Journal of Forecasting, *J. Forecast.* **34**, 303–314 (2015)
Published online 12 March 2015 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.2336

Modeling Compositional Time Series with Vector Autoregressive Models

PETRA KYNČLOVÁ,[1,2] PETER FILZMOSER[1]* AND KAREL HRON[2,3]
[1] *Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria*
[2] *Department of Mathematical Analysis and Applications of Mathematics, Palacký University, Olomouc, Czech Republic*
[3] *Department of Geoinformatics, Palacký University, Olomouc, Czech Republic*

ABSTRACT

Multivariate time series describing relative contributions to a total (like proportional data) are called compositional time series. They need to be transformed first to the usual Euclidean geometry before a time series model is fitted. It is shown how an appropriate transformation can be chosen, resulting in coordinates with respect to the Aitchison geometry of compositional data. Using vector autoregressive models, the standard approach based on raw data is compared with the compositional approach based on transformed data. The results from the compositional approach are consistent with the relative nature of the observations, while the analysis of the raw data leads to several inconsistencies and artifacts. The compositional approach is extended to the case when also the total of the compositional parts is of interest. Moreover, a concise methodology for an interpretation of the coordinates in the transformed space together with the corresponding statistical inference (like hypotheses testing) is provided. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS VAR model; compositional data; isometric log-ratio transformation; Granger causality

INTRODUCTION

Compositional data represent a special type of multivariate data that generally describe parts of a given whole (Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011). A D-part composition is defined as a vector $\mathbf{x} = (x_1, \ldots, x_D)^\top$ with strictly positive real components. They carry only relative information, which is given by the ratios between the components (parts). Most standard statistical methods assume that the analyzed data come from the real Euclidean space with Euclidean geometry, whereas the natural sample space of compositions is the simplex (Aitchison, 1986). Thus using classical statistical tools for modeling compositional data may lead to inadequate results.

Compositional time series (CTS) represent multivariate time series of compositions, often characterized by a constant sum constraint representation, at each time point t. Thus a CTS can be defined as the series $\{\mathbf{x}_t : t = 1, \ldots, n\}$, where $\mathbf{x}_t = (x_{1t}, \ldots, x_{Dt})^\top$ are elements of the simplex $\mathcal{S}^D$, the sample space of representations of compositional data to a chosen constant sum constraint κ. CTS are thus characterized by positive components $x_{1t}, \ldots, x_{Dt}$ with a constant sum at each time t (frequently the constant is taken as 1). This constraint forms in practice the crucial problem when modeling compositional time series by standard multivariate time series methods. From the methodological point of view, the problem with a statistical analysis of CTS using standard methods is caused by the specific geometry of compositional data, the Aitchison geometry on the simplex (Egozcue and Pawlowsky-Glahn, 2006), that accounts for inherent properties of compositional data (Egozcue, 2009).

Several approaches for modeling CTS have been introduced. The principal strategy is based on using log-ratio transformations. This procedure consists of transforming given CTS to the space of coordinates—the real vector space with Euclidean structure—to leave the Aitchison geometry and, practically, break the unit sum constraint of the original time series. Subsequently, standard multivariate time series methods can be applied to the transformed time series.

In the context of CTS, the most frequently used transformations have been additive log-ratio (alr) transformations (Aitchison, 1986; Mills, 2010; Barceló-Vidal *et al.*, 2011), although they lead to oblique coordinates with respect to Aitchison geometry. The reasonable alternative is represented by the isometric log-ratio (ilr) transformations (Egozcue *et al.*, 2005) that result in orthonormal coordinates. In (Bergmann, 2008) a particular choice of ilr coordinates was used in order to facilitate interpretation of the results; nevertheless, due to the apparent complexity of interpretation of the ilr coordinates, their systematic use for the analysis of CTS is still not fully accepted

* Correspondence to: Peter Filzmoser, Department of Statistics and Probability Theory, Vienna University of Technology, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria. E-mail: P.Filzmoser@tuwien.ac.at

(Barceló-Vidal *et al.*, 2011). Consequently, although several different approaches for analyzing CTS have been proposed (see Larrosa, 2005), even with a compositional vector integrated autoregressive moving average (VARIMA) model on the simplex (Barceló-Vidal *et al.*, 2011), compositional time series modeling does not appear to be extensively known.

According to Barceló-Vidal *et al.* (2011), the full compositional VARIMA model and estimation of the parameters do not depend on the specific log-ratio transformation used. However, restricted models, which are applied to facilitate the interpretation of parameters, lead to different compositional ARIMA models depending on the transformation applied to the data.

This paper is based on using a special choice of ilr transformation in order to facilitate a concise approach for the interpretation in coordinates. We focus on vector autoregressive (VAR) models, but an extension to more general models would be possible. The next section provides a general introduction to the geometry of compositional data, and the third section refers to special transformations in this context. The fourth section explains how the VAR model can be used for compositional data, and it also shows that the resulting final model and predictions do not depend on the particular choice of transformation. Further extensions concerning modeling both the relative and absolute information in the context of time series are contained in the fifth section. Practical examples in the sixth section highlight major differences of time series modeling and hypothesis testing when using untransformed or appropriately transformed data. The seventh section concludes.

THE SIMPLEX $\mathcal{S}^D$ AS A COMPOSITIONAL SPACE

The sample space of representations of D-part compositions to a chosen constant sum constraint is given by the simplex

$$\mathcal{S}^D = \left\{ (x_1, \ldots, x_D)^\top : x_i > 0, i = 1, \ldots, D; \sum_{i=1}^{D} x_i = \kappa \right\}$$

where κ is a positive constant. Owing to the relative character of compositional data, the specific choice of κ is not relevant; the information contained in the ratios between the compositional parts remains the same. The $(D-1)$-dimensional vector space structure on the simplex $\mathcal{S}^D$ is induced by the operations of perturbation and power transformation, defined for compositions $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ and $\alpha \in \mathbb{R}$ as

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, x_2 y_2, \ldots, x_D y_D)^\top, \quad \alpha \odot \mathbf{x} = \mathcal{C}\left(x_1^\alpha, x_2^\alpha, \ldots, x_D^\alpha\right)^\top$$

respectively. Here $\mathcal{C}(\cdot)$ denotes the closure operation that converts each compositional vector from $\mathbb{R}_+^D$ into its representation in $\mathcal{S}^D$. Using the opposite element of $\mathbf{y}$, $\mathbf{y}^{-1} = \mathcal{C}\left(y_1^{-1}, y_2^{-1}, \ldots, y_D^{-1}\right)^\top$, the inverse perturbation $\ominus$ can be defined as

$$\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus \mathbf{y}^{-1}$$

Additionally, the Aitchison inner product is defined for two compositions $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ as

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}$$

which induces the Euclidean vector space structure of the simplex $\mathcal{S}^D$. The inner product can be used to construct a norm and a distance in the simplex

$$\|\mathbf{x}\|_a^2 = \langle \mathbf{x}, \mathbf{x} \rangle_a, \quad d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a$$

The distance is known as the Aitchison distance and it holds important properties associated with compositional data, like invariance under perturbation, invariance under permutation of parts and subcompositional coherence (Pawlowsky-Glahn and Buccianti, 2011).

Compositions in $\mathcal{S}^D$ can be expressed as a perturbation-linear combination of compositional vectors, forming a basis or a generating system of $\mathcal{S}^D$ (with respect to Aitchison geometry). The corresponding coordinates of compositions (real vectors) thus result from transformations of $\mathcal{S}^D$ onto $\mathbb{R}^{D-1}$ or a hyperplane of $\mathbb{R}^D$. Because the coordinates are formed by logarithms of ratios (log-ratios), we refer to log-ratio transformations. The preferable representation of compositions is formed by their coordinates with respect to an orthonormal basis, leading to a one-to-one isometric mapping of Aitchison geometry on the simplex $\mathcal{S}^D$ to Euclidean geometry in real space $\mathbb{R}^{D-1}$. A brief review of frequently used log-ratio transformations is provided in the next section.

LOG-RATIO TRANSFORMATIONS OF COMPOSITIONS AND THEIR INTERPRETATION

Consider a composition $\mathbf{x} = (x_1, \ldots, x_D)^\top \in \mathcal{S}^D$. The alr transformation is a mapping from the simplex $\mathcal{S}^D$ to the real space $\mathbb{R}^{D-1}$, and it depends on the choice of denominator in the log-ratios forming the coordinates. Accordingly, alr transformations are defined as

$$\mathbf{y}^{(k)} = \mathrm{alr}_k(\mathbf{x}) = \left(\ln \frac{x_1}{x_k}, \ldots, \ln \frac{x_{k-1}}{x_k}, \ln \frac{x_{k+1}}{x_k}, \ldots, \ln \frac{x_D}{x_k} \right)^\top, \quad k = 1, \ldots, D \tag{1}$$

Although the alr transformations seem to be easily interpretable, they are not isometric, because their corresponding bases on the simplex are not orthonormal with respect to Aitchison geometry (Egozcue and Pawlowsky-Glahn, 2006).

The centered log-ratio (clr) transformation of $\mathbf{x} \in \mathcal{S}^D$ is defined as

$$\mathbf{z} = (z_1, \ldots, z_D)^\top = \mathrm{clr}(\mathbf{x}) = \left(\ln \frac{x_1}{g(\mathbf{x})}, \ldots, \ln \frac{x_D}{g(\mathbf{x})} \right)^\top \tag{2}$$

where $g(\mathbf{x})$ is the geometric mean of the parts of $\mathbf{x}$. This transformation is isometric and maps $\mathcal{S}^D$ into the subspace $V = \{\mathbf{z} \in \mathbb{R}^D : z_1 + \ldots + z_D = 0\}$ of $\mathbb{R}^D$. Thus the transformed composition lies on a hyperplane through the origin of $\mathbb{R}^D$, which is orthogonal to the vector of units $\mathbf{1}_D$. The clr transformation is closely connected to the ilr transformation. Assuming that the inverse clr transformation is an isometry of V onto $\mathcal{S}^D$, then an orthonormal basis in $\mathcal{S}^D$ can be derived from an orthonormal basis in V.

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_{D-1}\}$ be an arbitrary orthonormal basis of the space $V \subset \mathbb{R}^D$, then the vectors $\mathbf{e}_i = \mathrm{clr}^{-1}(\mathbf{v}_i)$, $i = 1, \ldots, D-1$, represent an orthonormal basis in the simplex $\mathcal{S}^D$. According to this apparent finding, we can define the ilr transformations as one-to-one mappings, assigning for a composition $\mathbf{x} \in \mathcal{S}^D$ coordinates with respect to a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_{D-1}\}$ on the simplex, i.e.

$$\mathbf{u} = \mathrm{ilr}(\mathbf{x}) = (\langle \mathbf{x}, \mathbf{e}_1 \rangle_a, \ldots, \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_a)^\top \tag{3}$$

The ilr transformations represent an isometric isomorphism of vector spaces. Thus, for $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ and $\alpha, \beta \in \mathbb{R}$,

$$\mathrm{ilr}(\alpha \odot \mathbf{x} \oplus \beta \odot \mathbf{y}) = \alpha \cdot \mathrm{ilr}(\mathbf{x}) + \beta \cdot \mathrm{ilr}(\mathbf{y})$$

and also

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \langle \mathrm{ilr}(\mathbf{x}), \mathrm{ilr}(\mathbf{y}) \rangle, \quad d_a(\mathbf{x}, \mathbf{y}) = d(\mathrm{ilr}(\mathbf{x}), \mathrm{ilr}(\mathbf{y})), \quad \|\mathbf{x}\|_a = \|\mathrm{ilr}(\mathbf{x})\| = \|\mathbf{u}\|$$

The ilr coordinates can also be expressed as linear combinations of logarithms of parts whose coefficients add to zero. Considering the $D \times (D-1)$ matrix $\mathbf{V}$ with columns $\mathbf{v}_i = \mathrm{clr}(\mathbf{e}_i)$, the vector of ilr coordinates associated with the matrix $\mathbf{V}$ of a composition $\mathbf{x} \in \mathcal{S}^D$ with respect to $\mathbf{e}_i$, $i = 1, \ldots, D-1$, is

$$\mathbf{u}_\mathbf{V} = \mathrm{ilr}_\mathbf{V}(\mathbf{x}) = \mathbf{V}^\top \mathrm{clr}(\mathbf{x}) = \mathbf{V}^\top \log(\mathbf{x}) \tag{4}$$

where the matrix $\mathbf{V}$ is called the contrast matrix associated with the orthonormal basis $\mathbf{e}_i$, $i = 1, \ldots, D-1$ (Egozcue *et al.*, 2005).

Due to the relative character of compositional data and the dimension of the simplex (one less than the number of parts in a composition), a problem of interpretation of the orthogonal coordinates (also called balances) arises in the sense of their relation to the original compositional parts. This problem was solved by introducing the sequential binary partition procedure (Egozcue and Pawlowsky-Glahn, 2005), which involves splitting parts of a composition into separated groups so that balances representing the groups and the relations between the groups are constructed. A special choice of balances leads to coordinates

$$\mathrm{ilr}(\mathbf{x}) = (z_1, \ldots, z_{D-1})^\top, \quad z_j = \sqrt{\frac{D-j}{D-j+1}} \ln \frac{x_j}{\sqrt[D-j]{\prod_{l=j+1}^{D} x_l}}, \quad j = 1, \ldots, D-1 \tag{5}$$

Here, all the relative information (ratios) of part x_1 to the parts $x_2, \ldots, x_D$ is contained in the balance z_1 (Fišerová and Hron, 2011; Filzmoser *et al.*, 2012). Parts of the remaining subcomposition are represented by $z_2, \ldots, z_{D-1}$, nevertheless, they cannot be interpreted in the same way as z_1. This can be simply achieved by perturbing parts of the

original composition in equation (5) and considering the particular role of z_1. Finally, the inverse ilr transformation $\mathbf{x} = \text{ilr}^{-1}(\mathbf{z})$, where

$$x_1 = \exp\left(\sqrt{\frac{D-1}{D}}z_1\right),$$
$$x_i = \exp\left(-\sum_{j=1}^{i-1}\frac{1}{\sqrt{(D-j+1)(D-j)}}z_j + \sqrt{\frac{D-i}{D-i+1}}z_i\right), \qquad i = 2,\ldots,D-1,$$
$$x_D = \exp\left(-\sum_{j=1}^{D-1}\frac{1}{\sqrt{(D-j+1)(D-j)}}z_j\right)$$

is used to express the coordinates back on the simplex.

Coordinate representations given by different log-ratio transformations have linear relationships, because vectors $\text{alr}_k(\mathbf{x})$, $\text{clr}(\mathbf{x})$ and $\text{ilr}_\mathbf{V}(\mathbf{x})$ represent coordinates of the same composition $\mathbf{x}$ with respect to different bases of the Euclidean vector space $(\mathcal{S}^D, \oplus, \odot)$. In particular, consider two different orthonormal bases of V, $\{\mathbf{v}_1,\ldots,\mathbf{v}_{D-1}\}$ and $\{\mathbf{v}_1^*,\ldots,\mathbf{v}_{D-1}^*\}$, and the corresponding matrices $\mathbf{V}$ and $\mathbf{V}^*$:

$$\mathbf{V} = [\mathbf{v}_1 : \ldots : \mathbf{v}_{D-1}], \qquad \mathbf{V}^* = \left[\mathbf{v}_1^* : \ldots : \mathbf{v}_{D-1}^*\right] \tag{6}$$

A linear relationship between two ilr transformations of a composition $\mathbf{x} \in \mathcal{S}^D$ with respect to the different bases can then be defined as

$$\text{ilr}_\mathbf{V}(\mathbf{x}) = \mathbf{V}^\top\mathbf{V}^*\text{ilr}_{\mathbf{V}^*}(\mathbf{x}) \tag{7}$$

Other relations between log-ratio transformations can be found in Egozcue *et al.* (2005) and Barceló-Vidal *et al.* (2011).

Finally, let us introduce a (perturbation) matrix product in the simplex, defined for $\mathbf{A} \in \mathbb{R}_{D\times D}$ and $\mathbf{x} \in \mathcal{S}^D$ as

$$\mathbf{A} \boxdot \mathbf{x} = \mathcal{C}\left(\prod_{j=1}^{D} x_j^{a_{1j}}, \ldots, \prod_{j=1}^{D} x_j^{a_{Dj}}\right)^\top \tag{8}$$

This operation forms a linear transformation with respect to Aitchison geometry, but only if the rows of $\mathbf{A}$ add up to zero, i.e. $\mathbf{A}\mathbf{1}_D = \mathbf{0}_D$. Otherwise, the matrix product on the simplex is not scale invariant, i.e. $\mathbf{A} \boxdot \mathbf{x} \neq \mathbf{A} \boxdot (k\mathbf{x})$ for $k > 0$. Assuming the same restriction for the columns of $\mathbf{A}$, $\mathbf{A}^\top\mathbf{1}_D = \mathbf{0}_D$, then the function $\mathbf{x} \to \mathbf{A} \boxdot \mathbf{x}$ represents an endomorphism on the simplex $\mathcal{S}^D$. The matrix associated with the identity endomorphism is the so-called centering matrix $\mathbf{G}_D = \mathbf{I}_D - D^{-1}\mathbf{1}_D\mathbf{1}_D^\top$.

Let $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ and $\mathbf{y} = \mathbf{A} \boxdot \mathbf{x}$ be an endomorphism on $\mathcal{S}^D$. It is easy to see that this endomorphism can be expressed as

$$\text{clr}(\mathbf{y}) = \mathbf{A} \cdot \text{clr}(\mathbf{x})$$

in the space of clr coordinates (hyperplane of $\mathbb{R}^D$). By using the relationship between clr and ilr transformations, this results in

$$\text{ilr}_\mathbf{V}(\mathbf{y}) = \mathbf{A}_\mathbf{V} \cdot \text{ilr}_\mathbf{V}(\mathbf{x})$$

where the matrix $\mathbf{A}_\mathbf{V}$ is obtained from $\mathbf{A}$ as $\mathbf{A}_\mathbf{V} = \mathbf{V}^\top\mathbf{A}\mathbf{V}$. It can be shown that $\mathbf{A}$ is not the only matrix that corresponds to $\mathbf{A}_\mathbf{V}$ in this transformation. In fact, $\mathbf{A}_\mathbf{V}$ can also be expressed as $\mathbf{A}_\mathbf{V} = \mathbf{V}^\top\mathbf{A}_0\mathbf{V}$, where $\mathbf{A}_0 = \mathbf{V}\mathbf{A}_\mathbf{V}\mathbf{V}^\top = \mathbf{V}\mathbf{V}^\top\mathbf{A}\mathbf{V}\mathbf{V}^\top = \mathbf{G}_D\mathbf{A}\mathbf{G}_D$ (see Pawlowsky-Glahn and Buccianti, 2011). Accordingly, $\mathbf{A}$ and $\mathbf{A}_0$ represent the same linear transformation on the simplex $\mathcal{S}^D$, i.e. $\mathbf{A} \boxdot \mathbf{x} = \mathbf{A}_0 \boxdot \mathbf{x}$.

VAR MODEL FOR COMPOSITIONAL TIME SERIES

The VAR model

Let $\mathbf{x}_t = (x_{1t},\ldots,x_{Dt})^\top$ be a compositional vector measured at time t, $t = 1,\ldots,n$. Then $\mathbf{z}_t = (z_{1t},\ldots,z_{D-1,t})^\top$ represents coordinates of $\mathbf{x}_t$ obtained by using an ilr transformation (determined by a contrast matrix $\mathbf{V}$).

Here we consider a VAR model in reduced form with p lags, denoted by VAR(p), which is defined as

$$\mathbf{z}_t = \mathbf{c}_{\mathbf{V}} + \mathbf{A}_{\mathbf{V}}^{(1)}\mathbf{z}_{t-1} + \mathbf{A}_{\mathbf{V}}^{(2)}\mathbf{z}_{t-2} + \ldots + \mathbf{A}_{\mathbf{V}}^{(p)}\mathbf{z}_{t-p} + \boldsymbol{\epsilon}_t \tag{9}$$

where $\mathbf{c}_{\mathbf{V}}$ is a real vector, $\mathbf{A}_{\mathbf{V}}^{(i)}$ $(i = 1, \ldots, p)$ are parameter matrices, and $\boldsymbol{\epsilon}_t$ is the error component (see, for example, Lütkepohl, 2005). The error process is considered to be a zero mean white noise process with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. This means that the transformed observation $\mathbf{z}_t$ is modeled based on the p earlier observations $\mathbf{z}_{t-1}, \ldots, \mathbf{z}_{t-p}$. The VAR($p$) model can be equivalently expressed directly on the simplex as

$$\mathbf{x}_t = \mathbf{b} \oplus \left(\mathbf{A}^{(1)} \boxdot \mathbf{x}_{t-1}\right) \oplus \left(\mathbf{A}^{(2)} \boxdot \mathbf{x}_{t-2}\right) \oplus \ldots \oplus \left(\mathbf{A}^{(p)} \boxdot \mathbf{x}_{t-p}\right) \oplus \mathbf{w}_t \tag{10}$$

where $\mathbf{b}$ represents the compositional counterpart to $\mathbf{c}_{\mathbf{V}}$ and $\{\mathbf{w}_t\}$ is the white noise process on the simplex (see Barceló-Vidal *et al.*, 2011).

Let us consider two different ilr-transformed coordinates $\mathbf{z}_t$ and $\mathbf{z}_t^*$ $(t = 1, \ldots, n)$ for given compositional time series $\{\mathbf{x}_t : t = 1, \ldots, n\}$. Let $\mathbf{z}_t = \mathrm{ilr}_{\mathbf{V}}(\mathbf{x}_t)$ represent ilr coordinates of the composition $\mathbf{x}_t \in \mathcal{S}^D$ associated with the matrix $\mathbf{V}$, and $\mathbf{z}_t^* = \mathrm{ilr}_{\mathbf{V}^*}(\mathbf{x}_t)$ represent ilr coordinates associated with the matrix $\mathbf{V}^*$. Using the following relations:

$$\mathbf{z}_t^* = \mathbf{V}^{*\top}\mathbf{V}\mathbf{z}_t, \quad \mathbf{A}_{\mathbf{V}^*}^{(i)} = \mathbf{V}^{*\top}\mathbf{V}\mathbf{A}_{\mathbf{V}}^{(i)}\mathbf{V}^{\top}\mathbf{V}^*, \quad i = 1, \ldots, p$$

it can be shown that a VAR model for compositional time series does not depend on the concrete choice of the ilr transformation. In this case we say that two VAR(p) models, resulting from two different ilr transformations

$$\mathbf{z}_t = \mathbf{c}_{\mathbf{V}} + \mathbf{A}_{\mathbf{V}}^{(1)}\mathbf{z}_{t-1} + \mathbf{A}_{\mathbf{V}}^{(2)}\mathbf{z}_{t-2} + \ldots \mathbf{A}_{\mathbf{V}}^{(p)}\mathbf{z}_{t-p} \tag{11}$$

$$\mathbf{z}_t^* = \mathbf{c}_{\mathbf{V}^*} + \mathbf{A}_{\mathbf{V}^*}^{(1)}\mathbf{z}_{t-1}^* + \mathbf{A}_{\mathbf{V}^*}^{(2)}\mathbf{z}_{t-2}^* + \ldots \mathbf{A}_{\mathbf{V}^*}^{(p)}\mathbf{z}_{t-p}^* \tag{12}$$

are compositionally equivalent. This means that the final model on the simplex (10), obtained from using the inverse ilr transformation, is invariant to the choice of ilr transformation, and the same predictions are thus obtained (Barceló-Vidal *et al.*, 2011) (the equivalent properties also hold for alr and clr transformations). Moreover, within the log-ratio methodology, the obtained predictions can always be rescaled to a prescribed constant sum constraint without loss of information.

While for prediction purposes any of the above mentioned log-ratio transformations can be applied due to the compositional equivalence of VAR models, the role of an appropriate coordinate representation becomes crucial if also statistical inference (such as hypothesis testing) is considered. In the following sections we show how ilr coordinates (5) can be used for this purpose.

Model specification

The order of a VAR model, i.e. the number of lags p of VAR(p), is unknown in practice, but it can be chosen by using selection criteria. The general approach is to fit VAR(p) models for $p = 0, \ldots, p_{\max}$ and then choose that number of lags which minimizes the corresponding function of the given selection criterion.

In this paper, the Akaike information criterion (AIC), Hannan–Quin criterion (HQ), Schwarz criterion (SC) and final prediction error (FPE) are computed to choose the value p. They are defined as follows:

$$\mathrm{AIC}(p) = \ln\left|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}(p)\right| + \frac{2}{n}pK^2 \tag{13}$$

$$\mathrm{HQ}(p) = \ln\left|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}(p)\right| + \frac{2\ln\ln n}{n}pK^2 \tag{14}$$

$$\mathrm{SC}(p) = \ln\left|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}(p)\right| + \frac{\ln n}{n}pK^2 \tag{15}$$

$$\mathrm{FPE}(p) = \left[\frac{n + Kp + 1}{n - Kp - 1}\right]^K \left|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}(p)\right| \tag{16}$$

where $K = D - 1$ is the dimension of $\mathbf{z}_t$, n is the length of the time series, and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}(p)$ is the maximum likelihood estimator of the residual covariance matrix (see, for example, Lütkepohl, 2005). All the above criteria for model specification are invariant to the choice of ilr transformation, since the value of the determinant of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}(p)$ remains unchanged.

The above criteria have different properties: The AIC tends to asymptotically overestimate the real order with a positive probability. Conversely, HQ and SC yield consistent estimates of the order p, and under general conditions the estimated order converges in probability, if the true VAR order p is less than or equal to $p_{\max}$ (see, for example, Lütkepohl, 2005). In many cases, the choice of order depends on the objective of the analysis. For instance, if forecasting is the main aim, then the correct order of the VAR model is not needed. In this case it is reasonable to find a suitable model for prediction by choosing such an order that minimizes a measure of forecast precision. Note that the order of a VAR model can also be specified by sequential testing procedures, namely by testing zero restrictions on parameter matrices (see Lütkepohl, 2005).

Estimation of VAR models

The stationary VAR(p) model (9) can be written in the form of a matrix equation:

$$\mathbf{Y} = \mathbf{ZB} + \mathbf{E} \tag{17}$$

where $\mathbf{Y} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^\top$, the tth row of the $n \times [(D-1)p+1]$ matrix $\mathbf{Z}$ equals $\mathbf{Z}_t = \left(1, \mathbf{z}_{t-1}^\top, \ldots, \mathbf{z}_{t-p}^\top\right)^\top$ and $\mathbf{B} = [\mathbf{c}, \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(p)}]^\top$ contains the parameters. Assuming a sample of size n, $\mathbf{z}_1, \ldots, \mathbf{z}_n$, and p pre-sample values, $\mathbf{z}_{-p+1}, \ldots, \mathbf{z}_0$, the parameters $\mathbf{B}$ can be estimated separately for each equation (formed by the columns of $\mathbf{Y}$) by the ordinary least squares (OLS) method. If the regressors in all equations are the same (no restriction for parameters are imposed), the estimator is identical to the generalized least squares (GLS) estimator. This estimator is also identical to the maximum likelihood (ML) estimator (conditional on the size of a given initial pre-sample), if the VAR(p) process $\mathbf{z}_t$ is normally distributed and $\boldsymbol{\epsilon}_t$ (rows of the $n \times p$ error matrix $\mathbf{E}$) represent a white noise process, thus $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}), t = 1, \ldots, n$.

Such an estimator has the convenient asymptotic properties of standard estimators, and is consistent and asymptotically efficient. If the VAR(p) process is not stationary, or if restrictions are imposed on the parameters, the GLS estimator may be more beneficial (Lütkepohl, 2005).

Hypotheses testing

A frequent task in the context of multivariate time series analysis is testing for causality. For that reason, Granger causality was introduced (see, for example, Lütkepohl, 2005), which represents a statistical concept that is based on prediction. In other words, we are interested in testing whether one variable could help to improve predictions of the remaining observed variables. Considering our special choice (5) of ilr transformation, our aim is to test whether variability of z_1, which carries all relative information of the chosen compositional part x_1 to all parts $x_2, \ldots, x_D$ (up to a permutation of these parts), has an effect on the coordinates $z_2, \ldots, z_{D-1}$, representing the remaining subcomposition.

Consider our VAR(p) model (9) with no restrictions on the parameters. Using the operation of vectorization for the estimated parameters $\hat{\mathbf{B}}$, we obtain

$$\hat{\boldsymbol{\beta}} = \operatorname{vec}\left(\hat{\mathbf{B}}\right) = \operatorname{vec}\begin{pmatrix} \hat{c}^\top \\ \left(\hat{\mathbf{A}}^{(1)}\right)^\top \\ \vdots \\ \left(\hat{\mathbf{A}}^{(p)}\right)^\top \end{pmatrix}$$

Under general assumptions for VAR models, $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normally distributed with asymptotic covariance matrix

$$\widehat{\operatorname{avar}}\left(\hat{\boldsymbol{\beta}}\right) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} \otimes (\mathbf{Z}^\top \mathbf{Z})^{-1} \tag{18}$$

with

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \frac{1}{n-(D-1)p-1} \sum_{t=1}^{n} \left(\mathbf{z}_t - \mathbf{Z}\hat{\mathbf{B}}\right)\left(\mathbf{z}_t - \mathbf{Z}\hat{\mathbf{B}}\right)^\top \tag{19}$$

Considering the mentioned properties, we may test linear hypotheses for parameters of the general form $\mathbf{R}\hat{\boldsymbol{\beta}} = \mathbf{r}$ by using the Wald statistics:

$$\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)^\top \left\{\mathbf{R}\left[\widehat{\operatorname{avar}}\left(\hat{\boldsymbol{\beta}}\right)\right]\mathbf{R}^\top\right\}^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \sim F(q, n-(D-1)p-1) \tag{20}$$

where q is the number of parameters tested and $\mathbf{R}$ is a $q \times (D-1)^2 p + D - 1$ restricting matrix. The structure of the restricting matrix $\mathbf{R}$ depends on the particular tested null hypothesis. The elements of $\mathbf{R}$ are 0, when the corresponding parameter is not tested, and they are 1, when the significance of a parameter is tested (see, for example, Lütkepohl, 2005).

Without loss of generality, the main interest consists in testing Granger causality from z_1 to the remaining coordinates $z_2, \ldots, z_{D-1}$. However, the hypothesis can also be tested reversely, i.e. if the coordinates $z_2, \ldots, z_{D-1}$ have a significant effect on z_1. The particular null hypothesis about non-Granger causality generally depends on the concrete objective of the analysis and the structure of data.

$\mathcal{T}$ SPACES IN THE TIME SERIES CONTEXT

In this section we introduce an extension of the above considerations to the case, when both relative and absolute information are of interest. The concept of compositional data analysis as introduced above was based on the assumption that compositional data carry exclusively relative information, which is contained in the ratios between their parts. Nevertheless, in practice both relative and absolute information is often necessary to take into account in order to provide a reasonable output of a data analysis. The latter information is expressed by modeling absolute values of the sum of the original compositional parts (if this is not trivial, like for proportional representations of compositions summing to one). Although a careful treatment of relative information is required, provided by the log-ratio approach, the absolute values predictions may represent the final objective of multivariate time series analysis in the case of forecasting.

Formally, consider a vector with D strictly positive components, $\mathbf{x} \in \mathbb{R}_+^D$. For the log-ratio approach, the data can be expressed as closed observations with a constant sum κ, frequently with $\kappa = 1$, without loss of information (Aitchison, 1986). Thus all information about the total amount is ignored. This information can be incorporated by defining an extended vector space $\mathcal{T} = \mathbb{R}_+ \times \mathcal{S}^D$ that allows us to model the relative structure of the values and the absolute total sum as an additional variable jointly in one model (see Pawlowsky-Glahn *et al.*, 2013).

The vector $\tilde{\mathbf{x}} = [t(\mathbf{x}), \mathcal{C}(\mathbf{x})] = [t_x, \tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_D]$ is the element of $\mathcal{T} = \mathbb{R}_+ \times \mathcal{S}^D$, where $t(\mathbf{x})$ stands for the total sum, i.e. $t(\mathbf{x}) = \sum_{i=1}^{D} x_i$. In the time series context, often the logarithm of the total $t(\mathbf{x})$ will be taken. The compositions $\mathbf{x} = (x_1, \ldots, x_D)^\top$ are modeled by employing a log-ratio transformation: in our case, the suggested ilr transformation. Subsequently, statistical analysis can be performed (including Granger causality with the total variable). In the case of time series analysis, also back-transformation to the original values is often required for prediction purposes. Hence the log-total needs to be back-transformed by the exponential transformation, and the forecast compositions in proportions are multiplied by these values. An example in the next section will show the usefulness of this procedure.

ILLUSTRATIVE EXAMPLES

First example

We consider a dataset that contains compositional time series of monthly measured paper production shares in Europe from May 2004 to December 2009. The dataset was kindly provided by Statistics Austria. The time series is represented as proportions of the overall paper production per month in Austria (x_1), the eurozone countries without Austria (x_2), and EU countries not in the eurozone plus the remaining countries in Europe (x_3). Therefore, for each month, the values of the three categories sum to one. All analyses in this paper are carried out using the R enviroment, mainly packages *vars* (Pfaff, 2012) and *robCompositions* (Templ *et al.*, 2013).

The aim of this section is to compare the standard approach, when the VAR model is applied directly to the original time series, and the compositional approach based on using the ilr transformation (5) and applying the VAR model to the coordinates. Figure 1 shows plots of the raw (left) and ilr-transformed (right) data. The ilr variable z_1

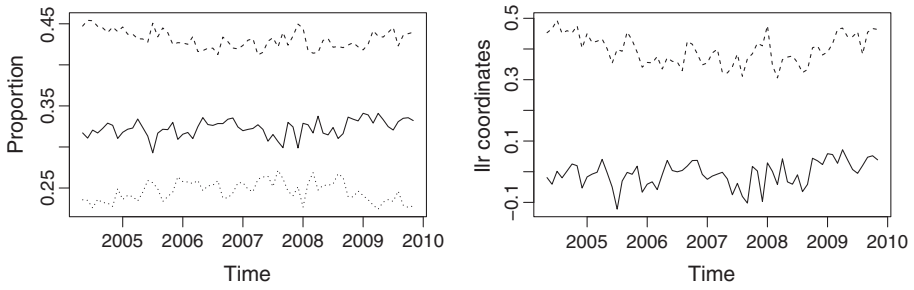

Figure 1. Left: raw untransformed paper production time series (solid line: Austria (x_1); dashed line: eurozone countries without Austria (x_2); dotted line: EU states not in the eurozone plus other countries in Europe (x_3)). Right: ilr-transformed time series (solid line z_1; dashed line z_2)

(solid line) represents the relative information of the Austrian paper production to the other two parts, while z_2 describes the relative information between x_2 and x_3.

The process of modeling time series consists of model specification, parameter estimation, and diagnostic checking of an assumed VAR(p) model. In this case, a seasonal VAR model is considered because of the monthly observed data. The number of lags is chosen based on using the model selection criteria mentioned earlier. In our example, just the proportional data are available and thus a singularity problem occurs when analyzing the raw data with the standard approach due to the unit sum constraint. This problem is usually 'circumvented' by omitting one variable, and applying the VAR model only to the remaining two parts; the values for the omitted part is supposed to be calculated afterwards as the complement to one. Nevertheless, the information contained in the last variable is dropped as the unit sum constraint is just a proper representation of compositional data, not their inherent property. Thus, to demonstrate the results of the standard approach, three models were built for the case of subsequently omitting variable x_1, x_2, and x_3, respectively, from the original time series. For the compositional approach, all three time series can be used. Using the mentioned model selection criteria, the resulting numbers of lags p for the original and the ilr-transformed time series are shown in Table I. The same numbers of lags are obtained here for all three cases of omitting one variable, but with no guarantee that the results will not differ in general. A VAR(1) model is selected for the standard and the compositional approach in order to allow a subsequent comparison of the results.

The undesirable effect of omitting one compositional part in our VAR(1) models for the original time series can be seen, for example, in the case of forecasting future values. For this purpose we use data from May 2004 to November 2009 for forecasting the (relative) values for December 2009. We apply all three possible VAR models for the original data as well as the ilr approach and compare the corresponding forecasts also with the true December 2009 composition. The results differ for the standard approach in each case of omitting one variable from the model, whereas for the compositional approach the predictions are always the same due to the invariance of using log-ratio transformations. The elimination of a variable from the standard analysis and its subsequent calculation as the complement to one might even cause the obtained predictions to be negative or equal to zero, or they could be greater than one. This cannot occur when using log-ratio transformations.

The relationships between the variables (original parts or the ilr coordinates, respectively) can be further analyzed by Granger causality analysis. In our example, the hypothesis tested using the ilr approach is $\mathrm{H_0} : a_{21}^{(1)} = 0$, which represents Granger non-causality between the coordinates defined by equation (5). With this transformation, the aim is to test the influence of the relative amount of paper production in Austria on other European countries, and also conversely. In other words, we are interested in testing if coordinate z_1 has no effect on z_2, i.e. whether relative information on x_1 (production of paper in Austria) does not influence the ratio to x_2 and x_3. In our case, the Granger non-causality of z_1 to z_2 is not rejected at the 0.05 significance level. This means that past values of z_1 probably do not contain information that is useful for predicting z_2. The relative information contained in the paper production of Austria is then not useful in forecasting the ratio to the other countries. Conversely, we can test whether $\mathrm{H_0} : a_{12}^{(1)} = 0$ (the ratio between x_2 and x_3 has no effect on the variable x_1). In this case, the null hypothesis is also not rejected at the 0.05 significance level.

Similar tests can also be carried out for the standard approach. However, the crucial problem consists in omitting one variable as the initial step of the analysis. Granger causality can be investigated only between two variables. The null hypothesis cannot be rejected in each case at the 0.05 significance level. The resulting p-values for reasonable combinations are summarized in Table II.

Table I. Resulting numbers of lags, using different model selection criteria, for the untransformed and the ilr-transformed data

Approach	AIC(p)	HQ(p)	SC(p)	FPE(p)
Untransformed	4	1	1	4
ilr-transformed	4	1	1	4

Table II. Testing Granger causality

Null hypothesis	p-value
z_1 does not Granger cause z_2	0.741
z_2 does not Granger cause z_1	0.203
x_1 does not Granger cause x_2	0.426
x_1 does not Granger cause x_3	0.291
x_2 does not Granger cause x_1	0.387
x_3 does not Granger cause x_1	0.387

Apparently, tests for Granger causality for the standard and compositional case are not comparable, because they produce results with completely different interpretations with respect to the original time series. Using the standard approach, one effect is always excluded from the complex analysis. According to the obtained results, one cannot state any information about the excluded variable and its influence on the remaining observations. Conversely, the log-ratio approach offers an overall analysis of all components. The testing is performed in the space of ilr coordinates, which enhances interpretability of the results. Thus, although the null hypothesis was not rejected in all mentioned cases, the ilr results are more reasonable due to modeling effects of all variables.

Second example

The second example consists of modeling a four-part compositional time series dataset. The data represent gross bonuses of metal production in Austria (in thousands of euros) considering white-collar workers (x_1), blue-collar workers (x_2), commercial apprentices (x_3) and industrial apprentices (x_4). The data are available monthly from January 2004 until November 2010 (see Figure 2). Our goal is to model the relative structure (relative contributions of the parts on total metal production) of the compositional time series. However, one can also be interested in predictions of the original absolute values (in thousands of euros), based on the multivariate (relative) structure of the compositional data and the total of the compositional parts. For interpretation purposes, we can define the total metal production as the sum of all compositional parts in (original) absolute values as

$$X_t = x_1 + x_2 + x_3 + x_4$$

and then investigate this total as an additional variable in coordinate representation of the compositional time series. Consequently, the predicted total is used to compute predictions in absolute numbers. The compositional approach using the methodology of $\mathcal{T}$ spaces (see previous main section) for modeling multivariate time series of compositions with a total is compared to the standard approach for the original data.

Initially, the VAR model is applied to the variables x_1, x_2, x_3, x_4 in the standard way. The metal production data from Austria are represented as monthly measured data; thus the seasonal effect is considered by including dummy variables in the model. The suggested numbers of lags according to the mentioned model selection criteria are summarized in Table III, and a lag of one is selected. In contrast to the first example, the time series observations were collected in absolute values (without any constant sum constraint), thus omitting a variable for the standard approach is not necessary.

Considering the compositional structure of the data, the theory of $\mathcal{T}$ spaces can be involved to perform a reasonable analysis taking the Aitchison geometry of compositional data into account. The isometric log-ratio transformation (5) is applied to the observed variables, and the total sum is taken as an additional variable in the model. The ilr-transformed data values and the log-transformed total sum X_t are displayed in Figure 3. Subsequently, the time series are investigated also by taking into account the monthly seasonality to choose the corresponding number of lags p (summarized in Table III). Finally, the VAR model with $p = 1$ lag is selected for both the standard and compositional analyses.

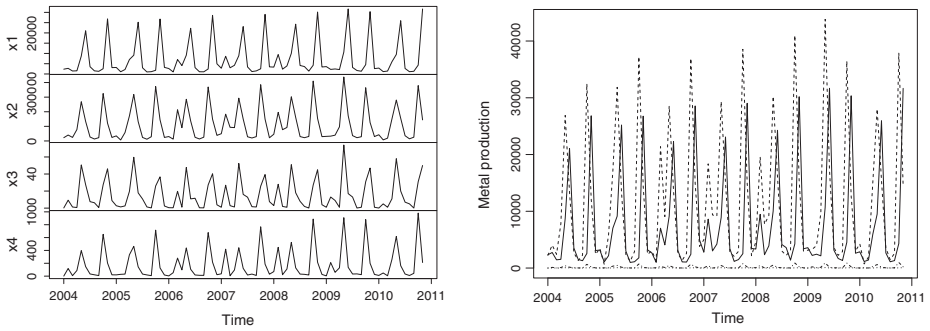

Figure 2. Time series of metal production in absolute numbers plotted separately (left) and jointly (right); x_1 is represented by a solid, x_2 by a dashed, x_3 by a dotted and x_4 by a dashed-dotted line, respectively

Table III. Resulting numbers of lags, using different model selection criteria, for the standard approach and the compositional approach using $\mathcal{T}$ spaces

Approach	AIC(p)	HQ(p)	SC(p)	FPE(p)
Standard	10	1	1	2
Compositional	10	1	1	1

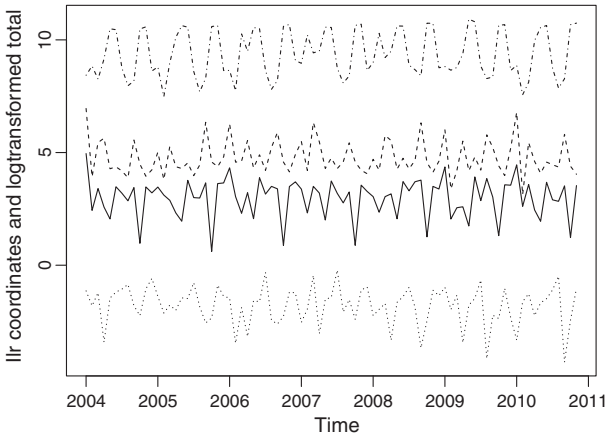

Figure 3. ilr-transformed time series (z_1 solid, z_2 dashed and z_3 dotted line) and log-transformed total X_t of metal production (dashed-dotted line)

As in the previous example, the time series can be tested for Granger causality in order to detect significant effects between variables, and between variables and the total, respectively. In the case of applying the compositional approach, hypothesis testing is carried out in the space of orthonormal coordinates. Consequently, we consider the ilr coordinates z_1, z_2, z_3 and log-transformed total X_t, where z_1 explains all relative information (ratios) concerning the original compositional part x_1 with respect to the remaining parts x_2, x_3 and x_4. The reasonable null hypothesis seems to be whether z_1 does not Granger cause z_2, z_3 and $\log(X_t)$. The resulting p-value is 0.526, which indicates that Granger non-causality cannot be rejected at the 0.05 significance level. In other words, the relative information of x_1, here gross bonuses of white-collar workers in metal production in Austria, compared to the other compositional parts (included in the coordinate z_1) does not have a significant effect for predicting the coordinates z_2, z_3 and $\log(X_t)$.

The previous null hypothesis focused only on all the relative information of part x_1 to the remaining parts. Nevertheless, one might also be interested in testing the relative information concerning x_2 (and x_3, x_4, respectively). This can be achieved by exchanging x_1 with another part of interest in the ilr transformation (5), and again focusing on the coordinate z_1. In that way we obtain significance for z_1, which accounts for all the relative information of x_4 to x_1, x_2, x_3, on the coordinates z_2, z_3 and the log-transformed total X_t. The corresponding p-value of 0.043 indicates that the relative information contained in x_4, representing industrial apprentices, has a significant effect on predicting time series of the other variables x_1, x_2, x_3 and the total. In other words, the development of gross bonuses of industrial apprentices influences the data structure of the other workers employed in the metal industry in Austria. Nevertheless, note that by omitting the total variable from the above test, the influence of z_1 on the (purely) relative structure of the other parts (represented by the variables z_2, z_3) is not significant.

The alternative possibility of investigating causality involves testing whether the additional total variable affects the observed variables x_1, x_2, x_3 and x_4. In the space of coordinates, we thus test whether $\log(X_t)$ does not Granger cause z_1, z_2, z_3. This particular null hypothesis is not rejected at the 0.05 significance level, where the resulting p-value of 0.115 obviously does not depend on a permutation of the compositional parts in equation (5), resulting in different interpretations of the coordinate z_1.

Testing Granger causality with the standard approach is definitely not comparable with the compositional one. Testing by applying compositional techniques is performed in the space of coordinates and the total is considered as an additional variable involved in the investigated model (in its log-transformed form). Nevertheless, the null hypothesis that x_4 does not Granger cause x_1, x_2, x_3 using the original values is rejected at the 0.05 significance level (based on the p-value $6.16 \cdot 10^{-6}$) as well. Similar hypotheses, whether one concrete variable has no influence on the other ones, cannot be rejected in all the remaining cases. We can conclude that past values of x_4—gross bonuses of industrial apprentices—could be very useful in predicting future values of the remaining variables.

Finally, it might be interesting to compare the accuracy of the predictions obtained by the different approaches. In the compositional case, a back-transformation to the simplex is required to attain the final predictions. As a measure of prediction accuracy we consider the root mean squared error of prediction (RMSEP):

$$\text{RMSEP} = \sqrt{\frac{1}{m}\sum_{t=1}^{m} \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|^2}$$

where m is the number of predicted values ahead. In our case, predictions for 24 subsequent months are made (from December 2009 to November 2011) to compare the accuracy of forecasting. The results show that the RMSEP for the

compositional case is 167.02, whereas when using the standard approach the RMSEP is 170.94. According to these results, the compositional approach results in slightly better predictive ability than the conventional one.

CONCLUDING REMARKS

Compositional time series are by definition multivariate, mostly represented by a constant sum constraint, and they carry only relative information. Since their sample space is the simplex rather than the real space with the usual Euclidean geometry, they need to be expressed in appropriate (preferably orthonormal) coordinates with respect to Aitchison geometry before VAR models are employed. We have proposed a specific ilr transformation to represent the compositional parts in orthonormal coordinates, which is preferable to other log-ratio transformations because it allows for a meaningful interpretation of the results in terms of the original compositional parts. Moreover, the particular choice of ilr transformation does not change the resulting predictions of the original compositional values.

Applying VAR models directly to the raw untransformed data may lead to inappropriate models that do not respect the compositional nature of the data. One may obtain artefacts like singularity of the data due to their constant sum constraint, or predictions outside the data range of proportional data (negative or larger than one). Omitting a variable in the case of time series with constant sum can even lead to different models and predictions, depending on which variable is omitted. Also, for compositional time series without a constant sum constraint, analysis of the raw data will not focus on the relative information inherent in the data, and may be driven by large values rather than by small ones, which might be of equal importance in a relative sense.

Compositional time series can also be represented as compositional data with a total. The main idea of modeling these time series is based on the theory of $\mathcal{T}$ spaces, which enables modeling the relative structure of variables with the absolute total sum together in one model. This approach is thus especially useful when both relative and absolute information is of interest for time series analysis. Furthermore, using the $\mathcal{T}$ spaces approach, the total is included in the model as separate information from the relative one (represented by orthonormal coordinates), so that Granger causality also between the coordinates and the total can be investigated. This fact could be of interest in many real-world problems.

ACKNOWLEDGEMENTS

The paper is supported by the Operational Program Education for Competitiveness—European Social Fund (project CZ.1.07/2.3.00/20.0170 of the Ministry of Education, Youth and Sports of the Czech Republic), and grant IGA PrF_2014_028, Mathematical Models of the Internal Grant Agency of Palacký University, Olomouc.

REFERENCES

Aitchison J. 1986. *The Statistical Analysis of Compositional Data*, Monographs on Statistics and Applied Probability. Chapman & Hall: London.

Barceló-Vidal C, Aguilar L, Martín-Fernández JA. 2011. Compositional VARIMA time series. In *Compositional Data Analysis: Theory and Applications*, Pawlowsky-Glahn V, Buccianti A (eds). Wiley: Chichester; 87–103.

Bergmann J. 2008. Compositional time series: an application. In *Proceedings of CoDaWork '08: 3rd Compositional Data Analysis Workshop*. University of Girona: Spain.

Egozcue JJ. 2009. Reply to 'On the Harker variation diagrams ...' by J. A. Cortés. *Mathematical Geosciences* **41**: 829–834.

Egozcue JJ, Pawlowsky-Glahn V. 2005. Groups of parts and their balances in compositional data analysis. *Mathematical Geosciences* **37**(7): 795–828.

Egozcue JJ, Pawlowsky-Glahn V. 2006. Simplicial geometry for compositional data. In *Compositional Data Analysis in the Geosciences: From Theory to Practice*, Buccianti A, Mateu-Figueras G, Pawlowsky-Glahn V (eds). Geological Society: London; 145–160.

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueraz G. 2005. Isometric logratio transformations for compositional data analysis. *Mathematical Geosciences* **35**(3): 279–300.

Fišerová E, Hron K. 2011. On interpretation of orthonormal coordinates for compositional data. *Mathematical Geosciences* **43**(4): 455–468.

Filzmoser P, Hron K, Reimann C. 2012. Interpretation of multivariate outliers for compositional data. *Computers and Geosciences* **39**: 77–85.

Larrosa JM. 2005. Compositional time series: past and present, Working Paper, EconWPA 0510002.

Lütkepohl L. 2005. *New Introduction to Multiple Time Series Analysis*. Springer: Berlin.

Mills T. 2010. Forecasting compositional time series. *Quality and Quantity* **44**: 673–690.

Pawlowsky-Glahn V, Buccianti A (eds). 2011. *Compositional Data Analysis: Theory and Applications*. Wiley: Chichester.

Pawlowsky-Glahn V, Egozcue JJ, Lovell D. The product space $\mathcal{T}$ (tools for compositional data with a total), In *Proceedings of CoDaWork'13; 5th Compositional Data Analysis Workshop*: Vorau, Austria, 2013.

Pfaff B. 2012. *vars: VAR Modelling*. R package version 1.5-0.

Templ M, Hron K, Filzmoser F. 2013. *robCompositions: Robust Estimation for Compositional Data*. R package version 1.6.3.

Authors' biographies:
Petra Kynčlová PhD student of statistics at the Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology. Her research interest is mainly focused on compositional data analysis and methods.

Peter Filzmoser Professor in statistics at the Institute of Statistics and Mathematical Methods in Economics of the Vienna University of Technology. He has focused on computational statistics, robust statistical estimation,and compositional data analysis.

Karel Hron Associate Professor of applied statistics at the Department of Mathematical Analysis and Applications of Mathematics of Palacky University, Olomouc. His research interests include compositional data analysis, multivariate and robust statistics.

Authors' addresses:
Petra Kynčlová, Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria and Department of Mathematical Analysis and Applications of Mathematics, Palacký University, Olomouc, Czech Republic.

Peter Filzmoser, Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria.

Karel Hron, Department of Mathematical Analysis and Applications of Mathematics, Palacký University, Olomouc, Czech Republic and Department of Geoinformatics, Palacký University, Olomouc, Czech Republic.

