



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Forecasting Compositional Time Series:
A State Space Approach**

**Ralph D. Snyder, J. Keith Ord, Anne B. Koehler, Keith
R. McLaren and Adrian Beaumont**

April 2015

Working Paper 11/15

Forecasting Compositional Time Series: A State Space Approach¹

Ralph D. Snyder, Monash University, Australia

J. Keith Ord, Georgetown University, USA

Anne B. Koehler, Miami University, USA

Keith R. McLaren, Monash University, Australia

Adrian Beaumont, Melbourne University, Australia

April 21, 2015

ABSTRACT

A method is proposed for forecasting composite time series such as the market shares for multiple brands. Its novel feature is that it relies on multi-series adaptations of exponential smoothing combined with the log-ratio transformation for the conversion of proportions onto the real line. It is designed to produce forecasts that are both non-negative and sum to one; are invariant to the choice of base series in the log-ratio transformation; recognize and exploit features such as serial dependence and non-stationary movements in the data; allow for the possibility of non-uniform interactions between the series; and contend with series that start late, finish early, or which have

¹ We wish to thank Ward's Auto Group (<http://wardsauto.com/data-center>) for making the data used in this paper available.

values close to zero. Relying on an appropriate multivariate innovations state space model, it can be used to generate prediction distributions in addition to point forecasts and to compute the probabilities of market share increases together with prediction intervals. A shared structure between the series in the multivariate model is used to ensure that the usual proliferation of parameters is avoided. The forecasting method is illustrated using data on the annual market shares of the major (groups of) brands in the U.S. automobile market, over the period 1961 – 2013.

KEY WORDS: Exponential smoothing; Proportions; Prediction intervals; Automobile sales; Market shares.

JEL classification: C22

1 INTRODUCTION

The need to forecast proportions arises in a wide variety of areas, such as business, economics, geology, and political science. Specific examples include the market shares of competing products, the proportions of jobs in different sectors of the economy, or the geochemical composition of a selected ore. In some cases, as with a rock sample, 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 an 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.

Aitchison (1986) is a key reference for compositional time series analysis. His analysis shows that we can draw on a wide variety of established and well understood time series methods by using the log-ratio transform to map the proportions onto the real line. This transformation has become relatively standard over the years (Aitchison & Egozcue, 2005; Brundson & Smith, 1998; Quintana & West, 1988) and is adopted in our paper. A key advantage is that it enables us to base the analysis on the normal probability distribution. Other approaches in the literature based on the Dirichlet distribution (Grunwald, Raftery & Guttorp, 1993) or on the hyper-spherical transformation (Mills, 2010) are not considered here.

The novel feature of this paper is the coupling of the log-ratio transformation with vector exponential smoothing (De Silva, Hyndman & Snyder, 2009; De Silva, Hyndman and Snyder, 2010; Hyndman, Koehler, Ord & Snyder, 2008). Such an approach has parallels in the vector

ARIMA models of Brundson & Smith (1998) and the more recent work of Barceló-Vidal, Aguilar & Martín-Fernández (2011). The rationale for this shift of emphasis lies in the nature of market share data. New brands may be launched or old brands may fade away. ARIMA models assume that series are stationary after suitable differencing and can notionally extend to the “infinite past”. State space models allow for the intrinsically non-stationary nature of a start-up and we are able to describe such phenomena with these models. The emphasis in this paper is on pure time series models without any explanatory variables. Such extensions are clearly possible, although as Brodie, Danahar, Kumar & 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 should hopefully be bested by the systematic development of time series models.

The emphasis on market shares as a basis for policy-making has not met with universal support. Armstrong & Green (2007) argue that a firm’s emphasis on increasing market share can come at the cost of overall profitability. We remain neutral on that issue and concentrate on the forecasting questions, believing that well-led companies would want to consider both market share and longer term profitability. We should also note 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 demonstrated by Kumar, Nagpalb & Venkatesan (2002).

The paper is structured as follows. The data on (grouped) company shares of the U.S. automobile market, which serves as the example for our empirical work, is described in the next section. The basic model and the transformations used to ensure that the non-negativity and adding up constraints are satisfied are introduced in Section 3. Maximum likelihood estimation, with

extensions to allow for series of unequal length, model selection and the construction of prediction intervals are considered in Section 4. The detailed analysis of the U.S. automobile data is undertaken in Section 5, while conclusions with discussion are provided in Section 6.

2 DESCRIPTION OF THE DATA

Annual U.S. total vehicle sales market share data (<http://wardsauto.com/data-center>), shown in Figure 1, exhibit many of the issues considered in this paper and is used to illustrate our proposed approach in Section 5. The series cover the period 1961–2013, but the sales for 2013 were not used when estimating the models, being retained to evaluate the results of the prediction method. The original series relate to all the principal manufacturers who sell in the American market, but 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: BMW, Daimler, Volkswagen and Other [1 series].

We selected these groupings with several factors in mind: the groups should be reasonably homogeneous in terms of market appeal and we did not want so many groups that the results were difficult to discern graphically. In addition, as we are particularly interested in new product entry, we wished to have some groups enter the market later than others (Japan and Korea). It should be noted that any analysis based upon log-ratios is not invariant under grouping of ‘brands’ but the effects will be minimized provided we select reasonably homogeneous groups.

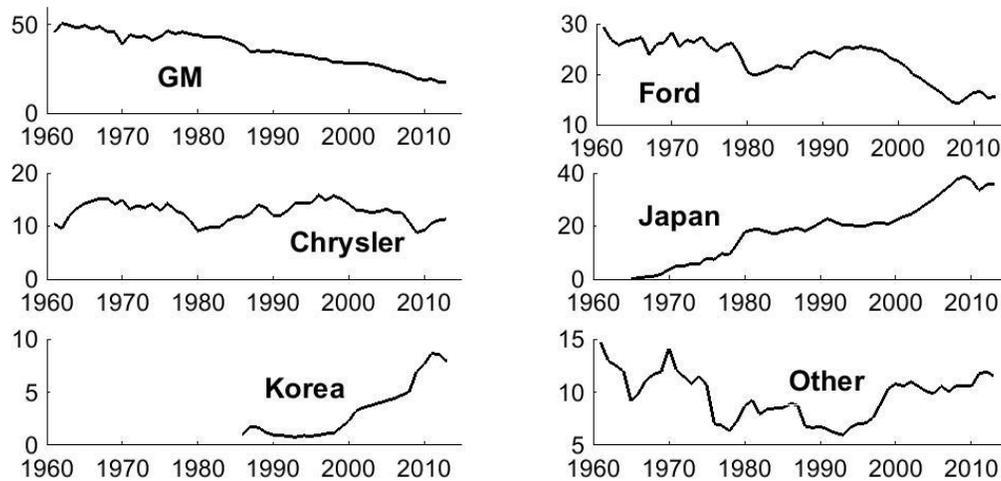


Figure 1. Plots of percentage market shares for annual automobile sales 1961 – 2013. Source: Ward’s Auto Group.

The non-stationary nature of these series is evident from Figure 1. The market shares for GM and Ford have steadily declined, whereas those for the Japanese and Korean manufacturers have increased. The series for Chrysler chronicles the decline and recovery of that company in the seventies and eighties, whereas the other series (mostly German manufacturers) reflects a decline and more recent resurgence.

As is to be expected, the series show strong serial dependence so that the random walk model is a viable choice, as noted by Brodie et al. (2001). However, things are not as straightforward as they appear, as the total shares must still add to one. For an individual company a binary “Us versus Them” model (Brodie et al., 2001) might be considered appropriate, 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, strong serial dependence, non-stationarity, non-uniform interactions among the series and new entries. By virtue of the grouping

we performed, we are not confronted by missing values or near-zeros in this example, but we also address those questions briefly.

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 generally denoted by $\{z_{it}; i = 0, 1, \dots, r; t = 1, 2, \dots, n\}$. However, it is convenient to temporarily drop the t subscript when considering the transformations.

One transformation is needed to convert the shares vector \mathbf{z} of dimension $r + 1$ into the unbounded vector $\mathbf{y} \in \mathbb{R}^r$ so 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 $i = 1, \dots, r$ where z_0 is the base series. For the moment it is assumed that none of the series contain zeros. As 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

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

We apply the log-ratio transformation in each period to give the unbounded r -vectors $\mathbf{y}_1, \dots, \mathbf{y}_n$ in

what shall be termed the log-ratio space; the GM series is used as the base. It transpires that the resulting series 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 Structure

Vector innovations state space models are used to represent the evolution of the transformed non-stationary time series. (For the moment it is assumed that all series have a common start and end date). In these models the series vector \mathbf{y}_t in typical period t is linearly related 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 measurement equation

$$\mathbf{y}'_t = \mathbf{w}'\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}'_t \quad (1)$$

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

$$\mathbf{X}_t = \mathbf{F}\mathbf{X}_{t-1} + \mathbf{g}\boldsymbol{\varepsilon}'_t \quad (2)$$

where $\mathbf{F} \in \mathbb{R}^{k \times k}$ and $\mathbf{g} \in \mathbb{R}^k$ are also fixed. The structural vectors and matrix \mathbf{w} , \mathbf{F} and \mathbf{g} may depend on a vector $\boldsymbol{\theta} \in \mathbb{R}^p$ of common parameters. The $r(r+1)/2$ potentially distinct elements of \mathbf{V} are additional parameters.

Such a model can always be reduced to the form of a vector ARIMA model, provided we keep in mind the finite start-up of some or all of the series, contrary to the usual ARIMA assumptions. The

details of this reduction for the univariate case are given in Hyndman et al. (2008, Chapter 11).

The multivariate model specified in equations (1) and (2) is composed of a univariate model for each 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 established in the Appendix that the shared structure is necessary and sufficient for maximum likelihood estimates 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. So, the comparison between two series indexed by i and j , exploiting the common \mathbf{w} in the measurement equation (1), is governed by the relationship $z_{it}/z_{jt} = \exp(\mathbf{w}'(\mathbf{x}_{i,t-1} - \mathbf{x}_{j,t-1}) + (\varepsilon_{it} - \varepsilon_{jt}))$ 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 (2) that $\mathbf{x}_{it} - \mathbf{x}_{jt} = \mathbf{F}(\mathbf{x}_{i,t-1} - \mathbf{x}_{j,t-1}) + \mathbf{g}(\varepsilon_{it} - \varepsilon_{jt})$. 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 to provide a sufficiently rich set of options for market share analysis. In all cases we retain the assumption $\boldsymbol{\varepsilon}_t \sim \text{NID}(0, \mathbf{V})$, with \mathbf{V} 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 explicitly consider this model in our empirical analysis. At another extreme is the multivariate random walk $\mathbf{y}_t = \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$ (referred to as the naïve model by Brodie et al., 2001). It is of interest because it is the simplest model to allow for both serial dependence and

non-stationarity.

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 embody 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 in Hyndman et al. (2008, Chapter 3). Ignoring the elements of \mathbf{V} , the common structure assumption means that the vector LLM and the vector LTM have just one and two parameters respectively, reducing the proliferation of parameters that normally accompanies multivariate models.

The following notation is employed in Table 1: $\ell_t \in \mathbb{R}^k$ 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.

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

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

Some restrictions must be imposed on the parameters to ensure forecastability (Hyndman et al., 2008, pp. 152-155). For the models we consider, forecastability corresponds to invertibility plus those situations where some of the smoothing parameters may be zero. But traditional constraints 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 actively debated.

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

Model	Traditional	Invertibility/Forecastability
Local level, LLM	$0 \leq \alpha \leq 1$	$0 \leq \alpha \leq 2$
Local trend, LTM	$0 \leq \alpha \leq 1$	$2\alpha + \beta \leq 4$
	$0 \leq \beta \leq \alpha$	$\alpha \geq 0, \beta \geq 0$

These particular models may be expressed in terms of the general notation established in the previous section. The formats are 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 it 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 it is again a boundary case when the traditional constraints are applied but an interior solution for the invertibility constraints. A final version of interest may be termed the local momentum model (LMM) where the rate of change has a local level structure, or $\nabla \mathbf{y}_t = \nabla \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - (1 - \beta)\boldsymbol{\varepsilon}_{t-1}$. Clearly the LMM is a special case of the LTM (when $\alpha = 1$), but its interpretation is of interest as it corresponds to a local level model for the second derivative, or acceleration.

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

Table 3. 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	ℓ_t	1	1	1	1
Local level	ℓ_t	1	1	α	α
Local trend	$\begin{bmatrix} \ell'_t \\ \mathbf{b}'_t \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$	$\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$
Local Momentum	$\begin{bmatrix} \ell_t \\ \mathbf{b}_t \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ \beta \end{bmatrix}$	$\begin{bmatrix} 1 \\ \beta \end{bmatrix}$

4 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 state equation parameters and the variance matrix elements respectively. We considered two different estimation procedures, both maximum likelihood-based and these are described in turn, based upon the requirements of the LTM.

4.1 A Vector ARIMA Approach

In this framework the problem is to estimate the parameters (α, β) and the $r(r+1)/2$ potentially distinct elements of the variance matrix \mathbf{V} . An advantage of the ARIMA approach appears to be that the unobservable random state vectors are absent and need not be estimated. However, direct specification of the unconditional likelihood is not possible because of the existence of the two unit roots (differences). It is instead necessary to base the likelihood function on the density

$p(\mathbf{y}_3, \dots, \mathbf{y}_n | \mathbf{y}_1, \mathbf{y}_2; \alpha, \beta, \mathbf{V})$, one that is conditioned on the first two observations. Equivalently,

this function may be represented as the density $p(\nabla^2 \mathbf{y}_3, \dots, \nabla^2 \mathbf{y}_n; \alpha, \beta, \mathbf{V})$ of second differences.

Thus, the effective sample size for each transformed series is $n - 2$.

The evaluation would be simplified by using the prediction decomposition of this density

$\prod_{t=3}^n p(\nabla^2 \mathbf{y}_t | \nabla^2 \mathbf{y}_3, \dots, \nabla^2 \mathbf{y}_{t-1}; \alpha, \beta, \mathbf{V})$. Options to do this include the Kalman filter. However,

each pass of the Kalman filter would require the specification of trial values of α, β and the

$r(r+1)/2$ elements of \mathbf{V} . By way of example, a search 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 4, $r = 5$, and there are 17 dimensions!

4.2 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 with the general exponential smoothing equations (in

the notation of Section 3.2):

$$\mathbf{e}'_t = \mathbf{y}'_t - \mathbf{w}'\mathbf{X}_{t-1} \quad (3)$$

and

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

The \mathbf{X}_t and \mathbf{e}_t which emerge for typical period t are conditioned on \mathbf{X}_0 and $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}$. The focus

is on the density $p(\mathbf{y}_1, \dots, \mathbf{y}_n | \mathbf{X}_0; \boldsymbol{\theta}, \mathbf{V})$ 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 sub-section: 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 $\prod_{t=1}^n p(\mathbf{y}_t | \mathbf{y}_1, \dots, \mathbf{y}_{t-1}, \mathbf{X}_0; \boldsymbol{\theta}, \mathbf{V})$ can in turn 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

sub-section, \mathbf{X}_0 must be indirectly observed 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

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

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

$$\hat{\mathbf{V}}(\boldsymbol{\theta}, \mathbf{X}_0) = \sum_{t=1}^n \mathbf{e}_t \mathbf{e}_t' / n. \quad (4)$$

Substituting this into the likelihood function gives its concentrated counterpart:

$$\mathcal{L}(\mathbf{X}_0, \boldsymbol{\theta}) = -\frac{nr}{2} \ln(2\pi) - \frac{n}{2} \ln|\hat{\mathbf{V}}| - \frac{nr}{2}. \quad (5)$$

We maximize (5) to obtain the maximum likelihood predictor of the seed state matrix \mathbf{X}_0 and

maximum likelihood estimate of the parameter vector $\boldsymbol{\theta}$. The effect is to also minimize the

estimated generalized variance $|\hat{\mathbf{V}}|$. 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 reduction in the search dimension is

from 17 to 12. In general, the advantage of the exponential smoothing approach over the ARIMA approach with a Kalman filter becomes much more pronounced for data sets with a larger number of series because the dimensionality is linear rather than quadratic in the number of series.

Finally, we observe that for longer series, a simple heuristic may be used to obtain good approximations for the starting values, thereby further reducing the dimensionality of the parameter space 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. So, in our application in Section 5, we persist with the use of maximum likelihood predictors of the seed states.

4.3 An Adjustment for the Sample Size

A curious difference between the approaches in the previous two sub-sections is the sample size. In the vector ARIMA approach we difference the series d times, where d is the number of unit roots. Thus, the effective sample size becomes $n - d$. In the state space approach, the conditioning on \mathbf{X}_0 means that the sample size appears to be n and formula (4) for the estimated variance matrix has a divisor of n . The question is whether this divisor really makes sense?

To get a better sense of the issue involved, consider the local level model. A form of degeneracy occurs when $\alpha = 1$, corresponding to the naïve model. Then the typical error vector becomes $\mathbf{y}_t - \mathbf{y}_{t-1}$, a vector which is independent of the seed state (now a vector designated by $\mathbf{x}_0 \in \mathbb{R}^r$). The only exception is the first error $\mathbf{y}_1 - \mathbf{x}_0$. The sum of squared errors, suitably scaled by the inverse of the variance matrix \mathbf{V} , is minimized when $\hat{\mathbf{x}}_0 = \mathbf{y}_1$. The optimized sum of squared errors is left with

$n - 1$ terms. The adjective optimized is important here. Remember that the formula (4) for the variance matrix was derived for a given value of \mathbf{X}_0 , not necessarily its optimized value.

Generalizing, it appears that the optimization of the seed states leads to d linear restrictions on the errors. Returning to the LLM when $\alpha = 1$, the need to start with the second error term is obvious. However, even when $\alpha \neq 1$ the first observed error term is considerably reduced because of the dependence on $\hat{\alpha}$. In general, the bias will be reduced if we estimate the variance matrix by

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{X}}_0) = \sum_{t=1}^n \mathbf{e}_t \mathbf{e}_t' / (n - d). \quad (6)$$

Note the change of notation from $\hat{\mathbf{V}}(\boldsymbol{\theta}, \mathbf{X}_0)$ in (4) to $\hat{\mathbf{V}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{X}}_0)$ in (6).

4.4 Accounting for Different Series Lengths

Earlier it was noted that we needed to take into account all of the following features: a constrained total (to one), strong serial dependence, non-stationarity, non-uniform interactions among the series and new entries. The state space formulation in Section 3 accounts for all these factors, but does not explicitly enable us to address the issue of new entries. 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.

The first step is to ensure that the error terms for Japan and Korea make no contribution to the estimate (6) of the variance matrix in the periods 1961-1964 and 1961-1985 respectively. This is done quite simply by forcing their ‘errors’ to zero for these run-in periods. More specifically, we set

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

Corresponding adjustments to sample sizes are required so that the estimator for the typical element

of \mathbf{V} becomes $\hat{v}_{ij} = \sum_{t=1}^n e_{it}e_{jt} / \min\{n_i - d, n_j - d\}$ where n_i and n_j designate the sample sizes of series i and series j .

Differences in the start-up dates of the series also imply that not all the seed state vectors are coincident in time. In our example, GM, Ford, Chrysler and Others have 1960 as their origin, whereas for Japan and Korea, the origins are 1964 and 1985 respectively. A general approach to this issue requires a framework that minimizes the extra effort required from a coding perspective.

Fortunately, this difficulty may be resolved by 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 the series so that each has a seed state vector in this year. Those of Japan and Korea may be thought of as artificial seed states. General exponential smoothing is applied from 1961 to 2012 in the usual way, except (7) is now used instead of (3) to determine the errors. This means the seed states for Japan and Korea in 1964 and 1985 are deterministically related to their artificial counterparts in 1960. Optimization of the artificial seed states for 1960 then leads to appropriate optimized values for the 1964 and 1985 seed states.

The expression (5) for the log-likelihood also needs to be adapted so that only those series observed at time t contribute to the likelihood function. To this end, let $\mathbf{D}_t \in \mathbb{R}^{r \times r}$ 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 most readily seen by mentally 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

$\mathbf{D}_t = [0_{(r-r_t) \times r_t}, I_{r_t \times r_t}]$ 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 associated $r_t \times r_t$ variance matrix $\mathbf{V}_t = \mathbf{D}_t \mathbf{V} \mathbf{D}_t'$. The log-likelihood function must be amended to

$$\mathcal{L}(\mathbf{X}_0, \boldsymbol{\theta}) = -\frac{nr}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^n \ln |\hat{\mathbf{V}}_t| - \frac{nr}{2} \text{ where } \hat{\mathbf{V}}_t = \mathbf{D}_t \hat{\mathbf{V}} \mathbf{D}_t'.$$

An alternative approach that produces essentially the same results but is less easy to implement computationally, is to use equation (6) to estimate the variance matrix combined with indicator variables to show the starting point of each series.

4.5 Zero or Near-zero Values

The data set we used was constructed to avoid problems with near-zero shares. However, such concerns often arise in the analysis of market shares when a product is in decline or simply not attracting customers. Indeed, the previous discussion of new entrants could be viewed as an extreme case of the near-zero problem. Fry, Fry and McLaren (2000) proposed a solution to this problem that involved a small perturbation of the (near) zero values to avoid numerical difficulties with the log-ratio transformations.

4.6 Model Selection

The Akaike information criterion (Akaike, 1974) is used for model comparisons and selection. It is defined in our context by $\text{AIC} = -2\mathcal{L}(\hat{\mathbf{X}}_0, \hat{\boldsymbol{\theta}}) + 2(rd + p) + r(r + 1)$. The second and third terms

represent the penalty based on rd for the number of seed states, p for the number of state equation parameters and $r(r+1)/2$ for the number of potentially distinct elements in the variance matrix.

The number of terms in the variance matrix does not vary and so has no impact on the model selection process in the present framework, but such terms should be included to allow for cases with restricted versions of the variance.

4.7 Model Checking

The usual panoply of plots for residuals may be employed to check the validity of the selected model but, for the most part these analyses are univariate in nature. A multivariate check may be formulated by considering the joint distribution of $\mathbf{y}_t = \{y_{1t}, \dots, y_{rt}\}'$ and letting $\hat{\mathbf{y}}_t$ denote the fitted value of \mathbf{y}_t . The quadratic form $Q_t = (\mathbf{y}_t - \hat{\mathbf{y}}_t)' \hat{\mathbf{V}}^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_t)$ is asymptotically distributed as χ_r^2 when the errors are multivariate normal, so that when the null hypothesis of a valid model holds, approximately 90 percent of the vectors of residuals should satisfy the inequalities

$$Q_t < \chi_{0.9}^2(r_t). \quad (8)$$

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

4.8 Prediction

Point predictions for the log-ratios follow from equations (1) and (2) in the usual way. The general expression for the h -step-ahead forecasts is:

$$\hat{\mathbf{y}}'_{t+h|t} = \mathbf{w}' \mathbf{F}^{h-1} \mathbf{X}_t. \quad (9)$$

The assumption of multivariate normality then provides a prediction region using the chi-square distribution to produce a 90 percent interval $\chi_{0.05}^2(r_t) < Q_t < \chi_{0.95}^2(r_t)$. Unfortunately, for $r > 2$, the quadratic form cannot be unscrambled to produce intervals for the individual market shares.

Point predictions and prediction intervals for the market shares are obtained by simulation. The method is described for only one step ahead, that is for period $n + 1$ given observations through time n . The one-step intervals are likely to be of greatest interest, but the method may be extended to cover longer forecast horizons.

The steps in the simulation method are as follows:

1. Randomly generate R vectors from the log-ratio space prediction distribution $N(\hat{\mathbf{y}}_{n+1|n}, \hat{\mathbf{V}})$ for period $n + 1$ and apply the logistic transformation to give a sample of composite vectors $\mathbf{z}_{n+1}^{(1)}, \dots, \mathbf{z}_{n+1}^{(R)}$.
2. Estimate pertinent statistics from the simulated sample as follows:
 - a. Point predictions with sample averages.
 - b. Prediction intervals with appropriate sample quantiles.
 - c. Market share increase probabilities with the proportion of simulated sample values in excess of current observed market shares.

5 RESULTS FOR AUTOMOBILES DATA

5.1 Estimation and Model Selection

The U.S. automobiles data described in Section 2 was 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 constraints sets (traditional and invertibility).

The estimation results are summarized in Table 4. The AICs indicate that the vector local trend models are likely 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 previous interest in the random walk model (c.f. Brodie et al., 2001), the relatively poor performance in this example is noteworthy. The AIC results also indicate that use of the invertibility restrictions on the parameters leads to better predictions (as judged by 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 is an interesting model in its own right. In effect, the first differences follow a local level model. Accordingly, predictions must be exponentially weighted averages of the differenced series. A first difference may be interpreted as actual growth (in the transformed data) and the b_t the structural growth rate. Overall, the margin of difference for the two LTM versions is slight and the simpler interpretation of the local momentum model might override the slight statistical differences.

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

Table 4. Summary of estimation results

Model	Restrictions	Alternative Description	AIC	alpha	beta	coverage ^b (per cent)
-------	--------------	-------------------------	-----	-------	------	----------------------------------

LLM	traditional	random walk	-1871	1.00		82
LLM	invertibility	SES ^a	-1896	1.25		85
LTM	traditional	local momentum	-1919	1.00	0.19	80
LTM	invertibility	LES ^a	-1923	1.11	0.16	83

Notes:

- a. SES and LES denote (multivariate) simple exponential smoothing and linear exponential smoothing respectively.
- b. Coverage of nominal 90 per cent intervals for the chi-square statistic.

5.2 Model Diagnostics

The coverages of the nominal 90 per cent semi-infinite intervals (closed from above) for the chi-square error statistic (8) are also shown in the final column of Table 4. In all cases the coverage is somewhat below the nominal levels. However, the plot of the Q-statistics for the local trend model in Figure 2 suggests that this model has “settled down” after the Korean entry in 1986 and the later results are in line with expectations.

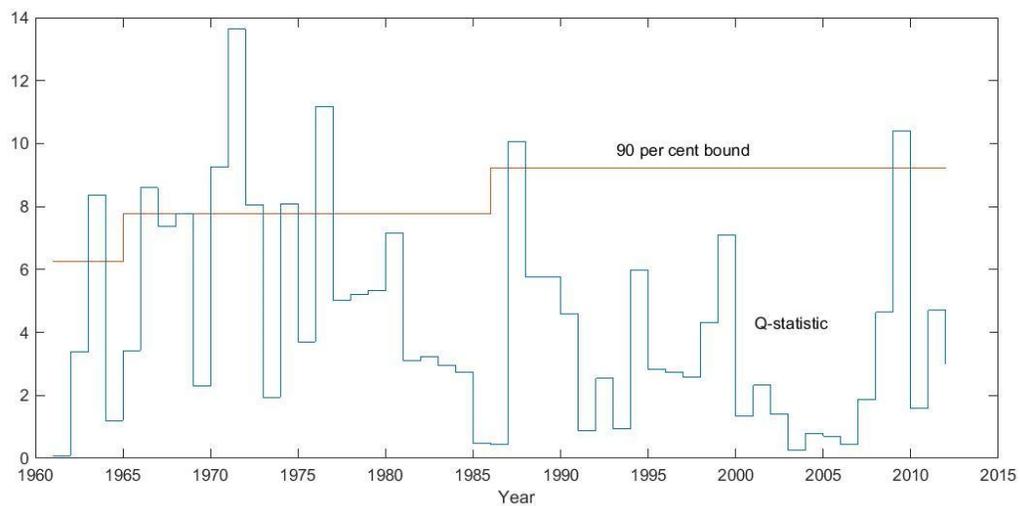


Figure 2. Plot of Q-statistics and 90 per cent upper limits for the local trend model: 1961-2012.

5.3 Forecasts and Prediction Intervals for the Best Local Trend Model

The point predictions and 90 per cent prediction intervals for 2013 from the best fitting local trend model are listed in Table 5. The point predictions are simple averages of the simulated sample of proportions. The lower and upper limits of the prediction intervals are the 5th and 95th percentiles. The observed shares in 2013 have also been listed: they all lie within the prediction intervals.

Table 5. 90 percent prediction intervals for 2013 percentage shares for the local trend model

Market Segment	Lower Limits	Point Predictions	Upper Limits	Actual Shares
GM	14	16	19	18
Ford	13	15	17	16
Chrysler	9	11	13	11
Japan	30	37	43	36
Korea	7	9	13	8
Other	10	12	14	11

Too much should not be made of one year's results but Table 6 indicates that, overall, the U.S. manufacturers fared slightly better than expected, as did the Japanese. These modest gains in market share came at the expense of Korean and other manufacturers.

Table 6. Predicted probabilities (percentages) of an increase from the local trend model, together with actual market shares (percentages)

Market Segment	Increase Probability	Actual Shares		
		2012	2013	Increase
GM	26	17.6	17.5	no
Ford	32	15.2	15.7	yes
Chrysler	48	11.2	11.3	yes
Japan	55	35.6	36.0	yes
Korea	64	8.5	7.9	no
Other	57	11.9	11.5	no

The median absolute percentage error (MAPE) is another way of summarizing forecast performance. Defined as the median of the absolute percentage prediction errors

$100|z_{i,2013} - \hat{z}_{i,2013}|/z_{i,2013}$ for the series $i = 0, 1, \dots, 5$, it turned out to be about 5.3 per cent.

6 DISCUSSION

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

Other features of compositional data are the entry and exit of ‘brands’ into the market and the existence of near-zero shares. These are handled with extensions described in Sections 4.4 and 4.5.

The state space approach provides a convenient mechanism for dealing with such issues, something that is not readily available in conventional ARIMA modeling. We also described model selection procedures using the AIC, adapted to allow for different series lengths.

The estimation procedure described in Section 4 uses the innovations state space approach, which enables us to by-pass the more numerically costly ARIMA approach with a Kalman filter .

Although the total number of parameters is large, we are able to concentrate the likelihood to reduce the dimensionality to {starting values + smoothing parameters}. Provided the series is not too short and the parameter α is not too close to zero, heuristic starting values could be used so that the number of parameters is down to one or two and standard non-linear methods are easily applied to obtain the parameter estimates.

The generation of point forecasts and the computation of prediction intervals pose a problem because the standard approach based upon multivariate normal errors provides intervals only for the ratios of shares. This difficulty is resolved using the simulation procedure described in Section 4.8.

Finally, the methods developed in the paper are demonstrated using annual sales data in the U.S. automobiles market. In this case, the local momentum model is found to provide a better fit according to AIC, relative to models involving only local levels. Typically, interest focuses upon changing 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 this paper may encourage the release of further examples.

APPENDIX: INVARIANCE TO REBASING

A feature of the multivariate model (1)-(2) 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 needed to ensure that 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 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 $i=1, \dots, r$, $\tilde{y}_{1t} = \ln(z_{0t}/z_{1t})$ and $\tilde{y}_{it} = \ln(z_{it}/z_{1t})$ for $i=2, \dots, r$. Then $\tilde{\mathbf{y}}_t$ and \mathbf{y}_t are linearly related by

$$\tilde{\mathbf{y}}_t = \mathbf{A}\mathbf{y}_t \tag{A10}$$

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.

To explore the issue of invariance and the necessity of common structural features, a more general model than (1)-(2) is needed. So, the typical series y_{it} is now assumed to be governed by a state space model with 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 0 and a time invariant variance.

The primary focus is on a special case where $\mathbf{w} = \mathbf{1}$, a vector of ones. It is called here a normalized

innovations state space model. Examples include the local level and local trend models used in the study. A non-normalized model can usually be converted to normalized form by rescaling the associated states by their coefficients— see below for further details.

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

$$\mathbf{y}'_t = \mathbf{1}'\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}'_t \quad (\text{A11})$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}\boldsymbol{\varepsilon}_t \quad (\text{A12})$$

where $\mathbf{X}_t = [\mathbf{x}_{1t} \ \dots \ \mathbf{x}_{rt}]$, $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$, now $\mathbf{F} = \text{diag}(\mathbf{F}_1, \dots, \mathbf{F}_r)$ and $\mathbf{G} = \text{diag}(\mathbf{g}_1, \dots, \mathbf{g}_r)$. The $\boldsymbol{\varepsilon}_t$ are serially uncorrelated normally distributed random vectors with mean $\mathbf{0}$ and a time invariant variance matrix \mathbf{V} . 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 a block diagonal structure.
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 $\mathbf{F}_i = \mathbf{F}(\boldsymbol{\theta}_i)$ and $\mathbf{g}_i = \mathbf{g}(\boldsymbol{\theta}_i)$ where \mathbf{F} and \mathbf{g} are functions with the properties: (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 directly modelled through the states but through contemporaneously correlated innovations.

This normalized state space model is said here to have a block-diagonal structure, 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). Its sub-models are said to have a common structure if $\mathbf{F}_1 = \dots = \mathbf{F}_r$ and $\mathbf{g}_1 = \dots = \mathbf{g}_r$, something that occurs when $\boldsymbol{\theta}_1 = \dots = \boldsymbol{\theta}_r$.

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

$$\tilde{\mathbf{y}}'_t = \mathbf{1}'\tilde{\mathbf{X}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}'_t \quad (\text{A13})$$

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{F}}\tilde{\mathbf{x}}_{t-1} + \tilde{\mathbf{G}}\tilde{\boldsymbol{\varepsilon}}_t \quad (\text{A14})$$

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

Theorem 2

The normalized models (A11)-(A12) and (A13)-(A14) 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 establishing with Equation (A10), that both multivariate models are compatible if and only if

$$\tilde{\boldsymbol{\varepsilon}}_t = \mathbf{A}\boldsymbol{\varepsilon}_t \quad (\text{A15})$$

$$\tilde{\mathbf{X}}_t = \mathbf{X}_t\mathbf{A}' \quad (\text{A16})$$

$$\tilde{\mathbf{F}} = \mathbf{B}\mathbf{F}\mathbf{B} \quad (\text{A17})$$

$$\tilde{\mathbf{G}} = \mathbf{B}\mathbf{G}\mathbf{A} \quad (\text{A18})$$

$$\tilde{\mathbf{V}} = \mathbf{A}\mathbf{V}\mathbf{A}' \quad (\text{A19})$$

where $\mathbf{B} = \mathbf{A} \otimes \mathbf{I}$ arises when (A16) is vectorized to give $\tilde{\mathbf{x}}_t = \mathbf{B}\mathbf{x}_t$. It is then established that when

\mathbf{F} and \mathbf{G} are block-diagonal, $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{G}}$ can be block diagonal if and only if $\mathbf{F}_{11} = \dots = \mathbf{F}_{rr}$ and

$\mathbf{G}_{11} = \dots = \mathbf{G}_{rr}$. To illustrate this aspect of the proof, consider the right hand side of condition (A18)

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, (A18) 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 conditions (A17) and

(A18) for any positive value of r .

Corollary 1

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

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 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 and so the likelihood functions for both models are equal when $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$, $\tilde{\mathbf{X}}_0 = \mathbf{X}_0 \mathbf{A}'$ and $\tilde{\mathbf{V}} = \mathbf{A} \mathbf{V} \mathbf{A}'$. 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 normalized form as defined above. In general, not all measurement equation coefficients must equal one and sometimes they may be zero as in common additive seasonal models. The above theory can be adapted for this context by extending the definition of a normalized 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-normalized 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-normalized model can be converted to extended normalized 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 $w_{iq} = 0$ where $q = 1, \dots, r$ is 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 normalized form $y_{it} = \boldsymbol{\delta}'(\mathbf{W}_i \mathbf{x}_{i,t-1}) + \varepsilon_{it}$ and

$(\mathbf{W}_i \mathbf{x}_{it}) = (\mathbf{W}_i \mathbf{F}_i \mathbf{W}_i^{-1})(\mathbf{W}_i \mathbf{x}_{i,t-1}) + (\mathbf{W}_i \mathbf{g}_i) \varepsilon_{it}$. 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.

Anderson, B. D. O. & Moore, J. B. (1979), Optimal filtering Englewood Cliffs, New Jersey, Prentice-Hall.

Armstrong, J. S. & Green, K. C. (2007). Competitor-oriented objectives: The myth of market share. International Journal of Business, 12, 115-134.

Barceló-Vidal, C., Aguilar, L. & Martín-Fernández, J. A. (2011). Compositional VARIMA Time Series. In Pawlowsky-Glahn, V & Buccianti, A. (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 Armstrong, J. S. (ed.) Principles of Forecasting. Boston and Dordrecht, Kluwer, pp. 597-611.

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. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22, 679–688.

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.

Quintana, J. M. & West, M. (1988). Time series analysis of compositional data. In: Bernardo, J. M., 31

de Groot, M. H., Lindley, D. V., & Smith, A. F. M. (eds.), Bayesian Statistics 3, pp. 747–756, New York, Oxford University Press.

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.