International Journal of Forecasting 33 (2017) 502–512

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting compositional time series: A state space approach

Ralph D. Snyder [a], J. Keith Ord [b,*], Anne B. Koehler [c], Keith R. McLaren [a], Adrian N. Beaumont [d]

[a] Monash Business School, Monash University, Clayton, Victoria 3800, Australia
[b] McDonough School of Business, Georgetown University, Washington, DC 20057, USA
[c] Farmer School of Business, Miami University, Oxford, OH 45056, USA
[d] School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria 3010, Australia

ARTICLE INFO

Keywords:
Log ratio transformation
Market shares
Maximum likelihood estimation
Model invariance
Multi-series models
New products
Prediction distributions
US automobiles sales
Vector exponential smoothing

ABSTRACT

A framework for the forecasting of composite time series, such as market shares, is proposed. Based on Gaussian multi-series innovations state space models, it relies on the log-ratio function to transform the observed shares (proportions) onto the real line. The models possess an unrestricted covariance matrix, but also have certain structural elements that are common to all series, which is proved to be both necessary and sufficient to ensure that the predictions of shares are invariant to the choice of base series. The framework includes a computationally efficient maximum likelihood approach to estimation, relying on exponential smoothing methods, which can be adapted to handle series that start late or finish early (new or withdrawn products). Simulated joint prediction distributions provide approximations to the required prediction distributions of individual shares and the associated quantities of interest. The approach is illustrated on US automobile market share data for the period 1961–2013.

1. Introduction

The need for forecasts of proportions arises in a wide variety of areas, such as business, economics, demography, and political science. Specific examples include the market shares of competing products, the proportions of jobs in different sectors of the economy, and the age composition of a population. In some cases, only measurements of the proportions are available; in others, such as market share data, both total sales and proportions are available, although the analysis typically focuses on the latter. Statistical methods for the analysis of data on proportions are known as compositional time series methods, and one essential component is some form of transformation to ensure that the specified random variables are non-negative and defined on a simplex so that they sum to one.

The monograph by Aitchison (1986) is a key reference for compositional data analysis. His analysis shows that we can draw on a wide variety of established and well-understood statistical methods by using the log-ratio transformation to map the proportions onto the real line. This transformation has become relatively standard over the years for both cross-sectional and time series analysis (Aitchison & Egozcue, 2005; Brundson & Smith, 1998; Quintana & West, 1988), and is adopted in our paper. A

* Corresponding author.
E-mail addresses: ralph.snyder@monash.edu (R.D. Snyder), Keith.Ord@georgetown.edu (J.K. Ord), abkoehler@gmail.com (A.B. Koehler), keith.mclaren@monash.edu (K.R. McLaren), abeaumont@unimelb.edu.au (A.N. Beaumont).

http://dx.doi.org/10.1016/j.ijforecast.2016.11.008
0169-2070/

key advantage of the transformation is that it enables us to develop the analysis on the whole real line, allowing parameter estimation to be based on least squares (minimum generalized variance) arguments. Of course, the construction of prediction intervals requires explicit distributional assumptions, with the multivariate normal distribution being the most common choice.

Once the log-ratio transformation has been applied, other forms of analysis become feasible, such as functional data analysis (cf. Hyndman & Booth, 2008), which enables the consideration of more general time-dependent mean and variance structures. This paper does not consider this possibility or the other approaches in the literature, based on the Dirichlet distribution (Grunwald, Raftery, & Guttorp, 1993) or the hyper-spherical transformation (Mills, 2010).

The novel feature of this paper is its coupling of the log-ratio transformation with linear innovations state space models and the associated technique of vector exponential smoothing (De Silva, Hyndman, & Snyder, 2009, 2010; Hyndman, Koehler, Ord, & Snyder, 2008). The rationale for this shift in emphasis lies in the nature of market share data, where new brands may be launched or old brands fade away. State space models can be structured easily to allow for the intrinsically non-stationary nature of a start-up, so that varying numbers of series ("births" and "deaths") may be considered. In contrast, vector ARIMA models (Barceló-Vidal, Aguilar, & Martín-Fernández, 2011; Brundson & Smith, 1998) assume that series are stationary after suitable differencing, and can notionally extend to the "infinite past". These authors do not consider births or deaths.

One issue with log-ratio time series models when there are three or more brands lies in the choice of the base series. When there are no restrictions on the parameter space, the results are not invariant to the choice of this base. One of the contributions of this paper is the identification of necessary and sufficient conditions for the forecasts to be invariant to the choice of base.

The emphasis in this paper is on pure time series models. Extensions that include explanatory variables are clearly possible, although as Brodie, Danaher, Kumar, and Leeflang (2001) observe, "causal models should only be used when input variables can be forecast with reasonable accuracy". In the absence of such information, they note that the naïve random walk model works well over short horizons, a scheme that one would hope could be bested by the systematic development of time series models. It should also be emphasized that we are not advocating the exclusive use of time series models for the analysis of market shares, but recognize that a combination of methods often adds value, as was demonstrated by Kumar, Nagpalb, and Venkatesan (2002).

The paper is structured as follows. The data on (grouped) company shares of the US automobile market, which serves as the example for our empirical work, are described in the next section. The basic model and the transformations that are used to ensure that the non-negativity and adding up constraints are satisfied are introduced in Section 3. Prediction distributions are developed in Section 4. Estimation procedures, with extensions to allow for series of unequal lengths due to births and deaths, are considered in Section 5, along with model selection issues and the construction of prediction intervals. A detailed analysis of market shares in the US automobile market is undertaken in Section 6, including two cases of new entrants into the overall market. Conclusions with discussion are provided in Section 7. The necessary and sufficient constraints on the parameter space that are required for base series invariance are derived in the Appendix.

2. Description of the data

The annual log-ratios of US total vehicle sales market share data (http://wardsauto.com/data-center), shown in Fig. 1, exhibit many of the issues considered in this paper and are used in Section 5 to illustrate our proposed approach. The original shares series related to 31 principal manufacturers who sell, or used to sell, in the American market. For the purpose of discussion, we consolidated them into six series as follows:

American: GM, Ford, Chrysler [3 series]

Japanese: Honda, Isuzu, Mazda, Mitsubishi, Nissan, Subaru, Suzuki and Toyota [1 series]

Korean: Hyundai and Kia [1 series]

Other (principally German): BMW, Daimler, Volkswagen and Other [1 series].

We selected these groupings with several factors in mind: they should be reasonably homogeneous in terms of market appeal; they should be small enough in number that they can be discerned in graphs; and they should not hide the entry of new products (Japan and Korea). It should be noted that any analysis that is based upon log-ratios will not be invariant under groupings of 'brands', but the effects will be minimized as long as we select reasonably homogeneous groups.

A fundamental issue concerning the use of log-ratio models needs to be addressed at this point. Hierarchical time series modeling (e.g., Hyndman, Ahmed, Athanasopoulos, & Shang, 2011) relies upon linear structures for combining elements (e.g., combine different options to define the sales of a particular model of automobile, then aggregate across all models of car produced by the manufacturer, then possibly across cars and trucks, and so on). In principle, the granularity of the data can be made as fine or as coarse as the analysis requires. However, the level of aggregation cannot be changed once the log-ratio transformation has been applied. Thus, the hierarchical and compositional approaches are complementary, in that hierarchical methods can be used to forecast component series within a given level of aggregation, and compositional methods are then used to compare entities at that level of aggregation.

The non-stationary nature of the transformed series is evident from Fig. 1. The market share for GM has declined steadily, meaning that the relative market shares of the other manufacturing groups have increased. The Japanese and Korean manufacturers have increased their shares relative to GM. The series for Chrysler chronicles the relative decline and recovery of that company in the seventies and eighties, whereas the "Other" series (mostly German manufacturers) reflects a decline followed by a more recent resurgence.

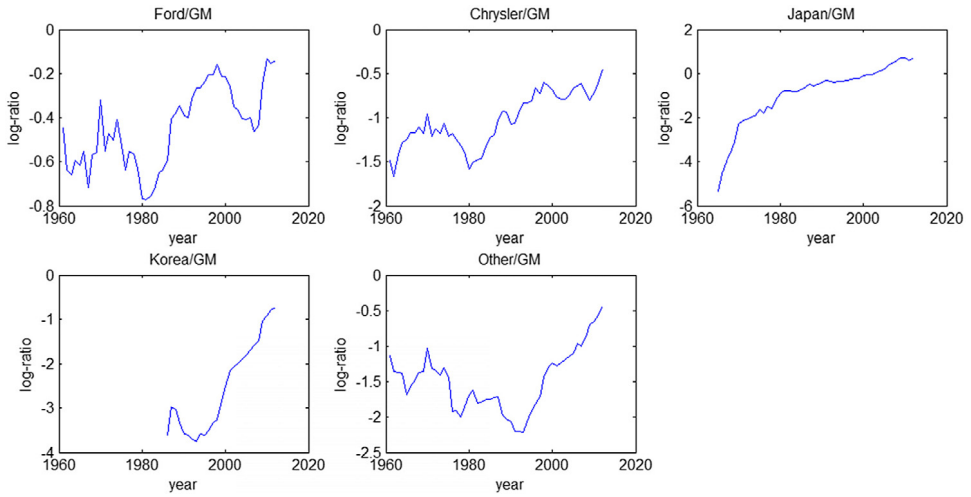

Fig. 1. Plots of log-ratios of percentage market shares for annual automobile sales, 1961–2013, using GM as the base series for each case.
Source: Ward's Auto Group.

As is to be expected, the series show a strong serial dependence, meaning that the random walk model is a viable choice, as was noted by Brodie et al. (2001). However, things are not as straightforward as they appear, as the total shares must still add to one and the prediction intervals must remain in the interval (0, 1). A binary logistic "Us versus Them" model (Brodie et al., 2001) might be considered appropriate for an individual company, but it fails to recognize movements among competitors that could have major consequences. Thus, we seek an approach that will handle all of: constrained total shares equal to one, non-negative shares, strong serial dependence, non-stationarity, non-uniform interactions among the series, and new entries. The grouping that we performed ensured that we are not confronted by missing values or near-zeros in this example, but we address those questions briefly as well.

3. Models and transformations

3.1. Data transformations

It is initially envisaged that there are $r+1$ time series, all of equal length n. An adequate representation of the shares data involves both a series index i and a time subscript t. The $r+1$ shares at time t are denoted by $\{z_{it};\ i = 0, 1, \ldots, r;\ t = 1, 2, \ldots, n\}$, where $z_{it} \geq 0$ and $\sum_i z_{it} = 1$. However, it is convenient to drop the t subscript temporarily when considering the transformations.

One transformation is needed to convert the shares vector $\mathbf{z}$ of dimension $r+1$ into the unbounded vector $\mathbf{y} \in \mathbb{R}^r$ in order to ensure that that the constraint on the sum is preserved without creating a singular distribution. The preferred transformation is the log-ratio (Aitchison, 1986) $y_i = \ln(z_i/z_0)$ for $i = 1, \ldots, r$, where z_0 is the base series. For the moment, it is assumed that none of the series contain zeros. As was noted by Brundson and Smith (1998), the subsequent analysis must be invariant to the choice of base series.

The inverse of the log-ratio transformation is the logistic function

$$z_i = \begin{cases} \dfrac{\exp(y_i)}{1 + \sum_{j=1}^{r} \exp(y_j)} & i = 1, \ldots, r \\ z_0 = 1 - \sum_{j=1}^{r} z_j & i = 0. \end{cases} \tag{1}$$

The log-ratio transformation is applied in each period in order to obtain the unbounded r-vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ in what will be termed the log-ratio space. The resulting series generally retain many of the features of the original series, such as non-stationarity and serial dependence.

3.2. Innovations state space models for time series with common structures

Vector innovations state space models are used to represent the evolution of the transformed non-stationary time series (it is assumed for the moment that all series have common start and end dates). In these models, the series vector $\mathbf{y}_t$ in typical period t is related linearly to a random matrix $\mathbf{X}_{t-1} \in \mathbb{R}^{k\times r}$ of state variables, representing the latent states from the end of the previous period. This relationship is given by the row vector of measurement equations

$$\mathbf{y}_t' = \mathbf{w}'\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t', \tag{2}$$

where $\mathbf{w} \in \mathbb{R}^k$ is a fixed vector and the $\boldsymbol{\varepsilon}_t \in \mathbb{R}^r$ are serially independent $N(\mathbf{0}, \mathbf{V})$ innovations vectors. The variance matrix $\mathbf{V}$ has non-zero covariances, reflecting contemporaneous dependencies. The evolution of the states is governed by the recurrence relationship

$$\mathbf{X}_t = \mathbf{F}\mathbf{X}_{t-1} + \mathbf{g}\boldsymbol{\varepsilon}_t', \tag{3}$$

where $\mathbf{F} \in \mathbb{R}^{k\times k}$ and $\mathbf{g} \in \mathbb{R}^k$ are also fixed. The structural vectors $\mathbf{w}$ and $\mathbf{g}$ and the matrix $\mathbf{F}$ may depend on a vector

Table 1
The local level and local trend state space models and their reduced forms.

Model	Specification	Reduced form
Local level (LLM)	$\mathbf{y}_t = \boldsymbol{\ell}_{t-1} + \boldsymbol{\varepsilon}_t$ $\boldsymbol{\ell}_t = \boldsymbol{\ell}_{t-1} + \alpha\boldsymbol{\varepsilon}_t$	$\nabla\mathbf{y}_t = -(1-\alpha)\,\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t$
Local trend (LTM)	$\mathbf{y}_t = \boldsymbol{\ell}_{t-1} + \mathbf{b}_{t-1} + \boldsymbol{\varepsilon}_t$ $\boldsymbol{\ell}_t = \boldsymbol{\ell}_{t-1} + \mathbf{b}_{t-1} + \alpha\boldsymbol{\varepsilon}_t$ $\mathbf{b}_t = \mathbf{b}_{t-1} + \beta\boldsymbol{\varepsilon}_t$	$\nabla^2\mathbf{y}_t = -(\alpha-1)\,\boldsymbol{\varepsilon}_{t-2} - (2-\alpha-\beta)\,\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t$

$\boldsymbol{\theta} \in \mathbb{R}^p$ of common parameters. The $r(r+1)/2$ distinct elements of $\mathbf{V}$ are additional parameters.

Such a model can always be reduced to the *form* of a vector ARIMA model, provided that we keep in mind the finite start-up times of some or all of the series, contrary to the usual ARIMA assumptions. The details of this reduction for the univariate case are given by Hyndman et al. (2008, Chapter 11).

The multivariate model specified in Eqs. (2) and (3) consists of a univariate model for each log-ratio series. These univariate sub-models would be independent except that it is assumed that they share common structural features through $\{\mathbf{w}, \mathbf{F}, \mathbf{g}\}$, and have correlated innovations.

It is critical that the models be structured so that the results do not depend upon the choice of the base series. We call this the *fundamental requirement of log-ratio models*. It is established in the Appendix that the shared (common) structure $\{\mathbf{w}, \mathbf{F}, \mathbf{g}\}$ is necessary and sufficient for the maximum likelihood (or generalized variance) estimators of the parameters to be invariant to a change in the base series used with the log-ratio transformation. Moreover, it ensures that pairwise comparisons of series can be made independently of other series, including the base series. Thus, the comparison between two series indexed by i and j, exploiting the common $\mathbf{w}$ in the measurement equation (Eq. (2)), is governed by the relationship $z_{it}/z_{jt} = \exp\left(\mathbf{w}'\left(\mathbf{x}_{i,t-1} - \mathbf{x}_{j,t-1}\right) + \left(\varepsilon_{it} - \varepsilon_{jt}\right)\right)$, where $\mathbf{x}_{i,t-1}$ and $\mathbf{x}_{j,t-1}$ are columns i and j of the $\mathbf{X}_{t-1}$ matrix. Moreover, it can be inferred from Eq. (3) that $\mathbf{x}_{it} - \mathbf{x}_{jt} = \mathbf{F}\left(\mathbf{x}_{i,t-1} - \mathbf{x}_{j,t-1}\right) + \mathbf{g}\left(\varepsilon_{it} - \varepsilon_{jt}\right)$. Only the states and errors for the two series enter these relationships.

3.3. Special cases of the general model

Several special cases of the general model are now considered, in order to provide a sufficiently rich set of options for our market share analysis. We retain the assumption $\boldsymbol{\varepsilon}_t \sim \text{NID}(0, \mathbf{V})$ in all cases, where $\mathbf{V}$ is unrestricted for the stochastic error terms. Conceptually, the simplest starting point in any multivariate time series analysis is the global means model $\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\mu} \in \mathbb{R}^r$ is a vector of constant means. It allows for neither serial dependence nor non-stationarity, but is relevant insofar as it illustrates a case where the starting values (the means) are of critical importance and must be estimated efficiently. We do not consider this model explicitly in our empirical analysis. At another extreme is the multivariate random walk $\mathbf{y}_t = \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$ (referred to by Brodie et al., 2001 as the naïve model). It is of interest because it is the simplest model that allows for both serial dependence and non-stationarity.

Table 2
Parameter restrictions for the local level and local trend models.

Model	Traditional	Invertibility/forecastability
Local level, LLM	$0 \le \alpha \le 1$	$0 \le \alpha \le 2$
Local trend, LTM	$0 \le \alpha \le 1$ $0 \le \beta \le \alpha$	$2\alpha + \beta \le 4$ $\alpha \ge 0,\ \beta \ge 0$

The models used in this paper, called the vector local level model (LLM) and the vector local trend model (LTM), are shown in Table 1, together with their VARIMA analogues. They include the elementary global means and random walk models, but provide a much richer framework for the analysis of multivariate time series. Useful background on the univariate versions of these models is provided by Hyndman et al. (2008, Chapter 3). Ignoring the elements of $\mathbf{V}$, the common structure required for invariance to the choice of the base series means that the vector LLM and the vector LTM have just one and two structural parameters respectively, reducing the proliferation of parameters that normally accompanies multivariate models. However, the variance matrix is unrestricted. In addition, there are also the starting values (k for LLM and $2k$ for LTM).

The following notation is employed in Table 1: $\boldsymbol{\ell}_t \in \mathbb{R}^k$ for a vector of levels; $\mathbf{b}_t \in \mathbb{R}^k$ for a vector of underlying growth rates; α, β for the common parameters; and ∇ for the difference operator.

Various restrictions must be imposed on the parameters in order to ensure forecastability (Hyndman et al., 2008, pp. 152–155). For the models that we consider, forecastability corresponds to invertibility together with those situations where some of the smoothing parameters may be zero, but the traditional constraints that are applied in exponential smoothing studies may also be used. Both sets of constraints are shown in Table 2, but the traditional constraints are more restrictive in these cases (though not in general), and the choice of constraint set continues to be debated actively; see for example Hyndman et al. (2008, Chapter 10).

These particular models can be expressed in terms of the general notation established in the previous section, with the formats being summarized in Table 3. The constant means model is a limiting case of the LLM ($\alpha = 0$) and will rarely be of practical interest, but is worth noting because the initial conditions determine the forecasts for all future time periods. The random walk is also a special case of the LLM (when $\alpha = 1$), and is again a boundary case when the traditional constraints are applied, but an interior solution for the invertibility constraints. Reflecting what is now common practice in forecasting, the random walk will be used for benchmarking purposes. A special case of LTM may be termed the *local momentum model*

Table 3
The matrices and vectors associated with particular models.

Model	$\mathbf{x}_t$	$\mathbf{w}'$	$\mathbf{F}$	$\mathbf{g}$	$\boldsymbol{\theta}$
Constant means	$\boldsymbol{\mu}$	1	1	0	0
Random walk	$\boldsymbol{\ell}_t$	1	1	1	1
Local level	$\boldsymbol{\ell}_t$	1	1	α	α
Local trend	$\begin{bmatrix}\boldsymbol{\ell}_t' \\ \mathbf{b}_t'\end{bmatrix}$	$\begin{bmatrix}1 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}\alpha \\ \beta\end{bmatrix}$	$\begin{bmatrix}\alpha \\ \beta\end{bmatrix}$
Local momentum	$\begin{bmatrix}\boldsymbol{\ell}_t' \\ \mathbf{b}_t'\end{bmatrix}$	$\begin{bmatrix}1 & 1\end{bmatrix}$	$\begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}$	$\begin{bmatrix}1 \\ \beta\end{bmatrix}$	$\begin{bmatrix}1 \\ \beta\end{bmatrix}$

(LMM), where the rate of change has a local level structure, or $\nabla \mathbf{y}_t = \nabla \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - (1-\beta)\boldsymbol{\varepsilon}_{t-1}$. It is obtained when $\alpha = 1$, but its interpretation is of particular interest because it corresponds to a local level model for the second derivative, or acceleration.

The structures of each of these models are summarized in Table 3. We do not consider the constant means model in the empirical section, but the other four models are all estimated and compared.

4. Prediction

Prediction distributions form the primary focus of this section because other forms of potential interest, such as point predictions and prediction intervals, can then be derived from them once they are known. In the log-ratio space, they are multivariate normal distributions. By applying the vec operator to the state space equations (Eqs. (2) and (3)), it can be established that their means and variances can be calculated recursively for future periods $t = n+1, \ldots, n+h$ using the relationships:

$$\boldsymbol{\mu}^y_{t|n} = \mathbb{W}\boldsymbol{\mu}^x_{t-1|n}$$
$$\boldsymbol{\mu}^x_{t|n} = \mathbb{F}\boldsymbol{\mu}^x_{t-1|n}$$
$$\mathbf{V}^y_{t|n} = \mathbb{W}\mathbf{V}^x_{t-1|n}\mathbb{W}' + \mathbf{V} \quad \text{and}$$
$$\mathbf{V}^x_{t|n} = \mathbb{F}\mathbf{V}^x_{t-1|n}\mathbb{F}' + \mathbb{G}\mathbf{V}\mathbb{G}',$$

where

1. $\boldsymbol{\mu}^x_{t|n}$ and $\mathbf{V}^x_{t|n}$ are the conditional mean and variance of the vectorised states $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$;
2. $\boldsymbol{\mu}^y_{t|n}$ and $\mathbf{V}^y_{t|n}$ are the conditional mean and variance of the log-ratios vector $\mathbf{y}_t$;
3. $\mathbb{W} = \mathbf{I} \otimes \mathbf{w}'$, $\mathbb{F} = \mathbf{I} \otimes \mathbf{F}$, and $\mathbb{G} = \mathbf{I} \otimes \mathbf{g}$, where $\mathbf{I} \in \mathbb{R}^{r \times r}$ is an identity matrix and $\otimes$ is the Kronecker product operator; and
4. $\mathbf{V}^x_{n|n}$ is approximated by the zero matrix $\mathbf{0} \in \mathbb{R}^{r \times r}$.

Despite the prediction distributions of market shares being related to the multivariate normal distributions in log-ratio space, explicit formulas for the prediction intervals of individual shares are not available when there are three or more series. However, these intervals can be approximated from simulated samples, with the method for typical period $t > n$ being:

1. Randomly generate P vectors from the log-ratio space prediction distribution $N\left(\boldsymbol{\mu}^y_{t|n}, \mathbf{V}^y_{t|n}\right)$, apply the logistic transformation (Eq. (1)) to each of them, and then concatenate the resulting column vectors of shares to form a matrix $\mathbf{Z}_t \in \mathbb{R}^{r \times P}$.
2. Each row of $\mathbf{Z}_t$ is a simulated sample of size P from a marginal distribution of the shares for a particular market segment. Pertinent statistics can be calculated as follows:
   a. Point prediction with the row average.
   b. A β-percent prediction interval with the $\beta/2$ and $100 - \beta/2$ row percentiles.
   c. The proportion of row values that are in excess of the current market share as an estimate of the probability of an increase in the market share.

An exception to the need for simulated prediction intervals involves the chi-square statistic in the log-ratio space. The assumption of multivariate normality provides a prediction region by using the chi-square distribution to produce a one-sided 90% interval $Q_t < \chi^2_{0.90}(r_t)$; see Section 5.6. Unfortunately, the quadratic form cannot be unscrambled to produce intervals for the individual market shares for $r > 2$.

5. Estimation

Looking at the local trend model as the most general case under consideration, the r series could involve as many as $rk + p + r(r+1)/2$ parameters, where the three terms correspond to the counts of the starting values, the smoothing parameters and the variance matrix elements respectively. We consider two different maximum likelihood-based estimation procedures in the next two subsections, illustrated in terms of the LTM.

5.1. A vector ARIMA approach

The problem in this framework is the estimation of the parameters (α, β) and the $r(r+1)/2$ potentially distinct elements of the variance matrix $\mathbf{V}$. One advantage of the ARIMA approach is that the unobservable random state vectors appear to be absent and need not be estimated. However, a direct specification of the unconditional likelihood is not possible because of the existence of the two unit roots (differences). Instead, it is necessary to base the likelihood function on the density $p(\mathbf{y}_3, \ldots, \mathbf{y}_n | \mathbf{y}_1, \mathbf{y}_2; \alpha, \beta, \mathbf{V})$, which is conditioned on the first two observations. Equivalently, this function may be represented as the density $p\left(\nabla^2\mathbf{y}_3, \ldots, \nabla^2\mathbf{y}_n; \alpha, \beta, \mathbf{V}\right)$ of second differences. Thus, the effective sample size for each transformed series is $n-2$.

The evaluation can be simplified by using the prediction decomposition of this density $\prod_{t=3}^{n} p\left(\nabla^2\mathbf{y}_t | \nabla^2\mathbf{y}_3, \ldots, \nabla^2\mathbf{y}_{t-1}; \alpha, \beta, \mathbf{V}\right)$. Options for doing this include the

Kalman filter. However, each pass of the Kalman filter requires the specification of trial values of α, β and the $r\,(r+1)/2$ elements of $\mathbf{V}$. By way of example, a search that is designed to find the maximum likelihood estimates would need to iterate over a $(2+r\,(r+1)/2)$-dimensional parameter space. In the context of our empirical example in Section 6, $r = 5$; therefore, there are 17 dimensions!

5.2. The innovations state space approach

Consider the general innovations state space model again. Given a value for the seed state matrix $\mathbf{X}_0$, the errors can be calculated recursively using the general exponential smoothing equations (in the notation of Section 3.2):

$$\mathbf{e}_t' = \mathbf{y}_t' - \mathbf{w}'\mathbf{X}_{t-1} \tag{4}$$

and

$$\mathbf{X}_t = \mathbf{F}\mathbf{X}_{t-1} + \mathbf{g}\mathbf{e}_t'.$$

The $\mathbf{X}_t$ and $\mathbf{e}_t$ that emerge for typical period t are conditioned on $\mathbf{X}_0$ and $\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}$. The focus is on the density $p\,(\mathbf{y}_1, \ldots, \mathbf{y}_n|\mathbf{X}_0; \boldsymbol{\theta}, \mathbf{V})$ of the *entire* sample, but the existence of unit roots means that, as before, we must condition on something, in this case the seed state. Exponential smoothing in this context is the analogue of the Kalman filter in the previous subsection: it allows us to create a prediction decomposition of the likelihood function, albeit one that is conditioned on the seed state $\mathbf{X}_0$. This decomposition $\prod_{t=1}^{n} p\,(\mathbf{y}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}, \mathbf{X}_0; \boldsymbol{\theta}, \mathbf{V})$ in turn can be written in terms of the one-step-ahead prediction errors $\prod_{t=1}^{n} p\,(\mathbf{e}_t|\mathbf{X}_0; \boldsymbol{\theta}, \mathbf{V})$. Unlike the sample values $\mathbf{y}_1$ and $\mathbf{y}_2$ in the previous subsection, $\mathbf{X}_0$ must be observed indirectly by estimation.

However, we may produce a concentrated version of the likelihood as follows. Given a trial value for $\mathbf{X}_0$, the log of the conditional likelihood can be written as

$$\mathfrak{L}\,(\boldsymbol{\theta}, \mathbf{V}, \mathbf{X}_0) = -\frac{nr}{2}\ln(2\pi) - \frac{n}{2}\ln|\mathbf{V}| - \frac{1}{2}\sum_{t=1}^{n}\mathbf{e}'\mathbf{V}^{-1}\mathbf{e}.$$

For given values of $(\boldsymbol{\theta}, \mathbf{X}_0)$, the maximum likelihood estimator for $\mathbf{V}$ is

$$\hat{\mathbf{V}}\,(\boldsymbol{\theta}, \mathbf{X}_0) = \sum_{t=1}^{n}\mathbf{e}_t\mathbf{e}_t'/n. \tag{5}$$

Substituting this into the log-likelihood function gives its concentrated counterpart:

$$\mathfrak{L}\,(\mathbf{X}_0, \boldsymbol{\theta}) = -\frac{nr}{2}\ln(2\pi) - \frac{n}{2}\ln\left|\hat{\mathbf{V}}\right| - \frac{nr}{2}. \tag{6}$$

We maximize Eq. (6) so as to obtain the maximum likelihood estimates of the seed state matrix $\mathbf{X}_0$ and the parameter vector $\boldsymbol{\theta}$. This also has the effect of minimizing the estimated generalized variance $\left|\hat{\mathbf{V}}\right|$.

The benefit of this approach to maximum likelihood estimation is that it reduces the dimensionality of the search space from $p + r\,(r+1)/2$ to $rk + p$. In our case study with a local trend, $r = 5$, $k = 2$ and $p = 2$, so that the search dimension drops from 17 to 12. In general, the advantage of the exponential smoothing approach over the ARIMA approach with a Kalman filter is much more pronounced for data sets with larger numbers of series, because the dimensionality is linear rather than quadratic in the number of series.

Finally, for longer series, we observe that a simple heuristic may be used to obtain good approximations of the starting values, thereby reducing the dimensionality of the parameter space further to only one or two. Such heuristics are often used in single-series applications (see Hyndman et al., 2008, pp. 71–73, for further discussion), but are unreliable when α and β are small. Our application in Section 6 uses maximum likelihood predictors of the seed states.

5.3. Accounting for different series lengths

It was noted earlier that there is a need to take into account all of the following features: a constrained total (to one), non-negativity, strong serial dependence, non-stationarity, non-uniform interactions among the series, and new entries. The state space formulation in Section 3 accounts for all of these factors, but does not address the issue of new entrants explicitly. The auto share time series for Japan and Korea are shorter than the others, and the model formulation must be adapted to account for this.

To ensure that the error terms for Japan and Korea make no contribution to the estimate in Eq. (5) of the variance matrix in the periods 1961–1964 and 1961–1985 respectively, the 'errors' for these run-in periods are forced to zero using the formula

$$e_{it} = \begin{cases} y_{it} - \mathbf{w}'\mathbf{x}_{i,t-1} & \text{if series } i \\ & \text{is observed in period } t \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Corresponding adjustments to the sample sizes are also required, so that the estimator for the typical element of $\mathbf{V}$ becomes

$$\hat{v}_{ij} = \sum_{t=1}^{n} e_{it}e_{jt}/\min\left\{n_i, n_j\right\}, \tag{8}$$

where n_i and n_j designate the sample sizes of series i and series j.

The differences in the series' start-up dates also imply that not all of the seed state vectors are coincident in time. In our example, GM, Ford, Chrysler and Others have 1960 as their origin, whereas the origins for Japan and Korea are 1964 and 1985 respectively. A general approach to this issue requires a framework which minimizes the extra effort required from a coding perspective. Fortunately, this difficulty may be resolved using the following approach, which is described in the context of the automobile sales series but is quite general in application.

Make 1960 the origin of all of the series, so that each has a seed state vector in this year. Those for Japan and Korea can be thought of as *artificial* seed states. General exponential smoothing is applied in the usual way between 1961 and 2003, except that Eq. (7) is now used to determine the errors instead of Eq. (4). This means that the seed states for Japan and Korea in 1964 and 1985 are related deterministically to their artificial counterparts in

1960. The optimization of the artificial seed states for 1960 then leads to appropriate optimised values for the 1964 and 1985 seed states.

The generalized variance, used as the fitting criterion in Section 5.2, also needs to be adapted so that only those series that are observed at time t contribute. To this end, let $\mathbf{D}_t \in \mathbb{R}^{r_t \times r}$ be a matrix that selects the r_t-vector $\tilde{\mathbf{y}}_t$ of *observed* series values from the vector $\mathbf{y}_t$ using the formula $\tilde{\mathbf{y}}_t = \mathbf{D}_t \mathbf{y}_t$. The structure of the matrix $\mathbf{D}_t$ can be seen most readily by ordering the series into the $(r - r_t)$ series that are not observed at time t, followed by the r_t series that are observed. Then $\mathbf{D}_t = [\mathbf{0}_{r_t \times (r-r_t)}, \mathbf{I}_{r_t \times r_t}]$ and the one-step ahead prediction errors are given by the r_t vector $\tilde{\mathbf{e}}_t = \mathbf{D}_t \mathbf{e}_t$, with the associated $r_t \times r_t$ variance matrix $\mathbf{V}_t = \mathbf{D}_t \mathbf{V} \mathbf{D}_t'$. The natural logarithm of the generalized variance can then be written as:

$$\hat{v} = \sum_{t=1}^{n} \ln \left| \hat{\mathbf{V}}_t \right| .$$

In the current context of series with unequal lengths, and for reasons similar to those outlined by Schmidt (1977), a conclusive link between the minimum generalized variance and maximum likelihood estimators has proven analytically intractable thus far. However, our investigation of this matter has provided strong hints that the two types of estimates are indeed identical, and, in the special case of series of equal lengths, as indicated in Section 5.2, the two types of estimates are the same for series of equal lengths.

5.4. Zero or near-zero values

The approach in Section 5.3 is designed explicitly to cover new entrants (and, by implication, can also be applied in reverse when a series drops permanently to zero). A different situation arises when a series dips temporarily below a small threshold but then recovers (e.g., non-supplies in a strike, or health-related issues depressing sales). In such circumstances, it is straightforward to apply a modified version of the solution proposed by Fry, Fry, and McLaren (2000), involving a small perturbation of the (near) zero values so as to avoid numerical difficulties with the log-ratio transformations. The procedure is as follows:

- At time period t, determine how many of the $r+1$ series have values that fall below a pre-specified threshold, say τ. Denote the number of such values by m_t.
- If $m_t = 0$, retain the observed values for that time period.
- If $m_t > 0$, define the adjusted values $z_{it}^* = \tau$ if $z_{it} \leq \tau$ and $z_{it}^* = (1 - \tau m_t) z_{it} / S_t$ if $z_{it} > \tau$ where $S_t = \sum_{z_{it} > \tau} z_{it}$.

Note that the sum of the adjusted values in each time period is still 1.0, and the ratios of the adjusted values for values that are not deemed to be near zero are unchanged. Clearly, this approach is ad-hoc, but it appeared to work well with other series that we considered where near-zero values were an issue.

5.5. Model selection

The Akaike information criterion (Akaike, 1974) is used for model comparisons and selection. Often, the small sample variant AICc is preferred; however, it is more model-specific than the AIC, and is currently unknown for innovations state space models. In our context, the AIC is defined as AIC# $= \hat{v} + 2\,(rk + p + 0.5r\,(r+1))$. This replaces the optimised likelihood in the traditional definition with the optimised generalized variance. The # is used to signal that it may not correspond to the AIC exactly in this particular circumstance, because the link between the optimised likelihood and optimised generalized variance, as noted in Section 5.3, has not yet been established for series of unequal lengths.

In the above formula for AIC#, the penalty includes rk for the number of seed states, p for the number of *free* smoothing parameters and $r(r+1)/2$ for the number of distinct elements in the variance matrix. As the number of elements in the variance matrix does not vary here, it has no impact on the model selection process in the present framework; however, such terms should be included in order to allow comparisons with other possible cases with restricted versions of the variance.

5.6. Model checking

The usual panoply of plots for residuals may be employed to check the validity of the selected model, but these analyses are generally univariate in nature. A multivariate check may be formulated by considering the joint distribution of $\tilde{\mathbf{y}}_t$ and letting $\hat{\tilde{\mathbf{y}}}_t$ denote the fitted value of $\tilde{\mathbf{y}}_t$. The quadratic form $Q_t = (\tilde{\mathbf{y}}_t - \hat{\tilde{\mathbf{y}}}_t)' \hat{\mathbf{V}}_t^{-1} (\tilde{\mathbf{y}}_t - \hat{\tilde{\mathbf{y}}}_t)$ is distributed asymptotically as $\chi^2\,(r_t)$ when the errors are multivariate normal, so that when the null hypothesis of a valid model holds, approximately $100(1-\alpha)\%$ of the vectors of residuals should satisfy the inequalities

$$Q_t < \chi^2_{1-\alpha}(r_t). \tag{9}$$

The degrees of freedom depend upon the number of series that are actually observed at time t.

6. Results for the automobile data

6.1. Estimation and model selection

The US automobile data described in Section 2 were analyzed using the methods developed in the previous sections. We considered four of the possibilities shown in Table 2, based on two models (local level and local trend) and the two constraint sets (traditional and invertibility). The final ten years of data were reserved as a holdout sample for prediction evaluation purposes.

The estimation results are summarized in Table 4. The models are parsimonious in having only one or two smoothing parameters; however, it should be kept in mind that this constraint is both necessary and sufficient if we are to obtain results that are invariant to the choice of base series, as demonstrated in the Appendix. The AICs indicate that the vector local trend models are likely

Table 4
Summary of the estimation results.

Models	Restrictions	Alternative description	Log generalized variance	AIC#	Alpha	Beta	Coverage[b]
RWM	None	Random walk	−745.94	−705.94			0.83
LLM	Traditional	SES[a]	−771.97	−731.97	1.00		0.81
LLM	Stability	SES[a]	−778.21	−736.21	1.20		0.84
LTM	Traditional	TES[a]	−799.95	−747.95	1.00	0.21	0.81
LTM	Stability	TES[a]	−799.97	−747.97	1.01	0.21	0.81

[a] SES and TES denote (multivariate) simple exponential smoothing and trend-corrected exponential smoothing, respectively.
[b] Coverage refers to nominal 90% one-sided intervals for the chi-square statistic.

to produce better forecasts than the vector local level models. It should be noted that the optimum LLM and LTM schemes with the 'traditional' restrictions produced the random walk and local momentum models respectively. Given the previous interest in the random walk model (cf. Brodie et al., 2001), its relatively poor performance in this example is noteworthy. The AIC results also indicate that the use of the invertibility restrictions on the parameters leads to better predictions (minimizing the AIC) than the traditional exponential smoothing restrictions.

The fitted local trend model with the traditional constraints has $\hat{\alpha} = 1$. A local trend model with $\alpha = 1$ can be rewritten as $\nabla \mathbf{y}_t = \mathbf{b}_{t-1} + \boldsymbol{\varepsilon}_t$, where $\mathbf{b}_t = \mathbf{b}_{t-1} + \beta \boldsymbol{\varepsilon}_t$. This model is of interest in its own right. In effect, the first differences follow a local level model. Accordingly, the predictions must be exponentially weighted averages of the differenced series. A first difference may be interpreted as the *actual* growth (in the transformed data) and $\mathbf{b}_t$ as the *structural* growth rate. Overall, the margin of difference for the two LTM versions is small, and the simpler interpretation of the local momentum model might override the slight statistical differences.

In all cases, large values of α mean that the prediction intervals widen substantially as the forecast horizon increases, but this is a function of the length of the time period used in the example. We could reasonably expect much smaller α values in conventional market share studies that look at monthly or even weekly data.

6.2. Model diagnostics

The coverages of the nominal 90% semi-infinite intervals (closed from above) for the chi-square error statistic in Eq. (9) are also shown in the final column of Table 4. The coverage is somewhat below the nominal level in all cases, but the plot of the Q-statistics for the local trend model in Fig. 2 suggests that this model "settled down" after the Korean entry in 1986, with the later results being in line with expectations.

6.3. Forecasts and prediction intervals for the best local trend model

The point predictions and 80% prediction intervals from the best fitting local trend model are illustrated in Fig. 3. The point predictions are simple averages of the simulated sample of proportions. The lower and upper limits of the prediction intervals are the 10th and 90th percentiles. The observed shares are also shown, and all lie well within the prediction intervals.

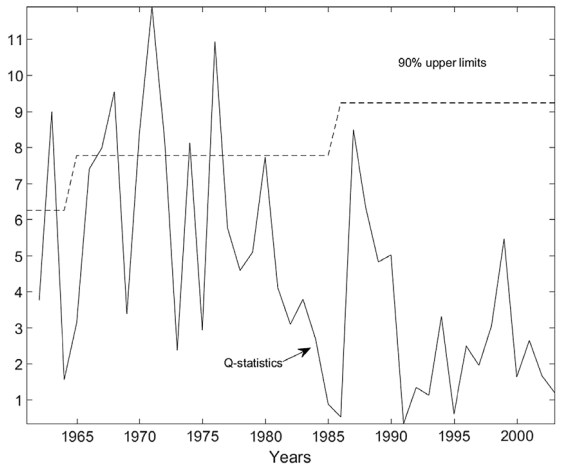

Fig. 2. Plot of Q-statistics and 90% upper limits for the local trend model: 1961–2003. The shifts in the confidence limits are due to the Japanese and Korean entries into the market.

The point forecasts are quite good but the 80% prediction intervals are too wide. This may be due to a decline in volatility in the later part of the series.

One question that is often asked is whether a particular manufacturer is likely to increase its share of the market in future years. Fig. 4 provides estimates of the probability of such an increase for each group in 2004.

7. Discussion

This paper has developed a state space approach for the forecasting of compositional time series that is invariant to the choice of the base series in the log-ratio transformation and satisfies the constraints that the predicted proportions must be non-negative and sum to one. The associated models may be extended to include seasonal patterns and explanatory variables, provided that common parameters are specified for the state variables, as is the case in the basic models defined in Table 1. The coefficients for the explanatory variables will be different in each equation, as those inputs have differential effects on shares (e.g., falling oil prices boost the sales of SUVs). For non-seasonal schemes in particular, the notion of local momentum may be useful for describing changes in shares, and it also requires one fewer parameter, which could be important when only short series are available.

Other features of compositional data are the entry into and exit from the market of 'brands', and the existence of zero or near-zero values. These are handled using

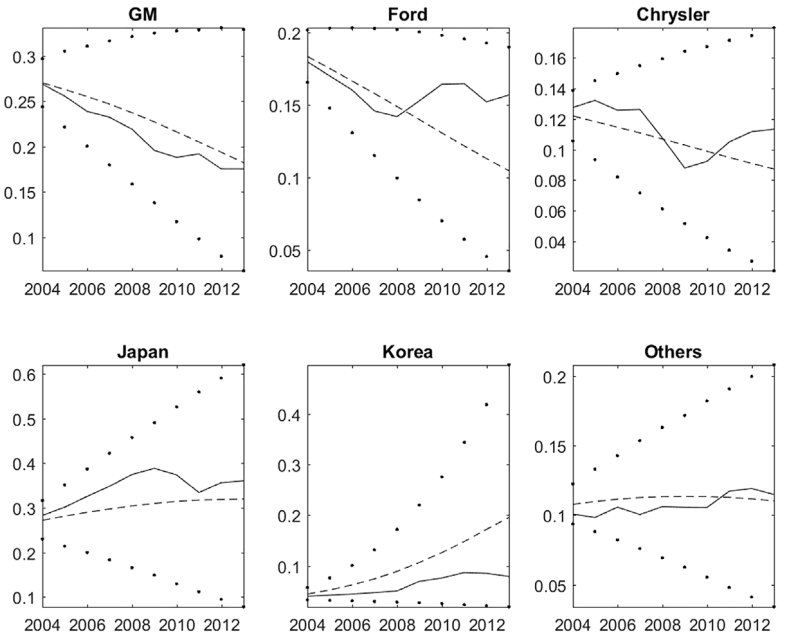

Fig. 3. One- to 10-year-ahead forecasts using the LTM models with the invertibility constraints. The solid lines are actual shares, the dashed lines are predictions, and the dotted lines are the 80% prediction limits.

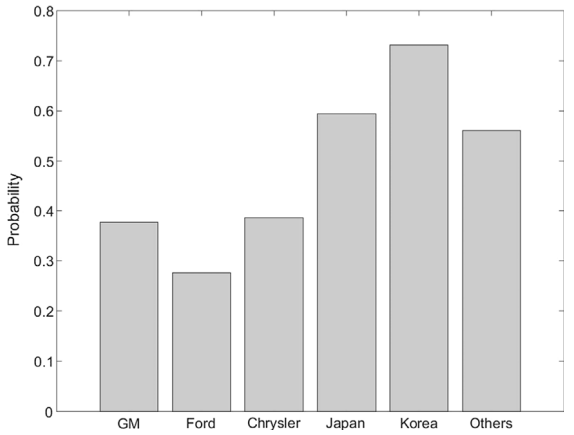

Fig. 4. Probabilities of increased market shares, by manufacturer, for 2004 (one step ahead).

the extensions described in Sections 5.3 and 5.4. The state space approach provides a convenient mechanism for dealing with such issues. The births and deaths of series may be incorporated using a redefined generalized variance criterion, while zero or near-zero values are typically occasional values from which the series recovers, and can be dealt with using a local perturbation technique. We have also described model selection procedures using the AIC, adapted to allow for different series lengths.

The approach could also be extended to include estimation error. The methods outlined by Ord, Koehler, and Snyder (1997) could be adapted to the context of this paper if the additional accuracy seemed warranted.

The estimation procedure described in Section 5 uses the innovations state space approach, which enables us to bypass the numerically more costly ARIMA approach based upon a Kalman filter. Although the total number of parameters is large, we are able to concentrate the likelihood in order to reduce the dimensionality to {starting values + smoothing parameters} when the series are of equal length. Provided that the series are not too short and the parameter α is not too close to zero, heuristic starting values can be used to bring the number of parameters down to one or two, and standard non-linear methods are easy to apply in order to obtain the parameter estimates.

Generating point forecasts and computing prediction intervals poses a problem because the standard approach, based upon multivariate normal errors, only provides intervals for the ratios of shares. This difficulty is resolved by using the simulation procedure described in Section 4.

Finally, the methods developed in this paper are demonstrated using annual sales data from the US automobile market. In this case, the local momentum and local trend models are found to provide better fits than models involving only local levels, according to the AIC. Typically, the interest focuses on changes in market shares, a dynamic that the local level models are unable to describe. Databases describing movements in market shares across multiple brands do not seem to be readily available for academic research purposes, but we hope that this paper will encourage the release of further examples.

Acknowledgments

Our thanks to Stephan Kolassa of SAP (Switzerland) for suggesting some improvements to an earlier draft of the paper. We wish to thank Ward's Auto Group for making the data used in this paper available.

Appendix. Invariance to rebasing

A feature of the multivariate model in Eqs. (2)–(3) is that it is built from univariate state space models which share common structural features ($\mathbf{w}$, $\mathbf{F}$, $\mathbf{g}$). The issue explored in this appendix is whether this common structure is necessary to ensure that the predictions of market shares are invariant to the choice of base series in the log-ratio transformation.

Theorem 1. *Let $\mathbf{y}_t \in \mathbb{R}^r$ and $\tilde{\mathbf{y}}_t \in \mathbb{R}^r$ be the log-ratio series obtained with different base series z_{0t} and z_{1t} respectively. This means that $y_{it} = \ln(z_{it}/z_{0t})$ for $i = 1, \ldots, r$, $\tilde{y}_{1t} = \ln(z_{0t}/z_{1t})$, and $\tilde{y}_{it} = \ln(z_{it}/z_{1t})$ for $i = 2, \ldots, r$. Then $\tilde{\mathbf{y}}_t$ and $\mathbf{y}_t$ are related linearly by*

$$\tilde{\mathbf{y}}_t = \mathbf{A}\mathbf{y}_t, \tag{A.10}$$

where $\mathbf{A} = \begin{bmatrix} -1 & & & \\ -1 & 1 & & \\ \vdots & & \ddots & \\ -1 & & & 1 \end{bmatrix}$. *The matrix $\mathbf{A} \in \mathbb{R}^{r\times r}$ is independent of any model parameters, has a unit determinant, and is self-inverse, in the sense that $\mathbf{A} = \mathbf{A}^{-1}$.*

See Aitchison (1986) for a proof of essentially the same result.

A more general model than that in Eqs. (2)–(3) is needed for an exploration of the issue of invariance and the necessity of common structural features. Thus, the typical series y_{it} is now assumed to be governed by a state space model with the measurement and transition equations given by $y_{it} = \mathbf{w}_i'\mathbf{x}_{i,t-1} + \varepsilon_{it}$ and $\mathbf{x}_{it} = \mathbf{F}_i\mathbf{x}_{i,t-1} + \mathbf{g}_i\varepsilon_{it}$, respectively, where $\mathbf{x}_{it} \in \mathbb{R}^k$ is a state vector and ε_{it} is an innovation. The structural elements $\mathbf{w}_i \in \mathbb{R}^k$, $\mathbf{F}_i \in \mathbb{R}^{k\times k}$ and $\mathbf{g}_i \in \mathbb{R}^k$ are now indexed by i. The ε_{it} are serially independent normal random variables, with a mean of zero and a time-invariant variance.

The primary focus here is on a special case where $\mathbf{w}_i = \mathbf{1}$, a vector of ones. It is referred to here as a *normalised* innovations state space model. Examples include the local level and local trend models used in this study. A non-normalised model can usually be converted to normalised form by rescaling the associated states by their coefficients; see below for further details.

The multivariate version of the normalised innovations state space model, formed by stacking the normalised sub-models, has the form

$$\mathbf{y}_t' = \mathbf{1}'\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t' \tag{A.11}$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}\boldsymbol{\varepsilon}_t, \tag{A.12}$$

where $\mathbf{X}_t = \begin{bmatrix}\mathbf{x}_{1t} & \cdots & \mathbf{x}_{rt}\end{bmatrix}$, $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$, and now $\mathbf{F} = \text{diag}(\mathbf{F}_1, \ldots, \mathbf{F}_r)$ and $\mathbf{G} = \text{diag}(\mathbf{g}_1, \ldots, \mathbf{g}_r)$. The $\boldsymbol{\varepsilon}_t$ are serially uncorrelated normally distributed random vectors with mean $\mathbf{0}$ and a time-invariant variance matrix $\mathbf{V}$. The key features of this multivariate model are:

1. Each series depends on its own state vector and no others.
2. The matrices $\mathbf{F}$ and $\mathbf{G}$ have block diagonal structures.
3. The typical sub-model i has its own parameter vector $\boldsymbol{\theta}_i$ which potentially determines some of the elements of $\mathbf{F}_i$ and $\mathbf{g}_i$. The implied relationships are written as $\mathbf{F}_i = F(\boldsymbol{\theta}_i)$ and $\mathbf{g}_i = g(\boldsymbol{\theta}_i)$, where F and g are functions with the properties that (a) they are common to all sub-models; and (b) the $\boldsymbol{\theta}_i$ are identifiable from $\mathbf{F}_i$ and $\mathbf{g}_i$.
4. Interdependencies between series are not modelled through the states directly, but through contemporaneously correlated innovations.

This normalised state space model is said to have a *block-diagonal structure* here, although this should not be taken to mean that the variance matrix $\mathbf{V}$ is block diagonal. It is a time series analogue of a *seemingly unrelated regression* (Zellner, 1962), and its sub-models are said to have a *common structure* if $\mathbf{F}_1 = \cdots = \mathbf{F}_r$ and $\mathbf{g}_1 = \cdots = \mathbf{g}_r$, something that occurs when $\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_r$.

The block diagonal normalised state space model for the rebased time series is written as

$$\tilde{\mathbf{y}}_t' = \mathbf{1}'\tilde{\mathbf{X}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t' \tag{A.13}$$

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{F}}\tilde{\mathbf{x}}_{t-1} + \tilde{\mathbf{G}}\tilde{\boldsymbol{\varepsilon}}_t, \tag{A.14}$$

where the tildes are used to distinguish associated quantities from their counterparts for $\mathbf{y}_t$.

Theorem 2. *The normalised models in Eqs. (A.11)–(A.12) and (A.13)–(A.14) for $\mathbf{y}_t$ and $\tilde{\mathbf{y}}_t$ are compatible if and only if their sub-models share a common structure.*

A proof of this theorem begins by temporarily ignoring the block-diagonal property and using Eq. (A.10) to establish that the two multivariate models are compatible if and only if

$$\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{A}\boldsymbol{\varepsilon}_t \tag{A.15}$$

$$\tilde{\mathbf{X}}_t = \mathbf{X}_t\mathbf{A}' \tag{A.16}$$

$$\tilde{\mathbf{F}} = \mathbf{B}\mathbf{F}\mathbf{B} \tag{A.17}$$

$$\tilde{\mathbf{G}} = \mathbf{B}\mathbf{G}\mathbf{A} \tag{A.18}$$

$$\tilde{\mathbf{V}} = \mathbf{A}\mathbf{V}\mathbf{A}', \tag{A.19}$$

where $\mathbf{B} = \mathbf{A} \otimes \mathbf{I}$ arises when Eq. (A.16) is vectorised to give $\tilde{\mathbf{x}}_t = \mathbf{B}\mathbf{x}_t$. It is then established that if $\mathbf{F}$ and $\mathbf{G}$ are block diagonal, $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{G}}$ can only be block diagonal if $\mathbf{F}_{11} = \cdots = \mathbf{F}_{rr}$ and $\mathbf{G}_{11} = \cdots = \mathbf{G}_{rr}$. To illustrate this aspect of the proof, consider the right hand side of Eq. (A.18) for $r = 2$:

$$\begin{bmatrix}-\mathbf{I} & \mathbf{0}\\ -\mathbf{I} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{g}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{g}_2\end{bmatrix}\begin{bmatrix}-1 & 0\\ -1 & 1\end{bmatrix} = \begin{bmatrix}\mathbf{g}_1 & \mathbf{0}\\ \mathbf{g}_1 - \mathbf{g}_2 & \mathbf{g}_2\end{bmatrix}.$$

The block-diagonal property prevails if and only if $\mathbf{g}_2 = \mathbf{g}_1$. Moreover, Eq. (A.18) implies that $\tilde{\mathbf{g}}_1 = \mathbf{g}_1$ and $\tilde{\mathbf{g}}_2 = \mathbf{g}_2$. The general result follows by applying this same basic logic to Eqs. (A.17) and (A.18) for any positive value of r.

Corollary 1. *The two multivariate models are compatible if and only if $\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_r = \tilde{\boldsymbol{\theta}}_1 = \cdots = \tilde{\boldsymbol{\theta}}_r$.*

This is a consequence of Theorem 2 and the identification property of the common functions F and g.

Theorem 3. *Let* $\boldsymbol{\theta}$ *and* $\tilde{\boldsymbol{\theta}}$ *be the common unknown parameters for the multivariate models for* $\mathbf{y}_t$ *and* $\tilde{\mathbf{y}}_t$ *respectively. Then, the conditional maximum likelihood estimates, designated with a circumflex, satisfy the following invariance conditions:* $\hat{\tilde{\boldsymbol{\theta}}} = \hat{\boldsymbol{\theta}}$, $\hat{\tilde{\mathbf{X}}}_0 = \hat{\mathbf{X}}_0\mathbf{A}'$ *and* $\hat{\tilde{\mathbf{V}}} = \mathbf{A}\hat{\mathbf{V}}\mathbf{A}'$.

This theorem follows because $\mathbf{A}$ has a unit Jacobian determinant, meaning that the likelihood functions for the two models are equal when $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$, $\tilde{\mathbf{X}}_0 = \mathbf{X}_0\mathbf{A}'$ and $\tilde{\mathbf{V}} = \mathbf{AVA}'$. The respective maximum likelihood estimates are then related by the same formulas. This property implies that the choice of base series does not matter.

An innovations state space model need not always possess a normalised form as defined above. In general, measurement equation coefficients need not always equal one, and may sometimes be zero, as in common additive seasonal models. The above theory can be adapted for this context by extending the definition of a normalised form to involve measurement equations given by $y_{it} = \boldsymbol{\delta}'\mathbf{x}_{i,t-1} + \varepsilon_{it}$, where $\boldsymbol{\delta}$ is a vector of ones and zeros. A non-normalised model has a measurement equation $y_{it} = \mathbf{w}_i'\mathbf{x}_{i,t-1} + \varepsilon_{it}$ with $w_i = w(\boldsymbol{\theta}_i)$, where w is a common function which may include a constant function. This non-normalised model can be converted to extended normalised form by introducing the diagonal matrix $\mathbf{W}_i \in \mathbb{R}^{k\times k}$, where $w_{iqq} = w_{iq}$ if $w_{iq} \neq 0$ and $w_{iqq} = 1$ if $w_{iq} = 0$, with $q = 1, \ldots, r$ being the row (column) index. The use of the common w function ensures that any zeros in $\mathbf{w}_i$ are in the same locations for all i. The measurement and transition equations are transformed to the extended normalised form $y_{it} = \boldsymbol{\delta}'\left(\mathbf{W}_i\mathbf{x}_{i,t-1}\right) + \varepsilon_{it}$ and $\left(\mathbf{W}_i\mathbf{x}_{it}\right) = \left(\mathbf{W}_i\mathbf{F}_i\mathbf{W}_i^{-1}\right)\left(\mathbf{W}_i\mathbf{x}_{i,t-1}\right) + \left(\mathbf{W}\mathbf{g}_i\right)\varepsilon_{it}$, and the above theory then carries over to this transformed model.

References

Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman and Hall.

Aitchison, J., & Egozcue, J. J. (2005). Compositional data analysis: where are we and where should we be heading? *Mathematical Geology*, *37*, 829–850.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Barceló-Vidal, C., Aguilar, L., & Martín-Fernández, J. A. (2011). Compositional VARIMA time series. In V. Pawlowsky-Glahn, & A. Buccianti (Eds.), *Compositional data analysis: theory and applications*. Chichester, UK: John Wiley & Sons, Ltd., (Chapter 7).

Brodie, R. J., Danaher, P. J., Kumar, V., & Leeflang, P. S. H. (2001). Econometric models for forecasting market share. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 597–611). Boston and Dordrecht: Kluwer.

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

De Silva, A., Hyndman, R., & Snyder, R. (2009). A multivariate innovations state space Beveridge–Nelson decomposition. *Economic Modelling*, *26*, 1067–1074.

De Silva, A., Hyndman, R., & Snyder, R. (2010). The vector innovation structural time series framework: a simple approach to multivariate forecasting. *Statistical Modelling*, *10*(4), 353–374.

Fry, J. M., Fry, T. R. L., & McLaren, K. R. (2000). Compositional data analysis and zeros in micro data. *Applied Economics*, *32*, 953–959.

Grunwald, G. K., Raftery, A. E., & Guttorp, P. (1993). Time series of continuous proportions. *Journal of the Royal Statistical Society: Series B*, *55*, 103–116.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, *55*, 2579–2589.

Hyndman, R. J., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, *24*(3), 323–342.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing*. Berlin: Springer.

Kumar, V., Nagpalb, A., & Venkatesan, R. (2002). Forecasting category sales and market share for wireless telephone subscribers: a combined approach. *International Journal of Forecasting*, *18*, 583–603.

Mills, T. C. (2010). Forecasting compositional time series. *Quality and Quantity*, *44*, 673–690.

Ord, J. K., Koehler, A. B., & Snyder, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, *92*(440), 1621–1629.

Quintana, J. M., & West, M. (1988). Time series analysis of compositional data. In J. M. Bernardo, M. H. de Groot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics, vol. 3* (pp. 747–756). New York: Oxford University Press.

Schmidt, P. (1977). Estimation of seemingly unrelated regressions with unequal numbers of observations. *Journal of Econometrics*, *5*(3), 365–377.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*, 348–368.

Ralph D. Snyder is an adjunct associate professor in the Department of Econometrics and Business Statistics at Monash University, Australia. Having extensive publications on business forecasting and inventory management, he has played a leading role in the establishment of the class of innovations state space models for exponential smoothing.

J. Keith Ord is a professor emeritus in the McDonough School of Business, Georgetown University, Washington DC. He has published extensively on forecasting and applied statistics and is a co-author of Kendall's Advanced Theory of Statistics. He is a fellow of the American Statistical Association and the International Institute of Forecasters.

Anne B. Koehler is a professor emerita of Miami University, Oxford, Ohio. She has numerous publications, many of which are on forecasting models for seasonal time series and exponential smoothing methods. She is a fellow of the International Institute of Forecasters and the Decision Sciences Institute.

Keith R. McLaren is a professor emeritus in the Department of Econometrics and Business Statistics at Monash University, Australia. His publications include papers in applied econometrics, with a special emphasis on using compositional data analysis to impose regularity on consumer demand systems expressed as budget shares.

Adrian N. Beaumont recently completed his Ph.D. at the University of Melbourne on aspects of state space models and exponential smoothing. He has published a number of papers in the IJF on forecasting and provides regular poll and election analysis for The Conversation, an Australian consortium between editors and academics which provides web based informed news analysis and commentary.