

Dirichlet ARMA models for compositional time series



Tingguo Zheng^{a,b,*}, Rong Chen^c

^a Department of Statistics, School of Economics, Xiamen University, Xiamen, Fujian 361005, China

^b Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China

^c Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA

ARTICLE INFO

Article history:

Received 1 July 2015

Available online 4 April 2017

AMS 2000 subject classifications:

primary 62M10

secondary 62H12

Keywords:

Compositional data

Dirichlet distribution

Gaussian pseudo-likelihood

Multivariate time series

UK gross final expenditure series

Vector ARMA model

ABSTRACT

A compositional time series is a multivariate time series in which the observation vector at each time point is a set of proportions that sum to 1. Traditionally, such time series are modeled by taking a log-ratio transformation of the observations and then modeling them with a Gaussian vector autoregressive moving average (ARMA) model. In this paper, a new class of models is proposed by assuming that the proportions follow a time-varying Dirichlet distribution, and that the corresponding time-varying parameters, after a proper transformation, assume an ARMA-type of dynamic structure. The new model is referred to as the Dirichlet autoregressive moving average (DARMA) model. Under this model, after a proper transformation, the original data follow a vector ARMA model with a martingale difference sequence as its noise series. Two specific transformations are studied under the DARMA framework. Estimation procedures are developed and their numerical properties are investigated. Simulation studies and real examples are presented to demonstrate the properties of the proposed models, and comparisons are made with the existing modeling approaches.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Compositional time series are multivariate time series which at time t are non-negative proportions that sum to [1]. Such time series are often encountered in biology, business, environmental sciences, demography, ecology, economics, geology, and political science. For example, the dynamics of market shares of a certain product among all competitors is an important factor in developing the marketing and development strategies of the companies. The change in the composition of the gross domestic product (GDP) for different economic sectors reveals important signals for economic policy makers.

Since the pioneering work of Aitchison [2,3], researchers have shown an increasing interest in modeling compositional time series; see, e.g., [11,20,31]. The main difficulty is the unit-sum constraint, which severely complicates statistical analysis. The standard multivariate techniques, such as the vector autoregressive moving average (VARMA), are no longer applicable.

Data transformation is one of the main tools used in modeling compositional time series. Specifically, the compositional data are first transformed to follow the normal distribution, and then a time series model is assumed for the transformed series. The major advantage of this approach is that under the normal assumption, most existing standard modeling, estimation, and forecasting tools are readily available. Various forms of transformation have been proposed in the literature.

* Corresponding author at: Department of Statistics, School of Economics, Xiamen University, Xiamen, Fujian 361005, China.
E-mail addresses: zhengt@xmu.edu.cn (T. Zheng), rongchen@stat.rutgers.edu (R. Chen).

Brunsdon [10], Smith and Brunsdon [35], and Brunsdon and Smith [11] proposed the alr-VARMA model using the additive log-ratio (alr) transformation. Quintana and West [31] and Brandt, Monroe and Williams [8] proposed the clr-VAR model using the centered log-ratio (clr) transformation. Bergman [6] proposed the ilr-VAR model using the isometric log-ratio (ilr) transformation. The well-known Box–Cox transformation has also been used; see, e.g., [1,3,7]. Integrated VARMA models have been studied as well [4]. This type of time series modeling, which is based on transformation and a distributional assumption on innovations, is often called the innovations-based approach. In addition, some studies employ the linear Gaussian state space model for modeling the transformed compositions, see, e.g., [7,33,34]. The key assumption of these models is, of course, the ability to find a normal or Gaussian transformation.

In terms of applications, the alr-VARMA model has been most frequently and successfully used for modeling and forecasting series, such as unemployment rates [11], trends in obesity [28], expenditure shares [29], and performance data in cricket [29].

An alternative approach is based on the distributional assumption of the original data, which is known as the data-based approach. This is a more direct approach that is often easier to interpret and understand. The most natural distribution for compositional data is the Dirichlet distribution. Although the Dirichlet distribution has a strong implied dependence structure and has been deemed an inappropriate candidate for modeling independent compositional data [3], it was found to be useful when used as a conditional distribution. With or without covariates, the varying coefficients can accommodate flexible covariance/correlation structures; see, e.g., [9,12,13,24]. For a compositional time series, Grunwald et al. [20] developed a state-space model to model the compositional data directly under the constraints, where the observations that are conditional on the unobserved state are assumed to have a Dirichlet distribution and the state follows the Dirichlet conjugate distribution. A generalization of the Dirichlet distribution was also proposed to allow for dependence between the components [14].

In this paper, we propose a new class of time series models for compositional data under a conditional Dirichlet distribution, with time-varying coefficients. It is a data-based model and can be seen as a special case of the multivariate version of the Martingalized-GARMA (M-GARMA) model used by Zheng et al. [38]. Specifically, with a conditional Dirichlet assumption on the compositional data at each time point, the time-varying distributional parameters are modeled by a vector-ARMA-type process through a link function. Therefore, it is referred to as the Dirichlet autoregressive moving average (DARMA) model. In fact, under this model, the original data, after a transformation that is determined by the link function, follow a vector ARMA model in which the noise process is a martingale difference sequence (MDS). Obviously, it is closely related to the transformation to normality when using the vector-ARMA model approach, but has a more solid foundation and fewer restrictions.

To exploit the DARMA model, we further consider two specific link functions corresponding to additive log-ratio and centered log-ratio data transformations. We will refer to them as alr-DARMA and clr-DARMA. The relationship between the link function and the transformation is important as it is a key component in evaluating the likelihood function for estimation. We will show that, under the DARMA framework, these two transformations are equivalent models with different parametrizations.

The rest of this paper is organized as follows. Section 2 starts with a brief review of the Dirichlet distribution and its properties, and then introduces the DARMA models and its two special cases, the alr-DARMA and clr-DARMA models. The link functions associated with the two specific DARMA models and the equivalence of these two models are also discussed. Section 3 discusses three estimation approaches for the DARMA models. In Section 4, we perform two simulation studies to investigate the finite-sample properties of the proposed estimators, and compare the performances of the proposed DARMA models with the existing methods for modeling compositional time series. Finally, in Section 5 we analyze the share data of UK gross final expenditure using the DARMA models.

2. The model

2.1. Dirichlet distribution

Let $\mathbf{y} = (y_1, \dots, y_K)^\top$ be a $K \times 1$ positive random vector that satisfies that $y_1 + \dots + y_K = 1$ and $y_i \in (0, 1)$ for all $i \in \{1, \dots, K\}$. The random vector \mathbf{y} follows a Dirichlet distribution with positive parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^\top$, denoted $\mathbf{y} \sim \text{Dir}(\boldsymbol{\alpha})$, if the probability density function of \mathbf{y} is

$$f(\mathbf{y} | \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^K y_i^{\alpha_i - 1},$$

where $B(\boldsymbol{\alpha}) = \prod_{i=1}^K \Gamma(\alpha_i) / \Gamma(\sum_{i=1}^K \alpha_i)$ is the multinomial beta function serving as the normalizing constant and Γ is Euler's gamma function. The Dirichlet distribution, which is a multivariate generalization of the Beta distribution, is subject to degeneracy since $y_1 + \dots + y_K = 1$.

Moreover, $\tau = \alpha_1 + \dots + \alpha_K$ can be interpreted as the inverse scale parameter or the concentration parameter. The expectation of each component is $E(y_i) = \alpha_i / \tau$, the variance is $\text{var}(y_i) = \{\alpha_i(\tau - \alpha_i)\} / \{\tau^2(\tau + 1)\}$, and the covariance for $i \neq j$ is $\text{cov}(y_i, y_j) = (-\alpha_i \alpha_j) / \{\tau^2(\tau + 1)\}$.

2.2. Dirichlet ARMA model

Suppose that $\mathbf{y}_t = (y_{1t}, \dots, y_{Kt})^\top$ is a compositional time series consisting of a K -dimensional vector of non-negative components such that $y_{1t} + \dots + y_{Kt} = 1$ for each t . Let $\mathcal{F}_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \dots\}$ be the σ -field generated by all the information up to t . We assume that the conditional distribution of \mathbf{y}_t follows a Dirichlet distribution:

$$\mathbf{y}_t \mid \mathcal{F}_{t-1} \sim \text{Dir}(\boldsymbol{\alpha}_t), \tag{1}$$

where $\boldsymbol{\alpha}_t = (\alpha_{1t}, \dots, \alpha_{Kt})^\top$ has strictly positive components. Moreover, the scale parameter τ_t is defined as $\tau_t = \alpha_{1t} + \dots + \alpha_{Kt}$. The next step is to link the current parameter $\boldsymbol{\alpha}_t$ with the past information in \mathcal{F}_{t-1} .

Based on the M-GARMA framework of Zheng et al. [38], with a properly selected K -dimensional link function h and its companion function $g(\boldsymbol{\alpha}_t) = E\{h(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\}$, we assume that the reparametrized vector of the time-varying parameters, $\boldsymbol{\eta}_t = g(\boldsymbol{\alpha}_t)$, follows a multivariate ARMA-type process, viz.

$$\boldsymbol{\eta}_t = A_0 + \sum_{j=1}^p A_j h(\mathbf{y}_{t-j}) + \sum_{j=1}^q B_j \boldsymbol{\eta}_{t-j}, \tag{2}$$

where p and q are non-negative integers, A_0 is a constant vector, A_1, \dots, A_p and B_1, \dots, B_q are the coefficient matrices associated with lag variables. In fact, the quantity $\boldsymbol{\eta}_t$ is a one-step-ahead prediction of $h(\mathbf{y}_t)$ given the past values $\{h(\mathbf{y}_{t-1}), \dots, h(\mathbf{y}_{t-p}), \boldsymbol{\eta}_{t-1}, \dots, \boldsymbol{\eta}_{t-q}\}$.

By adding $\boldsymbol{\varepsilon}_t = h(\mathbf{y}_t) - \boldsymbol{\eta}_t = h(\mathbf{y}_t) - g(\boldsymbol{\alpha}_t)$ to both sides of (2), we have

$$h(\mathbf{y}_t) = A_0 + \sum_{j=1}^m (A_j + B_j) h(\mathbf{y}_{t-j}) + \boldsymbol{\varepsilon}_t - \sum_{j=1}^q B_j \boldsymbol{\varepsilon}_{t-j}, \tag{3}$$

where $m = \max(p, q)$. Note that $E(\boldsymbol{\varepsilon}_t \mid \mathcal{F}_{t-1}) = 0$. Hence, the process $(\boldsymbol{\varepsilon}_t)$ is a MDS. This setting is very similar to the standard modeling approach using a VARMA model on a transformed time series. The major difference is that the standard approach assumes the noise to be Gaussian under the h transformation, while in the DARMA model, the noise process is a MDS that is determined by the conditional Dirichlet distribution assumption in (1) and the transformation h . More detailed discussions on the differences are presented in Section 2.4.

Using the terminology of M-GARMA in Zheng et al. [38], $g(\boldsymbol{\alpha}_t)$ and $h(\mathbf{y}_t)$ serve as the link function and the y -link function, respectively. Given the Dirichlet distribution assumption and the specific y -link function h , the link function g is completely determined. In general, $g \neq h$ unless h is the identity function. When modeling rates or proportional time series data, the identity link function is not ideal because it is difficult to provide feasible parameter conditions to ensure that all values of the conditional expectation are bounded between 0 and 1. We also note that an extension of the GARMA model by Benjamin et al. [5] in the compositional time series case would force $g(\boldsymbol{\alpha}_t) = h(\boldsymbol{\alpha}_t/\tau_t)$. The resulting noise sequence is no longer a MDS [32], which complicates the analysis and leads to difficulties in the investigation of the probabilistic properties of the series and asymptotic behavior of the estimators.

Our model is flexible enough to include deterministic covariates, such as trends and seasonal dummies, and exogenous variables with easy interpretations.

2.3. Two specific models

The choice of y -link function is the crux of building a DARMA model. Two log-ratio transformations, the additive log-ratio (alr) transformation and the centered log-ratio (clr) transformation, are traditionally used for compositional data [3], and have been used to transform compositional time series into a Gaussian process [8,11,32].

In this section, we investigate the use of these two transformations as the y -link function in a DARMA model.

2.3.1. The alr-DARMA model

If we denote $k = K - 1$, the additive log-ratio (alr) transformation,

$$\text{alr}(\mathbf{y}_t) = \left(\ln \frac{y_{1t}}{y_{Kt}}, \dots, \ln \frac{y_{kt}}{y_{Kt}} \right)^\top,$$

is a one-to-one transformation from the natural sample space, namely the simplex $\mathcal{S}^K = \{(y_{1t}, \dots, y_{Kt})^\top : y_{1t} > 0, \dots, y_{Kt} > 0; y_{1t} + \dots + y_{Kt} = 1\}$ to \mathbb{R}^k . In this transformation, the K th component y_{Kt} serves as the reference component. It should be noted that the analysis based on the transformation is permutation invariant; in other words, it is unaffected by the choice of common denominator.

The following proposition, whose proof is given in Appendix A.1, provides the mean and variance of $\text{alr}(\mathbf{y})$ when \mathbf{y} follows a Dirichlet distribution.

Proposition 1. Let $\mathbf{y} = (y_1, \dots, y_K)^\top \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$. The expectation of the log transformed variable, $\ln \mathbf{y} = (\ln y_1, \dots, \ln y_K)^\top$, and its covariance matrix are

$$E(\ln \mathbf{y}) = (\psi(\alpha_1) - \psi(\tau), \dots, \psi(\alpha_K) - \psi(\tau))^\top$$

and

$$\text{cov}(\ln \mathbf{y}) = \text{diag}\{\psi_1(\alpha_1), \dots, \psi_1(\alpha_K)\} - \psi_1(\tau) \mathbf{1}_K \mathbf{1}_K^\top,$$

where $\tau = \alpha_1 + \dots + \alpha_K$, ψ is the digamma function, ψ_1 is the trigamma function, $\mathbf{1}_K$ is a K -dimensional vector of 1's, and $\text{diag}\{\cdot\}$ is a diagonal matrix.

The expectation and covariance matrix of the additive log-ratio transformed variable,

$$\text{alr}(\mathbf{y}) = \left(\ln \frac{y_1}{y_K}, \dots, \ln \frac{y_k}{y_K} \right)^\top$$

are

$$\boldsymbol{\mu} = E\{\text{alr}(\mathbf{y})\} = (\psi(\alpha_1) - \psi(\alpha_K), \dots, \psi(\alpha_k) - \psi(\alpha_K))^\top$$

and

$$\text{cov}\{\text{alr}(\mathbf{y})\} = \text{diag}\{\psi_1(\alpha_1), \dots, \psi_1(\alpha_k)\} + \psi_1(\alpha_K) \mathbf{1}_k \mathbf{1}_k^\top,$$

where $k = K - 1$.

Let the alr transformation be the y -link function in the DARMA model. That is, $h(\mathbf{y}_t) = \text{alr}(\mathbf{y}_t)$ in (2). We have the following alr-DARMA model

$$\begin{aligned} \mathbf{y}_t \mid \mathcal{F}_{t-1} &\sim \text{Dir}(\boldsymbol{\alpha}_t) \quad \text{and} \quad \sum_{i=1}^K \alpha_{it} = \tau \\ \boldsymbol{\eta}_t &= A_0 + \sum_{j=1}^p A_j \text{alr}(\mathbf{y}_{t-j}) + \sum_{j=1}^q B_j \boldsymbol{\eta}_{t-j}, \end{aligned} \quad (4)$$

where

$$\boldsymbol{\eta}_t = g(\boldsymbol{\alpha}_t) = E\{h(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\} = (\psi(\alpha_{1t}) - \psi(\alpha_{Kt}), \dots, \psi(\alpha_{kt}) - \psi(\alpha_{Kt}))^\top. \quad (5)$$

The link function g in (5) is due to Proposition 1. Note that, given the k -dimensional $\boldsymbol{\eta}_t = (\eta_{1t}, \dots, \eta_{kt})^\top$, the K -dimensional vector of the time-varying parameters, $\boldsymbol{\alpha}_t = (\alpha_{1t}, \dots, \alpha_{Kt})^\top$, is unidentifiable since $k = K - 1$. In our model, we impose the constraint $\alpha_1 + \dots + \alpha_K = \tau$, where τ is the concentration parameter of the Dirichlet distribution. We assume that it is a time invariant unknown parameter to be estimated.

The likelihood function of the model given by (4) and (5) requires the inverse of the link function $g(\boldsymbol{\alpha}_t)$. With the constraint $\alpha_1 + \dots + \alpha_K = \tau$, we write $g(\boldsymbol{\alpha}_t)$ in (5) as

$$g(\boldsymbol{\alpha}_t) = \boldsymbol{\eta}_t = \left(\psi(\alpha_{1t}) - \psi\left(\tau - \sum_{i=1}^k \alpha_{it}\right), \dots, \psi(\alpha_{kt}) - \psi\left(\tau - \sum_{i=1}^k \alpha_{it}\right) \right)^\top. \quad (6)$$

This can be seen as a multivariate extension of the link function of the logit-Beta-M-GARMA model introduced in [38]. The following proposition shows that its inverse function is uniquely defined.

Proposition 2. Given $\tau > 0$ and a set of constant $\boldsymbol{\eta} = (\eta_1, \dots, \eta_k)^\top$, there is a unique set of positive numbers $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^\top$ that satisfy the system of equations: $g(\boldsymbol{\alpha}) = \boldsymbol{\eta}$ and $\alpha_1 + \dots + \alpha_K = \tau$, where the function g is specified in (6).

The proof for the existence and uniqueness of solving the above system is shown in Appendix A.2. According to Proposition 2, we can obtain the unique solution of $\boldsymbol{\alpha}_t = (\alpha_{1t}, \dots, \alpha_{Kt})^\top$ at each time t . We denote the inversion as $\boldsymbol{\alpha}_t = g^{-1}(\boldsymbol{\eta}_t)$ with the K th time-varying parameter $\alpha_{Kt} = \tau - (\alpha_{1t} + \dots + \alpha_{kt})$.

In addition, for the MDS noise process $\boldsymbol{\varepsilon}_t = h(\mathbf{y}_t) - g(\boldsymbol{\alpha}_t)$ in (3), the conditional covariance matrix given \mathcal{F}_{t-1} is the same as for $h(\mathbf{y}_t)$. Following Proposition 1, we have

$$V_t = \text{cov}\{\text{alr}(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\} = \text{diag}\{\psi_1(\alpha_{1t}), \dots, \psi_1(\alpha_{kt})\} + \psi_1(\alpha_{Kt}) \mathbf{1}_k \mathbf{1}_k^\top, \quad (7)$$

where $\mathbf{1}_k$ is a k -dimensional vector of 1's, $\text{diag}\{\cdot\}$ is a function that inserts a vector into the diagonal of a matrix, and ψ_1 is the trigamma function.

2.3.2. The clr-DARMA model

The centered log-ratio (clr) transformation is another popular transformation for compositional data. It is defined as:

$$\text{clr}(\mathbf{y}_t) = \left(\ln \frac{y_{1t}}{m(\mathbf{y}_t)}, \dots, \ln \frac{y_{Kt}}{m(\mathbf{y}_t)} \right)^\top,$$

where $m(\mathbf{y}_t)$ denotes the geometric mean of \mathbf{y}_t , i.e., $m(\mathbf{y}_t) = (y_{1t} \cdots y_{Kt})^{1/K}$.

Let $J = I_K - \mathbf{1}_K \mathbf{1}_K^\top / K$, where I_K is a K -dimensional identity matrix and $\mathbf{1}_K$ is a K -dimensional vector of 1's. It can easily be seen that $\text{clr}(\mathbf{y}_t) = J \ln \mathbf{y}_t$ and $\mathbf{1}_K^\top \text{clr}(\mathbf{y}_t) = 0$. Compared to the alr transformation, the clr transformation uses the geometric mean as the reference component, but introduces a singularity.

Setting the y -link function $h(\mathbf{y}_t) = \text{clr}(\mathbf{y}_t)$ in the DARMA model, the clr-DARMA model takes the form

$$\begin{aligned} \mathbf{y}_t \mid \mathcal{F}_{t-1} &\sim \text{Dir}(\boldsymbol{\alpha}_t) \quad \text{and} \quad \sum_{i=1}^K \alpha_i = \tau \\ \boldsymbol{\eta}_t &= A_0 + \sum_{j=1}^p A_j \text{clr}(\mathbf{y}_{t-j}) + \sum_{j=1}^q B_j \boldsymbol{\eta}_{t-j}, \end{aligned} \tag{8}$$

where $\boldsymbol{\eta}_t = \mathbf{g}(\boldsymbol{\alpha}_t) = E\{\text{clr}(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\}$. By Proposition 1, we have

$$\begin{aligned} \mathbf{g}(\boldsymbol{\alpha}_t) &= E\{\text{clr}(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\} = JE\{\ln(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\} \\ &= J(\psi(\alpha_{1t}) - \psi(\tau), \dots, \psi(\alpha_{Kt}) - \psi(\tau))^\top \\ &= (\psi(\alpha_{1t}) - \xi_t, \dots, \psi(\alpha_{Kt}) - \xi_t)^\top, \end{aligned}$$

where $\xi_t = \{\psi(\alpha_{1t}) + \dots + \psi(\alpha_{Kt})\} / K$.

Because $\mathbf{1}_K^\top J = \mathbf{0}$, therefore $\mathbf{1}_K^\top \boldsymbol{\eta}_t = \mathbf{0}$ for all t . We impose the following constraints on the parameter matrices A s and B s in (8): $\mathbf{1}_K^\top A_0 = \mathbf{0}$, $\mathbf{1}_K^\top A_i = \mathbf{0}_K^\top$, $\mathbf{1}_K^\top B_j = \mathbf{0}_K^\top$, where $\mathbf{0}_K$ is a vector of 0's. In addition, $\mathbf{1}_K^\top \text{clr}(\mathbf{y}_t) = 0$ for all t . Hence, we can re-parameterize model (8) with a lower-dimension equivalent representation, viz.

$$\boldsymbol{\eta}_t^* = A_0^* + \sum_{j=1}^p A_j^* \text{clr}^*(\mathbf{y}_{t-j}) + \sum_{j=1}^q B_j^* \boldsymbol{\eta}_{t-j}^* \tag{9}$$

where $\boldsymbol{\eta}_t^*$ and $\text{clr}^*(\mathbf{y}_{t-j})$ are the first k elements of $\boldsymbol{\eta}_t$ and $\text{clr}(\mathbf{y}_{t-j})$ respectively, and the coefficient matrices A_i^* and B_i^* are all $k \times k$ matrices. In fact, we have $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_t^{*\top}, -\mathbf{1}_k^\top \boldsymbol{\eta}_t^*)^\top$ and $\text{clr}(\mathbf{y}_t) = (\text{clr}^*(\mathbf{y}_t)^\top, -\mathbf{1}_k^\top \text{clr}^*(\mathbf{y}_t))^\top$.

The conditional variance matrix of $\text{clr}(\mathbf{y}_t)$ given \mathcal{F}_{t-1} is

$$\begin{aligned} V_t &= J(\text{diag}\{\psi_1(\alpha_{1t}), \dots, \psi_1(\alpha_{Kt})\} - \psi_1(\tau) \mathbf{1}_K \mathbf{1}_K^\top) J \\ &= J \text{diag}\{\psi_1(\alpha_{1t}), \dots, \psi_1(\alpha_{Kt})\} J. \end{aligned}$$

Note that this matrix is degenerated as $\mathbf{1}_K^\top V_t = \mathbf{0}_K^\top$ and its upper-left $k \times k$ submatrix is the conditional variance matrix of $\text{clr}^*(\mathbf{y}_t)$ given \mathcal{F}_{t-1} .

Again, as in the alr-DARMA case, we will need to find the solutions for $\boldsymbol{\alpha}_t$ under the system of equations $\mathbf{g}_t(\boldsymbol{\alpha}_t) = \boldsymbol{\eta}_t$ and $\alpha_1 + \dots + \alpha_K = \tau$. The systems of equations can be rewritten as:

$$\begin{cases} \psi(\alpha_{1t}) - \psi(\alpha_{Kt}) = \eta_{1t} - \eta_{Kt} \\ \psi(\alpha_{2t}) - \psi(\alpha_{Kt}) = \eta_{2t} - \eta_{Kt} \\ \vdots \\ \psi(\alpha_{kt}) - \psi(\alpha_{Kt}) = \eta_{kt} - \eta_{Kt} \\ \alpha_{1t} + \dots + \alpha_{Kt} = \tau, \end{cases}$$

where $k = K - 1$. This is equivalent to the system of equations under the alr-DARMA case, which has a unique solution as indicated in Proposition 2.

2.3.3. Equivalence of two log-ratio models

We next show the equivalence of the alr-DARMA and the reduced form of clr-DARMA models under (9). The following proposition, whose proof is trivial, gives the relationship between the clr transformation and alr transformation.

Proposition 3. Suppose \mathbf{y}_t is a $K \times 1$ vector of compositions. Let $\text{clr}^*(\mathbf{y}_t)$ be the first $k = K - 1$ components of the clr transformation of \mathbf{y}_t , and $\text{alr}(\mathbf{y}_t)$ be the alr transformation of \mathbf{y}_t . Then,

$$\text{clr}^*(\mathbf{y}_t) = P_k \text{alr}(\mathbf{y}_t) \quad \text{and} \quad \text{alr}(\mathbf{y}_t) = P_k^{-1} \text{clr}^*(\mathbf{y}_t), \tag{10}$$

where P_k is a $k \times k$ dimensional matrix

$$P_k = \frac{1}{K} \begin{bmatrix} K-1 & -1 & \cdots & -1 \\ -1 & K-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & K-1 \end{bmatrix} \quad \text{and} \quad P_k^{-1} = \begin{bmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{bmatrix}.$$

Substituting (10) into (9) results in the following representation of the clr-DARMA process:

$$P_k \boldsymbol{\eta}_t = E\{\text{clr}^*(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\} = A_0^* + \sum_{j=1}^p A_j^* P_k \text{alr}(\mathbf{y}_{t-j}) + \sum_{j=1}^q B_j^* P_k \boldsymbol{\eta}_{t-j}, \quad (11)$$

where $\boldsymbol{\eta}_t = E\{\text{alr}(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\}$. Multiplying both sides of (11) by P_k^{-1} yields

$$\boldsymbol{\eta}_t = A_0 + \sum_{j=1}^p A_j \text{alr}(\mathbf{y}_{t-j}) + \sum_{j=1}^q B_j \boldsymbol{\eta}_{t-j}, \quad (12)$$

where $A_0 = P_k^{-1} A_0^*$, $A_j = P_k^{-1} A_j^* P_k$ for all $j \in \{1, \dots, p\}$ and $B_j = P_k^{-1} B_j^* P_k$ for all $j \in \{1, \dots, q\}$. Model (12) is the same as the alr-DARMA model given by (4). Hence, the alr-DARMA model and the clr-DARMA model are equivalent, and the coefficient matrices are equivalent under the linear transformations. However, numerical stability in estimation may depend on the choice between the two models.

The alr and clr transformations are just two of the commonly used transformations. Many other y -link functions can be considered. For example, the cumulative logit transformations or the adjacent-category logit transformation (when components are ordered) can be used. With a chosen y -link h function and the induced link function g , the DARMA model ensures a vector ARMA structure with a MDS as its noise sequence.

2.4. Difference to the VARMA model after the log-ratio transformation

The standard approach to modeling compositional time series is to first apply a suitable transformation, and then use the Gaussian vector ARMA model. Specifically, using h as a transformation function, the transformed VARMA (t VARVA) model assumes that

$$h(\mathbf{y}_t) = A_0 + \sum_{j=1}^m (A_j + B_j) h(\mathbf{y}_{t-j}) + \boldsymbol{\varepsilon}_t - \sum_{j=1}^q B_j \boldsymbol{\varepsilon}_{t-j}, \quad (13)$$

where $m = \max(p, q)$, while the noise term $\boldsymbol{\varepsilon}_t$ is often assumed to be independent, identically distributed, and multivariate normal, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \Sigma)$. The alr and clr transformations are commonly used, which will be denoted as alr-VARMA and clr-VARMA. Following Proposition 3, the alr-VARMA and clr-VARMA models are also equivalent under a linear reparametrization.

There are several major differences between the DARMA model and the t VARMA model. First, the DARMA model directly assumes that the conditional distribution of the observed data \mathbf{y}_t follows a Dirichlet distribution. This is a data-based approach, which is a term used by Davis et al. [15] and Davis and Wu [16]. The t VARMA model is an innovation-based approach, which depends on a distributional (normal) assumption on the noise distribution of the ARMA model. The data-based approach is more natural and lends itself to easy and direct interpretation; the assumption is also more testable. The link functions reflect how the parameters in the conditional distribution relate to past information. Such a model structure is easy to verify through the model selection procedure. The innovation-based approach lacks interpretation and depends on the accuracy of the distribution assumption. While this assumption can be checked with residual analysis, it is often less quantifiable.

Second, the DARMA model directly models the structure of the time-varying parameters. One can estimate $\boldsymbol{\alpha}_t$ as a by-product of the model estimation; hence, its evolution can be graphically viewed and used to check whether any exogenous variables, including time trend and seasonality, should be included. It is also easier to impose individual modeling assumptions on each of the components. For the t VARMA model, such terms can be included, but are difficult to interpret.

Third, the t VARMA model introduces the covariance matrix Σ of the noise sequence. When a DARMA model is written in ARMA form (3), the noise sequence is a MDS with an induced covariance structure from the Dirichlet distribution assumption and the y -link function. The covariance structure is also time-varying. Hence, the t VARMA model has more flexibility in modeling the correlation between the components, but introduces more parameters in the model. It only allows homogeneous variance through time.

Fourth, the concentration parameter τ in the DARMA model can be time varying. For example, one may consider a GARCH type of structure in the form of

$$\tau_t = b\tau_{t-1} + a \sum_{i=1}^K \mathbf{y}_{t-1}^2 \quad \text{or} \quad \tau_t = b\tau_{t-1} + a \sum_{i=1}^K \alpha_{i,t-1} y_{i,t-1}.$$

To introduce such heteroscedasticity in the t VARMA model, a much more complicated structure is needed for the time-varying error covariance matrix.

Fifth, if one uses the Gaussian pseudo-likelihood to estimate the DARMA model under the representation (3), the resulting estimator is the same as the corresponding t VARMA model using the y -link function as the transformation; see Section 3 for details. Under mild conditions, such an estimator is consistent. Hence, the DARMA model and the t VARMA model often yield similar estimated AR and MA parameter matrices. However, our simulation results, which are discussed in Section 4, show that the maximum likelihood estimator under the DARMA model often has a smaller standard error than GMLE if the DARMA is the true data generating model.

3. Model estimation

In this section, we consider three approaches for parameter estimation of the proposed DARMA models. The first is the maximum likelihood estimation (MLE) based on the exact likelihood function; the second is the approximate MLE (AMLE) based on the approximate likelihood through an approximation of the link function; and the third is the Gaussian MLE (GMLE) of the AR and MA coefficient matrices using the Gaussian pseudo-likelihood estimation based on the representation of (3). The GMLE is the same as that under the t VARMA model.

The AMLE is a simple estimator for the DARMA model. It can also be viewed as the MLE for the DARMA model with an approximate link function for which the inverse of the approximate link function is analytically simple. GMLE is a commonly used estimator for complex models, especially non-Gaussian time series. It can be obtained quickly with existing software. In general, it also has good asymptotic properties. MLE is a “true” estimator without any approximation and, thus, has the best performance. However, MLE requires nonlinear optimization and may not be numerically stable. A good initial value will help significantly. AMLE and GMLE are good initial values.

3.1. MLE procedure

Let θ be the parameter vector containing all model parameters. For the alr-DARMA model and clr-DARMA model, it contains the AR and MA matrices and the concentration parameter τ .

Suppose the available data set is $\{\mathbf{y}_T, \dots, \mathbf{y}_1, \mathbf{y}_0, \dots, \mathbf{y}_{1-m}\}$, where $m = \max(p, q)$. The log-likelihood function conditional on the initial values $\{\mathbf{y}_0, \dots, \mathbf{y}_{1-m}\}$ is

$$L_T(\theta) = \sum_{t=1}^T \ell_t(\theta), \tag{14}$$

where

$$\ell_t(\theta) = \ln\{f(\mathbf{y}_t | \mathcal{F}_{t-1})\} = \ln \Gamma \left(\sum_{i=1}^K \alpha_{it} \right) - \ln \left\{ \sum_{i=1}^K \Gamma(\alpha_{it}) \right\} + \sum_{i=1}^K (\alpha_{it} - 1) \ln y_{it},$$

in which $\alpha_t = g^{-1}(\eta_t)$ can be obtained recursively under (3) given θ and a set of initial values $\eta_0, \dots, \eta_{1-q}$ if $q > 0$. As shown, $g^{-1}(\eta_t)$ is uniquely defined and can be obtained numerically for the equivalent alr-DARMA and clr-DARMA models. If another y -link function is used, careful study is needed to ensure that $g^{-1}(\eta_t)$ is clearly defined. If $g^{-1}(\eta_t)$ is not uniquely defined, the one that maximized $\ell_t(\theta)$ may be used. Finally, the MLE is obtained by maximizing (14) using nonlinear optimization procedures. When the MA order q is not zero, the initial value of $\eta_0, \dots, \eta_{1-q}$ can be treated as unknown parameters to be estimated, or it can be set to zero for simplicity.

The theory of Hall and Heyde [21] can be applied to study the asymptotic distribution of the MLE. For a one-dimensional M-GARMA model, Zheng et al. [38] provided sufficient conditions that ensure the asymptotic normality of the likelihood estimators. Since the DARMA model is a specific extension of the M-GARMA model, it is possible to carry out similar theoretical investigations. However, an investigation into concrete DARMA models is much more difficult due to the complexity of solving the vector α_t from a system of nonlinear equations. The solution $\alpha_t = g^{-1}(\eta_t)$ does not have an analytic expression, although it can be obtained numerically. The problem is under investigation but is beyond the scope of the current paper. We can use the estimated Fisher information matrix to obtain the standard error of our estimates.

3.2. AMLE procedure

The evaluation of the exact likelihood function requires numerically solving a system of nonlinear equations to obtain $g^{-1}(\eta_t)$. This is computationally intensive because the nonlinear optimization of the likelihood is not trivial and requires many such evaluations. Hence, a good initial value can significantly reduce the computational burden. For this purpose, we use an approximate link function \tilde{g} so that its inverse \tilde{g}^{-1} can be readily obtained. In this case, the likelihood function is an approximation of the true likelihood, and the resulting estimator is referred to as the approximate MLE or AMLE.

Following Zheng et al. [38], we approximate the link function using its first-order Taylor approximation. Specifically, under the approximation $E\{h(\mathbf{y}_t) | \mathcal{F}_{t-1}\} \approx h\{E(\mathbf{y}_t | \mathcal{F}_{t-1})\}$, we use $\tilde{g}(\alpha_t) = h(\mu_t) = h(\alpha_t/\tau)$, where $\mu_t = E(\mathbf{y}_t | \mathcal{F}_{t-1})$. By

replacing the link function g with this approximation, the model can be viewed as a multivariate extension of the generalized ARMA (GARMA) framework proposed by Benjamin et al. [5]. Under our framework, this model is treated as an approximation.

For the alr-DARMA model, the approximation becomes

$$\tilde{g}(\boldsymbol{\alpha}_t) = \text{alr}(\boldsymbol{\mu}_t) = \left(\ln \alpha_{1t} - \ln \left(\tau - \sum_{i=1}^k \alpha_{it} \right), \dots, \ln \alpha_{kt} - \ln \left(\tau - \sum_{i=1}^k \alpha_{it} \right) \right)^\top,$$

where $\tau = \alpha_{1t} + \dots + \alpha_{kt}$. Its inverse function is

$$\tilde{\alpha}_{it} = \tilde{g}^{-1}(\boldsymbol{\eta}_t) = \frac{\tau e^{\eta_{it}}}{1 + e^{\eta_{1t}} + \dots + e^{\eta_{kt}}}$$

for each $i \in \{1, \dots, k\}$, and $\tilde{\alpha}_{kt} = \tau (1 + e^{\eta_{1t}} + \dots + e^{\eta_{kt}})^{-1}$. This approximation makes computation much easier.

For the clr-DARMA model, the approximate solution for g is

$$\begin{aligned} \tilde{g}(\boldsymbol{\alpha}_t) &= \text{clr}(\boldsymbol{\mu}_t) = J \ln \boldsymbol{\mu}_t = J(\ln \alpha_{1t} - \ln \tau, \dots, \ln \alpha_{kt} - \ln \tau)^\top \\ &= (\ln \alpha_{1t} - \zeta_t, \dots, \ln \alpha_{kt} - \zeta_t)^\top, \end{aligned}$$

where $\tau = \alpha_{1t} + \dots + \alpha_{kt}$ and $\zeta_t = K^{-1} \sum_{i=1}^K \ln \alpha_{it}$. For this approximation, we still have $\mathbf{1}_K^\top \tilde{g}(\boldsymbol{\alpha}_t) = 0$. The inverse function $\tilde{g}^{-1}(\boldsymbol{\eta}_t)$ is then

$$\tilde{\alpha}_{it} = \frac{\tau e^{\eta_{it}}}{e^{\eta_{1t}} + \dots + e^{\eta_{kt}} + e^{\eta_{kt}}}$$

for each $i \in \{1, \dots, K\}$. Note that the AMLE of the clr-DARMA and alr-DARMA models are equivalent under linear reparametrization.

3.3. GMLE procedure

Representation (3) of the DARMA model is in a form of the t VARMA model, although the error process is a MDS. Following Yao and Brockwell [37] and Zheng et al. [38], we can use a Gaussian pseudo-likelihood approach to estimate the ARMA coefficients in the model, replacing the complicated noise MDS in (3) with a Gaussian white sequence with constant variance. This is equivalent to obtaining the standard MLE on the t VARMA model given by (13) directly. The estimator will be referred to as GMLE. Since this becomes a standard vector ARMA model estimation, many methods can be used [27,36].

In practice, we can further estimate the fixed parameter τ by replacing the AR and MA coefficient matrices in (14) with their corresponding GMLEs to optimize for τ . Since this is a one-dimensional optimization, the procedure is relatively simple because we can use the residuals from the t VARMA model \mathbf{e}_t^G to obtain $\boldsymbol{\eta}_t^G = h(\mathbf{y}_t) - \mathbf{e}_t^G$. Then $\boldsymbol{\alpha}_t$ can be recursively solved using a fixed τ and given $\boldsymbol{\eta}_t^G$, and the corresponding likelihood function can be evaluated.

Significant studies have been done on the asymptotic behavior of the GMLE for time series; see, e.g., [19,22,23,27,37]. These studies have proved the consistency under weak assumptions on the noise process, as well as obtained asymptotic normality under a conditionally homoscedastic martingale difference assumption on the linear innovations. In our model, although the noise process is a MDS, it is heteroscedastic. The approach used in Zheng et al. [38] may apply in our setting. A detailed investigation is ongoing.

4. Simulation studies

This section investigates the finite-sample performances of the proposed three estimators, MLE, AMLE, and GMLE, for the DARMA model. The performances of the proposed DARMA models are also compared with the existing methods used to fit a compositional data set generated from an independent model.

All estimates are obtained by applying a constraint optimization technique that uses the MaxSQPF algorithm, implementing a sequential quadratic programming technique; see Nocedal and Wright [30]. We also use the solver for systems of nonlinear equations (SolveNLE) in OxMetrics software [17] to recursively get the solution of $g^{-1}(\boldsymbol{\eta})$.

4.1. Finite-sample performances of the estimators under the DARMA model

Consider a trivariate alr-DAR(p) model with lag order $p = 1$. We simulate a time series of length T ($T = 100, 200$, or 500):

$$\mathbf{y}_t \mid \boldsymbol{\alpha}_t \sim \text{Dir}(\boldsymbol{\alpha}_t), \quad g(\boldsymbol{\alpha}_t) = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{alr}(\mathbf{y}_{t-1}), \quad \text{and} \quad \tau = \sum_{i=1}^3 \alpha_{it},$$

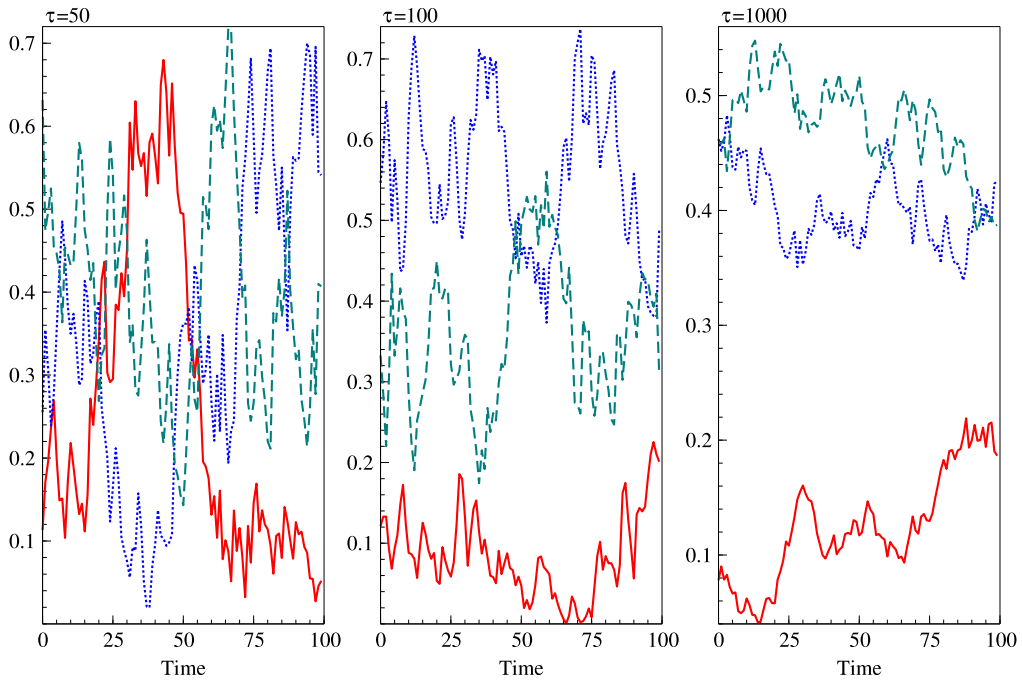


Fig. 1. Simulated series of compositions from the alr-DAR(1) process.

where $g(\alpha_t) = E\{\text{alr}(\mathbf{y}_t) \mid \mathcal{F}_{t-1}\}$ is defined by (6). The specific true parameter values are assigned as follows:

$$a_{10} = -0.07, \quad a_{20} = 0.01, \quad a_{11} = 0.95, \quad a_{12} = -0.05, \quad a_{21} = 0.01, \quad \text{and} \quad a_{22} = 0.95.$$

Three different values of τ are considered: $\tau = 50, 100,$ and 1000 . Fig. 1 shows a simulated series of $T = 100$ with three τ values. As expected, the series with a larger τ is relatively more stable and has a smaller unconditional variance.

We estimate the parameters using MLE, AMLE, and GMLE. Each setting is repeated 500 times. The means and standard errors of the estimates are presented in Table 1.

Several observations can be made from Table 1. First, GMLE performs worse than MLE when the scale parameter τ is small (large variance). The parameter estimates of GMLE are more biased than those of MLE. Moreover, the root mean squared errors (in parentheses) of the GMLE are evidently larger than those of MLE. Second, GMLE performs very well and the corresponding estimates are very close to those of MLE when the τ is large (small variance). Moreover, the GMLE becomes closer to MLE as the length of the time series increases. Third, AMLE performs poorly, especially when τ is small; the estimates are always biased due to the linear approximation. Finally, all estimates, including MLE, AMLE, and GMLE, of the parameter τ tend to be close. Of course, MLE requires nonlinear optimization, which takes a longer computational time and may occasionally become stuck in a local mode. Using GMLE as the initial value of MLE has shown to be very effective in saving computational time and improving accuracy.

4.2. DARMA model performance when tVARMA is the true model

We provide a numerical comparison of the performance of the proposed DARMA models with the existing traditional models when the true data generating model is a trivariate alr- $\text{VAR}(1)$ model.

We simulate a 2-dimensional time series (\mathbf{x}_t) from the following bivariate $\text{VAR}(1)$ process:

$$\mathbf{x}_t = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right),$$

with $\sigma_1 = 0.05, \sigma_2 = 0.05, \rho = 0.30,$ and

$$a_{10} = -0.07, \quad a_{20} = 0.01, \quad a_{11} = 0.95, \quad a_{12} = -0.05, \quad a_{21} = 0.01, \quad a_{22} = 0.95.$$

Based on \mathbf{x}_t , a compositional time series data (\mathbf{y}_t) , is obtained by setting, for each $t \in \{1, \dots, 500\}$,

$$y_{1t} = \frac{e^{x_{1t}}}{1 + e^{x_{1t}} + e^{x_{2t}}}, \quad y_{2t} = \frac{e^{x_{2t}}}{1 + e^{x_{1t}} + e^{x_{2t}}}, \quad y_{3t} = \frac{1}{1 + e^{x_{1t}} + e^{x_{2t}}}.$$

Table 1
Simulation results of the trivariate alr-DAR(1) model.

Parameter	True	T = 100			T = 200			T = 500		
		MLE	AMLE	GMLE	MLE	AMLE	GMLE	MLE	AMLE	GMLE
a_{10}	-0.07	-0.1119 (0.1136)	-0.1560 (0.1855)	-0.2452 (0.3029)	-0.0923 (0.0555)	-0.1276 (0.1056)	-0.2209 (0.2399)	-0.0741 (0.0263)	-0.1001 (0.0646)	-0.1991 (0.1572)
a_{20}	0.01	0.0108 (0.0775)	0.0130 (0.0794)	0.0069 (0.1056)	0.0102 (0.0415)	0.0124 (0.0395)	0.0089 (0.0544)	0.0123 (0.0236)	0.0147 (0.0225)	0.0109 (0.0343)
a_{11}	0.95	0.9047 (0.0701)	0.8044 (0.0989)	0.8237 (0.0918)	0.9270 (0.0433)	0.8245 (0.0785)	0.8437 (0.0520)	0.9423 (0.0257)	0.8391 (0.0651)	0.8449 (0.0782)
a_{12}	-0.05	-0.0610 (0.0801)	-0.0235 (0.1040)	-0.0469 (0.1797)	-0.0570 (0.0425)	-0.0271 (0.0536)	-0.0508 (0.1253)	-0.0526 (0.0210)	-0.0275 (0.0316)	-0.0393 (0.0894)
a_{21}	0.01	0.0122 (0.0439)	0.0143 (0.0428)	0.0153 (0.0519)	0.0116 (0.0220)	0.0129 (0.0213)	0.0132 (0.0253)	0.0118 (0.0120)	0.0128 (0.0114)	0.0125 (0.0175)
a_{22}	0.95	0.9050 (0.0606)	0.8679 (0.0578)	0.8809 (0.0589)	0.9248 (0.0358)	0.8879 (0.0332)	0.9041 (0.0376)	0.9414 (0.0199)	0.8989 (0.0190)	0.9158 (0.0437)
τ	50	52.141 (5.4658)	51.610 (6.0290)	49.731 (7.4510)	51.419 (3.8420)	51.061 (4.5749)	48.456 (7.0554)	50.560 (2.5311)	50.338 (3.5077)	47.234 (6.0223)
a_{10}	-0.07	-0.1150 (0.0905)	-0.1396 (0.1143)	-0.1696 (0.1424)	-0.0962 (0.0249)	-0.1189 (0.0800)	-0.1617 (0.1886)	-0.0779 (0.0257)	-0.0951 (0.0371)	-0.1404 (0.1153)
a_{20}	0.01	0.0176 (0.0615)	0.0178 (0.0612)	0.0165 (0.0689)	0.0137 (0.0371)	0.0144 (0.0362)	0.0141 (0.0388)	0.0121 (0.0181)	0.0129 (0.0176)	0.0125 (0.0207)
a_{11}	0.95	0.9073 (0.0655)	0.8615 (0.0716)	0.8717 (0.0663)	0.9271 (0.0403)	0.8772 (0.0568)	0.8899 (0.0520)	0.9426 (0.0211)	0.8947 (0.0302)	0.8980 (0.0521)
a_{12}	-0.05	-0.0660 (0.0762)	-0.0471 (0.0837)	-0.0561 (0.1147)	-0.0590 (0.0450)	-0.0427 (0.0484)	-0.0622 (0.1109)	-0.0538 (0.0207)	-0.0405 (0.0228)	-0.0493 (0.0463)
a_{21}	0.01	0.0178 (0.0380)	0.0181 (0.0375)	0.0187 (0.0395)	0.0135 (0.0229)	0.0139 (0.0224)	0.0141 (0.0237)	0.0117 (0.0107)	0.0121 (0.0105)	0.0121 (0.0120)
a_{22}	0.95	0.9057 (0.0548)	0.8920 (0.0527)	0.8979 (0.0531)	0.9274 (0.0321)	0.9128 (0.0306)	0.9218 (0.0315)	0.9423 (0.0175)	0.9272 (0.0162)	0.9374 (0.0190)
τ	100	104.80 (10.600)	104.81 (11.270)	103.19 (11.523)	101.96 (7.3299)	101.81 (8.6860)	100.15 (10.245)	100.84 (4.2393)	101.05 (4.3596)	98.568 (8.6143)
a_{10}	-0.07	-0.1336 (0.0817)	-0.1351 (0.0815)	-0.1357 (0.0825)	-0.0975 (0.0474)	-0.0991 (0.0473)	-0.0992 (0.0477)	-0.0797 (0.0236)	-0.0813 (0.0235)	-0.0808 (0.0243)
a_{20}	0.01	0.0152 (0.0534)	0.0152 (0.0533)	0.0152 (0.0534)	0.0148 (0.0286)	0.0148 (0.0285)	0.0149 (0.0287)	0.0115 (0.0153)	0.0115 (0.0153)	0.0116 (0.0154)
a_{11}	0.95	0.9035 (0.0596)	0.9000 (0.0593)	0.9020 (0.0602)	0.9302 (0.0349)	0.9265 (0.0347)	0.9289 (0.0351)	0.9431 (0.0178)	0.9394 (0.0176)	0.9423 (0.0182)
a_{12}	-0.05	-0.0719 (0.0862)	-0.0705 (0.0859)	-0.0727 (0.0879)	-0.0633 (0.0458)	-0.0619 (0.0457)	-0.0646 (0.0473)	-0.0558 (0.0234)	-0.0544 (0.0233)	-0.0566 (0.0245)
a_{21}	0.01	0.0165 (0.0378)	0.0165 (0.0378)	0.0166 (0.0379)	0.0147 (0.0208)	0.0147 (0.0208)	0.0148 (0.0209)	0.0115 (0.0110)	0.0115 (0.0110)	0.0115 (0.0111)
a_{22}	0.95	0.8943 (0.0553)	0.8933 (0.0552)	0.8938 (0.0551)	0.9237 (0.0313)	0.9226 (0.0313)	0.9233 (0.0313)	0.9399 (0.0172)	0.9388 (0.0171)	0.9397 (0.0172)
τ	1000	1051.7 (106.73)	1051.7 (106.73)	1051.3 (106.73)	1025.8 (74.069)	1025.8 (74.069)	1025.5 (74.051)	1008.9 (44.105)	1008.9 (44.104)	1008.8 (44.088)

Note: For each cell, the statistics given are based on 500 simulated samples, each consisting of a time series of length $T = 100, 200,$ and 500 . The mean and root mean squared error (in parentheses) for each estimator are shown.

The generated process (\mathbf{y}_t) follows the alr-VAR(1) process under (13). It is also equivalent to a clr-VAR(1) process with $\mathbf{x}_t^* = P_2 \mathbf{x}_t$. That is,

$$\text{clr}^*(\mathbf{y}_t) = \begin{bmatrix} a_{10}^* \\ a_{20}^* \end{bmatrix} + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} \text{clr}^*(\mathbf{y}_{t-1}) + \boldsymbol{\varepsilon}_t^*, \quad \boldsymbol{\varepsilon}_t^* \sim \mathcal{N} \left(0, \begin{bmatrix} \sigma_1^{*2} & \rho^* \sigma_1^* \sigma_2^* \\ \rho^* \sigma_1^* \sigma_2^* & \sigma_2^{*2} \end{bmatrix} \right).$$

The induced parameters are $\sigma_1^* = 0.0325, \sigma_2^* = 0.0325, \rho^* = -0.6579,$ and

$$a_{10}^* = -0.05, \quad a_{20}^* = 0.03, \quad a_{11}^* = 0.91, \quad a_{12}^* = -0.07, \quad a_{21}^* = 0.03, \quad a_{22}^* = 0.99.$$

We then estimate the generated data with the alr-DAR(1), clr-DAR(1), alr-VAR(1), and clr-VAR(1) models, respectively. Each simulation is repeated 500 times. Table 2 presents the simulation results. Both the means and standard errors of the estimates and the means of the residual sum of squares (mRSS) are reported. The mRSS for i th composition is defined as:

$$\text{mRSS}_i = \frac{1}{500} \sum_{\ell=1}^{500} \sum_{t=1}^T (y_{it}^{(\ell)} - \hat{\mu}_{it}^{(\ell)})^2,$$

Table 2
Simulation results of estimating and fitting the alr- $\text{VAR}(1)$ model.

Parameter	True	alr-DAR	alr-VAR	Parameter	True*	clr-DAR	clr-VAR
a_{10}	-0.07	-0.0813 (0.0216)	-0.0812 (0.0214)	a_{10}^*	-0.05	-0.0583 (0.0139)	-0.0582 (0.0137)
a_{20}	0.01	0.0122 (0.0215)	0.0122 (0.0215)	a_{20}^*	0.03	0.0352 (0.0138)	0.0352 (0.0137)
a_{11}	0.95	0.9418 (0.0162)	0.9418 (0.0160)	a_{11}^*	0.91	0.8981 (0.0229)	0.8982 (0.0226)
a_{12}	-0.05	-0.0543 (0.0155)	-0.0543 (0.0153)	a_{12}^*	-0.07	-0.0752 (0.0227)	-0.0751 (0.0225)
a_{21}	0.01	0.0121 (0.0158)	0.0121 (0.0158)	a_{21}^*	0.03	0.0330 (0.0239)	0.0329 (0.0238)
a_{22}	0.95	0.9400 (0.0178)	0.9399 (0.0177)	a_{22}^*	0.99	0.9836 (0.0255)	0.9835 (0.0254)
τ		2539.8 (124.74)		τ		2539.8 (124.86)	
σ_1	0.05		0.0498 (0.0016)	σ_1^*	0.0325		0.0324 (0.0010)
σ_2	0.05		0.0498 (0.0016)	σ_2^*	0.0325		0.0324 (0.0010)
ρ	0.30		0.2983 (0.0402)	ρ^*	-0.6579		-0.6585 (0.0248)
mRSS ₁		1.3330	1.3333			1.3330	1.3333
mRSS ₂		6.7172	6.7173			6.7172	6.7173
mRSS ₃		5.7336	5.7339			5.7336	5.7339

Note: For each cell, the statistics given are based on 500 simulated samples, each consisting of a time series of length $T = 500$. The mean and root mean squared error (in parentheses) of the parameter estimates and the mean of the residual sum of squares (mRSS $\times 10^{-2}$) estimated with different models are shown.

where $y_{it}^{(\ell)}$ is the ℓ th simulation of y_{it} at time t and $\widehat{\mu}_{it}^{(\ell)}$ is the corresponding fitted value. From the table, we can see that the estimates of the AR coefficient matrices under the alr-VAR (clr-VAR) model are very close to the estimates under the alr-DAR (clr-DAR) model. The residual sums of squares are also very close, although the DAR model has less total number of parameters. This shows that the DAR models perform well even though the true model is a t VAR model. The impact of the linear reparametrization in this example is minimal, although the variation in the estimates can be different.

5. Analyzing expenditure shares in the UK

In this application, we use the alr-DAR and alr-VAR models to fit a quarterly compositional time series of consumption (y_1), investment (y_2), government expenditure (y_3), and export (y_4) shares of the UK gross final expenditure from the first quarter of 1955 to the fourth quarter of 2013, a total of 236 observations. The data are seasonally adjusted and are shown in Fig. 2. The subperiod (1955:Q1–2005:Q4) of these data set has been analyzed by Mills [29] and Barceló-Vidal, Aguilar and Martín-Fernández [4].

For alr modeling, we use export y_{4t} as the reference component, and the y -link function is $\text{alr}(\mathbf{y}_t) = (\ln(y_{1t}/y_{4t}), \ln(y_{2t}/y_{4t}), \ln(y_{3t}/y_{4t}))^\top$. Based on the Bayesian information criteria (BIC), the order of the alr-DAR process is determined to be $p = 2$, which is consistent with that found by Mills [29]. The corresponding alr-VAR model on $\text{alr}(\mathbf{y}_t)$ assumes the form

$$\text{alr}(\mathbf{y}_t) = A_0 + A_1 \text{alr}(\mathbf{y}_{t-1}) + A_2 \text{alr}(\mathbf{y}_{t-2}) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, DRD),$$

where $D = \text{diag}(\sigma_1, \sigma_2, \sigma_3)$ and $R = (\rho_{ij})$ with $\rho_{ij} = \rho_{ji}$ and $\rho_{ii} = 1$ for all $i, j \in \{1, 2, 3\}$. We also tried the equivalent clr-VAR models and found that the impact of reparametrization is very small. Hence, the results will not be reported.

The estimation results are shown in Table 3. The top panel shows the estimates and their standard errors obtained from the estimated Fisher information matrices. The estimates reported under the alr-DAR model are exact MLEs. It can be seen that the parameter estimates of the AR coefficient matrices from the alr-DAR model are close to those from the alr-VAR model. Again, if the true model is indeed alr-DAR, then the estimate of the coefficient matrices under the alr-VAR model should be the same as GMLE under the true model of the alr-DAR, which should be close to MLE under alr-DAR. The standard errors of the alr-DAR model parameters are generally smaller than the corresponding ones under the alr-VAR model. The DAR models have an extra concentration parameter τ , while the t VAR models have error covariance matrices.

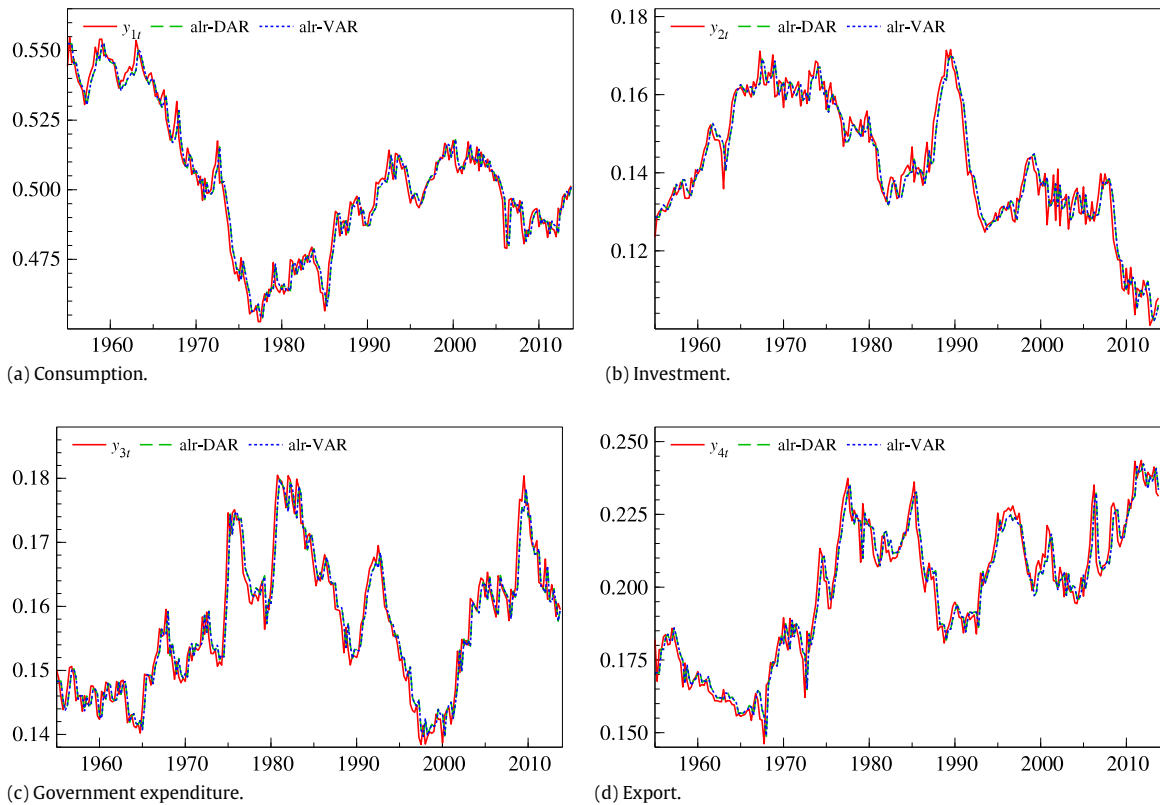


Fig. 2. The shares of UK gross final expenditure and the fitted values with alr-DAR and alr-VAR models.

In the bottom panel of [Table 3](#), we present several statistics to compare the different methods. In the first four rows, RSS_i stands for the residual sum of squares for the i th component, which is defined by

$$RSS_i = \sum_{t=1}^T (y_{it} - \hat{\mu}_{it})^2,$$

where $\hat{\mu}_{it}$ is the fitted value of y_{it} . [Fig. 2](#) plots the fitted values under the alr-DAR and alr-VAR models. For the DAR models, $\hat{\mu}_{it} = \hat{\alpha}_{it}/\hat{\tau}$ for all $i \in \{1, 2, 3, 4\}$. Based on the results of the RSS_i , the alr-DAR model performs slightly better than the alr-VAR model.

Finally, we perform several statistical tests to check the autocorrelation and normality of the ARMA residuals. The residual series of the alr-DAR is obtained as $\text{alr}(\mathbf{y}_t) - \hat{\boldsymbol{\eta}}_t$. It should behave as a martingale process according to (3), but should not be normally distributed. The residual series of the alr-VAR are based on the AR(2) model on the transformed series $\text{alr}(\mathbf{y}_t)$. [Fig. 3](#) plots the alr-DAR residuals. The vector Portmanteau test statistics of Li and McLeod [25] is based on the residual series, and is shown as $VQ(4)$ and $VQ(12)$ in [Table 3](#), using orders 4 and 12, respectively. There is no autocorrelation left in the residuals for both models, indicating that the selected lag order $p = 2$ is sufficient. In addition, the vector normality test statistic of Doornik and Hansen [18] for the residuals is shown as VN in the table. Apparently the normal assumption of the alr-VAR(2) is inappropriate, with the statistic 12.8 and a p -value of 0.047. The vector normality test also confirmed that the alr-DAR(2) residual process is not normally distributed, with the statistic 14.8 and a p -value of 0.02.

[Fig. 4](#) plots the conditional covariance matrices under the alr-DAR model based on (7). It is clearly not constant. Note that the alr-VAR model assumes constant conditional covariance.

6. Conclusions and discussion

This paper presented a new class of compositional time series models on the simplex, assuming a Dirichlet conditional distribution of the observations, with time varying Dirichlet parameters. The varying parameters, via a suitable link function, follow a vector ARMA type of structure. The framework is general in modeling compositional time series, and provides a cleaner and easier interpretation than the existing t VARMA models. Empirical study shows that when the underlying data generating mechanism is indeed a DARMA model, the estimation results from DARMA model are indeed better than that

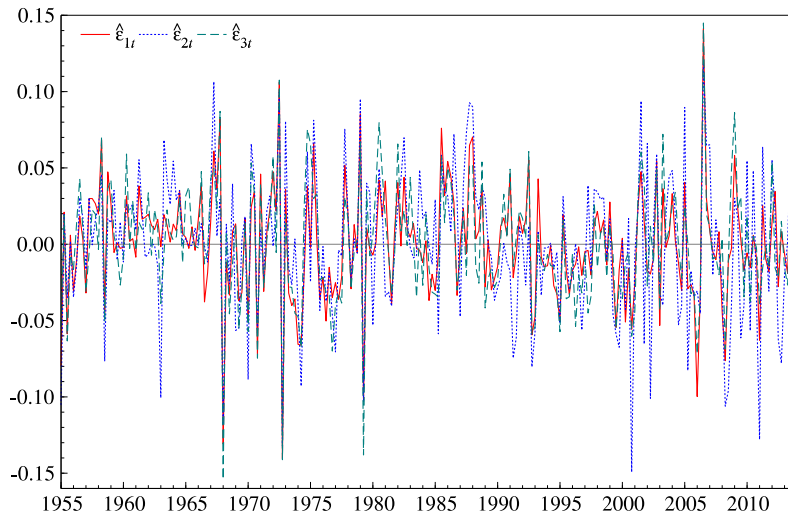


Fig. 3. Residuals of estimated alr-DAR model.

Table 3
Estimation results with alr-DAR and alr-VAR models.

	alr-DAR	alr-VAR
A_0	$\begin{bmatrix} -0.0234 \\ (0.0257) \\ -0.0869 \\ (0.0337) \\ -0.0223 \\ (0.0326) \end{bmatrix}$	$\begin{bmatrix} -0.0273 \\ (0.0317) \\ -0.0915 \\ (0.0408) \\ -0.0258 \\ (0.0344) \end{bmatrix}$
A_1	$\begin{bmatrix} 1.0259 & -0.1054 & -0.0549 \\ (0.1202) & (0.0562) & (0.0992) \\ 0.4376 & 0.5765 & -0.1824 \\ (0.1579) & (0.0747) & (0.1308) \\ 0.1392 & -0.1178 & 0.8638 \\ (0.1524) & (0.0715) & (0.1260) \end{bmatrix}$	$\begin{bmatrix} 1.0152 & -0.0964 & -0.0718 \\ (0.1483) & (0.0695) & (0.1229) \\ 0.4437 & 0.5804 & -0.2069 \\ (0.1910) & (0.0895) & (0.1583) \\ 0.1249 & -0.1046 & 0.8438 \\ (0.1610) & (0.0755) & (0.1335) \end{bmatrix}$
A_2	$\begin{bmatrix} 0.0214 & 0.1258 & -0.0496 \\ (0.1207) & (0.0569) & (0.0994) \\ -0.3767 & 0.4292 & 0.0539 \\ (0.1586) & (0.0756) & (0.1311) \\ -0.1347 & 0.1683 & -0.0100 \\ (0.1531) & (0.0725) & (0.1262) \end{bmatrix}$	$\begin{bmatrix} -0.0074 & 0.1162 & -0.0351 \\ (0.1490) & (0.0704) & (0.1230) \\ -0.3789 & 0.4257 & 0.0737 \\ (0.1919) & (0.0907) & (0.1585) \\ -0.1171 & 0.1545 & 0.0089 \\ (0.1618) & (0.0765) & (0.1336) \end{bmatrix}$
τ	8158.6 (2.9454)	
D		$\begin{bmatrix} 0.0360 & 0 & 0 \\ (0.0017) & & \\ 0 & 0.0464 & 0 \\ & (0.0021) & \\ 0 & 0 & 0.0392 \\ & & (0.0018) \end{bmatrix}$
R		$\begin{bmatrix} 1 & 0.7225 & 0.8740 \\ & (0.0312) & (0.0154) \\ 0.7225 & 1 & 0.6254 \\ (0.0312) & & (0.0398) \\ 0.8740 & 0.6254 & 1 \\ (0.0154) & (0.0398) & \end{bmatrix}$
$RSS_1 (10^{-3})$	5.1918	5.1962
$RSS_2 (10^{-3})$	3.6329	3.6348
$RSS_3 (10^{-3})$	1.8435	1.8455
$RSS_4 (10^{-3})$	7.5436	7.5565
Loglik	2881.82	1513.24
BIC	-5643.6	-2879.19
VQ(4)	34.918	38.540
VQ(12)	119.71	118.58
VN	14.907*	12.834*

* Indicates that the test statistic is significant at 5% levels. The standard deviation errors of the parameter estimates are reported in parentheses.

from a t VARMA model, though the ARMA parameter estimators of the t VARMA can be treated as the Gaussian pseudo-likelihood estimator. On the other hand, if the underlying data generating mechanism is actually a t VARMA model, fitting a DARMA model to the data, though more computationally costly, performs almost the same as that under the true model.

However, many questions remain. The probabilistic properties of the model, including stationary and ergodic conditions of the process is under investigation. Although representation (3) has a clear ARMA structure, the noise is a MDS with conditional heteroscedasticity. Its stationary and ergodic conditions are not trivial. The framework of Meyn and Tweedie [26] can be used, similar to the approach used by Zheng et al. [38]. The asymptotic properties of the MLE and GMLE need to

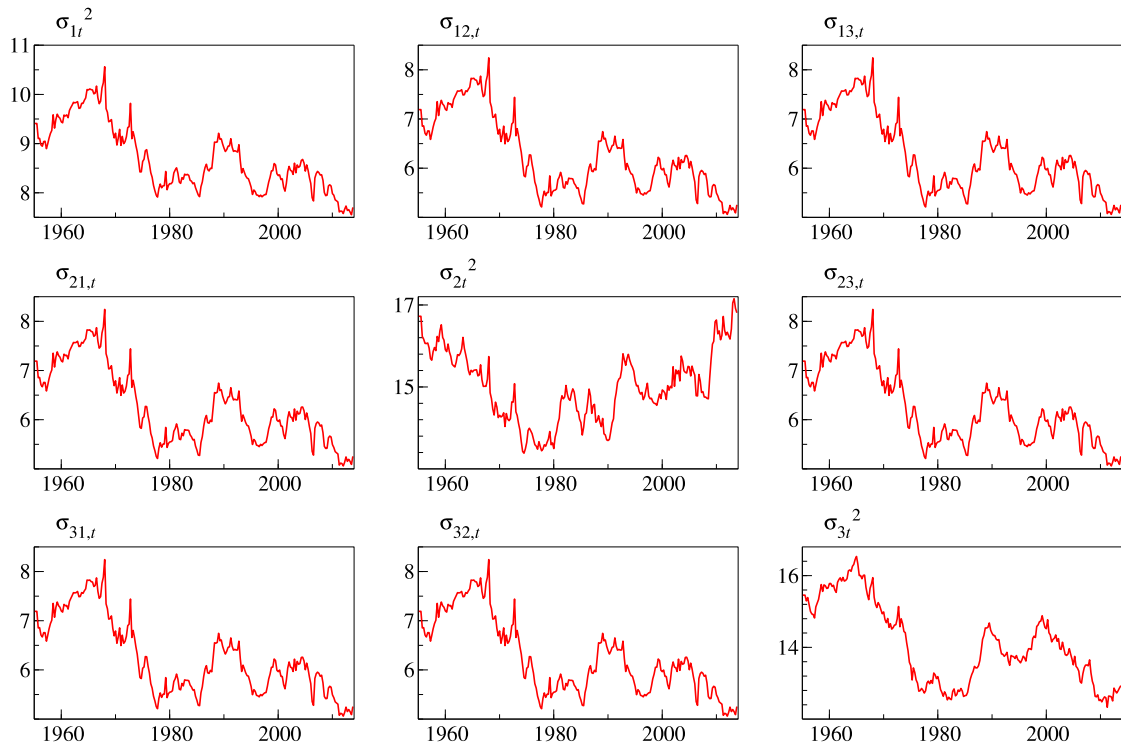


Fig. 4. Estimated conditional covariance matrices with the alr-DAR model.

be established. Another issue of the DARMA model is the relatively rigid conditional covariance structure of the Dirichlet distribution, although the conditional covariances change over time. One may use a generalized Dirichlet distribution [14] to increase flexibility.

Acknowledgments

The authors would like to thank the Editor, the Associate Editor and two anonymous reviewers for their careful reading, comments and suggestions to an earlier version of the paper. Zheng’s research was supported in part by the National Natural Science Foundation of China (No. 71371160), the Program for Changjiang Youth Scholar (2016) and the Program for New Century Excellent Talents in University (NCET-13-0509). Chen’s research was supported in part by National Science Foundation grants DMS-1513409 and DMS-1209085.

Appendix. Proofs

A.1. Proof of Proposition 1

We express the Dirichlet distribution in its exponential family representation:

$$f(\mathbf{y} \mid \boldsymbol{\alpha}) = H(\mathbf{y}) \exp\{\vartheta^\top T(\mathbf{y}) - A(\vartheta)\},$$

where $H(\mathbf{y}) = (y_1 \cdots y_K)^{-1}$, $\vartheta = (\alpha_1, \dots, \alpha_K)^\top$, $T(\mathbf{y}) = (\ln y_1, \dots, \ln y_K)^\top$,

$$A(\vartheta) = \sum_{i=1}^K \ln \Gamma(\alpha_i) - \ln \Gamma(\tau),$$

and $\tau = \alpha_1 + \cdots + \alpha_K$.

For a distribution in the exponential family, the expectation of $T(\mathbf{y})$ is $E\{T(\mathbf{y})\} = \partial A(\vartheta) / \partial \vartheta$ and its associated conditional variance is $\text{var}\{T(\mathbf{y})\} = \partial^2 A(\vartheta) / \partial \vartheta \partial \vartheta^\top$. Since $\tau = \alpha_1 + \cdots + \alpha_K$, we have

$$E(\ln y_i) = \frac{\partial A(\vartheta)}{\partial \vartheta_i} = \frac{\partial \ln \Gamma(\alpha_i)}{\partial \alpha_i} - \frac{\partial \ln \Gamma(\tau)}{\partial \tau} \frac{\partial \tau}{\partial \alpha_i} = \psi(\alpha_i) - \psi(\tau)$$

and its associated conditional covariance is

$$\text{cov}(\ln y_i, \ln y_j) = \frac{\partial^2 A(\vartheta)}{\partial \vartheta_i \partial \vartheta_j} = \frac{\partial \psi(\alpha_i)}{\partial \alpha_j} - \frac{\partial \psi(\tau)}{\partial \tau} \frac{\partial \tau}{\partial \alpha_j} = \psi_1(\alpha_i) \delta_{ij} - \psi_1(\tau),$$

where ψ is the digamma function, ψ_1 is the trigamma function, and δ_{ij} is the Kronecker delta satisfying that $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

Further, for all $i \in \{1, \dots, K - 1\}$, we have

$$E\{\ln(y_i/y_K)\} = E(\ln y_i - \ln y_K) = \psi(\alpha_i) - \psi(\alpha_K)$$

and the covariance between $\ln(y_i/y_K)$ and $\ln(y_j/y_K)$ is

$$\begin{aligned} &\text{cov}[\ln(y_i/y_K), \ln(y_j/y_K)] \\ &= \text{cov}(\ln y_i, \ln y_j) - \text{cov}(\ln y_i, \ln y_K) - \text{cov}(\ln y_K, \ln y_j) + \text{cov}(\ln y_K, \ln y_K) \\ &= \psi_1(\alpha_i) \delta_{ij} - \psi_1(\alpha_i) \delta_{iK} - \psi_1(\alpha_K) \delta_{Kj} + \psi_1(\alpha_K) \delta_{KK}. \end{aligned}$$

Since $i, j < K$, we have $\delta_{iK} = \delta_{Kj} = 0$ and $\delta_{KK} = 1$. Therefore,

$$\text{cov}\{\ln(y_i/y_K), \ln(y_j/y_K)\} = \psi_1(\alpha_i) \delta_{ij} + \psi_1(\alpha_K).$$

Putting these results into the vectors and the matrices, the conclusion follows. \square

A.2. Proof of Proposition 2

The system of equations $g(\alpha) = \eta$ and $\tau = \alpha_1 + \dots + \alpha_K$ can be rewritten as

$$\begin{cases} \psi(\alpha_1) - \psi(\tau - \alpha_1 - \dots - \alpha_k) = \eta_1 \\ \psi(\alpha_2) - \psi(\tau - \alpha_1 - \dots - \alpha_k) = \eta_2 \\ \vdots \\ \psi(\alpha_k) - \psi(\tau - \alpha_1 - \dots - \alpha_k) = \eta_k, \end{cases} \tag{A.1}$$

where ψ is a digamma function and $\alpha_k = \tau - \alpha_1 - \dots - \alpha_k$. We first show that there is a solution that satisfies the system of equations (existence), and then we prove that the solution is unique.

(i) Existence. Let $g^{(k)}(\alpha_k, \tau)$ be a k dimensional function on the domain of $\alpha_i > 0$ and $\tau = \alpha_1 + \dots + \alpha_K$. The i th function of $g^{(k)}(\alpha_k, \tau)$ is

$$g_i^{(k)}(\alpha_k, \tau) = \psi(\alpha_k) - \psi(\tau - \alpha_1 - \dots - \alpha_k).$$

We will show that the codomain of $g^{(k)}(\alpha_k, \tau)$ is the entire \mathbb{R}^k for any τ . For this, we use induction. For $k = 1$, $g_1^{(1)}(\alpha_1, \tau) = \psi(\alpha_1) - \psi(\tau - \alpha_1)$ is a continuous function. For any $\tau > 0$,

$$\lim_{\alpha_1 \rightarrow 0} g_1^{(1)}(\alpha_1, \tau) = \lim_{\alpha_1 \rightarrow 0} \psi(\alpha_1) - \psi(\tau) = -\infty,$$

and

$$\lim_{\alpha_1 \rightarrow \tau} g_1^{(1)}(\alpha_1, \tau) = \psi(\tau) - \lim_{\alpha_1 \rightarrow \tau} \psi(\tau - \alpha_1) = \infty.$$

Hence, the codomain of $g^{(k)}(\alpha_k, \tau)$ is \mathbb{R} for any $\tau > 0$. Assume at k , the codomain of $g^{(k)}(\alpha_k, \tau)$ is \mathbb{R}^k for any $\tau > 0$. Now consider $g^{(k+1)}(\alpha_{k+1}, \tau)$. For any $(\eta_1, \dots, \eta_k, \eta_{k+1}) \in \mathbb{R}^{k+1}$, $0 < c < \tau$, because the codomain of $g^{(k)}(\alpha_k, \tau - c)$ is \mathbb{R}^k . We can then find a set of $\alpha_k = (\alpha_1, \dots, \alpha_k)$ that satisfies $g^{(k)}(\alpha_k, \tau - c) = \eta_k$, where $\eta_k = (\eta_1, \dots, \eta_k)$ and $\alpha_1 + \dots + \alpha_k < \tau - c$. Since this solution depends on c , we denote it as $\alpha_k(c)$.

Let $s(c) = \alpha_1(c) + \dots + \alpha_k(c)$. By the uniqueness shown below, if there is a solution, the solution is unique. Therefore, $s(0)$ is a fixed finite number, the sum of the solutions of $g^{(k)}(\alpha_k, \tau) = \eta_k$, and $s(c) < \tau - c$. Let

$$g^*(c, \tau) = \psi(c) - \psi(\tau - s(c) - c).$$

We have

$$\lim_{c \rightarrow 0} g^*(c, \tau) = \lim_{c \rightarrow 0} \psi(c) - \psi\{\tau - s(0)\} = -\infty,$$

and when $c \rightarrow \tau$, $0 < \tau - s(c) - c < \tau - c \rightarrow 0$, hence $\psi\{\tau - s(c) - c\} \rightarrow -\infty$. Therefore,

$$\lim_{c \rightarrow \tau} g^*(c, \tau) = \psi(\tau) - \lim_{c \rightarrow \tau} \psi\{\tau - s(c) - c\} = \infty.$$

Since $g^*(c, \tau)$ is a continuous function, one can find a solution c^* such that $g^*(c^*, \tau) = \eta_{k+1}$. Let $\alpha_k = \alpha_k(c^*)$ and $\alpha_{k+1} = c^*$. We then have $g^{(k+1)}(\alpha_{k+1}, \tau) = \eta_{k+1}$, and the codomain of $g^{(k+1)}(\alpha_{k+1}, \tau)$ is \mathbb{R}^{k+1} for any τ .

(ii) Uniqueness. Suppose $\alpha^0 = (\alpha_1^0, \dots, \alpha_k^0)^\top$ is one solution of the system (A.1), and $\alpha^1 = (\alpha_1^1, \dots, \alpha_k^1)^\top$ is a different solution. Without loss of generality, we assume that $\alpha_1^0 < \alpha_1^1$. Hence, we have $\psi(\alpha_1^0) - \psi(\alpha_1^1) < 0$ due to the fact that the digamma function ψ is strictly increasing.

Based on the first equation of the system, we have $\psi(\alpha_1^0) - \psi(\tau - \alpha_1^0 - \dots - \alpha_k^0) = \eta_1$ and $\psi(\alpha_1^1) - \psi(\tau - \alpha_1^1 - \dots - \alpha_k^1) = \eta_1$. Consider the difference between the two equations, viz.

$$\psi(\tau - \alpha_1^0 - \dots - \alpha_k^0) - \psi(\tau - \alpha_1^1 - \dots - \alpha_k^1) = \psi(\alpha_1^0) - \psi(\alpha_1^1) < 0. \quad (\text{A.2})$$

Next, consider the i th equation of the system. Again, taking the difference of the two equations and using (A.2), we have

$$\psi(\alpha_i^0) - \psi(\alpha_i^1) = \psi(\tau - \alpha_1^0 - \dots - \alpha_k^0) - \psi(\tau - \alpha_1^1 - \dots - \alpha_k^1) < 0.$$

Since ψ is strictly increasing, we have $\alpha_i^0 < \alpha_i^1$ for all $i \in \{1, \dots, k\}$. However, (A.2) implies that $\tau - \alpha_1^0 - \dots - \alpha_k^0 < \tau - \alpha_1^1 - \dots - \alpha_k^1$, which results in a contradiction. Therefore, the solution of the system (A.1) is unique. \square

References

- [1] L. Aguilár Zuil, C. Barceló-Vidal, J.M. Larrosa, Compositional time series analysis: A review, in: Proceedings of the 56th Session of the International Statistical Institute, Lisboa, Portugal, August (2007), pp. 22–29.
- [2] J. Aitchison, The statistical analysis of compositional data (with discussion), *J. R. Stat. Soc. Ser. B Stat. Methodol.* 44 (1982) 139–147.
- [3] J. Aitchison, *The Statistical Analysis of Compositional Data*, Chapman and Hall, New York, 1986.
- [4] C. Barceló-Vidal, L. Aguilár, J.A. Martín-Fernández, Compositional VARIMA time series, in: V. Pawłowsky-Glahn, A. Buccianti (Eds.), *Compositional Data Analysis: Theory and Applications*, Wiley, Chichester, 2011, pp. 87–103.
- [5] M.A. Benjamin, R.A. Rigby, D.M. Stasinopoulos, Generalized autoregressive moving average models, *J. Amer. Statist. Assoc.* 98 (2003) 214–223.
- [6] J. Bergman, Compositional time series: An application, in: J. Daunis-i Estadella, J. Martín-Fernández (Eds.), *Proceeding of CoDaWork'08, The 3rd Compositional Data Analysis Workshop*, University of Girona, Girona, Spain, 2008.
- [7] A. Bhaumik, D.K. Dey, N. Ravishanker, A dynamic linear model approach for compositional time series analysis, Technical Report, University of Connecticut, Storrs, CT, 2003.
- [8] T.B. Brandt, B.L. Monroe, J.T. Williams, Time series models for compositional data, Working Paper, Indiana University, 1999.
- [9] J. Brehm, S. Gates, B. Gomez, A comparison of methods for compositional data analysis, in: Presented at the 1998 Political Methodology Society Annual Meeting, San Diego, CA, 1998.
- [10] T.M. Brunsdon, *The Time Series Analysis of Compositional Data* (Ph.D. thesis), University of Southampton, UK, 1987.
- [11] T.M. Brunsdon, T.M.F. Smith, The time series analysis of compositional data, *J. Off. Stat.* 14 (1998) 237–253.
- [12] G. Campbell, J.E. Mosimann, Multivariate methods for proportional shape, in: *ASA Proceedings of the Section on Statistical Graphics*, 1987, pp. 10–17.
- [13] G. Campbell, J.E. Mosimann, Modelling continuous proportional data with the Dirichlet distribution, Unpublished manuscript, 1987.
- [14] R.J. Connor, J.E. Mosimann, Concepts of independence for proportions with a generalization of the Dirichlet distribution, *J. Amer. Statist. Assoc.* 64 (1969) 194–206.
- [15] R.A. Davis, W.T.M. Dunsmuir, S.B. Streett, Observation-driven models for Poisson counts, *Biometrika* 90 (2003) 777–790.
- [16] R.A. Davis, R. Wu, A negative binomial model for time series of counts, *Biometrika* 96 (2009) 735–749.
- [17] J.A. Doornik, *Ox: An Object-Oriented Matrix Programming Language*, International Thomson Business Press, 2007.
- [18] J.A. Doornik, H. Hansen, An omnibus test for univariate and multivariate normality, *Oxford Bull. Econom. Statist.* 70 (2008) 927–939.
- [19] W.T.M. Dunsmuir, E.J. Hannan, Vector linear time series models, *Adv. Appl. Probab.* 8 (1976) 339–364.
- [20] G.K. Grunwald, A.E. Raftery, P. Guttorp, Time series of continuous proportions, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 55 (1993) 103–116.
- [21] P. Hall, C.C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press Inc., New York, 1980.
- [22] E.J. Hannan, M. Deistler, *The Statistical Theory of Linear Systems*, Wiley, New York, 1988.
- [23] E.J. Hannan, W.T.M. Dunsmuir, M. Deistler, Estimation of vector ARMAX models, *J. Multivariate Anal.* 10 (1980) 275–295.
- [24] R.H. Hijazi, R.W. Jernigan, Modeling compositional data using dirichlet regression models, *J. Appl. Probab. Statist.* 4 (2009) 77–91.
- [25] W.K. Li, A.I. McLeod, Distribution of the residual autocorrelations in multivariate ARMA time series models, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 43 (1981) 231–239.
- [26] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer, Berlin, 2005.
- [27] S. Meyn, R.L. Tweedie, *Markov Chains and Stochastic Stability*, second ed., Cambridge University Press, Cambridge, 2009.
- [28] T.C. Mills, Forecasting obesity trends in England, *J. Roy. Statist. Soc. Ser. A* 172 (2009) 107–117.
- [29] T.C. Mills, Forecasting compositional time series, *Qual. Quant.* 44 (2010) 673–690.
- [30] J. Nocedal, S. Wright, *Numerical Optimization*, Springer, New York, 1999.
- [31] J.M. Quintana, M. West, The time series analysis of compositional data, *Bayesian Stat.* 3 (1988) 747–756.
- [32] A.V. Rocha, F. Cribari-Neto, Beta autoregressive moving average models, *TEST* 18 (2009) 529–545.
- [33] D.B.N. Silva, *Modelling compositional time series from repeated surveys* (Ph.D. thesis), University of Southampton, 1996.
- [34] D.B.N. Silva, T.M.F. Smith, Modelling compositional time series from repeated surveys, *Surv. Methodol.* 27 (2001) 205–215.
- [35] T. Smith, T. Brunsdon, The time series analysis of compositional data, in: *Proceedings of the Survey Research Methods Section, American Statistical Association*, VA, 1989, pp. 26–32.
- [36] R.S. Tsay, *Multivariate Time Series Analysis with R and Financial Applications*, Wiley, New York, 2015.
- [37] Q. Yao, P.J. Brockwell, Gaussian maximum likelihood estimation for ARMA models I: Time series, *J. Time Series Anal.* 27 (2006) 857–875.
- [38] T. Zheng, H. Xiao, R. Chen, Generalized ARMA models with martingalized difference errors, *J. Econometrics* 189 (2015) 492–506.