Section 3.3 after introducing two notions of SN noises.

3.2 The Noise Distributions

Perhaps, the single most important reason for popularity of the normal white noise can be traced to the fact that if $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. $N(0, \sigma^2)$, then their *joint* distribution is multivariate normal, i.e. $\boldsymbol{\varepsilon}_n = (\varepsilon_1, \cdots, \varepsilon_n) \sim N_n(0, \sigma^2 I_n)$, and consequently every linear transformation of $\boldsymbol{\varepsilon}_n$ is normally distributed. Unfortunately, this closure property does not hold for i.i.d. SN random variables, since product of several density functions of the form (4) is not of the form (7). Thus, one needs either to modify the requirements of $\varepsilon_1, \cdots, \varepsilon_n$ being i.i.d. or the definition of multivariate SN in (7). We discuss the pros and cons of working with these possibilities in the next two subsections.

3.2.1 The SN Colored Noise

There are two ways to weaken the requirement of $\varepsilon_1, \cdots, \varepsilon_n$ being i.i.d., either take them to be *dependent* or *non-identically distributed*. For the first option, start with the simplest *jointly* SN distributed noise, i.e. $\boldsymbol{\varepsilon}_n \sim SN_n(0, \sigma^2 I_n, \boldsymbol{\alpha}_n, \tau)$. Then, it follows from (12) that

$$E(\boldsymbol{\varepsilon}_n) = \sigma\zeta_1(\tau)\boldsymbol{\delta}_n, Cov\ (\boldsymbol{\varepsilon}_n) = \sigma^2 I_n + \sigma^2\zeta_2(\tau)\boldsymbol{\delta}_n\boldsymbol{\delta}_n', \tag{25}$$

where

$$\boldsymbol{\delta}_n = (1 + \boldsymbol{\alpha}_n'\boldsymbol{\alpha}_n)^{-1/2}\boldsymbol{\alpha}_n, \tag{26}$$

so that the mean and covariance of $\boldsymbol{\varepsilon}_n$ satisfy neither (S1) nor (S2) so long as the entries of $\boldsymbol{\alpha}_n$ or $\boldsymbol{\delta}_n$ are distinct. One may rectify this problem by setting $\boldsymbol{\alpha}_n = \alpha 1_n$ where 1_n is the n-vector of 1's and α is a scalar. Then, (25) reduces to

$$E(\boldsymbol{\varepsilon}_n) = \sigma\delta\zeta_1(\tau)1_n, Cov(\boldsymbol{\varepsilon}_n) = \sigma^2 I_n + \sigma^2\delta^2\zeta_2(\tau)1_n 1_n', \tag{27}$$

which satisfy both (S1) and (S2) for a fixed n, where $\delta = \delta(n, \alpha) = \alpha(1 + n\alpha^2)^{-1/2}$. Note that this noise is *correlated*, since Cov $(\boldsymbol{\varepsilon}_n)$ is a compound symmetric matrix with

$$\rho = \frac{\delta^2\zeta_2(\tau)}{1 + \delta^2\zeta_2(\tau)},$$

which is always negative because $\zeta_2(\tau) < 0$. From here on we refer to

$$\boldsymbol{\varepsilon}_n \sim SN_n(0, \sigma^2 I_n, \alpha 1_n, \tau), \tag{28}$$

as a *(colored) SN-noise*.

The simplest example of non-identically distributed noise is obtained by taking ε_i's independent, only one having the SN distribution (6) and the rest normally distributed. Using the Wold decomposition and stochastic representation of such SN random variables, we show in Section 3.3.3. that these two distinct modifications of a white noise lead to the same class of SN stationary process.

3.2.2 The CSN White Noise

To overcome the lack of closure property of SN distributions, the class of multivariate closed skew-normal (CSN) distributions were introduced by González-Farías et al.(2004). Let $k \geq 1, \ell \geq 1$ be integers, $\mu \epsilon R^k, \nu \epsilon R^\ell$, D an arbitrary $\ell \times k$ matrix, Σ and Δ positive definite matrices of dimensions $k \times k$ and $\ell \times \ell$, respectively. Then, we say that a k-vector Y has a CSN-distribution, in symbol $Y \sim CSN_{k,\ell}(\mu, \Sigma, D, \nu, \Delta)$, if its density function is given by

$$f_{k,\ell}(y) = \phi_k(y; \mu, \Sigma)\Phi_\ell(D(y-\mu); \nu, \Delta)/\Phi_\ell(0; \nu, \Delta + D\Sigma D'), \tag{29}$$

where $\phi_k(\cdot\ ; \mu, \Sigma)$ and $\Phi_k(\cdot\ ; \mu, \Sigma)$ are the density and distribution functions of a k-dimensional normal with the indicated mean vector and covariance matrix. González-Farías et al. (2004) show that the CSN family subsumes many skew-normal distributions, including (7) which is obtained by taking $\ell = 1, \nu = 0$ and $\boldsymbol{\alpha} = \Delta^{-1/2}D'$. For a discussion of lack of identifiability of the class of CSN see Arellano-Valle and Azzalini (2004).

Now, we define the CSN white noise as the random variables

$$\varepsilon_1, \cdots, \varepsilon_n \; i.i.d. \sim CSN_{1,1}(0, \sigma^2, \alpha, 0, 1). \tag{30}$$

From the closure property of such distributions(González-Farías et al. 2004) it follows that

$$\boldsymbol{\varepsilon} = (\varepsilon_1 \; \cdots, \varepsilon_n) \sim CSN_{n,n}(0, \sigma^2 I_n, \alpha\, I_n, 0, I_n), \tag{31}$$

and for given $\psi_1, \cdots, \psi_n$ we have

$$\sum_{j=1}^n \psi_j \varepsilon_j \sim CSN_{1,n}\left(0, \sigma^2 \sum_{j=1}^n \psi_j^2, D_\psi, 0, \Delta_\psi\right), \tag{32}$$

where

$$D_\psi = \alpha \left(\sum_{j=1}^n \psi_j^2\right)^{-1} (\psi_1, \cdots, \psi_n)' = \alpha \left(\sum_{j=1}^n \psi_j^2\right)^{-1} \boldsymbol{\psi}_n,$$

$$\Delta_\psi = I_n + \alpha^2\sigma^2 \left[I_n - \left(\sum_{j=1}^n \psi_j^2\right)^{-1} \boldsymbol{\psi}_n \psi_n'\right].$$

The fact that the second dimension in (31)-(32) grows with n when stacking up n i.i.d. CSN random variables or forming their linear combinations is of major concern and seems to be the price one has to pay to attain a closure property. Evidently, this has unpleasant computational implications when defining infinite order MA models as in (22) or CSN ARMA as in (39) below.

3.3 Construction of Stationary SN ARMA Models

3.3.1 The SN ARMA Models

In studying the distribution of data from ARMA models, the initial values $\mathbf{e}$ in (24) can be deterministic or stochastic. For the moment, we assume that $\mathbf{e}$ is stochastic and

$$\begin{pmatrix} \boldsymbol{\varepsilon}_n \\ \\ \mathbf{e} \end{pmatrix} \sim SN_{n+p+q}\left(0, \Omega, \boldsymbol{\alpha}, \tau\right), \tag{33}$$

for a scale matrix Ω. In most of what follows we take Ω to be the covariance matrix of $\begin{pmatrix} \boldsymbol{\varepsilon}_n \\ \mathbf{e} \end{pmatrix}$ corresponding to the Gaussian stationary case of (1), i.e.

$$\Omega = \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & Cov(\mathbf{e}) \end{pmatrix}, Cov(\mathbf{e}) = \sigma^2 \begin{pmatrix} \sigma^{-2}\Gamma_p & C' \\ C & I_q \end{pmatrix},$$

where Γ_p is the $p \times p$ matrix with the (i,j)th element $\gamma_{i-j} = Cov(X_i, X_j)$, and $\sigma^2 C$ has elements defined by $Cov(\varepsilon_{i-q}, X_{j-p}) = \sigma^2 \psi_{j-i-p+q}$ for $j-i-p+q \geq 0$, and 0 otherwise (Box et al. 1994, p. 293). The ψ_j's are the coefficients in the MA(∞) representation of $\{X_t\}$, that is

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\beta(z)}. \tag{34}$$

Then, from (24) and the closure property of SN under affine transformations (Capitanio et al. 2003, p. 143), it follows that

$$\mathbf{X}_n \sim SN_n(0, \Sigma_X, \boldsymbol{\alpha}_X, \tau),$$

where

$$\Sigma_X = A\Omega A',\ \boldsymbol{\alpha}_X = \frac{D_X \Sigma_X^{-1} B' \boldsymbol{\alpha}}{[1 + \boldsymbol{\alpha}'(R_\Omega - B\Sigma_X^{-1}B')\boldsymbol{\alpha}]^{1/2}}, \tag{35}$$

with $D_X^2 =$ diag $(\Sigma_X), B = D_X^{-1}\Omega A'$ and R_Ω is the correlation matrix of Ω.

The nonconstancy of the shape parameter $\boldsymbol{\alpha}_X$ in (35) suggests that $\mathbf{X}_n$ cannot, in general, be a segment of a strictly stationary SN process. To study the covariance stationarity of $\{X_t\}$, however, it is prudent to start with the MA(q) models for which the matrix A in (24) is much simpler, the scale matrix Ω in (32) is a multiple of the identity and the subsequent computations are more illuminating.

Example 3. SN MA$(\mathbf{q})$ **Model.** In (33) setting $p = 0$, we note that $\begin{pmatrix} \boldsymbol{\varepsilon}_n \\ \mathbf{e} \end{pmatrix} = \boldsymbol{\varepsilon}_{n+q}\ \Omega = \sigma^2 I_{n+q}$ and $\boldsymbol{\alpha} = \alpha 1_{n+q}$. Thus, it follows from (27) and (24) that

$$\begin{aligned} E(\mathbf{X}_n) &= \sigma\delta(n+q,\alpha)(\textstyle\sum_{j=0}^q \theta_j)1_n, \\ \text{Cov}(\mathbf{X}_n) &= A\left[\sigma^2 I_{n+q} + \sigma^2\delta^2\zeta_2(\tau)1_{n+q}1'_{n+q}\right]A' \\ &= \sigma^2 AA' + \sigma^2\delta^2\zeta_2(\tau)\left(\textstyle\sum_{j=0}^q \theta_j\right)^2 1_n 1'_n \\ &= \sigma^2\Gamma_{n,MA(q)} + \sigma^2\delta^2\zeta_2(\tau)\left(\textstyle\sum_{j=0}^q \theta_j\right)^2 1_n 1'_n, \end{aligned} \tag{36}$$

where $\Gamma_{n,MA(q)}$ stands for the $n \times n$ segment of the stationary covariance matrix of a normal MA(q) model. We note that Cov$(\mathbf{X}_n)$ is a Toeplitz matrix and, in fact, a rank-1 perturbation of the $MA(q)$ covariance matrix to which it reduces when either $\sum_{j=0}^q \theta_j = 0$ or $n+q \to \infty$. Since $\delta(n+q,\alpha) = \frac{\alpha}{[1+(n+q)\alpha^2]^{-1/2}} \to 0$, in the latter case. However, from (36) with $c = \delta^2\zeta_2(\tau)\left(\sum_{j=0}^q \theta_j\right)^2$, it follows that the autocorrelation function (ACF) of an SN MA(q) is of the form

$$\rho_k = \frac{\gamma_k + c}{\gamma_0 + c}, k = 0, 1, 2, \cdots, \tag{37}$$

which does not vanish for $k > q$, in contradistinction to the familiar Gaussian $MA(q)$ models. Next, we compute the shape parameter of $\mathbf{X}_n$. Since

$$\Sigma_X = A\Omega A' = \sigma^2 AA' = \Gamma_{n,MA(q)}, R_\Omega = I_{n+q},$$
$$D_X = \sigma\left(\textstyle\sum_{j=0}^q \theta_j^2\right)^{1/2} I_n, B = \sigma\left(\textstyle\sum_{j=0}^q \theta_j^2\right)^{-1/2} A',$$

we have from (35) that

$$\boldsymbol{\alpha}_X = d^{-1/2}\alpha\left(\textstyle\sum_{j=0}^q \theta_j\right)(AA')^{-1} 1_n, \tag{38}$$

where

$$d = 1 + \alpha^2\left[n + q - \left(\textstyle\sum_{j=0}^q \theta_j^2\right)^{-1}\left(\textstyle\sum_{j=0}^q \theta_j\right)^2 1'_n (AA')^{-1} 1_n\right].$$

Thus, the entries of $\boldsymbol{\alpha}_X$ in (38) are generally distinct and the components of $\mathbf{X}_n$ are not identically distributed. A notable exception is when the MA parameters satisfy $\sum_{j=0}^q \theta_j = 0$, in which case $\boldsymbol{\alpha}_X = 0$ and $\mathbf{X}_n$ has a multivariate normal distribution even when the noise does not.

Example 4. SN ARMA $(\mathbf{p}, \mathbf{q})$ **Model.** For $p > 0$, as before we set the scale matrix Ω in (33) to be the covariance of $\begin{pmatrix} \boldsymbol{\varepsilon}_n \\ \mathbf{e} \end{pmatrix}$ in the normal stationary case. From (35), using this Ω and a general shape vector $\boldsymbol{\alpha}$, one could compute $\boldsymbol{\alpha}_X, E(\mathbf{X}_n)$ and $Cov(\mathbf{X}_n)$ as in (36)-(38). Unfortunately, these formulas are complicated and not as illuminating as those in Example 3. However, more transparent results can be obtained using the Wold decomposition of SN ARMA models described in Section 3.3.3, see (42).

3.3.2 The CSN ARMA Models

In this section, we introduce CSN ARMA models using the fact that these are affine transformations of CSN white noises. In particular, for a CSN MA(q) we have $\mathbf{X}_n = A\boldsymbol{\varepsilon}_{n+q}$, and from the closure property of CSN under affine transformations (González-Farías et al., 2004), it follows from (31) that

$$\mathbf{X}_n \sim CSN_{n,n+q}(0, \sigma^2 AA', D_A, 0, \Delta_A), \tag{39}$$

where

$$D_A = \alpha A'(AA')^{-1}, \Delta_A = I_{n+q} + \alpha^2\sigma^2(I_{n+q} - H),$$

with $H = A'(AA')^{-1}A$, the projection matrix on the row space of A.

In principle, one could use (39) for prediction, and estimation of the MA parameters $\theta_1, \cdots, \theta_q$, σ^2 and the skewness parameter α. However, a concern in using the CSN ARMA models is the tremendous increase in computations when n or q is large, as this involves computing distribution functions of $(n+q)$-dimensional normals at several points. The nature of this distribution when $n+q \to \infty$, and the associated computational complexity need further research and are the key ingredients in determining the usefulness of these models. It seems the choice between SN and CSN ARMA models may eventually boil down to the trade-off between the computational complexity and the uneasy feeling of working with a colored SN noise leading to ARMA models with ACFs like (37) which do not converge to zero for large lags.

Next, using the representation (6) we show that SN ARMA models are sums of independent Gaussian ARMA models and a multiple of a half-normal random variable, so that they are much closer to Gaussian ARMA models than are the CSN ARMA models.

3.3.3 The Wold and Spectral Decompositions of SN Stationary Processes

The reason for the peculiar form of the ACF of an SN MA(q) in (36)-(37) can be traced to the exchangeability of the SN colored noise in (28) and hence the presence of a common random component. Using the stochastic representation of SN random variables, we single out and identify this common component as a multiple of a half-normal variable and present a Wold decomposition of stationary solutions of SN ARMA models. The latter reveals the close connection between SN and Gaussian ARMA models which is used to compute the covariance functions and the infinite predictors of SN stationary models.

Let us start with the stochastic representation of n SN variables ε_i as

$$\varepsilon_i = \delta_i |Z_i| + (1-\delta_i^2)^{1/2}\, \varepsilon_i',\ i = 1, \cdots, n, \tag{40}$$

where Z_i and ε_i' are independent $N(0,1)$ random variables and δ_i's are fixed with $|\delta_i| \leq 1$. We note that when all the δ_i and Z_i are equal, then the ε_i's in (40) are, indeed, the colored SN noise in (28); and when all δ_i's are zero except for one, then the ε_i's are the non-identically distributed noise of Section 3.2.1. Substituting from (40) in (1), gives the process

$$X_t = \sum_{j=0}^{q} \theta_j \delta_j |Z_j| + \sum_{k=1}^{p} \beta_k X_{t-k} + \sum_{j=0}^{q} \theta_j \sqrt{1-\delta_j^2}\, \varepsilon_{t-j}', \tag{41}$$

whose marginal distribution is not SN unless all the Z_j's are the same which we assume from here on. Existence of a covariance stationary solution of (41) is guaranteed if all the roots of its AR characteristic polynomial lie outside the unit circle. Then, the Wold decomposition of the stationary solution of (41) and a corresponding decomposition of its covariance function (Pourahmadi, 2001, pp. 65-66) are given by

$$\begin{cases} X_t = \left(\sum_{j=0}^{q} \theta_j \delta_j\right) |Z| + \sum_{j=0}^{\infty} \psi_j\, \varepsilon_{t-j}' \\ \\ \text{Cov}\,(X_{t+k}, X_t) = \left(\sum_{j=0}^{q} \theta_j \delta_j\right)^2 \sqrt{\frac{2}{\pi}} + \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \end{cases} \tag{42}$$

where the ψ_j's are obtained from (34) replacing θ_j by $\theta_j \sqrt{1-\delta_j^2},\ j = 0, 1, \cdots, q$. Note that the ACF in (42) does not converge to zero as $k \to \infty$, unless $\sum_{j=0}^{q} \theta_j \delta_j = 0$ in which case $\{X_t\}$ is a Gaussian ARMA process. Thus, (42) decomposes an SN ARMA process in terms of a Gaussian ARMA process and a static half-normal process.

The Wold decomposition of a general SN stationary process $\{X_t\}$ is closely related to that of a Gaussian stationary process. This can be seen best through its spectral representation (Pourahmadi, 2001, pp. 30-31):

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dZ(\lambda),$$

where the process $\{Z(\lambda); -\pi < \lambda \leq \pi\}$ has *orthogonal increments* (Brockwell and Davis, 1991, Section 4.6; Pourahmadi, 2001, pp. 168-169). To allow some generality, we take $\{Z(\lambda)\}$ to have the form

$$dZ(\lambda) = |Z|dF(\lambda) + g(\lambda)dZ'(\lambda),$$

which is slightly different from (40), $F(\cdot)$ is a bounded increasing function, $g(\cdot)$ is a function on $(-\pi, \pi]$ and $\{Z'(\lambda)\}$ is a Gaussian orthogonal increment process and independent of Z. It is easy to verify that the process $\{Z(\lambda); -\pi < \lambda \leq \pi\}$ has orthogonal increments with finite-dimensional SN distributions (7). Its Fourier transform gives the SN process

$$X_t = |Z| \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda) \; + \; \int_{-\pi}^{\pi} e^{it\lambda} g(\lambda) \, dZ'(\lambda), \tag{43}$$

which is not stationary unless dF has a point mass, say one, at a fixed point λ_0, and the function g is so that the second integral in (43) is well-defined. Then,

$$X_t = |Z| e^{i\lambda_0 t} + X'_t, \tag{44}$$

where

$$X'_t = \int_{-\pi}^{\pi} e^{it\lambda} g(\lambda) dZ'(\lambda), t = 0, \pm 1, \cdots,$$

is a Gaussian stationary process. Using the Wold decomposition of $\{X'_t\}$ in (44) one obtains the most general linear representation of SN stationary processes (Pourahmadi, 2001, p.163).

4 Prediction and Correlation Curves

Computing the finite predictors and their prediction variances for a nonnormal time series $\{X_t\}$ are known to be challenging problems (Wiener, 1958; Shepp et al. 1980). In this section we present explicit formulas for these in the case of SN stationary time series and discuss the need for studying local dependence or correlation curve since their prediction variances are heteroscedastic.

4.1 Prediction of SN Stationary Processes

We start with the linear prediction of a second-order stationary time series $\{X_t\}$ to introduce the notation and ingredients needed for the nonlinear prediction. Let $\hat{X}_{n+1}$ be the linear least-squares predictor of X_{n+1} based on its finite past $\mathbf{X}_n = (X_1, \cdots, X_n)$ and σ^2_{n+1} be its prediction error variance. For computing $\hat{X}_{n+1}$ and σ^2_{n+1}, we need Γ_{n+1}, the $(n+1) \times (n+1)$ Toeplitz covariance matrix of $(\mathbf{X}_n, X_{n+1})$, and $\mu_X = E(X_t)$ the mean of the stationary series $\{X_t\}$. In fact, it is known (Pourahmadi, 2001, Chap. 7) that

$$\hat{X}_{n+1} = \mu_X + \boldsymbol{\gamma}'_n \Gamma_n^{-1} (\mathbf{X}_n - \mu_X \mathbf{1}_n)$$

$$\sigma^2_{n+1} = Var(X_{n+1} - \hat{X}_{n+1}) = \gamma_0 - \boldsymbol{\gamma}'_n \Gamma_n^{-1} \boldsymbol{\gamma}_n,$$

where $\boldsymbol{\gamma}_n = (\gamma_1, \cdots, \gamma_n)$.

In computing the finite nonlinear predictor and its prediction error variance for an SN stationary series, we need the full knowledge of the following distribution:

$$(\mathbf{X}_n, X_{n+1}) \sim SN_{n+1}(\boldsymbol{\mu}_{n+1}, \Sigma_{n+1}, \boldsymbol{\alpha}_{n+1}, \tau),$$

where $\boldsymbol{\mu}_{n+1} = \mu_X \mathbf{1}_{n+1} - \zeta_1(\tau) D_{n+1}\delta_{n+1}$ and $\Sigma_{n+1} = \Gamma_{n+1} - \zeta_2(\tau) D_{n+1}\delta_{n+1}\delta'_{n+1} D_{n+1}$. Using (14)-(17) with $k = n+1, k_1 = n, k_2 = 1, Y_1 = \mathbf{X}_n$ and $Y_2 = X_{n+1}$, explicit expressions for the optimal predictor $E(X_{n+1}|\mathbf{X}_n = \mathbf{x})$ and its prediction error variance $Var(X_{n+1}|\mathbf{X}_n = \mathbf{x})$ can be obtained from (18). To illustrate these, in the rest of this section we focus on the SN MA(q) model of Example 3 and compute $\alpha_{1(2)}, \alpha_{2.1}$ and $\tau_{2.1}$, and express them in terms of the past data and quantities familiar from the linear prediction. We note that from (36)-(38), for an MA(q) model we have

$$\sum\nolimits_{n+1} = \sigma^2 AA' = \sigma^2 \sum_{j=0}^{q} \theta_j^2 \cdot R = \sigma^2 \sum_{j=0}^{q} \theta_j^2 \cdot \begin{pmatrix} R_{11} & R_{12} \\ R'_{21} & 1 \end{pmatrix},$$

where R is the related correlation matrix, and

$$\boldsymbol{\alpha}_{n+1} = d^{-1/2}\, \alpha(\sum_{j=0}^{q} \theta_j)(AA')^{-1}\mathbf{1}_{n+1}, \boldsymbol{\alpha}_{n+1} = \begin{pmatrix} \boldsymbol{\alpha}_n \\ \alpha_{n+1} \end{pmatrix}.$$

Then,

$$\begin{aligned} \alpha_{1(2)} &= (1+\alpha_{n+1}^2)^{-1/2}\, (\boldsymbol{\alpha}_n + \alpha_{n+1} R_{11}^{-1} R_{12}) \\ &= (1+\alpha_{n+1}^2)^{-1/2}\, (\boldsymbol{\alpha}_n + \alpha_{n+1}\mathbf{a}_n), \end{aligned}$$

where $\mathbf{a}_n = R_{11}^{-1} R_{12}$ is the vector of finite linear predictor coefficients for a time series with covariance matrix $\sum_{n+1}$ (Pourahmadi, 2001, Section 7.1),

$$\alpha_{2.1} = \alpha_{n+1}\sigma_{n+1}(\sigma^2 \sum_{j=0}^{q} \theta_j^2)^{-1/2},$$

and

$$\tau_{2.1} = \tau\sqrt{1+\alpha'_{1(2)} R_{11}\alpha_{1(2)}} + (\sigma^2 \textstyle\sum_{j=0}^{q} \theta_j^2)^{-1/2}\, (1+\alpha_{n+1}^2)^{-1/2}$$
$$(\boldsymbol{\alpha}_n + \alpha_{n+1}\, \mathbf{a}_n)'(\mathbf{X}_n - \mu_1).$$

Now, the predictor in (18) is computed by replacing $\mu_{2.1}$ by $\hat{X}_{n+1}$ and using $\tau_{2.1}$ given above. The latter expressed in terms of the finite linear predictor of a time series with the covariance matrix $\sum_n$ rather than Γ_n can also be written in terms of the finite linear predictor of $\{X_t\}$ with the right covariance matrix Γ_n. To do this we note that since $\Gamma_n = \sum_n + \sigma^2(\sum_{j=0}^{q} \theta_j^2)\zeta_2(\tau)\delta_n\delta'_n$ is a rank-one perturbation of $\sum_n$, its inverse can be computed in terms of $\sum_n^{-1}$ (Pourahmadi, 2001, p.155).

4.2 Correlation Curves

The correlation curve as a local measure of association is useful for assessing departure from normality, linearity of regression and constancy of variance (Bjerve and Doksum, 1993). Since these features are present in SN distributions, it is expected that the correlation curve here could play a role similar to that of the Galton-Pearson correlation coefficient for the multivariate normal data.

Consider the class of nonlinear heteroscedastic regressions with $\mu(x) = E(Y|X = x)$ and $\sigma^2(x) = \text{Var}\,(Y|X = x)$, where we refer to the derivative of the mean function $\beta(x) = \mu'(x)$, as the regression slope or the local regression coefficient. Then, the correlation curve between X and Y is defined (Bjerve and Doksum, 1993) by

$$\rho(x) = \frac{\sigma_X \beta(x)}{[\sigma_X^2 \beta^2(x) + \sigma^2(x)]^{1/2}},$$

where $\sigma_X^2 = Var(X)$. It is easy to rewrite $\rho(x)$ as

$$\pm \left(1 + [\sigma_X \beta(x)/\sigma(x)]^{-2}\right)^{-1/2},$$

where the sign $\pm$ is the same as the sign of $\beta(x)$, so that the correlative curve is an increasing function of the standardized regression slope $\frac{\sigma_X \beta(x)}{\sigma(x)}$. The correlation curve is calibrated so that it lies in the range $[-1, 1]$, equals ρ in the bivariate normal case and has the universal property of a correlation coefficient in being invariant under linear transformations of X and Y, and hence is scale-free and comparable across cases. It is shown by Bjerve and Doksum (1993) that this correlation curve also satisfies other axioms of a universal correlation measure, except that it is not symmetric in X and Y. Next, we compute the correlation curve for a few specific distributions.

Example 5. Bivariate Skew-Normal. From the calculations in Example 1, it follows that

$$\frac{\sigma_1 \beta(y_1)}{\sigma(y_1)} = \frac{\rho\sigma_2 + \alpha_{1(2)} C \zeta_2(\tau_{2.1})}{\sqrt{\sigma_2^2(1-\rho^2) + C^2 \zeta_2(\tau_{2.1})}}$$

$$= \frac{1 + \frac{3}{\sqrt{5}}\zeta_2(\tau_{2.1})}{\sqrt{3 + \frac{9}{4}\zeta_2(\tau_{2.1})}}.$$

The corresponding plots of the standardized regression slope and the correlation curve as a function of the fixed variable y_1 are given in Figure 3.

Example 6. Bivariate Elliptically Contoured. Since the regression is linear in this case, using $m(\cdot)$ from Example 2 we have

$$\frac{\sigma_1 \beta(y_1)}{\sigma(y_1)} = \frac{\rho}{\sqrt{(1-\rho^2)(1 + {}^{y_1^2}/_{\nu\sigma_1^2})}},$$

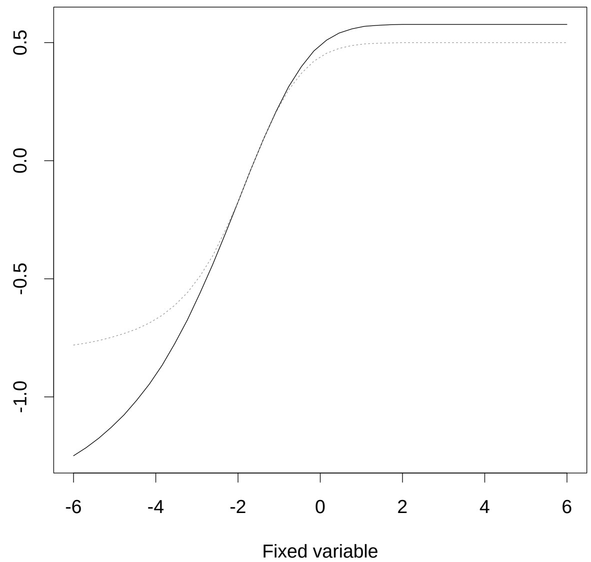

Figure 3: The standardized regression slope (solid) and the correlation curve (dotted) for Ex. 1

for (a), but the formula for $\sigma^2(y_1)$ for (b) has a much more complicated formula.

Example 7. Partial Correlation Curve. For $k_1 = k - 2, k_2 = 2, \mathbf{Y}_1 = (Y_2, \cdots, Y_{k-1}), \mathbf{Y}_2 = (Y_1, Y_k)$, the correlation coefficient computed from the 2×2 matrix in (16) gives the partial correlation curve between Y_1, Y_k after adjusting for the intervening variables in $\mathbf{Y}_1$ (Pourahmadi, 2001, Section 7.3). Note that unlike the normal case, the partial correlation curves for skew-normal and elliptically contoured distributions depend on the fixed variables y_1.

Since contours of SN distributions are not ellipsoidal, local measures of association like the (partial) correlation curves seem to be relevant in the model-formulation stage of SN ARMA models. An interesting open problem here is that of comparing the performance of these correlation curves to the traditional ACF and PACF used for model formulation of normal ARMA models discussed next.

5 The Model-Fitting Process for SN ARMA Time Series

For normal ARMA (p, q) models, there is a well-developed iterative three-stage model-fitting process cycling through model formulation, estimation and diagnostic (Box et al., 1994). In this section we discuss prospects of developing such a model-fitting process for the SN ARMA models. It is expected that each stage of the process is more challenging due to the presence of the skewness parameter α.

5.1 Model Formulation

The model formulation stage is concerned with identifying the orders p and q of the ARMA (p, q) model (1) using the time series data $X_1, \cdots, X_n$. Traditionally, this is done by comparing the theoretical patterns of the autocorrelation and partial autocorrelation functions (ACF and PACF) of ARMA (p, q) models with their sample counterparts, so far as their cut-off or decay to zero are concerned. For example, the fact that the theoretical ACF of an MA(q) model is zero after the lag q, can be used to determine q by scanning the sample ACF. However, this correlogram-based method may not work for SN ARMA models, because their ACFs have a pattern of zero different from those of the normal ARMA models. For example, it is evident from (37) that the ACF of an SN MA(q) is not zero after the lag q, but stays above (below) zero.

Fortunately, an **invariance property** of the SN distributions first noted by Genton et al. (2001), Loperfido (2001, 2004) and Genton (2004) implies that the distributions of the sample autocovariances of $X_1, \cdots, X_n$,

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(X_{t+k} - \bar{X})(X_t - \bar{X}),\ 0 \le k \le n-1,$$

with $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$, are the same for data from multivariate SN and normal distributions. Thus, the sample ACF can be used to identify the SN ARMA orders p and q as in the case of normal data. Furthermore, since the sample ACF and PACF are even functions of the data, their distributions are as well invariant to the choice of the parent distributions (Genton, 2004, pp.82-83). Therefore, correlogram-based model formulation techniques can be used verbatim in the case of SN ARMA models so long as the threshold parameter τ in (7) is set to zero. It should be noted that without this modification the invariance property does not hold. In a sense this requirement conflicts with the strong need to keep τ in (18) when forecasting is the ultimate goal of modeling the data. More research is needed to fully understand the role of the invariance property in the model formulation, estimation and diagnostics of SN ARMA models and their many extensions (Genton, 2004).

5.2 Model Estimation and Diagnostics

Once the order (p, q) of an SN ARMA model is known, the model estimation phase is concerned with estimating the parameters

$$\theta = (\beta_1, \cdots, \beta_p, \theta_1, \cdots, \theta_q, \sigma^2, \alpha).$$

The likelihood function $L(\theta|X_1, \cdots, X_n)$ can be written using (24) or (33)-(35). As expected (Azzalini, 1985; Pewsey, 2000), the major difficulty here is that of finding the maximum likelihood estimator (MLE) of the skewness parameter α. However, the coupling of α with the ARMA parameters makes the task even more daunting. For examples of coupling of these parameters in the case of SN MA (q) model, see (34), and for CSN MA(q), see (39). A prudent way to circumvent this difficulty is to employ an iterative method, alternating between the ML estimation of the ARMA parameters (Box et al. 1994, Chap.7) and the skewness parameter. Much work remains to be done on understanding the impact of estimation α on the ARMA parameters and vice versa.

Since the MLE of the parameters θ are not explicit functions of the data, it is not known whether the ACF and PACF of the model residuals are even functions of the data. Consequently, in the diagnostic stage when using the popular portmanteau lack-of-fit test (Box et al., 1994, p.314), one cannot appeal to the invariance property of SN distributions. Research on the estimation and diagnostic problems for ARMA models are currently in progress and will appear elsewhere.

5.3 Nonlinearity, Dependence and Estimation of α

Estimation of the skewness parameter is by far one of the most challenging problems in the literature of SN distributions (Azzalini, 1985; Pewsey, 2000; Genton, 2004). Fortunately, the appearance of a skewness parameter in the TAR model in (3) gives α a physical meaning and links the new difficult problem of estimation of α to the familiar problem of estimation of an autoregression coefficient.

It is well-known that the least squares estimator of α in (3) is given by

$$\hat{\alpha} = \frac{\sum_{t=2}^{n} X_t |X_{t-1}|}{\sum_{t=2}^{n} |X_{t-1}|^2}, \tag{45}$$

so that, perhaps, this is the only reasonable estimator of the skewness parameter having a closed form. Under the conditions of Theorem 2.2 in (Chan & Tsay, 1998) it can be shown that for n large

$$\sqrt{n}(\hat{\alpha} - \alpha) \sim AN(0,\ 1 - \alpha^2). \tag{46}$$

How good is this asymptotic distribution? The results of a small simulation study reported in Table 1 show that the sample mean and SD of the least squares estimator of the TAR parameter α are in good agreement with those in (46).

n	$\hat{\alpha}$	$SD(\hat{\alpha})$	$.8/\sqrt{n}$
200	0.5924	0.0559	0.0566
400	0.5979	0.0400	0.0400
600	0.6013	0.0322	0.0327
1000	0.5986	0.0260	0.0253
1600	0.6017	0.0206	0.0200
2000	0.5994	0.0181	0.0179

Table 1. Sample means and standard deviations of the least squares estimator of α in the TAR model (3) with $\alpha = 0.6$, based on 500 iterations.

Note that the estimator (45) of a skewness parameter is arrived at rather indirectly by trapping it in the model (3). The happy ending here seems to suggest that introduction of some sort of "nonlinearity" and "dependence" might be helpful in estimating the skewness parameter. The use of order statistics in estimation of α in Balakrishnan (2004) also introduces these two features. Are appearances of these features incidents or essential part of any good estimation procedure for estimating the skewness parameters?

6 Discussion and Future Work

Using dependent SN noise we have introduced a class of stationary time series models whose finite-dimensional distributions belong to the family of SN distributions studied by Azzalini and Dalla Valle (1996) and Capitanio et al. (2003). These time series are shown to be close to normal stationary time series, in that they can be decomposed as the sum of independent normal stationary series and a static half-normal series. Yet, their finite predictors are nonlinear and heteroscedastic with features similar to predictors of threshold AR models. The class of CSN stationary series (Gonzales-Farías et al. 2004) seems to be more general and natural for time series analysis, but could require heavier computation with increasing sample sizes. Statistical and computational issues related to fitting these two classes of models to data remain to be studied and these could shed some light on the question of which class should be preferred in data analysis. It seems correlogram-based model-formulation could be used equally well for the two classes, however, maximum likelihood estimation of the parameters of the CSN series could be more computationally intensive due to the presence of distribution functions of high-dimensional normal vectors. The predictor and variance functions for CSN stationary ARMA models is currently under study. The framework developed in this paper can be used to study general SN and CSN stationary random fields and vector-valued processes.

Acknowledgements. I would like to thank Peng Liu and participants of the Workshop on Skew-Elliptical Distributions (CIMAT, Mexico, Dec. 6-7, 2004) for many discussions and inputs. The work was supported by the NSF grant DMS-0307055.

References

Andel, J. and Barton, T. (1986). A note on threshold AR(1) model with Cauchy innovations. *J. Time Series Analysis*, **7**, 1-5.

Andel, J., Netuka, I. and Svara, K. (1984). On threshold autoregressive processes. Kybernetika, **20**, 89-106.

Anderson, T.W. (2003). *An Introduction to Multivariate Statistics*, 3rd ed., Wiley, New York.

Arellano-Valle, R.B. and Azzalini, A. (2004). On the unification of families of skew-normal distributions. Preprint

Arnold, B.C. and Beaver, R.J. (2000). Hidden truncation models. *Sankhyā*, **62**, 23-35.

Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scan. J. Statist.*, **12**, 171-178.

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

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *J. Roy. Statist. Soc. B*, **61**, 579-602.

Balakrishnan, N. (2004). Inference from skew-normal under censoring. Preprint.

Bjerve, S. and Doksum, K.A. (1993). Correlation curves: measures of association as functions of covariates. *Ann. of Statist.*, **21**, 890-902.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd Ed. Prenctice Hall.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd Ed. Springer.

Cambanis, S., Fotopoulos, S.B. and He, L. (2000). On the conditional variance for scale mixtures of normal distributions. *J. of Multivariate Analysis*, **74**, 163-192.

Capitanio, A., Azzalini, A. and Stanghellini, E. (2003). Graphical models for skew-normal variates. *Scand. J. Statist.* 30, 129-144.

Chan, K.S. and Tsay, R.S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, **85**, 413-426

Dalla Valle, A. (2004). The skew-normal distribution. In M.G. Genton (Ed.), *Skew elliptical distributions and their applications*, Chap. 1, (pp. 2-24). Chapman & Hall.

Fotopoulos, S.B. (2004). Tempered distributions and their application in computing conditional moments for mixture of normal distributions. *Statist. and Probab., Letters*, to appear.

Genton, M.G. (2004). *Skew-Elliptical Distributions And Their Applications: A Journey Beyond Normality*. Chapman & Hall/CRC, Boca Raton.

Genton, M.G., He, L. and Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statist. & Prob. Lett.*, **51**, 319-325.

González-Farías, G., Dominguez-Molina, J.A. and Gupta, A.K. (2004). Additive properties of skew-normal random vectors. *J. Statist. Planning and Inference*, **126**, 521-534.

Gordon, R.D. (1941). Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large value of the argument. *Ann. Math. Statist.*, **12**, 364-366.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. of Statistics*, **21**, 157-178.

Henze, N. (1986). A probabilistic representation of the 'skew-normal' distribution. *Scand. J. Statist.*, **13**, 271-275.

Jacobs, P.A. and Lewis, P.A.W. (1977). A mixed autoregressive-moving average exponential sequence and point process, EARMA (1,1). *Adv. Appl. Probab.* **9**, 87-104.

Lawrance, A.J. and Lewis, P.A.W. (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables. *J. Roy. Statist. Soc. B*, **47**, 165-183.

Loges, W. (2004). The stationary marginal distribution of a threshold AR(1) process. *J. of Time Series*, **25**, 103-125.

Loperfido, N. (2001). Quadratic forms of skew-normal random vectors. *Statist. & Prob. Lett.* **54**, 381-387.

Loperfido, N. (2004). Generalized skew-normal distributions. In M.G. Genton (Ed.), Skew-elliptical distributions and their applications, Chap. 4 (pp. 65-80). Chapman & Hall/CRC.

McKenzie, E. (2003). Discrete variate time series. In *Handbook of Statistics*, **21**, 573-606, D.N. Shanbhag and C.R. Rao, eds., Elsevier.

Petruccelli, J. D. and Woolford, S.W. (1984). A threshold AR(1) model. *J. Appl. Prob.*, **21**, 270-286.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *J. of Applied Statistics*, **27**, 859-870.

Pourahmadi, M. (1988). Stationarity of the solution of $X_t = A_t X_{t-1} + \varepsilon_t$ and analysis of non-Gaussian dependent random variables. *J. of Time Series Analysis*, **9**, 225-239.

Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*. Wiley, New York.

Shepp, L.A., Slepian, D., and Wyner, A.D. (1980). On prediction of moving average processes. *Bell Syst. Tech. J.*, **59**, 367-415.

Tarami, B. and Pourahmadi, M. (2003). Multivariate t autoregressions: Innovations, prediction variances and exact likelihood equations. *J. of Time Series Analysis*, **24**, 739-754.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Publications.

Wiener, N. (1958). *Nonlinear problems in random theory*. MIT Press, Cambridge, MA.

Wu, W.B. and W. Min (2004). On linear processes with dependent innovations. Technical Report, Dept. of Statistics, University of Chicago.

Xia, Y., Tong, H., Li, W.K. and Zhu. L.X. (2002). An adaptive estimation of dimension reduction space (with discussion). *J. of Royal Statist. Soc. B*, **64**, 363-410.