Modeling covariance matrices via partial autocorrelations

M.J. Daniels [a,*], M. Pourahmadi [b]

[a] *Department of Statistics, University of Florida, United States*

[b] *Department of Statistics, Texas A&M University, United States*

ARTICLE INFO

ABSTRACT

We study the role of partial autocorrelations in the reparameterization and parsimonious modeling of a covariance matrix. The work is motivated by and tries to mimic the phenomenal success of the partial autocorrelations function (PACF) in model formulation, removing the positive-definiteness constraint on the autocorrelation function of a stationary time series and in reparameterizing the stationarity-invertibility domain of ARMA models. It turns out that once an order is fixed among the variables of a general random vector, then the above properties continue to hold and follow from establishing a one-to-one correspondence between a correlation matrix and its associated matrix of partial autocorrelations. Connections between the latter and the parameters of the modified Cholesky decomposition of a covariance matrix are discussed. Graphical tools similar to partial correlograms for model formulation and various priors based on the partial autocorrelations are proposed. We develop frequentist/Bayesian procedures for modelling correlation matrices, illustrate them using a real dataset, and explore their properties via simulations.

1. Introduction

Positive-definiteness and high-dimensionality are two major obstacles in modeling the $p \times p$ covariance matrix Σ of a random vector $Y = (Y_1, \ldots, Y_p)'$. These can partially be alleviated using various decompositions which, in increasing order of effectiveness, are the variance–correlation [1], spectral [2] and Cholesky [3,4] decompositions. Only the latter has the unique distinction of providing an unconstrained and statistically interpretable reparameterization of a covariance matrix, but at the expense of imposing an order among the entries of Y. Three close competitors are, (i) the covariance selection models [5,6] based on full partial correlations obtained from Σ^{-1}, which are statistically interpretable, but constrained, (ii) the logarithm of eigenvalues and logit of Givens angles [7,2] and (iii) the matrix-logarithm models [8]. The latter two are based on an unconstrained, but not necessarily interpretable reparameterization of Σ.

We present yet another unconstrained and statistically interpretable reparameterization of Σ using the notion of partial autocorrelation from time series analysis [9,10, Chap. 7], which, like the Cholesky decomposition, also imposes an order among the entries of Y; this reparamaterization is also ideal for models that directly include correlation matrices, instead of covariance matrices, including multivariate probit models [11] and copulas [12]. For covariance matrices, we start with the decomposition $\Sigma = DRD$ or the variance–correlation strategy [1] and reduce the problem to and focus on reparameterizing a correlation matrix $R = (\rho_{ij})$ in terms of a simpler symmetric matrix $\Pi = (\pi_{ij})$ where $\pi_{ii} = 1$ and for $i < j$, π_{ij} is the *partial autocorrelation* between Y_i and Y_j adjusted for the *intervening* (not the remaining) variables. We note that unlike R and the matrix of full partial correlations (ρ^{ij}), Π has a simpler structure in that it is not required to be positive-definite

and hence its entries are free to vary in the interval $(-1, 1)$. Furthermore, using the Fisher z transform Π can be mapped to the matrix $\tilde{\Pi}$ where the off-diagonal entries of the latter take values in the entire real line $(-\infty, +\infty)$. The process of going from a constrained R to a real symmetric matrix $\tilde{\Pi}$ is reminiscent of finding a link function in the theory of generalized linear models [13]. Therefore, the analogues of graphical and analytical machineries developed in the contexts of regression and the Cholesky decomposition in Pourahmadi [3,14] and references therein, can be brought to the service of modeling correlation matrices. In the sequel, to emphasize the roles, the properties and the need for (time-) order with a slight abuse of language we refer to Π as the partial autocorrelation function (PACF) of Y or Σ, just as in time series analysis.

Compared with the long history of the use of the PACF in time series analysis [15–20], research on establishing a one-to-one correspondence between a general covariance matrix, its PACF and connecting the latter to the entries of the Cholesky factor of the former has a rather short history. An early work in the Bayesian context is due to Eaves and Chang [21], followed by Zimmerman [22] and Pourahmadi [3,10, p. 102] and Daniels and Pourahmadi [23] for longitudinal data, and Dégerine and Lambert-Lacroix [24] for the time series setup. For a general random vector, Kurowicka and Cooke [25,26] and Joe [27] have relied on graph-theoretical and standard multivariate techniques, respectively. The origins of a fundamental determinantal identity involving the PACF, unearthed recently by these three groups of researchers, can be traced to a notable and somewhat neglected paper of Yule [28, equ. 25] and in the literature of time series in connection with the Levinson–Durbin type algorithms. It plays a central role in Joe's [27] method of generating random correlation matrices with distributions independent of the order of the indices, and we use it effectively in introducing priors for the Bayesian analysis of correlation matrices. For a similar application in time series analysis, see [20].

Correlation matrices themselves are accompanied by additional challenges. The constraint of diagonal elements fixed (at one) complicates both reparameterizations, decompositions and computations. Other than the partial autocorrelation parameterization proposed here, there are no unconstrained parameterizations currently in the statistical literature for a correlation matrix. In addition, recent advances in Bayesian computations for correlation matrices (i.e., sampling with Markov chain Monte Carlo algorithms) rely on augmenting the correlation matrix with a diagonal scale matrix to create a covariance matrix (i.e., parameter expansion algorithms). The strategy is to then sample the inverse of this covariance matrix from a Wishart distribution and then transform back to the correlation matrix; see, e.g., [29,30]. However, these approaches do not easily extend to structured correlation matrices (as we will discuss here).

The outline of the paper is as follows. In Section 2, we review the recent results in reparameterizing a correlation matrix via PACF and the Cholesky decomposition. We use the latter to derive a remarkable identity expressing determinant of R as a simple function of the partial autocorrelations. This identity obtained by Dégerine, Lambert-Lacroix [24], Joe [27] and Kurowicka and Cooke [26], plays a fundamental role in introducing prior distributions for the correlation matrix R which is independent of the order of indices used in defining the PACF. The role of a generalized partial correlogram in formulating parsimonious models for R is discussed and illustrated using Kenward's [31] cattle data. In Section 3, we introduce new priors for correlation matrices based on this parameterization, examine their properties and relation to other priors that have appeared in the literature [1], present a simple approach to sample from the posterior distribution of a correlation matrix, and do some simulations to examine the behavior of these new priors. Section 4 provides guidance on the use of these models and tools in applications in behavior and social sciences. Section 5 summarizes the findings and provides directions for future work.

2. Reparameterizations of a correlation matrix

Modeling correlation matrices and simulating random or "typical" correlation matrices are of central importance in various areas of statistics [11,1,32,12], engineering and signal processing [33], social and behavior sciences [34], finance [35] and numerical analysis [36]. An obstacle in dealing with a correlation matrix R is that all its diagonal entries are the same and equal to one.

In this section, first we reparameterize correlation/covariance matrices of a general random vector $Y = (Y_1, \ldots, Y_p)'$ in terms of the *partial autocorrelations* between Y_j and Y_{j+k} adjusted for the intervening variables. Then, using the concept of regression which is implicit in introducing the partial autocorrelations and the Cholesky decomposition of matrices, we point out the connections among the PACF, the generalized autoregressive parameters and the innovation variances of Y introduced in Pourahmadi [3].

2.1. Reparameterization in terms of partial autocorrelations

The notion of PACF is known to be indispensable in the study of stationary processes and situations dealing with Toeplitz matrices such as the Szegö's orthogonal polynomials, trigonometric moment problems, geophysics, digital signal processing and filtering [37,10], identification of ARMA models, the maximum likelihood estimation of their parameters [19] and simulating a random or "typical" ARMA model [20]. The one-to-one correspondence between the stationary autocorrelation functions $\{\rho_k\}$ and their PACF $\{\pi_k\}$ [17,18] makes it possible to remove the positive-definiteness constraint on $\{\rho_k\}$, and work with $\{\pi_k\}$ which are free to vary over the interval $(-1, 1)$ independently of each other.

We parameterize a (non-Toeplitz) correlation matrix $R = (\rho_{ij})$ in terms of the lag-1 correlations $\pi_{i,i+1} = \rho_{i,i+1}$, $i = 1, \ldots, p-1$ and the partial autocorrelations $\pi_{ij} = \rho_{ij|i+1,\ldots,j-1}$ for $j - i \geq 2$, or the matrix $\Pi = (\pi_{ij})$. This allows swapping the constrained matrix R by the simpler matrix Π with ones on the diagonal and where for $i \neq j$, the π_{ij}'s can vary freely in

the interval $(-1, 1)$. The key idea behind this reparameterization is the well-known recursion formula [44, p. 41] see also (1) below, for computing partial correlations in terms of the marginal correlations (ρ_{ij}). It also lies at the heart of Kurowicka and Cooke [25,26] and Joe [27] approaches to constructing Π; however the recursive Levinson–Durbin algorithm used by Dégerine and Lambert-Lacroix [24] will be used in our presentation in Section 2.3.

Following Joe [27], for $j = 1, \ldots, p - k, k = 1, \ldots, p - 1$, let $r_1'(j, k) = (\rho_{j,j+1}, \ldots, \rho_{j,j+k-1})$, $r_3'(j, k) = (\rho_{j+k,j+1}, \ldots, \rho_{j+k,j+k-1})$, and $R_2(j, k)$ be the correlation matrix corresponding to the components $(j+1, \ldots, j+k-1)$. Then, the partial autocorrelations between Y_j and Y_{j+k} adjusted for the intervening variables, denoted by $\pi_{j,j+k} \equiv \rho_{j,j+k|j+1,\ldots,j+k-1}$, are computed using the expression

$$\pi_{j,j+k} = \frac{\rho_{j,j+k} - r_1'(j, k)R_2(j, k)^{-1}r_3(j, k)}{[1 - r_1'(j, k)R_2(j, k)^{-1}r_1(j, k)]^{1/2}[1 - r_3'(j, k)R_2(j, k)^{-1}r_3(j, k)]^{1/2}}. \tag{1}$$

In what follows and in analogy with R, it is convenient to arrange these partial autocorrelations in a matrix Π with $(j, j+k)$th entry $\pi_{j,j+k}$. Note that the function $g(\cdot)$ in (1) that maps a correlation matrix R into the partial autocorrelation matrix Π, is indeed invertible, so that solving (1) for $\rho_{j,j+k}$, one obtains

$$\rho_{j,j+k} = r_1'(j, k)R_2(j, k)^{-1}r_3(j, k) + D_{jk}\pi_{j,j+k}, \tag{2}$$

where D_{jk} is the denominator of the expression in (1). Then, the formulae (1) and (2) clearly establish a one-to-one correspondence between the matrices R and Π. In the sequel, with a slight abuse of language and following the tradition in times series analysis we refer to the matrix Π or $\pi_{j,j+k}$ viewed as a function of (j, k), as the partial autocorrelation function (PACF) of Y.

Evidently, when R is a stationary (Toeplitz) correlation matrix, then $\pi_{j,j+k}$ depends only on the lag k, see (1). Consequently, Π is a stationary (Toeplitz) matrix. Fortunately, the converse is also true and follows from (2). For ease of reference, we summarize these observations in Lemma 1 in Section 2.3. A correlation matrix R is stationary (Toeplitz) if and only if its associated PACF Π is a stationary (Toeplitz) matrix.

Moreover, for a stationary correlation matrix, R reduces precisely to the celebrated Levinson–Durbin formula [10, Theorem 7.3] for computing the PACF recursively.

2.2. An alternative reparameterization: Cholesky decomposition

Next, we present an alternative reparameterization of a covariance matrix via its Cholesky decomposition or the idea of autoregression for the underlying random vector.

Consider a mean-zero random vector Y with the positive-definite covariance matrix $\Sigma = (\sigma_{st})$. For $1 \leq t \leq p$, let $\hat{Y}_t$ be the linear least-squares predictor of Y_t based on its predecessors $Y_1, \ldots, Y_{t-1}$ and let $\varepsilon_t = Y_t - \hat{Y}_t$ be its prediction error with variance $\sigma_t^2 = \text{Var}(\varepsilon_t)$. Then there are unique scalars ϕ_{tj} so that $\hat{Y}_t = \sum_{j=1}^{t-1} \phi_{tj}Y_j$ or

$$Y_t = \sum_{j=1}^{t-1} \phi_{tj}Y_j + \varepsilon_t, \quad t = 1, \ldots, p. \tag{3}$$

Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p)'$ be the vector of successive uncorrelated prediction errors with $\text{Cov}(\varepsilon) = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2) = D$. Then, (3) rewritten in matrix form becomes $\varepsilon = TY$, where T is a unit lower triangular matrix with 1's on the main diagonal and $-\phi_{tj}$ in the (t, j)th position for $2 \leq t \leq p, j = 1, \ldots, t - 1$. Note that $\sigma_t^2 = \text{Var}(\varepsilon_t)$ is different from $\sigma_{tt} = \text{Var}(Y_t)$. However, when the responses are independent, then $\phi_{tj} = 0$ and $\sigma_t^2 = \sigma_{tt}$, so that the matrices T and D gauge the "dependence" and "heterogeneity" of Y, respectively.

Computing covariances using $\varepsilon = TY$, it follows that

$$T\Sigma T' = D, \quad |\Sigma| = \prod_{t=1}^{p} \sigma_t^2. \tag{4}$$

The first factorization in (4), called the modified Cholesky decomposition of Σ, makes it possible to swap the $p(p + 1)/2$ constrained parameters of Σ with the unconstrained set of parameters ϕ_{tj} and $\log \sigma_t^2$ of the same cardinality. In view of the similarity of (3) to a sequence of varying order autoregressions, we refer to the parameters ϕ_{tj} and σ_t^2 as the *generalized autoregressive parameters* (GARP) and *innovation variances* (IV) of Y or Σ [3]. A major advantage of (4) is its ability to guarantee the positive-definiteness of the estimated covariance matrix given by $\hat{T}^{-1}\hat{D}\hat{T}'^{-1}$ so long as the diagonal entries of $\hat{D}$ are positive.

It should be noted that imposing structures on Σ will certainly lead to constraints on T and D in (4). For example, a correlation matrix R with 1's as its diagonal entries is structured with possibly $p(p - 1)/2$ distinct parameters. In this case, certain entries of T and D are either known, redundant or constrained. In fact, it is easy to see that the diagonal entries of the matrix D for a correlation matrix are monotone decreasing with $\sigma_1^2 = 1$. For this reason and others, it seems more prudent to rely on the ordered partial correlations when reparameterizing a correlation matrix R as in Section 2.1, than using its Cholesky decomposition.

2.3. A multiplicative determinantal identity: Partial autocorrelations

First, we study the role of partial autocorrelations in measuring the reduction in prediction error variance when a variable is added to the set of predictors in a regression model. Using this and the second identity in (4) we obtain a fundamental determinantal identity expressing $|\Sigma|$ in terms of the partial autocorrelations and diagonal entries of Σ. Joe [27] and Kurowicka and Cooke [26] had obtained this identity using determinantal recursions and graph-theoretical methods based on (1), respectively. An earlier and a slightly more general determinantal identity for covariance matrices in the context of nonstationary processes was given by Dégerine and Lambert-Lacroix [24, p. 54] using an analogue of the Levinson–Durbin algorithm.

For u and v two distinct integers in $\{1, 2, \ldots, p\}$, let L be a subset of $\{1, 2, \ldots, p\} \setminus \{u, v\}$ and $\pi_{uv|L}$ stand for the partial correlation between Y_u and Y_v adjusted for Y_ℓ, $\ell \epsilon L$. We denote the linear least squares predictor of Y_u based on Y_ℓ, $\ell \epsilon L$ by $\hat{Y}_{u|L}$, and for v an integer Lv stands for the union of the set L and the singleton $\{v\}$.

Lemma 1. *Let $Y = (Y_1, \ldots, Y_p)'$ be a mean-zero random vector with a positive-definite covariance matrix Σ. Then, with u, v and L as above, we have*

$$(a) \quad \hat{Y}_{u|Lv} = \hat{Y}_{u|L} + \alpha_{uv}\left(Y_v - \hat{Y}_{v|L}\right), \qquad \alpha_{uv} = \pi_{uv|L}\sqrt{\frac{\mathrm{Var}(Y_u - \hat{Y}_{u|L})}{\mathrm{Var}(Y_v - \hat{Y}_{v|L})}}. \tag{5}$$

$$(b) \quad \mathrm{Var}\left(Y_u - \hat{Y}_{u|Lv}\right) = \left(1 - \pi_{uv|L}^2\right)\mathrm{Var}\left(Y_u - \hat{Y}_{u|L}\right). \tag{6}$$

Proof. (a) Let $sp\{Y_u; u\epsilon L\}$ stand for the linear subspace generated by the indicated random variables. Since $Y_v - \hat{Y}_{v|L}$ is orthogonal to $sp\{Y_u; u\epsilon L\}$ it follows that

$$sp\{Y_u; u\epsilon Lv\} = sp\{Y_u; u\epsilon L\} \oplus sp\{Y_v - \hat{Y}_{v|L}\},$$

and from the linearity of the orthogonal projection we have

$$\hat{Y}_{u|Lv} = \hat{Y}_{u|L} + \alpha_{uv}(Y_v - \hat{Y}_{v|L}),$$

where α_{uv}, the regression coefficient of Y_u on $Y_v - \hat{Y}_{v|L}$ is given by

$$\alpha_{uv} = \frac{\mathrm{Cov}(Y_u, Y_v - \hat{Y}_{v|L})}{\mathrm{Var}(Y_v - \hat{Y}_{v|L})}.$$

Since $\hat{Y}_{u|L} \epsilon\ sp\{Y_i; i\epsilon L\}$ is orthogonal to $Y_v - \hat{Y}_{v|L}$, the numerator of the expression above can be replaced by $\mathrm{Cov}(Y_u - \hat{Y}_{u|L}, Y_v - \hat{Y}_{v|L})$, so that

$$\alpha_{uv} = \frac{\mathrm{Cov}(Y_u - \hat{Y}_{u|L}, Y_v - \hat{Y}_{v|L})}{\sqrt{\mathrm{Var}(Y_u - \hat{Y}_{u|L})\mathrm{Var}(Y_v - \hat{Y}_{v|L})}} \cdot \sqrt{\frac{\mathrm{Var}(Y_u - \hat{Y}_{u|L})}{\mathrm{Var}(Y_v - \hat{Y}_{v|L})}}$$

$$= \pi_{uv|L}\sqrt{\frac{\mathrm{Var}(Y_u - \hat{Y}_{u|L})}{\mathrm{Var}(Y_v - \hat{Y}_{v|L})}}.$$

(b) From the first identity in (5), it is immediate that

$$Y_u - \hat{Y}_{u|Lv} = Y_u - \hat{Y}_{u|L} - \alpha_{uv}\left(Y_v - \hat{Y}_{v|L}\right),$$

or

$$Y_u - \hat{Y}_{u|Lv} + \alpha_{uv}(Y_v - \hat{Y}_{v|L}) = Y_u - \hat{Y}_{u|L}. \tag{7}$$

Computing variances of both sides of (7) and using the fact that $Y_u - \hat{Y}_{u|Lv}$ is orthogonal to $Y_v - \hat{Y}_{v|L}$, it follows that

$$\mathrm{Var}\left(Y_u - \hat{Y}_{u|Lv}\right) = \mathrm{Var}\left(Y_u - \hat{Y}_{u|L}\right) - \alpha_{uv}^2\mathrm{Var}\left(Y_v - \hat{Y}_{v|L}\right).$$

Now, substituting for α_{uv} from (a), the desired result follows. $\quad\square$

Next, we use the recursion (6) to express the innovation variances or the diagonal entries of D in terms of the partial correlations or the entries of π. Similar expressions for ϕ_{tj}'s, the entries of T, are not available. The next theorem sheds some light on this problem. Though the expressions are recursive and ideal for computation (Levinson–Durbin algorithm), they are not as explicit or revealing. The approach we use here is in the spirit of the Levinson–Durbin algorithm [10, Corollary 7.4] as extended by Dégerine and Lambert-Lacroix [24] to nonstationary processes.

Theorem 1. *Let* $Y = (Y_1, \ldots, Y_p)$ *be a mean-zero random vector with a positive-definite covariance matrix* Σ *which can be decomposed as in* (4).

(a) *Then, for* $t = 2, \ldots, p$; $j = 1, \ldots, t - 1$, *we have*

$$\sigma_t^2 = \sigma_{tt} \prod_{j=1}^{t-1} (1 - \pi_{jt}^2),$$

(b) $|\Sigma| = \left(\prod_{t=1}^{p} \sigma_{tt} \right) \prod_{t=2}^{p} \prod_{j=1}^{t-1} (1 - \pi_{jt}^2)$.
(c) *For* $t = 2, \ldots, p$; $L = \{2, \ldots, t - 1\}$

$$\phi_{t1} = \pi_{1t} \sqrt{\frac{\text{Var}(Y_t - \hat{Y}_{t|L})}{\text{Var}(Y_1 - \hat{Y}_{1|L})}}.$$

(d) $\phi_{tj} = \phi_{tj|L} - \phi_{t1}\phi_{1,t-j|L}$, *for* $j = 2, \ldots, t - 1$,
   where $\phi_{tj|L}$ *and* $\phi_{1,t-j|L}$ *are, respectively, the forward and backward predictor coefficients of* Y_t *and* Y_1 *based on* $\{Y_k; k \epsilon L\}$, *defined by*

$$\hat{Y}_{t|L} = \sum_{j=2}^{t-1} \phi_{tj|L} Y_j, \qquad \hat{Y}_{1|L} = \sum_{j=2}^{t-1} \phi_{t,t-j} Y_j.$$

Proof. (a) follows from the repeated use of (6) with $u = t$ and $v = j, j = 1, \ldots, t - 1$. (b) follows from (4) and (a). □

Part (c) proved first in [10, p.102] shows that only the entries of the first column of T are multiples of the partial autocorrelations appearing in the first column of π. However, for $j > 1$, since ϕ_{tj} is a multiple of the partial correlation between Y_j and Y_t adjusted for $\{Y_i; i \in [1, t) \setminus \{j\}\}$, (see Lemma 1(a)), it is not of the form of the entries of π. Note that these observations are true even when Y is stationary or Σ is Toeplitz [10, Lemma 7.8 and Theorem 7.3]. In the search for a connection with the entries of π, it is instructive to note that the nonredundant entries of $T^{-1} = (\theta_{tj})$ can be interpreted as the generalized moving average parameters (GMAP) or the regression coefficient of ε_j when Y_t is regressed on the innovations $\varepsilon_t, \ldots, \varepsilon_j, \ldots, \varepsilon_1$, see [10, p.103]. An alternative interpretation of θ_{tj} as the coefficient of Y_j when Y_t is regressed on $Y_j, Y_{j-1}, \ldots, Y_1$ is presented in [38, Sec. 2.2]. Consequently, θ_{tj} is a multiple of the partial correlation between Y_t and Y_j adjusted for $\{Y_1, \ldots, Y_{j-1}\}$. We hope these connections and working with partial correlations will offer similar advantage to working with the GARP in terms of the autoregression interpretation [10, Sections 3.5.3 and 3.5.4].

2.4. An attractive property of the PACF parameterization

Parsimonious modeling of the GARP of the modified Cholesky decomposition often relies on exploring for structure as a function of lag; for example, fitting a polynomial to the GARP as a function of lag [3]. For such models, the GARP, in a sense, have different interpretations within lag; i.e., the lag 1 coefficient from the regression of Y_3 on (Y_2, Y_1) is the Y_2 coefficient, when another variable, Y_1 is also in the model; however, for the regression of Y_2 on Y_1, the lag 1 coefficient is the Y_1 coefficient with no other variable in the model. So, for a given lag k, the lag k coefficients all come from conditional regressions where the number of variables conditioned on are different. However, by construction, the lag k partial autocorrelations are all based on conditional regressions where the number of conditioning variables are the same, always conditioning on $k - 1$ intervening variables. This will facilitate building models for the partial autocorrelations as a function of lag. We discuss such model building in the next section.

2.5. Parsimonious modeling of the PACF

In this section, we use the generalized partial correlogram, i.e. the plot of $\{\pi_{j,j+k}; j = 1, \ldots, p-k\}$ versus $k = 1, \ldots, p-1$, as a graphical tool to formulate parsimonious models for the PACF in terms of the lags and other covariates. If necessary, we transform its range to the entire real line using z transform. Such a modeling environment is much simpler and avoids working with the complex constraints on the correlation matrix R [1,39] or the matrix of full partial correlations constructed from Σ^{-1} [6]; note that the full partial correlations (ρ^{ij}) are defined as the correlation between two components conditional on *all* the other components.

Note that the partial autocorrelations $\pi_{j,j+k}$ between successive variables Y_j and Y_{j+k} are grouped by their lags $k = 1, \ldots, p - 1$, and heuristically, $\pi_{j,j+k}$ gauges the conditional (in)dependence between variables k units apart conditional

Table 1
Cattle Data. Sample correlations (below the main diagonal), sample PACF (above the main diagonal) and sample variances (along the main diagonal).

106	0.82	0.07	−0.24	0.03	0.01	0.16	−0.06	0.26	−0.22	0.19
0.82	155	0.91	0.03	0.02	−0.23	−0.17	0.01	−0.01	−0.07	−0.25
0.76	0.91	165	0.93	0.07	−0.04	−0.12	0.01	0.09	0.21	0.03
0.66	0.84	0.93	185	0.94	0.23	−0.18	−0.2	−0.22	0.02	0.27
0.64	0.80	0.88	0.94	243	0.94	−0.04	0.07	−0.23	−0.08	0.16
0.59	0.74	0.85	0.91	0.94	284	0.93	0.56	−0.3	−0.09	−0.24
0.52	0.63	0.75	0.82	0.87	0.93	306	0.93	0.35	−0.24	−0.18
0.53	0.67	0.77	0.84	0.89	0.94	0.93	341	0.97	0.15	−0.28
0.52	0.60	0.71	0.77	0.84	0.90	0.93	0.97	389	0.96	0.2
0.48	0.58	0.70	0.73	0.80	0.86	0.88	0.94	0.96	470	0.98
0.48	0.55	0.68	0.71	0.77	0.83	0.86	0.92	0.96	0.98	445

Table 2
Fitted PACF from the least-squares fit of a cubic polynomial to the sample PACF (first row) and the fitted Fisher's z transform of PACF from the least-squares fit of an exponential function of the lags (second-row).

Lags	1	2	3	4	5	6	7	8	9	10
Fitted PACF	0.89	0.24	−0.09	−0.19	−0.15	−0.02	0.11	0.16	0.07	−0.26
Fitted z transf.	−1.64	−0.49	−0.07	.09	0.15	0.17	0.18	0.18	0.18	0.18

on the intervening variables, so one expects it to be smaller for larger k. In the Bayesian framework, this intuition suggests putting shrinkage priors on the partial autocorrelations that shrink the matrix Π toward certain simpler structures [40].

2.6. Data illustration

To illustrate the capabilities of the generalized correlograms in revealing patterns, we use the cattle data [31] which consists of $p = 11$ bi-weekly measurements of the weights of $n = 30$ cows. Table 1, displays the sample (partial) correlations for the cattle data in the lower (upper) triangular segment and the sample variances are along the main diagonal. It reveals several interesting features of the dependence in the data that the commonly used profile plot of the data cannot discern. For example, note that all the correlations are positive, they decrease monotonically within the columns (time-separation), they are not constant (nonstationary) within each subdiagonal. In fact, they tend to increase over time (learning effect). Furthermore, the partial autocorrelations of lags 2 or more are insignificant except for the entries 0.56 and 0.35.

Fig. 1 presents the generalized correlograms corresponding to the sample correlation matrix of the data, the full partial correlations, the generalized partial correlogram and the Fisher's z transform of the PACF. Note that the first two correlograms suggest linear and quadratic patterns in the lag k, but in fitting such models one has to be mindful of the constraints on the coefficients so that the corresponding fitted correlation matrices are positive definite. Details of fitting such models and the ensuing numerical results can be found in [10]. The generalized partial correlogram in (c) reveals a cubic polynomial in the lags, i.e. $\pi_{j,j+k} = \gamma_0 + \gamma_1 k + \gamma_2 k^2 + \gamma_3 k^3$; in fitting such models the only constraint to observe is that the entries of the matrix Π are required to be in $(-1, 1)$. However, the Fisher z transform of the entries of Π are unconstrained and Fig. 1(d) suggests a pattern that can be approximated by an (exponential) function $\alpha + \beta \exp(-k)$, $k = 1, \ldots, p - 1$, with no constraints on (α, β) or another cubic polynomials in the lags. The least-squares fits of a cubic polynomial and an exponential function to the correlograms in Fig. 1(c)–(d) are summarized in Table 2. Note that fitting such models to the PACF amounts to replacing the entries of the kth subdiagonal of the matrix Π by a single number and hence R is approximated by a stationary (Toeplitz) matrix, see Lemma 1 and Theorem 2 below. In addition, this parameterization also allows the marginal variances to be similarly modelled parsimoniously (as a function of time) similar to the modelling of the prediction variances in [3]. The maximum likelihood estimation of the parameters and their asymptotic properties will be pursued in a future work.

3. Priors for R via the partial autocorrelations

In addition to the advantages for formulating parsimonious models, the unconstrainedness of the PACF suggests some approaches for constructing priors for R using independent linearly transformed Beta priors on $(-1, 1)$ for the PACF.

3.1. Independent priors on partial autocorrelations

Given that each partial autocorrelation is free to vary in the interval $(-1, 1)$, we may construct priors for R derived from independent priors on the PACF. A natural option would be a uniform distribution on the space of Π, i.e., a uniform distribution on the $p(p-1)/2$-dimensional hypercube; we denote this prior as the independent uniform (IU) prior. This can be shown to induce the following prior on the correlation matrix R:

$$p(R) \propto \left[\prod_{k=1}^{p-1} \prod_{j=1}^{p-k} (1 - \pi_{j,j+k}^2)^{p-1-k} \right]^{-1/2}. \tag{8}$$
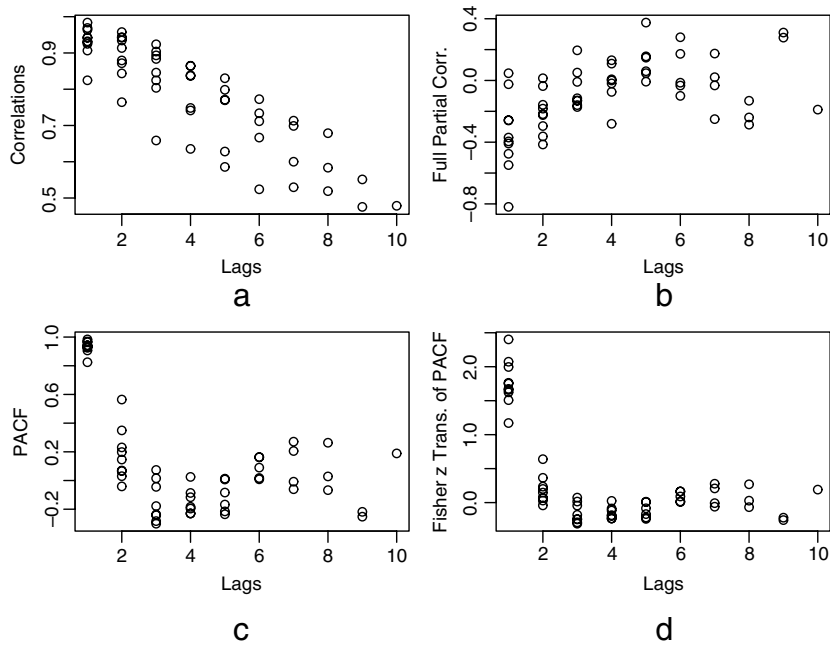
Fig. 1. (a) Generalized sample correlogram for the cattle data, (b) Generalized inverse correlogram, (c) Generalized partial correlogram, (d) Plot of Fisher z transform of the PACF.

One can express the prior in (8) in terms of the marginal correlations, $\rho_{j,j+k}$ by plugging in for $\pi_{j,j+k}$ from (1). This prior induces a particular behavior on the marginal correlations. Specifically, the priors on the (marginal) correlations $\rho_{j,j+k}$ become gradually more peaked at zero as the lag k grows. As an illustration of this behavior, Fig. 2 shows the histograms based on 10,000 simulations from the uniform prior on the partial autocorrelations for $p = 5$. As the dimension p of the correlation matrix grows, the priors become more peaked at zero for larger lags. This can also be seen by examining the prior probability of being in some interval, say $[-.5, .5]$, as a function of the lag. For $p = 15$, the (averaged) probabilities, ordered from lag 1 to lag 14, are respectively, (.50, .59, .65, .70, .73, .76, .78, .80, .82, .83, .84, .85, .86, .87). This would appear to be desirable behavior for longitudinal data which typically exhibits serial correlation decaying with increasing lags.

It is also evident from Fig. 2 that the priors for $\rho_{j,j+k}$ with k fixed, appear to be the same (see the subdiagonals). We state this observation more formally in the following theorem; see also Lemma 1.

Theorem 2. *If the partial autocorrelations, π_{jk} have independent stationary priors, i.e*

$$p(\pi_{jk}) = p(\pi_{il}) \quad \text{if } |j - k| = |i - l|, \tag{9}$$

(or the priors are the same along the subdiagonals of Π), then the marginal priors on the correlations ρ_{jk} are also stationary, i.e.

$$p(\rho_{jk}) = p(\rho_{il}) \quad \text{if } |j - k| = |i - l|. \tag{10}$$

(or the priors are the same along the corresponding subdiagonal of R).

Theorem implies that independent "stationary" priors on Π induce "stationary" priors on the marginal correlations. Most priors we introduce here satisfy this property.

More generally, independent linearly transformed Beta priors on the interval $(-1, 1)$ for partial autocorrelations are a convenient and flexible way to specify a prior for a correlation matrix R. These priors, denoted by Beta(α, γ), have the density

$$p(\rho) = \frac{1}{2\beta(\alpha, \gamma)} \left(\frac{1+\rho}{2}\right)^{\alpha-1} \left(\frac{1-\rho}{2}\right)^{\gamma-1}. \tag{11}$$

Interestingly, the uniform prior on the correlation matrix [1] corresponds to the following stationary Beta priors on the partial autocorrelations:

$$\pi_{i,i+k} \sim \text{Beta}(\alpha_k, \alpha_k), \tag{12}$$

where $\alpha_k = 1 + \frac{1}{2}(p - 1 - k)$; see [27]. We will refer to this prior as Barnard Beta (BB). As noted in [1], such a prior on R results in the marginal priors for each of the marginal correlations being somewhat peaked around zero (same peakedness for *all* ρ_{jk}). Also note that the priors become more peaked as p grows.
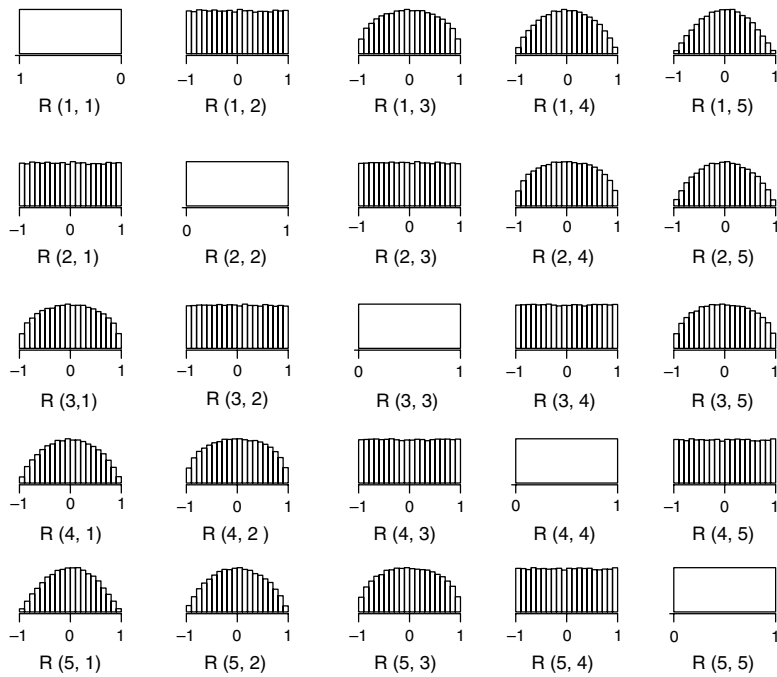
Fig. 2. Marginal priors on ρ_{jk} from independent uniform priors on the partial correlations, π_{jk}. The subplots are arranged as the matrix R.

In general, priors for the correlation matrix proportional powers of the determinant of the correlation matrix,

$$p(R) \propto |R|^{\alpha_p - 1} \tag{13}$$

are constructed by setting $\alpha_k = \alpha_p + \frac{1}{2}(p - 1 - k)$ in (12). Priors so constructed are proper, so improper priors like Jeffreys' for a correlation matrix in a multivariate normal model, $\pi(R) = |R|^{-(p+1)/2}$, are not special cases.

3.2. Shrinkage behavior of the BB and IU priors

The IU priors on the PACF induce desirable behavior for longitudinal (ordered data) by 'shrinking' higher lag correlations toward zero. The Beta priors in (12), which induce a uniform prior for R (BB priors) place a uniform $(-1, 1)$ prior on the lag $p - 1$ partial autocorrelations and shrink the other partial autocorrelations toward zero with the amount of shrinkage being *inversely* proportional to lag. This induces the desired behavior on the marginal correlations, making their marginal priors equivalent, but it is counter-intuitive for ordered/longitudinal data with serial correlation; in addition, the shrinkage of the lag one partial autocorrelations for the BB prior increases with p (recall the form in (12)). In such data, we would expect lower lag correlations to be less likely to be zero and higher lag correlations to be more likely to be zero. Thus, the independent uniform priors are likely to be a good default choice for the partial autocorrelations in terms of inducing desirable behavior on the marginal correlations and not counter-intuitively shrinking the partial autocorrelations. We explore this shrinkage behavior further via some simulations in Section 3.5.

In addition, we expect many of the higher lag partial autocorrelations to be close to zero for longitudinal data with serial correlation via conditional independence (see Table 1). To account for this, we could (aggressively) shrink the partial autocorrelations toward zero (with the shrinkage increasing with lag) using shrinkage priors similar to those proposed in [23,40] or by creating such priors based on the Beta distributions proposed here. We are currently exploring this.

3.3. Some other priors for a correlation matrix

Other priors for R have been proposed in the literature which cannot be derived based on independent priors on the partial autocorrelations. For example, the prior on R that induces marginal uniform $(-1, 1)$ priors on the ρ_{ij}'s [1] has the form:

$$p(R) = |R|^{p(p-1)/2-1} \prod_{i=1}^{p} |R[-i, -i]|^{-(p+1)/2} \tag{14}$$

where $R[-i, -i]$ is the submatrix of R with the ith row and column removed and $|R[-i, -i]| = [R]_{ii}^{-1}|R|$. Such a prior might not be a preferred one for longitudinal (ordered) data where the same marginal priors on all correlations (irrespective of lag) may not be the best default choice.

Eaves and Chang [21] derived some related reference priors for the set of partial correlations, $\pi_{1,j}$ for $j = 2, \ldots, p$; however, their priors are not natural for longitudinal data. Chib and Greenberg [11] specified a truncated multivariate normal distribution on the marginal correlations. Liechty et al. [41] placed normal distributions on the marginal correlations with the goal of grouping the marginal correlations into clusters. The latter two priors along with those in [6] for the full partial correlations (ρ^{ij}) are highly constrained given that they model the marginal (or full partial) correlations directly.

3.4. Bayesian computing

An additional issue with modeling the correlation matrix is computational. Our development here will focus on cases without covariates in the correlation matrix (this will be left for future work) under the class of independent priors on the PACF discussed in Section 3.1. The proposal here might be viewed as an alternative to the PX-RPMH algorithm in [30] that explicitly exploits the fact that we are modeling the partial autocorrelations themselves (a computational comparison will be left for future work). However, our approach will naturally allow structures in the partial autocorrelations which cannot be done when using current versions of the PX-RPMH (or similar) algorithms; for example if the partial autocorrelations are zero or constant within lag, as the correlation matrix is then highly constrained.

In the following, we assume the data, $\{Y_i : i = 1, \ldots, n\}$ are independent, normally distributed p-vectors with mean $X_i \beta$ and with covariance matrix $\Sigma = R$ (a correlation matrix). A natural way to sample the partial autocorrelations is via a Gibbs sampling algorithm in which we sample from the full conditional distributions of each of the partial autocorrelations. Given that the full conditional distributions of the partial autocorrelations are not available in closed form there are several options to sample them. We explore a simple one next.

We propose to use an auxiliary variable approach to sample each partial autocorrelation. Define the likelihood for the partial autocorrelation π_{jk} as $L(\Pi)$ and the prior as $p(\pi_{jk})$. As in [42], introduce a positive latent variable U_{jk} such that

$$L(\Pi)p(\pi_{jk}) = \int_0^\infty I\{u_{jk} < L(\Pi)\}p(\pi_{jk})\mathrm{d}u_{jk}. \tag{15}$$

To sample π_{jk}, we can proceed in two steps,

1. Sample $U_{jk} \sim Unif(0, L(\Pi))$.
2. Sample $p(\pi_{jk})$ constrained to the set $\{\pi_{jk} : L(\Pi) > U_{jk}\}$.

Truncated versions of the priors proposed here, linearly transformed Beta distributions (of which the uniform is a special case), can easily be sampled using the approach in [43]. The truncation region for step 2, given that the domain of π_{jk} is bounded, can typically be found quickly numerically. The likelihood evaluations needed to find the truncation interval can be made simpler by using the determinant identity derived in Theorem 1.

In the following, we list several facts about the likelihood that are useful in computing the marginal posteriors of individual π_{jk}.

Fact 1. If we factor $R^{-1} = CPC$, where P is a correlation matrix and C is a diagonal matrix, the elements of P are the full partial correlations, ρ^{ij} [44, Chapter 15].

Fact 2. To isolate the likelihood contribution of $\pi_{j,j+k}$, we can factor the entire multivariate normal distribution into $p(y_j, \ldots, y_{j+k})p(y_l : l < j \text{ or } l > j+k \mid y_j, \ldots, y_{j+k})$. The (l, k) entry of inverse of the correlation matrix, $R[j : j+k]$ for the first factor is related to the partial autocorrelation of interest (recall Fact 1).

Fact 3. Using the determinantal identity in Theorem 1(b), the determinant of submatrices of R in terms of partial autocorrelations can be written as a function of the partial autocorrelations,

$$|R[j : j+k]| = \prod_{l=1}^{k} \prod_{i=j}^{j+k-l} (1 - \pi_{i,i+l}^2). \tag{16}$$

Extensions of these computational procedures to modelling the correlation matrix when the matrix of interest is a covariance matrix is straightforward (see, e.g., [30]).

3.5. Simulations

We now conduct some simulations in a longitudinal setting to

1. examine the mixing behavior of the auxiliary variable sampler here and
2. compare the risk of the IU prior to the BB prior, the standard default prior for a correlation matrix.

In terms of the mixing of the Markov chain, the auxiliary variable sampler on the partial autocorrelations works quite well, with the lag correlation in the chain dissipating quickly. For example, for $p = 5$, $n = 25$, the lag correlations for each partial autocorrelation was negligible by lag 10. Similar results were seen for other p/n combinations.

For our simulation, we consider three true matrices representing typical serial correlation, an AR(1) with lag 1 correlation of .8 and one with lag correlation .6. Both these matrices have all partial autocorrelation beyond lag 1 equal to 0. We also

Table 3

Results for $p = 5$. Each row corresponds to $n = 10, 25, 50, 100$. IU: independent uniform priors; BB: Barnard Beta priors. LL: log likelihood loss; SEL-P: squared error loss on Fisher's z-transformation of the partial autocorrelations; SEL-M: squared error loss on Fisher's z-transformation of the marginal correlations. Full: matrix with lag 1–3 partial autocorrelations equal to (.8, .4, .1) with the rest zero.

R	LL		SEL-P		SEL-M	
	IU	BB	IU	BB	IU	BB
AR(.8)	2.4	1.6	1.4	1.1	1.7	1.1
	.69	.51	.48	.39	.55	.38
	.27	.23	.20	.18	.22	.17
	.11	.10	.09	.08	.09	.07
AR(.6)	1.1	.99	.89	.86	.98	.89
	.48	.45	.41	.39	.46	.44
	.23	.21	.19	.18	.19	.17
	.11	.11	.10	.09	.10	.10
Full	2.7	1.9	1.8	1.3	2.8	1.8
	.68	.50	.46	.35	.61	.38
	.25	.20	.19	.16	.21	.15
	.12	.11	.10	.09	.08	.07

Table 4

Results for $p = 10$. Each row corresponds to $n = 20, 50, 100$. IU: independent uniform priors; BB: Barnard Beta priors. LL: log likelihood loss; SEL-P: squared error loss on Fisher's z-transformation of the partial autocorrelations; SEL-M: squared error loss on Fisher's z-transformation of the marginal correlations. Full: matrix with lag 1–3 partial autocorrelations equal to (.8, .4, .1) with the rest zero.

R	LL		SEL-P		SEL-M	
	IU	BB	IU	BB	IU	BB
AR(.8)	5.0	2.8	3.3	2.3	5.2	2.7
	1.6	1.1	1.2	.88	1.7	.97
	.60	.48	.49	.42	.52	.36
AR(.6)	2.4	2.2	2.1	2.1	2.2	1.9
	.97	.85	.86	.79	.88	.71
	.49	.45	.45	.42	.42	.36
Full	6.4	3.5	4.3	2.7	11.0	4.8
	1.8	1.1	1.4	.94	2.9	1.3
	.69	.52	.55	.45	.90	.50

considered a matrix that had more non-zero partial autocorrelations with lags 1, 2, 3, 4 equal to (.8, .4, .1, 0) respectively; this corresponds to the lag 1–4 marginal correlations being equal to (.8, .78, .73, .68). We consider two size matrices/sample size combinations, $p = 5$ with $n = 10, 25, 50, 100$ and $p = 10$ with sample sizes, $n = 20, 50, 100$. We also consider several loss functions, log likelihood (LL) loss, $\text{tr}(\hat{R}R^{-1}) - \log |\hat{R}R^{-1}| - p$, with Bayes estimator the inverse of the posterior expectation of R^{-1} and squared error loss on Fisher's z-transform of the partial autocorrelations, π_{jk} (SEL-P) and the marginal correlations, ρ_{jk} (SEL-M), with Bayes estimator the posterior mean (of the z-transformed correlations).

For $p = 5$, the risk reductions from the IU prior are clear from Table 3, with percentage reductions as large as 30% for $n = 10$, 25% for $n = 25$ and 15% for $n = 50$. For $p = 10$, the risk reductions from the uniform prior are clear from Table 4, with percentage reductions as large as 50%. The largest risk reductions were for loss SEL-M (squared error loss on the Fisher's z-transform on the marginal correlations). The lower risk reductions for the first order autoregressive covariance matrices are related to only the lag 1 partial autocorrelations being non-zero; so the shrinkage of the BB prior for all the other partial autocorrelations is not unreasonable. Examination of squared error loss for the partial autocorrelations by lag indicates large reductions for the lag 1 partial autocorrelations and smaller increases for the other lag partial autocorrelations.

In addition, the estimates of the first order lag correlation show large differences (not shown). For AR (.8) with $p = 10$ and $n = 50$, the means were .75 under the IU prior and about .70 under the BB prior with larger discrepancies for $p = 5$.

Some of the risk reductions from using the IU priors on the partial autocorrelations are small. However, they come at no computational cost (unlike some priors for covariance matrices proposed in the literature) and are consistent with prior beliefs about partial autocorrelations representing serial correlation. The BB prior is not a good default choice due to its dampening effects on the important lower order partial autocorrelations. Further risk improvement might be expected through the use of more targeted shrinkage [23].

4. Partial correlations in the behavior and social sciences

The models and priors for partial correlations are extremely important for many applications involving longitudinal and functional data in the behavior and social sciences. In particular, modeling longitudinal data using structural equation and factor analytic models (i.e., latent variable models in general) typically require careful modeling of correlation matrices see e.g., [39] as do multivariate probit models [11,45,34]. The tools here provide both a general class of methods for using the partial autocorrelations that allow parsimonious modeling of correlations via regression modeling and sensible priors on

correlations within such models which is often essential in small to medium sized datasets. Such modeling takes on even more importance in the presence of incomplete data [46]. In addition, the uniform priors on the partial autocorrelations recommended in Section 3 provide no additional computational challenges over standard priors for a correlation matrix. Future work will illustrate these methods more fully in applications.

5. Discussion

Using the variance–correlation separation strategy, modeling a covariance matrix is reduced to that of its correlation matrix R which has the additional constraint that all its diagonal entries must be equal to one. Though the Cholesky decomposition can handle the positive-definiteness, it cannot be applied directly when there are additional constraints such as stationarity or constancy along diagonals [3, Sec. 2.6], zero entries [47] and separable covariance structures [48]. The reparameterization in terms of partial autocorrelations is shown to work well in the face of an additional constraint. It requires ordering the variables which is not a problem for longitudinal and functional data, but might be difficult to justify for other situations. Related work on trying to 'order' data that does not have a natural ordering can be found in [49,50]. The long history and successful use of the PACF in the time series literature provide valuable graphical and analytical tools which can be generalized to the nonstationary setup.

Given the conditioning structure of the partial autocorrelations, we expect many of them to be zero (see Table 1). Thus, it would be natural to adapt the approach in [6] to zero out the partial autocorrelations. We might expect computational simplifications given that the PACF are free to vary independently in $[-1, 1]$ unlike the full partial correlations. In addition, when constructing priors for the probability of a partial autocorrelation being zero, the lack of exchangeability of the partial autocorrelations (vs. the full partial correlations) given that they condition on different numbers of variables (i.e., only the intervening variables) must be taken into account; such issues have been addressed in [34] in a related setting.

We will explore the computational efficiency of other proposals for Bayesian computing in future work, including sampling all the partial autocorrelations together. In addition, we will derive strategies for Bayesian inference when modeling Fisher's z transform of the partial autocorrelations as a function of covariates.

Acknowledgments

References

[1] J. Barnard, R. McCulloch, X. Meng, Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage, Statistica Sinica 10 (2000) 1281–1312.
[2] R. Yang, J.O. Berger, Estimation of a covariance matrix using the reference prior, Annals of Statistics 22 (1994) 1195–1211.
[3] M. Pourahmadi, Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation, Biometrika 86 (1999) 677–690.
[4] Z. Chen, D. Dunson, Random effects selection in linear mixed models, Biometrics 59 (2003) 159–182.
[5] A.P. Dempster, Covariance selection, Biometrics 28 (1972) 157–175.
[6] F. Wong, C.K. Carter, R. Kohn, Efficient estimation of covariance selection models, Biometrika 90 (2003) 809–830.
[7] M. Daniels, R. Kass, Nonconjugate Bayesian estimation of covariance matrices in hierarchical models, Journal of the American Statistical Association 94 (1999) 1254–1263.
[8] T.Y.M. Chiu, T. Leonard, K.W. Tsui, The matrix-logarithm covariance model, Journal of the American Statistical Association 91 (1996) 198–210.
[9] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis-Forecasting and Control, revised 3rd ed., Prentice Hall, NJ, 1994.
[10] M. Pourahmadi, Foundations of Time Series Analysis and Prediction Theory, Wiley, New York, 2001.
[11] S. Chib, E. Greenberg, Analysis of multivariate probit models, Biometrika 85 (1998) 347–361.
[12] M. Pitt, D. Chan, R. Kohn, Efficient Bayesian inference for Gaussian copula regression models, Biometrika 93 (2006) 537–554.
[13] P. McCullagh, J.A. Nelder, Generalized Linear Models, 2nd ed., Chapman & Hall, London, 1989.
[14] M. Pourahmadi, Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance–correlation parameters, Biometrika 94 (2007) 1006–1013.
[15] M.H. Quenouille, Approximate tests of correlation in time series, Journal of Royal Statistical Society, Series B 11 (1949) 68–84.
[16] H.E. Daniels, The approximate distribution of serial correlation coefficients, Biometrika 43 (1956) 169–185.
[17] O. Barndorff-Nielsen, G. Schou, On the parameterization of autoregressive models for partial autocorrelation, Journal of Multivariate Analysis 3 (1973) 408–419.
[18] F.L. Ramsey, Characterization of the partial autocorrelation function, Annals of Statistics 2 (1974) 1296–1301.
[19] R.H. Jones, Maximum likelihood fitting of ARMA models to time series with missing observations, Technometrics 22 (1980) 389–395.
[20] M.C. Jones, Randomly choosing parameters from the stationarity and invertibility region of autoregressive-moving average models, Applied Statistics 36 (1987) 134–138.
[21] D. Eaves, T. Chang, Priors for ordered conditional variances and vector partial correlation, Journal of Multivariate Analysis 41 (1992) 43–55.
[22] D.L. Zimmerman, Viewing the correlation structure of longitudinal data through a PRISM, The American Statistician 54 (2000) 310–318.
[23] M.J. Daniels, M. Pourahmadi, Bayesian analysis of covariance matrices and dynamic models for longitudinal data, Biometrika 89 (2002) 553–566.
[24] S. Dégerine, S. Lambert-Lacroix, Partial autocorrelation function of a nonstationary time series, Journal of Multivariate Analysis 89 (2003) 135–147.
[25] D. Kurowicka, R. Cooke, A parameterization of positive definite matrices in terms of partial correlation vines, Linear Algebra and its Applications 372 (2003) 225–251.
[26] D. Kurowicka, R. Cooke, Completion problem with partial correlation vines, Linear Algebra and its Applications 418 (2006) 188–200.
[27] H. Joe, Generating random correlation matrices based on partial correlations, Journal of Multivariate Analysis 97 (2006) 2177–2189.
[28] G.U. Yule, On the theory of correlation for any number of variables treated by a new system of notation, Royal Society Proceedings 79 (1907) 85–96.
[29] C. Liu, Comment on The art of data augmentation (Pkg: p1-111), Journal of Computational and Graphical Statistics 10 (2001) 75–81.

[30] X. Liu, M.J. Daniels, A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization, Journal of Computational and Graphical Statistics 15 (2006) 897–914.
[31] M.G. Kenward, A method for comparing profiles of repeated measurements, Biometrics 44 (1987) 959–971.
[32] J. Pan, G. MacKenzie, On modelling mean-covariance structures in longitudinal studies, Biometrika 90 (2003) 239–244.
[33] R.B. Holmes, On random correlation matrices, SIAM Journal on Matrix Analysis and Applications 12 (1991) 239–272.
[34] X. Liu, M.J. Daniels, B. Marcus, Joint models for the association of longitudinal binary and continuous processes with application to a smoking cessation trial, Journal of the American Statistical Association (2009).
[35] R.F. Engle, Dynamic conditional correlation: A simple class of multivariate GARCH models, Journal of Business and Economics 20 (2002) 339–350.
[36] P.I. Davies, N.J. Higham, Numerically stable generation of correlation matrices and their factors, BIT 40 (2000) 640–651.
[37] H.J. Landau, Maximum entropy and the moment problem, Bulletin of the American Mathematical Society 16 (1987) 47–77.
[38] N. Wermuth, D.R. Cox, G.M. Marchetti, Covariance chains, Bernoulli 12 (2006) 841–862.
[39] M. Daniels, S.-L. Normand, Longitudinal profiling of health care units based on mixed multivariate patient outcomes, Biostatistics 7 (2006) 1–15.
[40] M.J. Daniels, R.E. Kass, Shrinkage estimators for covariance matrices, Biometrics 57 (2001) 1173–1184.
[41] J.C. Liechty, M.W. Liechty, P. Muller, Bayesian correlation estimation, Biometrika 91 (2004) 1–14.
[42] P. Damien, J. Wakefield, S.G. Walker, Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables, Journal of the Royal Statistical Society, Series B: Statistical Methodology 61 (1999) 331–344.
[43] P. Damien, S.G. Walker, Sampling truncated normal, Beta, and Gamma densities, Journal of Computational and Graphical Statistics 10 (2001) 206–215.
[44] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, 1984.
[45] C. Czado, Multivariate regression analysis of panel data with binary outcomes applied to unemployment data, Statistical Papers 41 (2000) 281–304.
[46] M.J. Daniels, J.W. Hogan, Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis, Chapman & Hall (CRC Press), 2008.
[47] S. Chaudhuri, M. Drton, T.S. Richardson, Estimation of a covariance matrix with zeros, Biometrika 94 (2007) 199–216.
[48] N. Lu, D.L. Zimmerman, The likelihood ratio test for a separable covariance matrix, Statistics and Probability Letters 73 (2005) 449–457.
[49] M.L. Stein, Z. Chi, L.J. Welty, Approximating likelihoods for large spatial data sets, Journal of Royal Statistical Society, Series B 50 (2004) 275–296.
[50] J.O. Berger, J.M. Bernardo, On the development of reference priors, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Bayesian Statistics 4: Proceedings of the Fourth Valencia Meeting, Clarendon Press, 1992, pp. 35–49.