Survey Methodology, December 2001
Vol. 27, No. 2, pp. 205-215
Statistics Canada, Catalogue No. 12-001

Modelling Compositional Time Series from Repeated Surveys

D.B.N. Silva and T.M.F. Smith [1]

Abstract

A compositional time series is defined as multivariate time series in which each of the series has values bounded between zero and one and the sum of the series equals one at each time point. Data with such characteristics are observed in repeated surveys when a survey variable has a multinomial response but interest lies in the proportion of units classified in each of its categories. In this case, the survey estimates are proportions of a whole subject to a unity-sum constraint. In this paper we employ a state space approach for modelling compositional time series from repeated surveys taking into account the sampling errors. The additive logistic transformation is used in order to guarantee predictions and signal estimates bounded between zero and one which satisfy the unity-sum constraint. The method is applied to compositional data from the Brazilian Labour Force Survey. Estimates of the vector of proportions and the unemployment rate are obtained. In addition, the structural components of the signal vector, such as the seasonals and the trends, are produced.

Key Words: Additive logistic transformation; Compositional time series; Kalman Filter; Labour force survey; Repeated surveys; State space models.

1. Introduction

All surveys are multivariate and multipurpose, and most are longitudinal, repeating the same questions over time. There are two broad classes of repeated surveys, those with overlapping first stage units and those with no overlap of first stage units. Both designs admit a longitudinal macro-analysis of population aggregates but only the former allows a micro-analysis and the estimation of gross flows or some other similar unit level dynamic process. In this paper we explore the time series analysis of a multivariate vector of population aggregates, a macro-analysis, while taking into account the influence of the sampling errors of the survey using disaggregated data.

Denote by $\boldsymbol{\theta}_t = (\theta_{1t}, ..., \theta_{M+1,t})'$ a vector of population quantities of interest at time t, and assume that observations are made at equally spaced time intervals $t = 1, 2, ..., T$. Let $\boldsymbol{y}_t = (y_{1t}, ..., y_{M+1,t})'$ represent a survey-based estimate of $\boldsymbol{\theta}_t$ based on data collected at time t. Repeated surveys produce time series $\{\boldsymbol{y}_t\}$ comprising estimates of the unknown target series $\{\boldsymbol{\theta}_t\}$. Focussing on the unknown population vector $\boldsymbol{\theta}_t$, it is natural to imagine that knowledge of $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{t-1}$ conveys useful information about $\boldsymbol{\theta}_t$ but without implying that it is perfectly predictable from $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{t-1}$. One way of representing this situation is by considering $\boldsymbol{\theta}_t$ to be a random variable which evolves stochastically in time following a certain time series model, as first proposed for univariate survey analysis by Blight and Scott (1973), Scott and Smith (1974) and Scott, Smith and Jones (1977). The survey estimates $\boldsymbol{y}_t$ and $\boldsymbol{\theta}_t$ can then be written as:

$$\boldsymbol{y}_t = \boldsymbol{\theta}_t + \boldsymbol{e}_t \tag{1}$$

where $\{\boldsymbol{\theta}_t\}, \{\boldsymbol{y}_t\}$ and $\{\boldsymbol{e}_t\}$ are random processes and $\boldsymbol{e}_t = (e_{1t}, ..., e_{M+1,t})'$ are the sampling errors such that $E(\boldsymbol{e}_t|\boldsymbol{\theta}_t) = \boldsymbol{0}$ and $V(\boldsymbol{e}_t|\boldsymbol{\theta}_t) = \Sigma_t$.

The | early work of Scott *et al.* (1977) was concerned with univariate $\{y_t\}$ and distinguished different forms for the data available on $\{e_t\}$. If the only data available to the analyst are the population aggregate estimates $\{y_t\}$ then this is termed a secondary analysis and the examples in Scott *et al.* (1977) are based on a secondary analysis of survey data. If the individual data records are available, then variances and covariance can be estimated directly from the data and this is called a primary analysis. In addition, in the case of a rotating panel survey, elementary estimates (based on data from a set of units that join and leave the survey at the same time) can be used to estimate the covariance structure of the sampling errors. Subsequent work by Jones (1980) used a primary analysis to measure the structure of the sampling noise whereas Binder and Hidiroglou (1988), Binder and Dick (1989), Pfeffermann, Burck and Ben-Tuvia (1989), Pfeffermann and Burck (1990), Pfeffermann (1991), Binder, Bleuer and Dick (1993), Pfeffermann and Bleuer (1993), Pfeffermann, Bell and Signorelli (1996), Pfeffermann, Feder and Signorelli (1998) and Harvey and Chung (2000) employed an elementary analysis.

The time series analysis of survey data also requires that the signal process be modelled. In the early works it was assumed that $\{\theta_t\}$ was a stationary process and that $\{y_t\}$ was the superposition of two stationary processes therefore being itself stationary. Typically ARMA processes were assumed for $\{\theta_t\}$ and $\{e_t\}$, and hence for $\{y_t\}$. Binder and Hidiroglou (1988) wrote the processes in state space

1. D.B.N. Silva, Instituto Brasileiro de Geografia e Estatística, Escola Nacional de Ciências Estatísticas, Rua André Cavalcanti 106 – Rio de Janeiro, RJ Brazil, 20231.050. E-mail: denisesilva@ibge.gov.br; T.M.F. Smith, University of Southampton, Faculty of Mathematical Studies, Highfield, Southampton, S017 1BJ, United Kingdom. E-mail: tmfs@maths.soton.ac.uk.

form which led rapidly to the introduction of nonstationary processes for the signal $\{\theta_t\}$, and structural models involving trends and seasonals have been used since then.

The aim is to improve estimation of the unobservable signal and its components, but when the sampling errors are autocorrelated these autocorrelations can induce spurious trends which get confounded with the true signal trend, as pointed out by Tiller (1992) and Pfeffermann, Bell and Signorelli (1996). When the variation in the sampling errors is not taken into account, their autocorrelation structure may be absorbed into either the seasonal or the trend components, thus affecting the inference from the model.

A special case of interest in repeated surveys is when the univariate target parameter $\{\theta_t\}$ is proportion such as the unemployment rate. Unrestricted time series modelling of $\{\theta_t\}$ may lead to estimates outside the range $0 \le \theta_t \le 1$. Wallis (1987) used a logistic transformation to ensure that the estimates were bounded, however he failed to take into account the survey error. Pfeffermann (1991), Tiller (1992), Pfeffermann and Bleuer (1993), Pfeffermann, Bell and Signorelli (1996) fitted state space models to unemployment rate series taking into account survey errors but without using the logistic transformation to guarantee bounded estimates.

Most surveys are multivariate and there has been little work in the multivariate time series analysis of survey data. Brunsdon (1987) and Brunsdon and Smith (1998) analyse multivariate data from opinion polls taking into account the fact that the proportions are bounded and comprise a composition, but not allowing for the structure of the survey errors. This work provides useful insight into the modelling of time series of proportions. Compositional data have also been modelled using a state space approach, by Quintana and West (1988), Shephard and Harvey (1989) and Singh and Roberts (1992), but these authors also did not address the issue of modelling the autocovariance structure of the sampling errors when the observed compositions are obtained from repeated surveys.

The motivation for this work is that many variables investigated by statistical agencies have a multinomial response and interest lies in the estimation of the proportion of units classified in each of the categories. If this is the case, the vector of proportions sums to one and forms what is known as a composition. A compositional time series is therefore a multivariate time series comprising observations of compositions at each time point. We propose a class of multivariate state space models for compositional time series from repeated surveys, which takes into account the sampling errors and guarantees estimates satisfying the underlying constraints imposed by compositions. The procedure employs a signal-plus-noise structural model which yields seasonally adjusted series and estimates of the trend which satisfy the underlying sum constraint. The method is applied to compositional data from the Brazilian Labour Force Survey comprising estimates of the vector of proportions of labour market status. Estimates of Seasonally adjusted compositions, trends and unemployment rate series are produced.

2. A Framework for Modelling Compositional Data from Overlapping Surveys

We assume that $\{\boldsymbol{\theta}_t\}$ is multivariate and the components θ_{mt} form a composition, *i.e.*, $0 < \theta_{mt} < 1 \;\forall\, m, t$ and $\sum_{m=1}^{M+1} \theta_{mt} = 1$. In this case $\boldsymbol{y}_t$ is a vector of sample estimates, based on the cross-sectional data of time t and belongs to the Simplex:

$$\boldsymbol{S}^M = \left\{ \begin{array}{l} \boldsymbol{y}_t \colon 0 \le y_{mt} \le 1,\, m = 1, \ldots, M+1; \\ \qquad \displaystyle\sum_{m=1}^{M+1} y_{mt} = 1;\, t = 1, \ldots, T \end{array} \right\},$$

as in Brunsdon and Smith (1998). In addition, it is assumed that $\boldsymbol{y}_t$ is obtained from a survey with complex design and overlapping units between occasions. Since each of its components is subject to sampling errors, y_{mt} can be decomposed as:

$$y_{mt} = \theta_{mt} + e_{mt}, \quad m = 1, \ldots, M+1, \tag{2}$$

where θ_{mt} is the unknown population proportion assumed to follow a time series model, and e_{mt} is the sampling error. Considering the $M+1$ series simultaneously, (2) can be written in vector form as in equation 1. In addition, it is assumed that

$$\sum_{m=1}^{M+1} \theta_{mt} = \sum_{m=1}^{M+1} y_{mt} = 1 \quad \forall t, \tag{3}$$

which implies that $\sum_{m=1}^{M+1} e_{mt} = 0, \forall t$.

A compositional time series is a sequence of vectors $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{M+1,\,t})'$ each belonging to $\boldsymbol{S}^M$. Aitchison (1986) examined the difficulties of applying standard methods to modelling and analyzing compositions and suggested the use of transformations to map compositions from the Simplex $\boldsymbol{S}^M$ onto $\mathbb{R}^M$. One such transformation is the *additive logratio transformation* (a_M), defined in Aitchison (1986, page 113), which was first adopted in a time series context by Brunsdon (1987, page 75). The transformation is given by $\boldsymbol{v}_t = a_M(\boldsymbol{y}_t) = (v_{1t}, \ldots, v_{Mt})'$, with

$$v_{mt} = \log\left(\frac{y_{mt}}{y_{M+1,\,t}}\right), \quad m = 1, \ldots, M, \quad \forall t, \tag{4}$$

where *log* denotes the natural logarithm. Note that $y_{M+1,\,t} = 1 - \sum_{m=1}^{M} y_{mt}$, sometimes called the fill-up value, is used as the reference variable or category. The inverse transformation, known as the *additive logistic transformation*, is given by $\boldsymbol{y}_t = a_M^{-1}(\boldsymbol{v}_t) = (y_{1t}, \ldots, y_{M+1,\,t})'$ such that

$$y_{mt}=\begin{cases}\dfrac{\exp(v_{mt})}{1+\sum_{j=1}^{M}\exp(v_{jt})} & m=1,...,M, \quad \forall\, t,\\[2ex] \dfrac{1}{1+\sum_{j=1}^{M}\exp(v_{jt})} & m=M+1, \quad \forall\, t.\end{cases} \tag{5}$$

The state space modelling procedure for compositional time series is invariant to the choice of the reference variable (Silva 1996), and so any element $y_{mt} \neq y_{M+1,t}$ of $\boldsymbol{y}_t$ can be taken as the reference variable when applying the additive logistic transformation to the vector of survey estimates. When the logratios $\boldsymbol{v}_t$ are normally distributed the $M+1-$part composition has an additive logistic normal distribution as defined in Aitchison and Shen (1980). For compositional time series, Brunsdon (1987) recommended the use of Vector ARMA models (Tiao and Box 1981) for the transformed series.

We propose a procedure that not only provides predictions and filtered estimates that are bounded between zero and one and satisfy the unity-sum constraint, but also improves the estimation of the unobservable signal and its components, taking into account the sampling error.

Following Bell and Hillmer (1990), the model in (2) can be rewritten as:

$$y_{mt}=\theta_{mt}\left(1+\frac{e_{mt}}{\theta_{mt}}\right)=\theta_{mt}\,u_{mt}, \tag{6}$$

with

$$u_{mt}=\left(1+\frac{e_{mt}}{\theta_{mt}}\right)=(1+\tilde{u}_{mt}), \tag{7}$$

where $\tilde{u}_{mt}=e_{mt}/\theta_{mt}$ represents the relative sampling error of the estimated proportion.

Applying the additive logratio transformation defined in Aitchison (1986, page 113) to the vector $\boldsymbol{y}_t$, with components given in (2), produces a transformed vector $\boldsymbol{v}_t=a_M(\boldsymbol{y}_t)=(\boldsymbol{v}_{1t},...,\boldsymbol{v}_{Mt})'$ contained in $\mathbb{R}^M$. If $y_{M+1,t}$ is used as the reference variable, the transformed vector has as its m^{th} component:

$$v_{mt}=\log\left(\frac{y_{mt}}{y_{M+1,t}}\right)=\log\left(\frac{\theta_{mt}u_{mt}}{\theta_{M+1,t}u_{M+1,t}}\right)$$

$$=\log\left(\frac{\theta_{mt}}{\theta_{M+1,t}}\right)+\log\left(\frac{u_{mt}}{u_{M+1,t}}\right), \quad m=1,...,M. \tag{8}$$

From (8), a vector model for the transformed series can be written as:

$$\boldsymbol{v}_t=\boldsymbol{\theta}_t^*+\boldsymbol{e}_t^*, \tag{9}$$

with $\boldsymbol{v}_t=(v_{1t},...,v_{Mt})'$, $\boldsymbol{\theta}_t^*=(\theta_{1t}^*,...,\theta_{Mt}^*)'$ and $\boldsymbol{e}_t^*=(e_{1t}^*,...,e_{Mt}^*)'$, where $v_{mt}=\log(y_{mt}/y_{M+1,t})$, $\theta_{mt}^*=\log(\theta_{mt}/\theta_{M+1,t})$ and $e_{mt}^*=\log(u_{mt}/u_{M+1,t})$, for $m=1,...,M$. Note that model (9) has the same form as model (1).

To describe the survey data, model (9) must incorporate time series models for both $\{\boldsymbol{\theta}_t^*\}$ and $\{\boldsymbol{e}_t^*\}$. Hence a multivariate model for the transformed data will depend on the form of the time series models for $\{\boldsymbol{\theta}_t^*\}$ and $\{\boldsymbol{e}_t^*\}$.

The state space formulation for compositional data is examined in section 3, the model estimation is considered in section 4 and is illustrated using Brazilian Labour Force Survey data in section 5.

3. Modelling the Transformed Series

Our approach is based on assuming that the transformed series $\boldsymbol{v}_t=a_M(\boldsymbol{y}_t)$ has the signal plus noise structure in equation 9. We propose structural time series models for $\{\boldsymbol{\theta}_t^*\}$, as in Harvey (1989), and vector ARMA models (Tiao and Box 1981) for $\{\boldsymbol{e}_t^*\}$.

The transformed signal process $\{\boldsymbol{\theta}_t^*\}$ is assumed to follow the multivariate basic structural model, with each of the components $\{\theta_{mt}^*\}$ following a basic structural time series model (BSM) with possibly different parameters across the series. The cross-sectional relationship between the series is accounted for by the correlation structure of the system disturbances. The model for $\{\theta_{mt}^*\}, m=1,2,...,M$, is then given by:

$$\begin{cases}\theta_{mt}^*=L_{mt}^*+S_{mt}^*+I_{mt}^*,\\ L_{mt}^*=L_{m,t-1}^*+R_{m,t-1}^*+\eta_{mt}^{(l)},\\ R_{mt}^*=R_{m,t-1}^*+\eta_{mt}^{(r)},\\ S_{mt}^*=-\sum_{j=1}^{11}S_{m,t-1}^*+\eta_{mt}^{(s)},\end{cases} \tag{10}$$

where L_{mt}^* is the trend/level component of the signal θ_{mt}^*, R_{mt}^* is the corresponding change in the level, S_{mt}^* is the seasonal component and I_{mt}^* is an irregular component. For each component, the disturbances $\eta_{mt}^{(l)}, \eta_{mt}^{(r)}, \eta_{mt}^{(s)}$, and the irregular I_{mt}^*, are assumed to be mutually uncorrelated normal deviates with mean zero and variances $\sigma_{m_l}^2$, $\sigma_{m_r}^2, \sigma_{m_s}^2, \sigma_{m_I}^2$, respectively. That is, the $M\times 1$ vector disturbances $\boldsymbol{\eta}_t^{(l)}, \boldsymbol{\eta}_t^{(r)}, \boldsymbol{\eta}_t^{(s)}$ and $\boldsymbol{I}_t^*$, are mutually uncorrelated in all time periods. In addition, the irregulars I_{mt}^*, $I_{j(t-h)}^*$, with $m\neq j, h=\ldots,-2,-1,0,1,2,\ldots$, are assumed to be correlated when $h=0$, but uncorrelated for $h\neq 0$ and $\boldsymbol{I}_t^*$ has covariance matrix Σ_I. The same happens with the system disturbances $\eta_{mt}^{(a)}, \eta_{j(t-h)}^{(a)}, a = l, r, s,$ which are also correlated when $h = 0,$ but uncorrelated for $h \neq 0,$ with covariance matrices $\sum_l, \sum_r, \sum_s.$ At each time $t,$ the correlation structure between the components of the composition is summarized by $\sum_I$ and a block diagonal matrix with the blocks being $\sum_l, \sum_r, \sum_s.$ Note that the relation between the series arises via the non-zero off-diagonal elements of the disturbance covariance matrices. The multivariate model (10) for $\{\boldsymbol{\theta}_t^*\}$ has the following state space formulation:

$$\begin{cases} \boldsymbol{\theta}_t^* = \boldsymbol{H}^{(\theta)}\boldsymbol{\alpha}_t^{(\theta)} + \boldsymbol{I}_t^*; \\ \boldsymbol{\alpha}_t^{(\theta)} = \boldsymbol{T}^{(\theta)}\boldsymbol{\alpha}_{t-1}^{(\theta)} + \boldsymbol{G}^{(\theta)}\boldsymbol{\eta}_t^{(\theta)}, \end{cases} \tag{11}$$

where $\boldsymbol{H}^{(\theta)} = [1010000000000] \otimes \boldsymbol{I}_M,$

$$\boldsymbol{\alpha}_t^{(\theta)} = [L_{1t}^* \ldots L_{Mt}^* R_{1t}^* \ldots R_{Mt}^* S_{1t}^* \ldots S_{Mt}^* \ldots S_{1,t-10}^* \ldots S_{M,t-10}^*]'$$

$$\boldsymbol{\eta}_t^{(\theta)} = (\eta_{1t}^{(l)} \ldots \eta_{Mt}^{(l)}\ \eta_{1t}^{(r)} \ldots \eta_{Mt}^{(r)}\ \eta_{1t}^{(s)} \ldots \eta_{Mt}^{(s)})',$$

$$\boldsymbol{G}^{(\theta)} = \begin{bmatrix} \boldsymbol{I}_3 \\ \cdots\ \cdots \\ \mathbf{0}_{10\times 3} \end{bmatrix} \otimes \boldsymbol{I}_M,$$

$$\boldsymbol{T}^{(\theta)} = \left[\begin{array}{cc|cccccc} 1 & 1 & & & \mathbf{0}_{2\times 11} & & & \\ 0 & 1 & & & & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & -1 & -1 & \cdots & -1 & -1 & \\ & & 1 & 0 & \cdots & 0 & 0 & \\ \mathbf{0}_{11\times 2} & & 0 & 1 & \cdots & 0 & 0 & \\ & & \vdots & \vdots & \vdots & \vdots & \vdots & \\ & & 0 & 0 & \cdots & 1 & 0 & \end{array}\right] \otimes \boldsymbol{I}_M.$$

The transformed survey error process $\{\boldsymbol{e}_t^*\}$ is assumed to follow an M–dimensional vector autoregressive moving average process (VARMA), defined by $\boldsymbol{\Phi}(B)\boldsymbol{e}_t^* = \boldsymbol{\Theta}(B)\boldsymbol{a}_t,$ with mean vector $E(\boldsymbol{e}_t^*) = \mathbf{0}$ and

$$\boldsymbol{\Theta}(B) = \boldsymbol{I} - \boldsymbol{\Theta}_1 B - \ldots - \boldsymbol{\Theta}_q B^q,$$

$$\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \ldots - \boldsymbol{\Phi}_q B^p,$$

where $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q$ are coefficient matrices and $\boldsymbol{a}_t$ is an M–dimensional white noise random vector with zero mean and covariance structure:

$$E(\boldsymbol{a}_t \boldsymbol{a}_{t-h}') = \begin{cases} \sum_a & h = 0, \\ \mathbf{0} & h \neq 0. \end{cases}$$

The cross-covariance matrix function for the VARMA process $\{\boldsymbol{e}_t^*\},$ (see Wei 1993, page 333), is given by:

$$\boldsymbol{\Gamma}_{e^*}(h) = \mathrm{COV}(\boldsymbol{e}_{t-h}^*, \boldsymbol{e}_t^*) = E(\boldsymbol{e}_{t-h}^*\ \boldsymbol{e}_t^{*\prime}),$$

where $\{\boldsymbol{\Gamma}_{e^*}(h)\}_{mj} = \gamma_{e^*mj}(h) = \mathrm{COV}(e_{m,t-h}^*, e_{jt}^*),$ and the cross-correlation function for the vector process is defined as:

$$\mathrm{P}_{e^*}(h) = \boldsymbol{D}_{e^*}^{-1/2}\Gamma_{e^*}(h)\boldsymbol{D}_{e^*}^{-1/2},$$

where

$$\boldsymbol{D}_{e^*} = \mathrm{diag}(\gamma_{e^*11}(0), \ldots, \gamma_{e^*MM}(0)).$$

The state space representation of VARMA models can be found in Reinsel (1993, section 7.2). The separate models for the transformed signal and sampling errors can be cast into a unique state space model, see Silva (1996, Chapter 8) for details.

4. Estimation from the Transformed Data

As in previous sections, we distinguish between the estimation of the structure of the surveys errors, the noise, and the estimation of the covariance of the basic structural model. Once these are obtained, we employ the Kalman filter to get estimates of the trend and seasonals which determine the signal. Before carrying out the signal extraction, the VARMA model for the survey errors must be identified.

The model specification for the error process $\{\boldsymbol{e}_t^*\}$ depends on the sampling design, particularly on the level of sample overlap between occasions, and also on data availability. Many authors have considered the problem of modelling the sampling error process in a univariate framework, see, for example, Scott and Smith (1974), Pfeffermann (1989, 1991), Bell and Hillmer (1990), Binder and Dick (1989), Tiller (1989, 1992), Pfeffermann and Bleuer (1993), Binder, Bleuer and Dick (1993), Pfeffermann, Bell and Signorelli (1996) and Pfeffermann, Feder and Signorelli (1998). However, in all of these cases the authors are working with the original data instead of the transformed data. After transformation, it is difficult to carry out a full primary analysis based on individual observations, see Silva (1996, Chapter 7).

Many repeated surveys are based on a rotating panel design in which K panels of sampling units are investigated at each survey round (time point) and panels are replaced in a systematic manner, according to the rotating pattern of the survey design. In these surveys, elementary design unbiased estimtes $\boldsymbol{y}_t^{(k)}, k = 1, \ldots, K,$ for the population parameter $\boldsymbol{\theta}_t,$ can be obtained from each rotation group. A rotation group is a set of sampling units that joins and leaves the sample at the same time.

In a two-stage survey, in which the primary sampling units (enumeration areas) remain in the sample for all survey occasions, the replacement of panels of households (second-stage units) is ordinarily carried out within geographical regions defined by mutually exclusive groups of enumeration areas. Note that a survey with K panels produces K streams of estimates, where a stream is a time series of all sample estimates based on samples from the same enumeration area, that is, is a time series of elementary estimates.

Pfeffermann, Bell and Signorelli (1996) and Pfeffermann, Feder and Signorelli (1998) show how to estimate the autocorrelation of the sampling error process for univariate data, before transformation, using the so-called pseudo-errors, defined as:

$$\tilde{\boldsymbol{e}}_t^{(k)} = \boldsymbol{y}_t^{(k)} - \boldsymbol{y}_t, \tag{12}$$

where $\boldsymbol{y}_t = 1/K \sum_{k=1}^{K} \boldsymbol{y}_t^{(k)}$. If there is no rotation bias, it follows that:

$$\tilde{\boldsymbol{e}}_t^{(k)} = \boldsymbol{e}_t^{(k)} - \boldsymbol{e}_t, \tag{13}$$

thus contrasts in $\boldsymbol{y}_t^{(k)}$ are contrasts in the panel sampling errors $\boldsymbol{e}_t^{(k)}$.

For the compositional case we apply, for each elementary estimate, the transformation $\boldsymbol{v}_t^{(k)} = a_m(\boldsymbol{y}_t^{(k)}) = (v_{1t}^{(k)}, ..., v_{Mt}^{(k)})'$ which has as its m^{th} component, $(m = 1, ..., M)$:

$$v_{mt}^{(k)} = \log\left(\frac{y_{mt}^{(k)}}{y_{M+1,t}^{(k)}}\right) = \log\left(\frac{\theta_{mt}}{\theta_{M+1,t}}\right) + \log\left(\frac{u_{mt}^{(k)}}{u_{M+1,t}^{(k)}}\right). \tag{14}$$

From (14), a vector model for the k^{th} series of transformed elementary estimates can be written as:

$$\boldsymbol{v}_t^{(k)} = \boldsymbol{\theta}_t^* + \boldsymbol{e}_t^{*(k)}, \tag{15}$$

with $\boldsymbol{e}_t^{*(k)} = (e_{1t}^{*(k)}, ..., e_{Mt}^{*(k)})'$ and $e_{mt}^{*(k)} = \log(u_{mt}^{(k)} / u_{M+1,t}^{(k)})$, for $(m = 1, ..., M)$. Hence, from (15), M–dimensional time series of transformed pseudo-errors can be constructed from deviations of the transformed rotation group estimates about their overall mean. The transformed pseudo-errors for the k^{th} rotation group are defined as:

$$\begin{aligned} \tilde{\boldsymbol{e}}_t^{*(k)} &= (\tilde{e}_{1t}^{*(k)}, ..., \tilde{e}_{Mt}^{*(k)})' = \boldsymbol{v}_t^{(k)} - \boldsymbol{v}_t \\ &= (v_{1t}^{(k)} - v_{1t}, ..., v_{Mt}^{(k)} - v_{Mt})', \end{aligned} \tag{16}$$

where $\boldsymbol{v}_t = 1/K \sum_{k=1}^{K} \boldsymbol{v}_t^{(k)}$. Note, in addition, that $\tilde{\boldsymbol{e}}_t^{*(k)} = \boldsymbol{e}_t^{*(k)} - \boldsymbol{e}_t^*$.

From (14) and (15), it becomes clear that the framework introduced by Pfeffermann, Bell and Signorelli (1996) can also be applied to the transformed model.

The cross-correlation matrices of the transformed sampling errors can be obtained by averaging the cross-covariances matrices of the transformed pseudo-errors as follows (for details see Silva 1996, Chapter 7):

$$\boldsymbol{P}_{e^*}(h) = \left[\sum_{k=1}^{K} D_{\tilde{e}^*}^{(k)}\right]^{-1/2} \left[\sum_{k=1}^{K} \boldsymbol{\Gamma}_{\tilde{e}^*}^{(k)}(h)\right] \left[\sum_{k=1}^{K} D_{\tilde{e}^*}^{(k)}\right]^{-1/2}, \tag{17}$$

where

$$\boldsymbol{\Gamma}_{\tilde{e}^*}^{(k)}(h) = \text{COV}(\tilde{\boldsymbol{e}}_{t-h}^{*(k)}, \tilde{\boldsymbol{e}}_t^{*(k)}) = E(\tilde{\boldsymbol{e}}_{t-h}^{*(k)} \tilde{\boldsymbol{e}}_t^{*(k)\prime}),$$

with

$$\{\boldsymbol{\Gamma}_{\tilde{e}^*}^{(k)}(h)\}_{mj} = \text{COV}(\tilde{e}_{m,t-h}^{*(k)}, \tilde{e}_{jt}^{*(k)}) = \gamma_{\tilde{e}^{*mj}}^{(k)}(h)$$

and

$$\boldsymbol{D}_{\tilde{e}^*} = \text{diag}(\gamma_{\tilde{e}^*11}^{(k)}(0), ..., \gamma_{\tilde{e}^*MM}^{(k)}(0)).$$

Once the correlation matrices $\boldsymbol{P}_{e^*}(h), h = 1, 2, ...$ have been estimated, a VARMA model to represent the transformed survey error process can be selected and estimates of the respective parameter matrices can be computed, provided the series of transformed pseudo-errors are available. Then, as described in section 3, a state space model for representing the transformed signal and sampling errors can be defined and the Kalman filter equations can be used to get filtered and smoothed estimates for the unobservable components. The application of the Kalman Filter requires the estimation of the unknown hyper-parameters (the covariance matrices $\sum_l, \sum_r, \sum_s, \sum_I, \sum_a$) and the estimation of the initial state vector and the respective covariance matrices.

Having addressed the issue of how to model the survey estimates in a compositional framework and how to identify the time series model for the transformed sampling errors, the following section presents the results of an empirical study using compositional data from the Brazilian Labour Force Survey.

5. Modelling Compositional Time Series in the Brazilian Labour Force Survey

The Brazilian Labour Force Survey (BLFS) collects monthly information about employment, hours of work, education and wages together with some demographic information. It classified the survey respondents, aged 15 and over, according to their employment status in the week prior to the interview into three main groups: employed, unemployed and not in the labour force, following the International Labour Organization (ILO) definitions. The survey targets the population living at the six major metropolitan areas in the country. The BLS is a two-stage sample survey in which the primary sampling units (psu) are the census enumeration areas (EA) and the second-stage units (ssu) are the households. The primary sampling units are selected with probabilities proportional to their sizes and then a fixed number of households is selected from each sampled EA by systematic sampling. All household members within the selected households are enumerated. The primary sampling units remain the same for a period of roughly 10 years (as in a master sample). New primary sampling units are selected when information from a new population census becomes available.

In addition, the BLFS is a rotating panel survey. For any given month the sample is composed of four rotation groups of mutually exclusive sets of primary sampling units. The rotation pattern applies to panels of second-stage units (households). Within each rotation group a panel of households stay in the sample for four successive months, is rotated out for the following 8 months and then is sampled again for another spell of four successive months. Each month one panel is rotated out of the sample. The substituting panel can be a new panel or one that has already been observed for the first four months period. Note that the 4-8-4 rotation pattern induces a complex correlation structure for the sampling errors over time and that there is a 75% overlap between two successive months.

The empirical work was carried out using data from the São Paulo metropolitan area covering the period from January 1989 to September 1993 (57 observations). The quantities of interest are the proportions of employed, unemployed and not in the labour force, and also the unemployment rate. Using the monthly individual observations, the series of sample estimates and their respective estimated standard errors were computed using data of each specific survey round and standard estimators. For each month, two sets of estimates were obtained. The direct sample estimates, derived from the complete data collected at a given month and four elementary estimates, each based on data from a single rotation group. The panel estimates are used to estimate the sampling error autocorrelations and to help to identify the time series model for the sampling errors.

In this study the observed composition has $M+1=3$ components and the time series is defined as the sequence of vectors $\boldsymbol{y}_t=(y_{1t}, y_{2t}, y_{3t})'$, where:

y_{1t} is the estimated proportion of unemployed persons in month t;

y_{2t} is the estimated proportion of employed persons in month t;

y_{3t} is the estimated proportion of persons not in the labour force in month t.

The model for the BLFS must incorporate the special features of the data. Firstly, it is a compositional time series belonging to the Simplex $\boldsymbol{S}^2$ at each time t. Secondly, the time series are subject to sampling errors. Following the approach in section 2, we first map the composition onto $\mathbb{R}^2$ using the additive logratio transformation with y_{3t} as the reference category. As $\boldsymbol{y}_t$ is a vector of sample estimates, it can be modelled as in equation 1 and the vector model for the transformed series is given by equation 9. Then, the transformed composition is modelled using a multivariate state space model that accounts for the autocorrelations between the sampling errors. Finally, the model based estimates are transformed back to the original space. Figure 1 displays the series of transformed compositions.

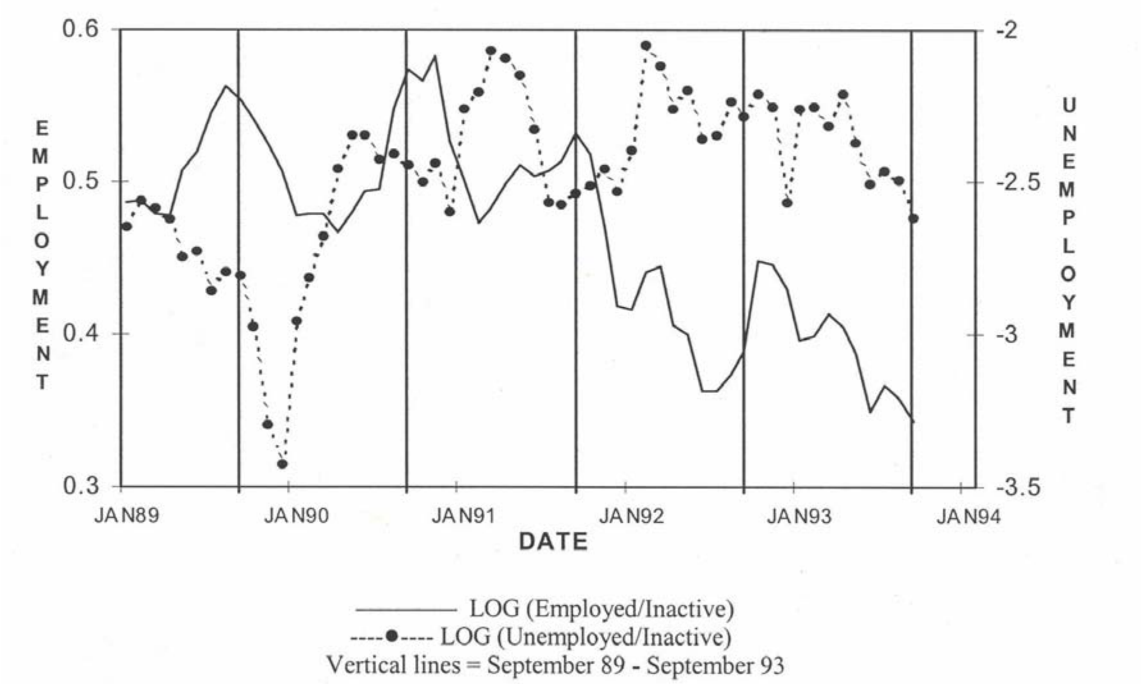

Figure 1. Brazilian Labour Force Series – SÃO PAULO Transformed Compositions.

The model for the transformed sample estimates $\boldsymbol{v}_t$ is composed of a bivariate model for the transformed signal $\boldsymbol{\theta}_t^*$, describing how the transformed population quantities evolve in time, and a bivariate model representing the time series relationship between transformed sampling errors $\boldsymbol{e}_t^*$. The Transformed signal process $\{\boldsymbol{\theta}_t^*\}$ is assumed to follow the bivariate basic structural model (equation 11) as described in section 3. As mentioned before, a VARMA model to represent the sampling error series was used. The correlation structure of the transformed sampling errors was estimated using the transformed pseudo-errors as in equation 16. In addition, estimates of the partial lag correlation matrices for $\{\boldsymbol{e}_t^*\}$ were computed using a recursive algorithm provided in Wei (1993, pages 359–362). A program in SAS-IML which gives the corresponding schematic representations (Tiao and Box 1981) and a statistical test to help establish the order of the vector process was developed. The form of the correlation matrices and the results for the statistical test, available in Silva (1996), indicate that a VAR(1), a VAR(2) or a VARMA(1,1) model could be used to represent the transformed sampling error process. In the event, the VARMA(1,1) was chosen because it yields smaller standard errors for estimates of the unemployment rate. The parameter estimates for this model were obtained from the relationship between the cross-covariance function and the parameter matrices given in Wei (1993, pages 346–347). The VARMA(1,1) fitted for $\{\boldsymbol{e}_t^*\}$ is given by:

$$\begin{bmatrix} e_{1t}^* \\ e_{2t}^* \end{bmatrix} = \begin{bmatrix} 0.7347 & 0.2414 \\ -0.9224 & -0.2072 \end{bmatrix} \begin{bmatrix} e_{1,t-1}^* \\ e_{2,t-1}^* \end{bmatrix}$$

$$- \begin{bmatrix} 0.3162 & 0.2590 \\ -0.7666 & -0.2749 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},$$

with

$$\hat{\sum}_a = \begin{bmatrix} 0.0001723 & 0.0003476 \\ 0.0003476 & 0.0051660 \end{bmatrix} \qquad (18)$$

Having put the combined model for the transformed survey estimates into the state space form, the Kalman Filter equations can be used to get filtered and smoothed estimates for the unobservable components. Note that the estimation of the model for the transformed sampling errors (equation 18) was implemented outside the Kalman Filter. The application of the Kalman Filter requires the estimation of the unknown hyperparameters (the covariances), the initial state vector and respective covariance matrix. Assuming that the disturbances $\boldsymbol{\eta}_t^{(\theta)}, \boldsymbol{a}_t$ and $\boldsymbol{I}_t$ are normally distributed, the log-likelihood function of the (transformed) observations can be expressed via the prediction error decomposition (for details see Harvey 1989). Estimates for the model covariances were obtained by maximum likelihood, applying a quasi-Newton optimization technique. A computer program to implement the maximization procedure was developed using the optimization routine NLPQN from SAS-IML.

The initialization of the Kalman filter was carried out using a combination of a diffuse and proper priors. Following this approach, the non-stationary components $(\boldsymbol{\alpha}^{(\theta)})'$ of the state vector were initialized with very large error variances and the respective components of the initial state vector were taken as zero. The stationary components $(e_{1t}^* \, e_{2t}^*)'$ were initialized by the corresponding unconditional mean and variance.

When fitting the model, the estimated covariance matrices obtained for the slope and seasonal components were very small and could be set to zero. This implies that the seasonals are assumed to be deterministic and that the slope is assumed to be fixed, giving rise to a local level model with a drift and non-stochastic seasonals for the signal. Indeed, as pointed out by Koopman, Harvey, Doornik and Shephard (1995, page 39), when the number of years considered in the analysis is small, it seems reasonable to fix the seasonals since there is not enough data to allow the estimation of a changing pattern. The fact that a fixed seasonal pattern is validated by the estimation process is a satisfactory feature of the modelling procedure. In addition, the estimated covariance matrix of the irregular component was also found to be very small (and hence undetectable) in comparison to the sampling error and so, as expected, in the presence of relatively large sampling errors, there was no need to include irregular components in the model for the transformed signal. The parameter estimates and respective asymptotic errors (displayed in parenthesis) are presented in Table 1.

Table 1
Estimates for the Hyperparameters and Standard Errors

Model	$\hat{\Sigma}_l \, x10^{-4}$ (2)	$\hat{\Sigma}_r = \hat{\Sigma}_s = \hat{\Sigma}_I$
BSM + VARMA(1, 1) (1)	$\begin{bmatrix} 2.78 & \mathbf{0.12} \\ (0.91) & \\ 1.95 & 87.0 \\ (3.55) & (27.10) \end{bmatrix}$	$\begin{bmatrix} 0 & - \\ 0 & 0 \end{bmatrix}$

(1) Local level model with drift and fixed seasonals for the signal.

(2) Upper-triangular contains correlation.

To evaluate the model performance, empirical distributions of the standardized residuals were compared with a standard normal distribution to verify the assumption that the innovations $(\boldsymbol{v}_t - \hat{\boldsymbol{v}}_{t|t-1})$ are normal deviates. Examination of corresponding normal plots revealed no departure from normality. In addition, we also computed the autocorrelations of the innovations, which were close to zero, further validating the model.

Predictions for y_{mt} and estimates for θ_{mt} are computed by applying the additive logistic transformation (equation 5) to predictions $\hat{\boldsymbol{v}}_{t|t-1}$ and smoothed estimates $\hat{\boldsymbol{\theta}}^*_{t|T}$ for the transformed series and signal, respectively. This transformation maps these estimates onto $\boldsymbol{S}^2$, guaranteeing that they satisfy the boundedness constraints.

Unfortunately, although $\hat{\boldsymbol{L}}^*_{t|T}$ and $\hat{\boldsymbol{S}}^*_{t|T}$ can be obtained from $\hat{\boldsymbol{\theta}}^*_{t|T}$, it is not straightforward to obtain estimates for the structural unobservable components of the original signal $\boldsymbol{\theta}_t$, such as $\hat{\boldsymbol{L}}_{t|T}$ and $\hat{\boldsymbol{S}}_{t|T}$, However, if a multiplicative model with no irregular component is assumed for $\{\theta_{mt}\}$, such that:

$$\theta_{1t} = L_{1t}S_{1t}, \theta_{2t} = L_{2t}S_{2t}, \theta_{3t} = L_{3t}S_{3t}, \tag{19}$$

where L_{mt} and S_{mt}, for $m = 1, 2, 3$ represent the trend and seasonal components of the unobservable signals, then applying an additive logratio transformation to $\boldsymbol{\theta}_t$ results in:

$$\begin{aligned}\log(\theta_{mt}/\theta_{3t}) &= \log\left(\frac{L_{mt}S_{mt}}{L_{3t}S_{3t}}\right)\\ &= \log\left(\frac{L_{mt}}{L_{3t}}\right) + \log\left(\frac{S_{mt}}{S_{3t}}\right), m = 1, 2.\end{aligned} \tag{20}$$

This can be rewritten as:

$$\theta^*_{mt} = L^*_{mt} + S^*_{mt}, \tag{21}$$

with $L^*_{mt} = \log(L_{mt}/L_{3t})$ and $S^*_{mt} = \log(S_{mt}/S_{3t})$. Thus, the use of a basic structural model for $\{\boldsymbol{\theta}^*_t\}$ corresponds to the case in which the underlying model for $\{\boldsymbol{\theta}_t\}$ decomposes the original signal into its trend and seasonal components in a multiplicative way. For deriving estimates, either filtered or smoothed, for L_{mt} note that:

$$\exp(L^*_{1t}) = L_{1t}/L_{3t}, \quad \exp(L^*_{2t}) = L_{2t}/L_{3t}. \tag{22}$$

To recover L_{1t}, L_{2t}, L_{3t}, in (22), it is necessary to assume an explicit relationship between these unobservable components based on model (19). By doing this, a third equation can be added to the system in (22) and an estimate of the original series components can be obtained. Note that the system has three unknowns for just two equations. In this case, it is quite natural to assume that the level components sum to one across the series, being also bounded between zero and one. Hence, trend estimates for the original series can be obtained solving:

$$\begin{cases}\exp(L^*_{1t}) & = L_{1t}/L_{3t},\\ \exp(L^*_{2t}) & = L_{2t}/L_{3t},\\ L_{1t} + L_{2t} + L_{3t} & = 1,\end{cases} \tag{23a}$$

which results in

$$\begin{aligned} L_{mt} &= \frac{\exp(L^*_{mt})}{1 + \sum_{k=1}^{2}\exp(L^*_{kt})}, \quad m = 1, 2;\\ L_{3t} &= \frac{1}{1 + \sum_{k=1}^{2}\exp(L^*_{kt})}.\end{aligned} \tag{23b}$$

As there is no irregular component in model (19) the seasonally adjusted figures are given by the trend estimates in (23). Therefore, the smoothed estimates for the trend of the original series of proportions are obtained by applying the additive logistic transformation to $\hat{\boldsymbol{L}}^*_{t|T}$. Consequently, estimates for the seasonal components of the original proportions can be computed as:

$$\hat{S}_{m,t|T} = \hat{\theta}_{m,t|T}/\hat{L}_{m,t|T}, \quad m = 1, 2, 3.$$

For labour force surveys, an important issues is the estimation of the unemployment rate series (as opposed to unemployment proportions) and also the production of the corresponding seasonally adjusted figures. Recall that θ_{1t} and θ_{2t} represent the unknown population proportions of unemployed and employed people, respectively. Using these proportions, the unknown unemployment rate at time is t defined as

$$R_1 = \frac{\theta_{1t}}{\theta_{1t} + \theta_{2t}} = \frac{1}{\left(1 + \dfrac{\theta_{2t}}{\theta_{1t}}\right)} = \left(\frac{\theta_{2t}}{\theta_{1t}} + 1\right)^{-1}. \tag{24}$$

Based on model (11), trend estimates for the unemployment rate can be obtained by simply replacing θ_{mt} by $L_{mt}, m = 1, 2,$ in equation 24. In conclusion, the methodology developed in this section provides signal (and trend) estimates that are bounded between zero and one and satisfy the unit-sum constraint. It also provides estimates for the seasonal and trend components of series comprising ratios of the original proportions which is a useful feature.

Figure 2 presents the design-based estimates and the model-dependent estimates for the proportion of unemployed persons, for the time period January 1989 to September 1993. The model-dependent estimates are the smoothed estimates which use all the data for the whole sample period. As can be seen from the graph, the signal estimates behave similarly to the design-based estimates although some of the sharp turning points in the series have been smoothed out.

Model-dependent trend estimates were obtained by fitting the basic structural model defined for the signal process when sampling error variation was modelled as a VARMA(1,1). These estimates were compared with the estimates produced by the familiar X-11 procedure. Figure 3 displays the trend produced for the unemployment rate series by both methods together with the estimates obtained by fitting a standard basic structural model which does not account for sampling error variation.

The trend produced by our model is smoother, suggesting that the model-dependent procedure succeeds in removing the fluctuations induced by the sampling errors. In addition, model-dependent estimates for the seasonal effects of the original compositions were also obtained from the multivariate modelling procedure which accounts for two very important features of the data, namely the compositional constraints and the presence of sampling errors.

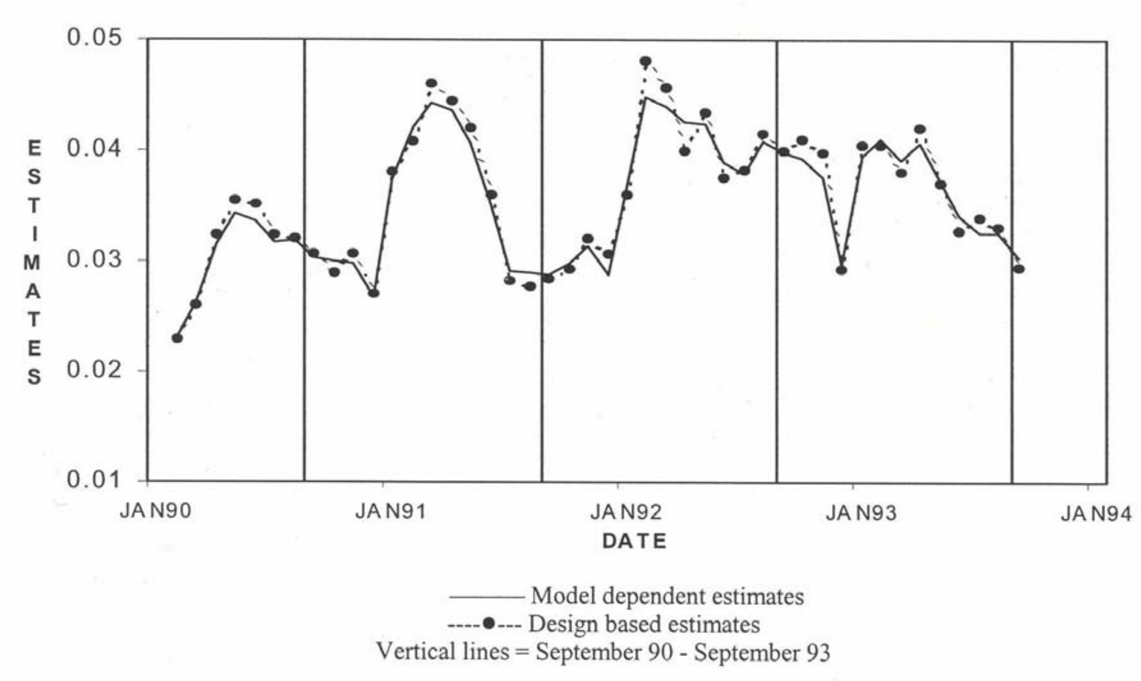

Figure 2. Brazilian Labour Force Series – SÃO PAULO design based and model dependent estimates proportion of unemployed persons.

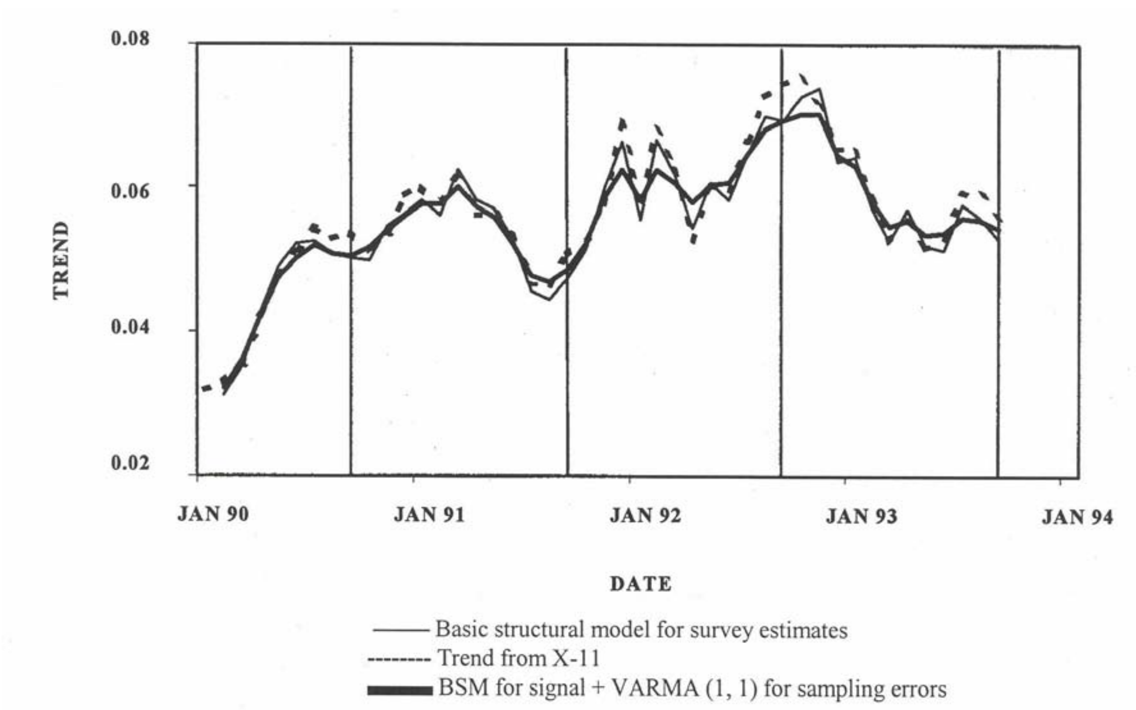

Figure 3. Brazilian Labour Force Series – SÃO PAULO trend estimates for the unemployed rates series.

6. Conlcusions

This paper proposes a state space approach for modelling compositional time series from repeated surveys. The important feature of the proposed methodology is that it provides bounded predictions and signal estimates of the parameters in a composition, satisfying the unity-sum constraint, while taking into account the sampling errors. This is accomplished by mapping the compositions from the Simplex onto Real space using the additive logratio transformation, modelling the transformed data employing multivariate state space models, and then applying the additive logistic transformation to obtain estimates in the original scale.

The empirical work using data from the Brazilian Labour Force Survey demonstrates the usefulness of this modelling procedure in a genuine survey situation, showing that it is possible to model the multivariate system and obtain estimates for all the relevant components. The results of the empirical work also show that smoother trends and fixed seasonals are obtained from a model which explicitly accounts for the sampling errors, when compared with estimates produced by X-11. In addition, because the model-dependent estimators combine past and current survey data, the standard deviations of these estimates are in general lower that the standard deviations of these estimates are in general lower that the standard deviations of the design-based estimators, as shown in Silva (1996, Chapter 8).

One drawback of the proposed procedure is that although confidence regions for the original compositional vector can be constructed based on the model-dependent estimates by using the additive logistic normal distribution, confidence intervals for the individual proportions are not readily available. Such intervals could be obtained from marginal distributions of the additive logistic normal distribution, but these can only be evaluated by integrating out some of the elements of the compositional vector and, as pointed out by Brundson (1987, page 135), this produces intractable expressions.

Under a state space formulation a wide variety of models is available to represent the multivariate signal and noise processes, which is a great benefit of this modelling procedure. The application of the method to different data sets is recommended. Further empirical research should also consider situations where the composition lies on a Simplex with dimensions higher than two and/or with compositions evolving close to the boundaries of the interval [0.1]. In addition, a better insight into the performance of the modelling procedure may be gained by applying the method to simulated data, for which the "true" underlying models are known. The models considered here can also be extended to incorporate rotation group bias effects and explanatory variables.

Acknowledgements

This research was supported by CAPES-Brazil and IBGE-Brazil and by a grant from the Economic and Social Research Council of the UK under its Analysis of Large and Complex Datasets Programme. The authors wish to thank the referees and Prof. Danny Pfeffermann for the suggestions that let to many improvements in the paper. Thanks are also due to Dr. Harold Mantel for his encouragement towards the preparation of the final version.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. New York: Chapman and Hall.

Aitchison, J., and Shen, S.M. (1980). Logistic-Normal distributions: Some properties and uses. *Biometrika*, 67, 261-272.

Bell, W.R., and Hillmer, S.C. (1990). The time series approach to estimation for repeated surveys. *Survey Methodology*, 16, 195-215.

Binder, D.A., and Hidiroglou, M.A. (1988). Sampling in time. In *Handbook of Statistics*, (Eds., P.R. Krishnaiah and C.R. Rao). Elsevier Science, 6, 187-211.

Binder, D.A., and Dick, J.P. (1989). Modelling and estimation for repeated surveys. *Survey Methdodology*, 15, 29-45.

Binder, D.A., Bleuer, S.R. and Dick, J.P. (1993). Time series methods applied to survey data. *Proceedings of the 49º International Statistical Institute Session*, 1, 327-344.

Blight, B.J.N., and Scott, A.J. (1973). A stochastic model for repeated surveys. *Journal of the Royal Statistical Society*, B, 35, 61-68,

Brunsdon, T.M. (1987). Time Series Analysis of Compositional Data. Unpublished Ph.D. Thesis. University of Southampton.

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

De Jong, P. (1988). The likelihood for a state space model. *Biometrika*, 75, 165-169.

De Jong, P. (1989). Smoothing and interpolation with the state space model. *Journal of the American Statistical Society*, 84, 1085-1088.

De Jong, P. (1991). The diffuse Kalman filter. *The Annals of Statistics*, 19, 1073-1083.

Fernandez, F.J.M., and Harvey, A.C. (1990). Seemingly unrelated time series equations and a test for homogeneity. *Journal of Business and Economic Statistics*, 8, 1, 71-81.

Gurney, M., and Daly, J.F. (1965). A multivariate approach to estimation in periodic sample surveys. *Proceedings of the American Statistical Association*, *Social Statistics Section*, 242-257.

Harrison, P.J., and Stevens, C.F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society*, B, 38, 205-47.

Harvey, A.C. (1986). Analysis and generalisation of a multivariate exponential smoothing model. *Management Science*, 32, 374-380.

Harvey, A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press. Cambridge.

Harvey, A.C. (1993). *Time Series Models*. Second Edition. Harvester Wheatsheaf. London.

Harvey, A.C., and Chung, C. (2000). Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistical Society*, A, 163, Part 3, 303-339.

Harvey, A.C., and Peters, S. (1984). Estimation Procedures for Structural Time Series Models. London School of Economics. Mimeo.

Harvey, A.C., and Shephard, N. (1993). Structural time series models. In *Handbook of Statistics*. (Eds. S. Maddala, C.R. Rao and H.D. Vinod). Elsevier Science Publishers, 11, 261-302.

IBGE (1980). Methodologia da Pesquisa Mensal de Emprego 1980. Relatórios Metodológicos. Fundação Instituto Brasileiro de Geografia e Estatística. Rio de Janeiro.

Jones, R.G. (1980). Best linear unbiased estimators for repeated surveys. *Journal of the royal Statistical Society*, B, 42, 221-226.

Koopman, S.J., Harvey, A.C., Doornik, J.A. and Shephard, N. (1995). *STAMP 5.0 – Structural Time Series Analyser, Modeller and Predictor*. Chapman & Hall. London

Mittnik, S. (1991). Derivation of the unconditional state covariance matrix for exact-likelihood estimation of ARMA models. *Journal of Economic Dynamics and Control*, 15, 731-740.

Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business and Economic Statistics*, 9, 163-177.

Pfeffermann, D., Burck, L. and Ben-Tuvia, S. (1989). A time series model for estimating housing price indexes adjusted for changes in quality. *Proceedings of the international Symposium on Analysis of Data in Time*, 43-55.

Pfeffermann, D., and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methdoology*, 16, 217-237.

Pfeffermann, D., and Bleuer, S.R. (1993). Robust joint modelling of labour force series of small areas. *Survey Methodology*, 19, 149-164.

Pfeffermann, D., Bell, P. and Signorelli, D. (1996). Labour force trend estimation in small areas. *Proceedings of the Annual Research Conference, Bureau of the Census*, 407-431.

Pfeffermann, D., Feder, M. and Signorelli, D. (1998). Estimation of autocorrelations of survey errors with applications to trend estimation in small samples. *Journal of Business and Economics Statistics*, 16, 339-348.

Quintana, J.M., and West, M. (1988). Time series analysis of compositional data. *Journal of Bayesian Statistics*, (Eds. J.H. Bernado, M.A. Degroot and A.F.M. Smith). Oxford University Press, 3.

Reinsel, G.C. (1993). *Elements of Multivariate Time Series Analysis*. Springer-Verlag.

SAS INSTITUTE INC. (1995). *SAS/IML Software*: *Changes and Enhancements through Release 6.11.* SAS Institute Inc. Cary, NC.

Scott, A.J., and Smith, T.M.F. (1974). Analysis of repeated surveys using time series methods to the analysis of repeated surveys. *International Statistical Review*, 45, 13-28.

Scott, A.J., Smith, T.M.F. and Jones, R.G. (1977). The application of time series methods to the analysis of repeated surveys. *International Statistical Review*, 45, 13-28.

Shephard, N.G., and Harvey, A.C. (1989). Tracking the Level of Support for the Parties During General Election Campaigns. Mimeo. Dept. of Statistics, London School of Economics.

Silva, D.B.N. (1996). Modelling compositional Time Series From Repeated Surveys. Unpublished Ph.D. Thesis. University of Southamptom. UK.

Singh, A.C., and Roberts, G.R. (1992). State space modelling of cross-classified time series of counts. *International Statistical Review*, 60, 321-335.

Smith, T.M.F., and Brunsdon, T.M. (1986). Time Series Methods for Small Areas. Unpublished Report. University of Southampton.

Tiao, G.C., and Box, G.E.P. (1981). Modelling multiple time series with applications. *Journal of the American Statistical Association*.

Tiller, R.B. (1989). A Kalman filter approach to labor force estimation using survey data. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 16-25.

Tiller, R.B. (1992). Time series modelling of sample data from the U.S. Current Population Survey. *Journal of Official Statistics*, 8, 2, 149-166.

Wallis, F. (1987). Time series analysis of bounded economic variables. *Journal of Time Series Analysis*, 8, 115-123.

Wei, W.W.S. (1993). *Time Series Analysis – univariate and multivariate methods*. Addison-Wesley.

West, M., and Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag.