Compositional Time Series: An Application

J. Bergman[1]

[1]Lund University, Lund, Sweden; *jakob.bergman@stat.lu.se*

Abstract

The composition of the labour force is an important economic factor for a country. Often the changes in proportions of different groups are of interest.

I this paper we study a monthly compositional time series from the Swedish Labour Force Survey from 1994 to 2005. Three models are studied: the ILR-transformed series, the ILR-transformation of the compositional differenced series of order 1, and the ILR-transformation of the compositional differenced series of order 12. For each of the three models a VAR-model is fitted based on the data 1994-2003. We predict the time series 15 steps ahead and calculate 95 % prediction regions. The predictions of the three models are compared with actual values using MAD and MSE and the prediction regions are compared graphically in a ternary time series plot.

We conclude that the first, and simplest, model possesses the best predictive power of the three models.

Key words: Compositional time series, VAR models, ILR, ternary time series plot

1 Introduction

The composition of the labour force is an important economic factor for a country. It is usually measured in absolute numbers, i.e. how many people that at given time was employed, how many that was unemployed, and how many that was not in the labour force. Often however it is not the absolute number of people that is of interest, but the relative. I this paper we present three simple time series models for the relative labour force data and examine their predictive power.

1.1 Compositions

A composition is a vector of non-negative components summing to a constant, usually 1, or put symbolically, a vector $\boldsymbol{x}$ such that

$$\boldsymbol{x} = (x_1, \dots, x_D)'; \quad x_1 > 0, \dots, x_D > 0; \quad x_1 + \cdots + x_D = 1.$$

Compositions arise in many different areas. The geochemical compositions of different rock specimens, the proportion of expenditures on different commodity groups in household budgets, and the party preferences in a party preference survey are all three different examples of compositions from different scientific areas.

The sample space of a composition $\boldsymbol{x}$ is referred to as the *Simplex*, $\mathcal{S}^D$. It has been known, though not always acknowledged, since the days of Pearson (1897), that normal statistical methods are not applicable to elements of the simplex, i.e. compositions. See Aitchison (1986, ch. 3), or the reprint 2003, for a detailed description of the problems with standard methods on compositions.

The major way, following the ideas of Aitchison, of resolving these problems has been through a logratio transformation from the D-part Simplex, $\mathcal{S}^D$, to the real space $\mathcal{R}^d$ or $\mathcal{R}^D$, where $d = D - 1$. The most popular transformations are the additive logratio transformation (ALR) and the centred logratio transformation (CLR) (Aitchison, 1986), and more recently the isometric logratio transformation (ILR) (Egozcue and others, 2003). The three transformations are related, see for instance Barceló-Vidal and others (2007).

1.2 The Geometry of the Simplex

The Simplex utilizes its own geometry, based on the Simplicial distance (Aitchison, 1986)

$$\Delta_S(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{D} \left(\ln \frac{x_i}{g(\boldsymbol{x})} - \ln \frac{y_i}{g(\boldsymbol{y})} \right)^2}$$

where $g(\boldsymbol{x}) = (x_1 x_2 \cdots x_D)^{1/D}$ is the geometric mean. The inner product of two compositions is thus, following Egozcue and others (2003), defined as

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_S = \sum_{i=1}^{D} \ln \frac{x_i}{g(\boldsymbol{x})} \ln \frac{y_i}{g(\boldsymbol{y})}.$$

The inner product can be used to construct the norm, $\| \cdot \|_S$, of a composition

$$\| \boldsymbol{x} \|_S^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle_S.$$

The Simplex also has the basic operation *perturbation*, $\oplus$, which is analogous to addition in the real space, and *inverse perturbation*, $\ominus$, which is analogous to subtraction (Aitchison, 1986). This means that the distance between two compositions also can be calculated as

$$\Delta_S(\boldsymbol{x}, \boldsymbol{y}) = \| \boldsymbol{x} \ominus \boldsymbol{y} \|_S \, .$$

1.3 The ILR Transformation

A D-part Simplex is spanned by a d-part basis. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$ be any orthonormal basis of $\mathcal{S}^D$. Such a basis can be constructed as

$$\boldsymbol{v}_i = \mathcal{C}\Big(\big(\underbrace{e^{\sqrt{\frac{1}{i(i+1)}}}, \ldots, e^{\sqrt{\frac{1}{i(i+1)}}}}_{i \text{ elements}}, e^{-\sqrt{\frac{i}{(i+1)}}}, 1, \ldots, 1\big)\Big)$$

where $\mathcal{C}(\boldsymbol{x}) = (x_1, \ldots, x_D)/\sum_{i=1}^{D} x_i$ is the closure operation. The ILR transformation of a composition $\boldsymbol{x}$ is then defined as

$$\operatorname{ilr}(\boldsymbol{x}) = \big(\langle \boldsymbol{x}, \boldsymbol{v}_1\rangle_S, \ldots, \langle \boldsymbol{x}, \boldsymbol{v}_d\rangle_S\big).$$

The names derives from the fact that the transformation retains distances:

$$\parallel \boldsymbol{x} \ominus \boldsymbol{y} \parallel_S = \parallel \operatorname{ilr}(\boldsymbol{x}) - \operatorname{ilr}(\boldsymbol{y}) \parallel$$

where $\parallel \cdot \parallel$ is the normal real space norm. (Egozcue and others, 2003)

Like the ALR transformation the ILR is a transformation from from $\mathcal{S}^D$ to $\mathcal{R}^d$, but unlike the ALR and the CLR the resulting components of the ILR transformation are hard to interpret. (The components of the ALR transformation are the logarithm of the relative magnitude of a component compared to a reference component and the components of the CLR transformation are the logarithm of the relative magnitude of a component compared to the geometric mean.) The ILR result in a much more complex vector. For example, if $\boldsymbol{x} = (x_1, x_2, x_3, x_4)'$ then the resulting vectors of the different transformations are the following

$$\begin{aligned}
\operatorname{alr}(\boldsymbol{x}) &= \left(\ln\frac{x_1}{x_4}, \ln\frac{x_2}{x_4}, \ln\frac{x_3}{x_4}\right)' \\
\operatorname{clr}(\boldsymbol{x}) &= \left(\ln\frac{x_1}{g(\boldsymbol{x})}, \ln\frac{x_2}{g(\boldsymbol{x})}, \ln\frac{x_3}{g(\boldsymbol{x})}, \ln\frac{x_4}{g(\boldsymbol{x})}\right)' \\
\operatorname{ilr}(\boldsymbol{x}) &= \left(\frac{1}{\sqrt{2}}\ln\frac{x_1}{x_2}, \frac{1}{\sqrt{6}}\ln\frac{x_1x_2}{x_3^2}, \frac{1}{\sqrt{12}}\ln\frac{x_1x_2x_3}{x_4^3}\right)'
\end{aligned}$$

where $g(\boldsymbol{x})$ is the geometric mean as before.

1.4 Compositional Time Series

A time series of compositions is referred to as a Compositional Time Series (CTS). Compositional time series arise in many different situations, e.g. party preference surveys, labour force surveys or pollution measurements.

Even though there has only been relatively few papers published about CTS, there has been several approaches to CTS. Larrosa (2005) and Aguilar Zuil and others (2007) have reviewed the different approaches to CTS.

The first to discuss and use an ALR approach to CTS seem to be Aitchison (1986) and Brunsdon (1987), which were followed by Smith and Brunsdon (1989) and Brunsdon and Smith (1998). In that approach the CTS is transformed with an ALR, and the transformed series is then analysed with standard models, e.g. VAR or VARMA.

There has also been some ideas on how to model the time series on the Simplex. Apart from Aitchison and Brunsdon, Billheimer and Guttorp (1995) and Billheimer and others (1998) have used autoregressive and conditional autoregressive models. Barceló-Vidal and others (2007) introduced a compositional ARIMA model.

2 Data

Data consists of a monthly series from the Swedish Labour Force Survey (AKU), conducted by Statistics Sweden, from 1976 to 2005 (see Persson and Henkel (2005) for a description of the Swedish Labour Force Survey). The labour force is divided into three groups: Employed, Unemployed and Not in the labour force. The time series and the ILR-transformed series are plotted in figure 1. As can be seen from the plot there was a major decrease in employment in the early 1990's during the Swedish fiscal crisis, constituting a structural change in the series. To avoid having to model the change, we only consider the series after 1994. This shortened series is plotted in Figure 2 together with the ILR-transformed series. In this shortened series there are no evident structural changes.

The time series from January 1994 up to December 2003 is used for modelling. The remaining time series (January 2004–March 2005) is used for model evaluation.

3 Analysis

We investigate three models.

$$\begin{aligned} \mathrm{I}: \boldsymbol{z}_t &= \mathrm{ilr}\,(\boldsymbol{x}_t) \\ \mathrm{II}: \boldsymbol{z}_t &= \mathrm{ilr}\,(\boldsymbol{x}_t \ominus \boldsymbol{x}_{t-1}) \\ \mathrm{III}: \boldsymbol{z}_t &= \mathrm{ilr}\,(\boldsymbol{x}_t \ominus \boldsymbol{x}_{t-12}) \end{aligned}$$

In the models the $\boldsymbol{x}$ is the untransformed time series defined in section 2. Model II and III are the equivalences of the ordinary difference operator of order 1 and 12 respectively. The transformed time series are thereafter modelled using traditional VAR modelling technique.

3.1 VAR models

We assume a vector autoregressive (VAR) model for simplicity. The observation at time t, $\boldsymbol{z}_t$, is assumed to depend on the m earlier observations. It is defined as

$$\boldsymbol{z}_t = \boldsymbol{\nu} + \boldsymbol{A}_1\boldsymbol{z}_{t-1} + \boldsymbol{A}_2\boldsymbol{z}_{t-2} + \cdots + \boldsymbol{A}_m\boldsymbol{z}_{t-m} + \boldsymbol{\epsilon}_t$$

where $\boldsymbol{\nu}$ is a constant and $\boldsymbol{\epsilon}_t$ is the error term. In model I and II we also add seasonal dummy variables to account for the seasonal variation in the series.

3.2 Order Selection Criteria

We use order selection criteria to choose the number of lags, m, to include in the different models. Four criteria are used: the Akaike Information (AIC), the Hannan-Quinn (HQ), the Schwartz (SC) and the Final Prediction Error (FPE) criteria. According to the different criteria, one should choose should the m that minimizes the respective functions

$$\begin{aligned} \mathrm{AIC}(m) &= \ln|\widetilde{\Sigma}_\epsilon(m)| + \frac{2mK^2}{T} \\ \mathrm{HQ}(m) &= \ln|\widetilde{\Sigma}_\epsilon(m)| + \frac{2\ln\ln T}{T}mK^2 \\ \mathrm{SC}(m) &= \ln|\widetilde{\Sigma}_\epsilon(m)| + \frac{\ln T}{T}mK^2 \\ \mathrm{FPE}(m) &= \left[\frac{T+Km+1}{T-Km-1}\right]^K |\widetilde{\Sigma}_\epsilon(m)| \end{aligned}$$

where K is the dimension of $\boldsymbol{z}_t$ (here $K = 2$), T is the length of the time series, and $|\widetilde{\Sigma}_\epsilon(m)|$ is the determinant of the ML-estimate of the residual covariance matrix. The criteria are based on

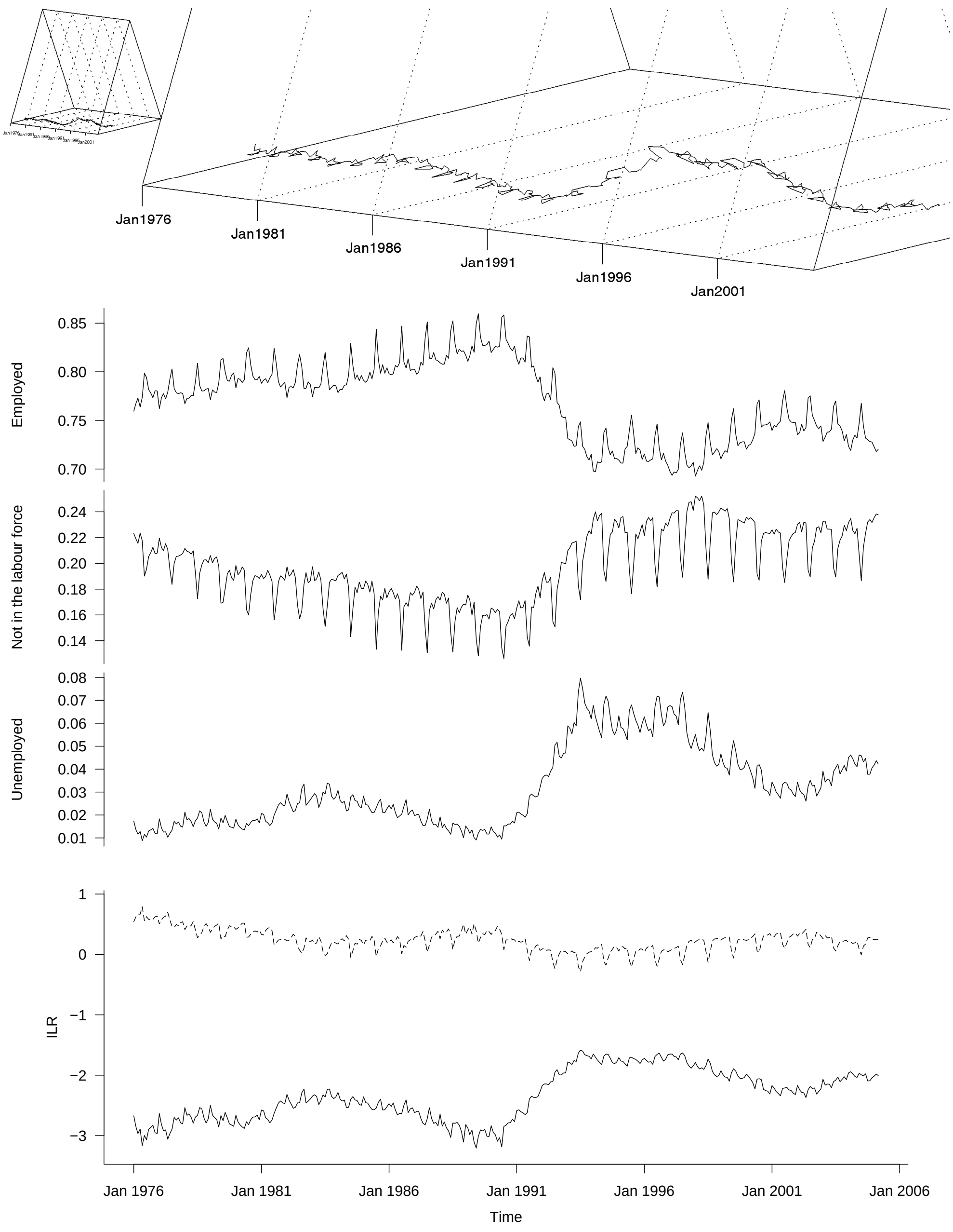

Figure 1: The top plot shows the time series in a ternary time series plot, where the top corner of the Simplex represents 100 % Unemployment, the bottom left corner 100 % Employment, and the bottom right corner that 100 % of the population are Not belonging to the labour force. The middle plot shows the three components of the time series in a standard time series plot. (Note that the vertical axis has been cut and has different scales.) The bottom plot shows the ILR-transformed series. (The second component of the transformed series is plotted with a dashed line.) In all three plots the structural change in the series during the early 1990's is clearly visible, as well as the seasonal pattern.

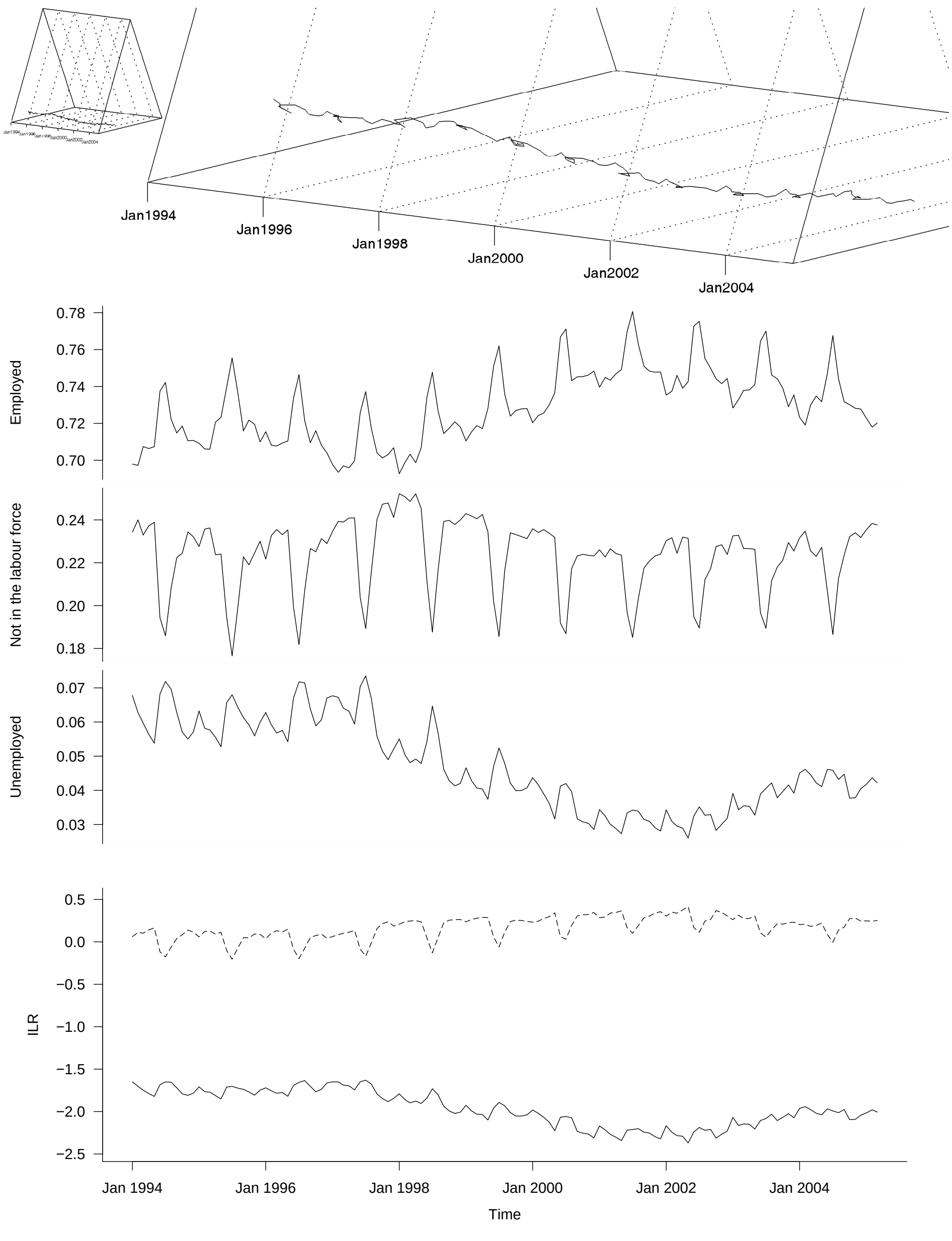

Figure 2: The top plot shows the time series from 1994 to 2005 in a ternary time series plot. The middle plot shows the three components of the time series in a standard time series plot. (Note that the vertical axis has been cut and has different scales.) The bottom plot shows the ILR-transformed series. (The second component of the transformed series is plotted with a dashed line.)

Table 1: Values for the four selection criteria for one to six lags for model I. The minimum value for each criterion is marked with bold face.

m	1	2	3	4	5	6
AIC(m)	-14.04	-14.12	-14.32	$\mathbf{-14.38}$	-14.37	-14.31
HQ(m)	-13.77	-13.81	-13.97	$\mathbf{-13.99}$	-13.94	-13.84
SC(m)	-13.37	-13.35	$\mathbf{-13.46}$	-13.42	-13.31	-13.15
FPE(m)	$7.99 \cdot 10^{-7}$	$7.39 \cdot 10^{-7}$	$6.06 \cdot 10^{-7}$	$\mathbf{5.73 \cdot 10^{-7}}$	$5.82 \cdot 10^{-7}$	$6.20 \cdot 10^{-7}$

Table 2: Values for the four selection criteria for one to six lags for model II. The minimum value for each criterion is marked with bold face.

m	1	2	3	4	5	6
AIC(m)	14.03	-14.28	$\mathbf{-14.33}$	-14.32	-14.27	-14.30
HQ(m)	-13.75	-13.97	$\mathbf{-13.98}$	-13.93	-13.84	-13.83
SC(m)	-13.35	$\mathbf{-13.51}$	-13.46	-13.36	-13.21	-13.14
FPE(m)	$8.12 \cdot 10^{-7}$	$6.31 \cdot 10^{-7}$	$\mathbf{6.01 \cdot 10^{-7}}$	$6.08 \cdot 10^{-7}$	$6.40 \cdot 10^{-7}$	$6.26 \cdot 10^{-7}$

different theoretical ideas and have slightly different properties. It can however be shown, that for $T \geq 16$, $\text{SC}(m) \leq \text{HQ}(m) \leq \text{AIC}(m)$. (Lütkepohl, 2005)

3.3 Prediction Regions

Predictions for 15 months ahead are calculated, and assuming bivariate normality $1-\alpha$ prediction regions $\boldsymbol{I}_{\boldsymbol{\mu}}$ are also calculated as

$$\boldsymbol{I}_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} : (\hat{\boldsymbol{z}}_h - \boldsymbol{\mu})' \boldsymbol{\Sigma}_{\hat{\boldsymbol{z}}_h}^{-1} (\hat{\boldsymbol{z}}_h - \boldsymbol{\mu}) \leq \chi^2_\alpha(2)\}$$

where $\hat{\boldsymbol{z}}_h$ is the prediction h steps ahead and $\boldsymbol{\Sigma}_{\hat{\boldsymbol{z}}_h}$ is the associated covariance matrix. (For model II and III, the $\boldsymbol{\Sigma}_{\hat{\boldsymbol{z}}_h}$ are also adjusted for the differencing.) The confidence regions $\boldsymbol{I}_{\boldsymbol{\mu}}$ are then transformed back to $\mathcal{S}^D$ using the inverse ILR transformation. The predictions are also transformed back as

$$\begin{aligned}
\text{I} : \hat{\boldsymbol{x}}_t &= \text{ilr}^{-1}(\hat{\boldsymbol{z}}_t) \\
\text{II} : \hat{\boldsymbol{x}}_t &= \text{ilr}^{-1}(\hat{\boldsymbol{z}}_t) \oplus \boldsymbol{x}_{t-1} \\
\text{III} : \hat{\boldsymbol{x}}_t &= \text{ilr}^{-1}(\hat{\boldsymbol{z}}_t) \oplus \boldsymbol{x}_{t-12}
\end{aligned}$$

The analyses are carried out using the R environment (R Development Core Team, 2008), primarily utilizing the `compositions` (van den Boogaart and others, 2006) and `vars` (Pfaff, 2008) packages. The ternary time series plots have been made in Matlab.

4 Results

The values of the model selection functions for the three models are presented in Tables 1-3. The minimum value for each criterion and model is marked with bold face. For model I the AIC, HQ and FPE all suggested four lags, and a VAR(4)-model with seasonal dummy variables was fitted to the data. Also for model II, three of four criteria agreed, recommending three lags. A VAR(3)-model with seasonal dummies was therefore fitted to the data. For the third model, two criteria (AIC and FPE) recommended five lags, and two criteria (HQ and SC) recommended three lags. As values of the functions for three and five lags are quite similar (compared to the differences

Table 3: Values for the four selection criteria for one to six lags for model III. The minimum value for each criterion is marked with bold face.

m	1	2	3	4	5	6
AIC(m)	−13.19	−13.29	−13.47	−13.49	**−13.51**	−13.43
HQ(m)	−12.90	−12.96	**−13.09**	−13.08	−13.05	−12.93
SC(m)	−12.47	−12.47	**−12.54**	−12.46	−12.37	−12.20
FPE(m)	$1.87 \cdot 10^{-6}$	$1.70 \cdot 10^{-6}$	$1.43 \cdot 10^{-6}$	$1.39 \cdot 10^{-6}$	$\mathbf{1.38 \cdot 10^{-6}}$	$1.49 \cdot 10^{-6}$

Table 4: Mean Absolute Deviations (MAD) and Mean Squared Errors (MSE) for the three models.

Model	I	II	III
MAD	0.0560	0.0571	0.1234
MSE	0.0038	0.0036	0.0233

between five and two lags), based on parsimony, a VAR(3)-model was chosen with no seasonal dummy variables.

The computations yield the following models

$$\begin{aligned}
\text{I} : \hat{\boldsymbol{z}}_t &= \begin{bmatrix} -0.127 \\ -0.045 \end{bmatrix} + \begin{bmatrix} 0.737 & -0.003 \\ -0.165 & 0.507 \end{bmatrix} \boldsymbol{z}_{t-1} + \begin{bmatrix} 0.202 & -0.111 \\ -0.109 & -0.010 \end{bmatrix} \boldsymbol{z}_{t-2} \\
&+ \begin{bmatrix} 0.420 & 0.481 \\ 0.043 & 0.226 \end{bmatrix} \boldsymbol{z}_{t-3} + \begin{bmatrix} -0.433 & -0.494 \\ 0.199 & 0.177 \end{bmatrix} \boldsymbol{z}_{t-4} + \hat{\boldsymbol{A}}_{SD}(t) \\
\text{II} : \hat{\boldsymbol{z}}_t &= \begin{bmatrix} -0.002 \\ 0.001 \end{bmatrix} + \begin{bmatrix} -0.220 & 0.031 \\ -0.135 & -0.446 \end{bmatrix} \boldsymbol{z}_{t-1} + \begin{bmatrix} 0.012 & -0.046 \\ -0.228 & -0.432 \end{bmatrix} \boldsymbol{z}_{t-2} \\
&+ \begin{bmatrix} 0.444 & 0.469 \\ -0.178 & -0.180 \end{bmatrix} \boldsymbol{z}_{t-3} + \hat{\boldsymbol{A}}_{SD}(t) \\
\text{III} : \hat{\boldsymbol{z}}_t &= \begin{bmatrix} 0.002 \\ 0.000 \end{bmatrix} + \begin{bmatrix} 0.545 & -0.204 \\ -0.071 & 0.462 \end{bmatrix} \boldsymbol{z}_{t-1} + \begin{bmatrix} 0.155 & -0.121 \\ -0.079 & 0.086 \end{bmatrix} \boldsymbol{z}_{t-2} \\
&+ \begin{bmatrix} 0.224 & 0.195 \\ 0.147 & 0.357 \end{bmatrix} \boldsymbol{z}_{t-3}
\end{aligned}$$

where $\hat{\boldsymbol{A}}_{SD}$ are the estimates of the seasonal dummy variables coefficients (not shown).

4.1 Predictions

For each of the three models predictions are calculated for the next 15 periods (months). To compare the accuracy of the predictions the Mean Absolute Deviations (MAD) and Mean Squared Errors (MSE) are calculated as

$$\text{MAD} = \frac{\sum_{i=1}^{n} \parallel \boldsymbol{x}_t \ominus \hat{\boldsymbol{x}}_t \parallel_S}{n}$$

$$\text{MSE} = \frac{\sum_{i=1}^{n} \parallel \boldsymbol{x}_t \ominus \hat{\boldsymbol{x}}_t \parallel_S^2}{n}$$

where n is the number of predictions. The values for three models are presented in Table 4. We note that model I and II yield almost the same MAD and MSE values, thus indicating the same predictive ability. Model III on the other hand seems to have much poorer predictive ability. The predictions and the actual values of the time series are presented in Table 5. In the table we can see that model III tend to overestimate the employment rate and underestimate the unemployment

Table 5: Actual and predicted values for the period January 2004 to March 2005. E is the percentage Employed, U the percentage Unemployed, and N is the percentage of the population Not in the labour force.

Time	Actual values			Predictions								
				Model I			Model II			Model III		
	E	U	N	E	U	N	E	U	N	E	U	N
Jan 2004	72.3	4.5	23.2	72.6	4.6	22.8	73.5	3.9	22.5	72.8	3.9	23.3
Feb 2004	71.9	4.6	23.5	72.4	4.4	23.3	72.6	4.6	22.8	73.3	3.4	23.3
Mar 2004	73.0	4.5	22.6	72.8	4.1	23.1	72.5	4.3	23.2	73.8	3.5	22.7
Apr 2004	73.5	4.2	22.3	72.9	4.1	23.0	73.0	4.0	23.0	73.8	3.5	22.7
May 2004	73.2	4.1	22.7	73.2	3.8	23.0	73.0	4.0	22.9	74.1	3.3	22.6
Jun 2004	74.7	4.6	20.7	75.7	4.7	19.6	73.4	3.8	22.8	76.5	3.9	19.7
Jul 2004	76.8	4.6	18.7	76.6	5.0	18.3	75.9	4.6	19.4	77.0	4.1	18.9
Aug 2004	74.4	4.3	21.3	74.4	4.8	20.8	76.9	4.9	18.2	74.6	4.2	21.2
Sep 2004	73.2	4.5	22.4	73.3	4.3	22.4	74.7	4.7	20.6	74.4	3.8	21.8
Oct 2004	73.0	3.8	23.2	73.4	4.1	22.6	73.7	4.2	22.1	73.9	4.0	22.1
Nov 2004	72.8	3.8	23.4	73.1	4.0	22.9	73.7	3.9	22.3	72.9	4.2	22.9
Dec 2004	72.8	4.0	23.2	73.1	4.1	22.8	73.5	3.9	22.6	73.5	3.9	22.5
Jan 2005	72.3	4.2	23.5	72.2	4.6	23.1	73.5	4.0	22.5	71.9	5.0	23.2
Feb 2005	71.8	4.4	23.8	72.3	4.3	23.3	72.7	4.5	22.8	72.1	4.4	23.5
Mar 2005	72.0	4.2	23.8	72.6	4.2	23.3	72.8	4.1	23.0	72.5	4.5	23.0

rate, whereas the other two models perform better. It is however also interesting to note that model II from August 2004 and onwards constantly overestimates the employment rate, underestimates the proportion not in the labour force, but gets the unemployment rate fairly accurate. Model I on the other hand only makes small, non-systematic errors.

To give the complete picture, the predictions and prediction regions are also plotted on a ternary time series plot in Figure 3. We note that model I and II manage to predict the true values very well, whereas model III clearly deviates, confirming the previous analyses. The plot though allows us to compare the prediction regions, which is hard to do non-graphically.

From the plot it is clearly visible that the first model has the smallest prediction regions, though the regions of model III is only slightly larger for the first twelve predictions. Model III though has less accurate predictions. Model II has the largest prediction regions, at least for predictions three steps or more ahead. The model thus has the poorest predictive power.

This indicates that the model of choice would be model I since it manages to get the predictions accurate and have a good predictive power.

5 Conclusions

We have in this paper demonstrated that the ILR-transformation can be utilized to analyse Compositional Time Series. Three different models have been constructed and analysed. The first and simplest model produced the best predictions with the best predictive power.

The models could of course have been improved if exogenous variables had been introduced. It has however not been our aim to perfectly model the temporal changes in the Swedish work force, but to examine to possibilities for time series modelling of compositional data.

In two of the models we used the difference operator, on the compositional time series. This produces a result that can be interpreted within the simplicial framework. It is however not clear how e.g. a cointegration model could be interpreted within the simplicial framework. And of course,

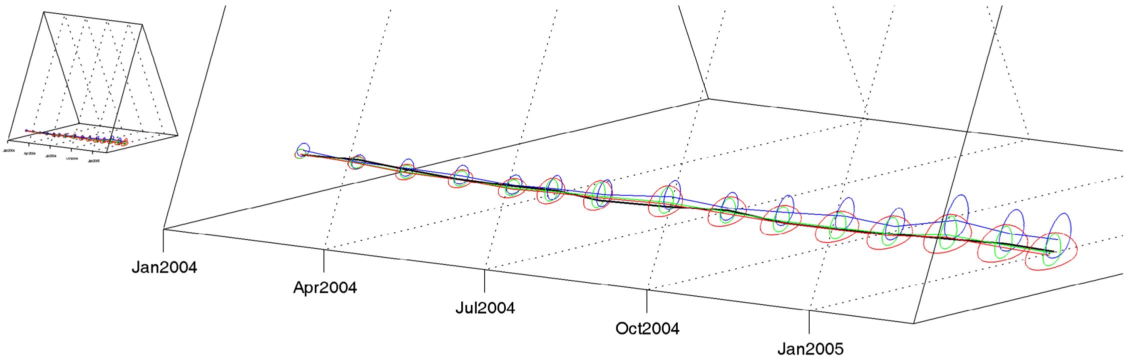

Figure 3: Predictions and confidence regions plotted in the simplex. The actual values are plotted with black, and the predictions of model I in green, of model II in red and of model III in blue. (The top corner of the Simplex represents 100 % Unemployment, the bottom left corner 100 % Employment, and the bottom right corner that 100 % of the population are Not belonging to the labour force.) As can been seen both the first and the second model manages to predict the time series very well. Model I though produces smaller confidence regions.

it is unsatisfactory that the coefficients of the model, $\boldsymbol{A}_k$, not easily interpreted.

References

Aguilar Zuil, L., C. Barceló-Vidal, and J. M. Larrosa (2007). Compositional time series analysis: A review. In *Proceedings of the 56th Session of the ISI (ISI 2007), Lisboa, August 22-29 (in press).*

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

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Caldwell, NJ: The Blackburn Press.

Barceló-Vidal, C., L. Aguilar, and J. Martín-Fernández (2007). Time series of compostional data: A first approach. In *Proceedings of the 22nd International Workshop of Statistical Modelling (IWSM 2007), Barcelona, July 2-6*, pp. 81–86.

Billheimer, D. and P. Guttorp (1995, Nov). Spatial models for discrete compositional data. Technical report, Dept. of Statistics, University of Washington, Seattle.

Billheimer, D., P. Guttorp, and W. F. Fagan (1998). Statistical analysis and interpretation of discrete compositional data. Technical Report Series 11, NRSCE.

Brunsdon, T. M. (1987). *Time series analysis of compositional data*. Ph. D. thesis, University of Southampton, Dept. of Mathematics.

Brunsdon, T. M. and T. M. F. Smith (1998). The time series analysis of compositional data. *Journal of Official Statistics 14*(3), 237–253.

Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology 35*(3), 279–300.

Larrosa, J. M. (2005, Oct). Compositional time series: Past and present. EconWPA. Econometrics 0510002.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin Heidelberg: Springer Verlag.

Pearson, K. (1897). Mathematical contributions to the theory of evolution.—on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London LX*, 489–498.

Persson, E. and M. Henkel (2005). Arbetskraftsundersökningarna 2004. Statistiska Meddelanden AM 12 SM 0510, Statistics Sweden.

Pfaff, B. (2008). *vars: VAR Modelling*. R package version 1.3-7.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Smith, T. M. F. and T. M. Brunsdon (1989). The time series analysis of compositional data. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 26–32.

van den Boogaart, K. G., R. Tolosana, and M. Bren (2006). *compositions: Compositional Data Analysis*. R package version 0.91-6.