Time Series of Continuous Proportions

Gary K. Grunwald; Adrian E. Raftery; Peter Guttorp

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 55, No. 1 (1993), 103-116.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281993%2955%3A1%3C103%3ATSOCP%3E2.0.CO%3B2-U

J. R. Statist. Soc. B (1993)
55, *No.* 1, *pp.* 103–116

Time Series of Continuous Proportions

By GARY K. GRUNWALD†,
University of Melbourne, Parkville, Australia

ADRIAN E. RAFTERY and PETER GUTTORP
University of Washington, Seattle, USA

[Received June 1989. Final revision September 1991]

SUMMARY

A vector of continuous proportions consists of the proportions of some total accounted for by its constituent components. An example is the proportions of world motor vehicle production by Japan, the USA and all other countries. We consider the situation where time series data are available and where interest focuses on the proportions rather than the actual amounts. Reasons for analysing such times series include estimation of the underlying trend, estimation of the effect of covariates and interventions, and forecasting. We develop a state space model for time series of continuous proportions. Conditionally on the unobserved state, the observations are assumed to follow the Dirichlet distribution, often considered to be the most natural distribution on the simplex. The state follows the Dirichlet conjugate distribution which is introduced here. Thus the model, although based on the Dirichlet distribution, does not have its restrictive independence properties. Covariates, trends, seasonality and interventions may be incorporated in a natural way. The model has worked well when applied to several examples, and we illustrate with components of world motor vehicle production.

Keywords: BAYESIAN FORECASTING; COMPOSITIONAL DATA; DIRICHLET DISTRIBUTION; RECURSIVE UPDATING; STATE SPACE MODEL

1. INTRODUCTION

Time series of proportions, or compositions, arise in many areas of application. Such series are characterized by components which are positive and sum to 1 at each time. Examples include the breakdown of household consumption by type of item in successive household budget surveys (Aitchison, 1982), market shares in successive time periods, proportions of time spent on different activities by individuals, groups or animals in successive time periods, changes in species composition in lakes due to environmental insults (Guttorp, 1990) and changes in the chemical composition of rock samples taken from successively deeper layers, corresponding to more distant time (Chayes, 1971). Although the data constitute a multivariate time series, standard techniques such as multivariate autoregressive integrated moving average (ARIMA) modelling (Tiao and Box, 1981) and Kalman filtering (Kalman, 1960) are not applicable because of the positivity and constant sum constraints.

As a specific example, which is analysed in Section 3, consider the composition of world motor vehicle production for 1947–87 shown in Table 1. The total has grown very rapidly and has obscured the relative changes in the three sources (Japan, the USA and all other countries combined), which are often of interest.

The top graph in Fig. 1 shows all the series on a single graph. At each year, the ordinates of the points represent the proportion of production from Japan (lower

†*Address for correspondence*: Department of Statistics, University of Melbourne, Parkville, Victoria 3052, Australia.

0035–9246/93/55103 $2.00

TABLE 1
World motor vehicle production and percentage change in US gross national product

Year	*Japan*	*USA*	*Other*	G_t	*Year*	*Japan*	*USA*	*Other*	G_t
1947	11	4796	1047	−2.8	1968	4086	10820	13670	4.1
1948	20	5221	1324	3.9	1969	4675	10206	15148	2.4
1949	29	6244	1771	0.0	1970	5289	8284	16114	−0.3
1950	32	8006	2540	8.5	1971	5811	10672	17198	2.8
1951	38	6757	2657	10.3	1972	6294	11311	18191	5.0
1952	39	5562	2719	3.9	1973	7083	12682	19395	5.2
1953	50	7349	3099	4.0	1974	6552	10072	18109	−0.5
1954	70	6537	3618	−1.3	1975	6942	8987	17532	−1.3
1955	69	9204	4470	5.6	1976	7842	11497	19002	4.9
1956	111	6919	4622	2.1	1977	8514	12703	19729	4.7
1957	182	7220	5126	1.7	1978	9269	12900	20131	5.3
1958	188	5121	6045	−0.8	1979	9636	11480	20408	2.5
1959	263	6724	6883	5.8	1980	11043	8010	19443	−0.2
1960	482	7905	7990	2.2	1981	11180	7943	18107	1.9
1961	814	6653	7742	2.6	1982	10732	6985	18396	−2.5
1962	991	8197	9010	5.3	1983	11112	9205	19433	3.6
1963	1284	9109	10318	4.1	1984	11465	10924	19383	6.8
1964	1702	9308	11004	5.3	1985	12271	11651	20357	3.4
1965	1876	11138	11528	5.8	1986	12260	11335	21638	2.8
1966	2286	10396	12293	5.8	1987	12249	10910	22521	3.4
1967	3146	9024	11997	2.9					

†Sources, Motor Vehicle Manufacturers Association of the U.S. (1988) and US Government (1989); units are thousands of vehicles.

point) and the cumulative proportion from Japan and the USA together (higher point). The main feature is the striking growth of the Japanese share, from around $\frac{1}{3}$% to about a third of the total. There are also bumps every 5 or so years, especially in the last half of the series. These can be seen in Fig. 1 to be correlated with the movement of the US economy, measured by the percentage change G_t in US gross national product (GNP) and shown in the bottom graph of Fig. 1.

For analysis, the data are best thought of as a time series of vectors, each of which has positive components summing to 1. This sample space is called the simplex, and for three-component proportions a graph of the series in this two-dimensional set (Fig. 2) is often useful. Again, the trend described above is evident.

In this paper we develop a methodology for modelling, forecasting, estimating trends and seasonal effects, deseasonalizing and assessing the effects of covariates and interventions on time series of continuous proportions. We take a state space approach to model the multivariate series directly in the simplex. An advantage of our method is the direct interpretability of the results in terms of the original proportions. The approach is fairly easily implemented and most of our results are exact. In the few cases where approximation is necessary we obtained good results with the accurate approximations of Tierney and Kadane (1986). One by-product of our work is the development of two new distributions on the simplex, the Dirichlet conjugate (DC) distribution (2.5) and the DCD distribution defined in Section 2.1. These are based on the Dirichlet distribution, but generalize it to allow for dependence between the components.

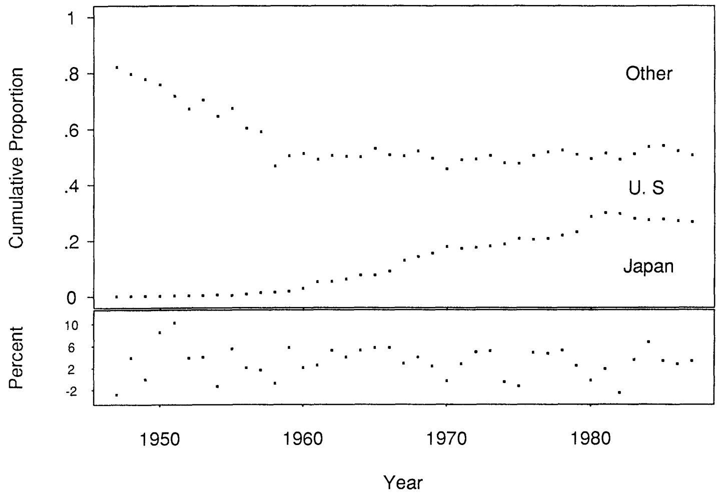

Fig. 1. Proportions of world motor vehicle production and percentage change in US gross national product

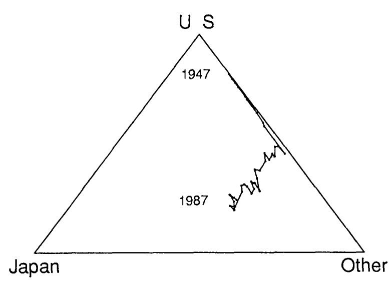

Fig. 2. Simplex plot of the proportions of world motor vehicle production

To our knowledge, the only other works concerning time series of continuous proportions in the multivariate setting ($d > 1$) are those of Smith and Brunsdon (1986), Brunsdon (1986) and Quintana and West (1988). These approaches are all based on the logistic normal distribution of Aitchison and Shen (1980) and Aitchison (1982, 1986) for the analysis of compositional data. In the first case, the authors apply multivariate ARIMA models (Box and Jenkins, 1976; Tiao and Box, 1981) to Aitchison's (1982) (asymmetric) log-ratios, whereas Quintana and West (1988) have used the symmetric log-ratios with their multivariate dynamic linear models.

Other proposals have been made by Azzalini (1984) and McKenzie (1985), who have studied time series of beta random variables, and by Wallis (1987) who has considered the use of the logistic transformation. However, these refer only to the univariate case.

A full comparison between these approaches and that proposed in this paper has yet to be made. However, the present approach appears to have the advantage of working with models that are based on the Dirichlet distribution, which many consider to be more natural for compositions, and of yielding results that are easily interpretable in terms of the odds ratios among the components of the composition.

One difficulty which arises no matter what approach is used is the problem of 0s. This is because in the Dirichlet distribution, as in the logistic normal, an exact 0 in one of the categories is an event of probability 0. In applications, however, we do encounter exact 0s, and they cannot be accommodated by any of the methods discussed here. One possible solution is to allow a singular component of the conditional distribution on the boundary of the simplex, with density proportional to the (non-singular) limiting conditional density at the boundary. This idea, which appears to be a refinement of an idea of Aitchison (1982) (end of section 7.4), may be useful more generally for the analysis of continuous proportions with 0s, outside the time series context.

In Section 2 we present the model and in Section 3 we illustrate its application to world motor vehicle production.

2. THE MODEL

In this section we review the Dirichlet distribution which describes the observations. We introduce and give some properties of the distribution conjugate to it, the DC distribution, which describes the state. We then define the state space model and show how it can incorporate covariates, trends, seasonality and interventions. Finally, we consider forecasting, estimation, model checking and model selection.

2.1. *Dirichlet Distribution and Dirichlet Conjugate Distribution*

Let $\mathbf{y} = (y_1, \ldots, y_{d+1})^{\mathrm{T}}$ be a vector of continuous proportions, namely a vector with positive components such that $\mathbf{y}^{\mathrm{T}}\mathbf{u}=1$ where $\mathbf{u} = (1, \ldots, 1)^{\mathrm{T}}$ is a $(d+1)$-vector of 1s. Then $\mathbf{y}$ follows the Dirichlet distribution if it has the density

$$p(\mathbf{y}\,|\,\boldsymbol{\alpha}) = D(\boldsymbol{\alpha})^{-1}\prod_{j=1}^{d+1} y_j^{\alpha_j-1}. \tag{2.1}$$

In density (2.1), $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{d+1})^{\mathrm{T}}$ where $\alpha_j > 0$ for $j=1, \ldots, d+1$ and

$$D(\boldsymbol{\alpha}) = \Gamma(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{u})^{-1}\prod_{j=1}^{d+1} \Gamma(\alpha_j)$$

is the Dirichlet function, a $(d+1)$-dimensional analogue of the beta function. We denote this situation by $\mathbf{y} \sim \mathrm{Dir}(\boldsymbol{\alpha})$. The sample space is the d-dimensional simplex $\mathbf{S}^d = \{\mathbf{y} \in \mathbf{R}_+^{d+1}:\mathbf{y}^{\mathrm{T}}\mathbf{u}=1\}$.

We write density (2.1) in exponential family form in the following way. Let $\mathbf{v}=\log \mathbf{y}$, $\bar{\nu}=\mathbf{v}^{\mathrm{T}}\mathbf{u}/(d+1)$ and $\mathbf{z}=\mathbf{v}-\mathbf{u}\bar{\nu}$. We call $\mathbf{z}$ the vector of symmetric log-ratios, and we write $\mathbf{z}=\mathrm{slr}(\mathbf{y})$. Also, let $\boldsymbol{\theta}=\boldsymbol{\alpha}/\tau$, where $\tau=\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{u}$, so that $\mathbf{y} \sim \mathrm{Dir}(\tau\boldsymbol{\theta})$. Then density (2.1) becomes

$$p(\mathbf{z}\,|\,\boldsymbol{\theta}, \tau) = \exp\{\tau\mathbf{z}^{\mathrm{T}}\boldsymbol{\theta}+\tau\bar{\nu}-\log D(\tau\boldsymbol{\theta})\}. \tag{2.2}$$

The sample space is $\mathbf{H}^d = \{\mathbf{z} \in \mathbf{R}^{d+1}:\mathbf{z}^{\mathrm{T}}\mathbf{u}=0\}$ and the parameter space is $(\boldsymbol{\theta}, \tau) \in$

$\mathbf{S}^d \times \mathbf{R}_+$. The purpose of this reparameterization is to separate the effects of location $\boldsymbol{\theta}$ and spread τ as far as possible.

The moments of the proportion vector $\mathbf{y}$ are

$$E[\mathbf{y}\,|\,\boldsymbol{\theta}, \tau] = \boldsymbol{\theta},$$

$$\text{var}[\mathbf{y}\,|\,\boldsymbol{\theta}, \tau] = \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}}/(\tau+1).$$

Thus $\boldsymbol{\theta}$ determines the location of the distribution of $\mathbf{y}$ in the simplex, and τ affects only the dispersion. By exponential family theory the moments of $\mathbf{z}$ are

$$E[\mathbf{z}\,|\,\boldsymbol{\theta}, \tau] = \boldsymbol{\Psi} - \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{u}/(d+1), \tag{2.3}$$

$$\text{var}[\mathbf{z}\,|\,\boldsymbol{\theta}, \tau] = \{(\boldsymbol{\Psi}'^{\mathrm{T}}\mathbf{u})(\mathbf{u}\mathbf{u}^{\mathrm{T}})/(d+1) + (d+1)\,\text{diag}(\boldsymbol{\Psi}') - \boldsymbol{\Psi}'\mathbf{u}^{\mathrm{T}} - \mathbf{u}\boldsymbol{\Psi}'^{\mathrm{T}}\}/(d+1). \tag{2.4}$$

In equations (2.3) and (2.4), $\boldsymbol{\Psi} = \psi(\tau\boldsymbol{\theta})$ and $\boldsymbol{\Psi}' = \psi'(\tau\boldsymbol{\theta})$, where we adopt the convention that $\psi(\mathbf{w}) = (\psi(w_1), \ldots, \psi(w_{d+1}))^{\mathrm{T}}$ and $\psi'(\mathbf{w}) = (\psi'(w_1), \ldots, \psi'(w_{d+1}))^{\mathrm{T}}$, $\mathbf{w}$ being any positive $(d+1)$-vector, ψ the digamma function $\psi(w) = \mathrm{d}\{\log \Gamma(w)\}/\mathrm{d}w$ and ψ' the trigamma function $\psi'(w) = \mathrm{d}\psi(w)/\mathrm{d}w$.

A family of conjugate prior distributions for $\boldsymbol{\theta}$, conditional on τ, is

$$p(\boldsymbol{\theta}\,|\,\sigma, \boldsymbol{\kappa}, \tau) \propto \exp[\sigma\{\tau\boldsymbol{\kappa}^{\mathrm{T}}\boldsymbol{\theta} - \log D(\tau\boldsymbol{\theta})\}], \tag{2.5}$$

where $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{d+1})^{\mathrm{T}}$. We denote this situation by $\boldsymbol{\theta} \sim \mathrm{DC}(\sigma, \boldsymbol{\kappa}, \tau)$; here DC stands for 'Dirichlet conjugate'. The state space is $\boldsymbol{\theta} \in \mathbf{S}^d$ and the parameter space is $(\sigma, \boldsymbol{\kappa}) \in \mathbf{R}_+ \times \mathbf{H}^d$. Because $\boldsymbol{\theta} \in \mathbf{S}^d$, this is a distribution on the simplex which does not appear to have been written down before. The mode $\hat{\boldsymbol{\theta}}$ of distribution (2.5) satisfies the equation

$$\psi(\tau\hat{\boldsymbol{\theta}}) - \psi(\tau)\mathbf{u} = \boldsymbol{\kappa}. \tag{2.6}$$

Using equation (2.6), $\hat{\boldsymbol{\theta}}$ is readily found by Newton–Raphson iteration.

If expressions (2.2) and (2.5) hold we say that $\mathbf{y}$ follows the compound Dirichlet-conjugate–Dirichlet (DCD) distribution, and we denote this by $\mathbf{y} \sim \mathrm{DCD}(\sigma, \boldsymbol{\kappa}, \tau)$. This is also a new distribution on the simplex. It follows from theorem 2 of Diaconis and Ylvisaker (1979) that, if $\mathbf{y} \sim \mathrm{DCD}(\sigma, \boldsymbol{\kappa}, \tau)$, then

$$E[\mathbf{z}\,|\,\sigma, \boldsymbol{\kappa}, \tau] = \boldsymbol{\kappa}, \tag{2.7}$$

so that $\boldsymbol{\kappa}$ determines the location of the DCD distribution, and σ and τ affect the dispersion in different ways. The DCD distribution is a mixture of Dirichlet distributions; τ is a common dispersion parameter for the individual Dirichlet distributions being mixed, while σ is a dispersion parameter for the DC mixing distribution.

2.2. *State Space Modelling*

State space modelling of time series goes back at least to Kalman (1960). For univariate normal observations, Harrison and Stevens (1976) developed a dynamic linear modelling strategy for a time series $w_1, w_2, \ldots$ which, in its simplest form, is based on the steady model

$$w_t = \phi_t + \epsilon_t, \tag{2.8a}$$

$$\phi_t = \phi_{t-1} + \delta_t, \tag{2.8b}$$

where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, $\delta_t \sim N(0, \sigma_\delta^2)$ and all the ϵ_t and δ_t are independent.

Equation (2.8) is a special case of the general state space model to which the Kalman filter applies and is based on the idea that w_t is made up of an unobserved random walk component ϕ_t and a noise component ϵ_t. The state ϕ_t may be recursively estimated by using the Kalman filter equations, leading to an updated posterior density for ϕ_t, $p(\phi_t | w^t)$, which is $N(m_t, C_t)$, and a predictive density for ϕ_{t+1}, $p(\phi_{t+1} | w^t)$, where $w^t = (w_1, \ldots, w_t)$. Equation (2.8) describes a random walk with observation error and with appropriate initialization is formally equivalent to an ARIMA(0, 1, 1) model, but seems more directly interpretable.

Smith (1979) considered the problem of generalizing model (2.8) to non-Gaussian situations. For the general case where $(w_t | \phi_t)$ has an exponential family distribution, analogies with the Gaussian Kalman filter led him to suggest as a non-Gaussian analogue of model (2.8) the power steady model, namely

$$\phi_t | w^t \sim \mathrm{CP}(\omega_t), \tag{2.9a}$$

$$p(\phi_{t+1} | w^t) \propto p(\phi_t | w^t)^k, \tag{2.9b}$$

where CP denotes the conjugate prior for the exponential family distribution of $(w_t | \phi_t)$ and $0 < k < 1$. He pointed out that model (2.9) satisfies the requirements that decisions made about the state ϕ_t at times t and $t+1$ be the same, and that the uncertainty associated with such decisions increases as time moves from t to $t+1$. Harvey and Fernandes (1989) wrote down the likelihood for model (2.9) and extended it to include explanatory variables in the Poisson, negative binomial and multinomial cases.

In model (2.9) the state transition distribution $p(\phi_{t+1} | \phi_t)$, analogous to equation (2.8b), is not defined except in some special cases, and Key and Godolphin (1981) have investigated some of the consequences of this. Smith and Miller (1986) and Smith (1990) have argued that this is not a defect of the model since equations (2.9) suffice to give the joint distribution of $(w_1, \ldots, w_n)$ and the predictive distribution of any set of future observations. Indeed, the correctness of an assumed form for the state transition distribution, such as equation (2.8b), cannot be verified from the data.

Model (2.9) applies to univariate observations, whereas our problem is multivariate. When w_t and ϕ_t are vectors, Smith (1981) proposed the symmetric multivariate power steady model, which remains defined by equations (2.9). However, for the Dirichlet observation distribution (2.1), when α is the state variable, this model has several unsatisfactory properties (Grunwald, 1987). For example, the dispersion of the forecast distribution *decreases* as forecasts are made further into the future. This is because equation (2.9b) increases the dispersion of the distribution of the state α, thus putting more weight on larger values of α, which correspond to less dispersed distributions of $\mathbf{y}$. These difficulties result from the attempt to estimate the dispersion at the same time as the location, a much more difficult problem that has required a large amount of effort even in the Gaussian case. Similar remarks apply to the generalizations of the symmetric multivariate power steady model proposed by Smith (1981, 1988).

Our solution to the problem is to use form (2.2) of the Dirichlet observation distribution instead of form (2.1) and to update the location $\boldsymbol{\theta}_t$ conditionally on the spread τ_t, which is updated separately. This appears to yield satisfactory results.

2.3. *Dirichlet Time Series Model*

Consider a time series $\{\mathbf{y}_t : t=1, \ldots, T\}$ of continuous proportions, where $\mathbf{y}_t = (y_{1t}, \ldots, y_{d+1,t})^{\mathrm{T}}$ $(t=1, \ldots, T)$. The basic assumption of the state space model is that there is an unobserved state $\boldsymbol{\theta}_t$ such that

$$(\mathbf{y}_t | \boldsymbol{\theta}_t, \tau_t) \sim \mathrm{Dir}(\tau_t \boldsymbol{\theta}_t). \tag{2.10}$$

Equation (2.10) is called the observation equation. The state $\{\boldsymbol{\theta}_t\}$ is assumed to evolve in time according to the steady model (2.9), namely

$$p(\boldsymbol{\theta}_{t+1} | \mathbf{D}_t) \propto \{p(\boldsymbol{\theta}_t | \mathbf{D}_t)\}^{\gamma} \qquad (0<\gamma<1). \tag{2.11}$$

In expression (2.11), the observed history $\mathbf{D}_t$ is defined recursively by $\mathbf{D}_t = \{I_t, \mathbf{D}_{t-1}\}$ where, for $t \geqslant 1$, $I_t = \{\mathbf{y}_t$, all other relevant information available at time t but not $t-1\}$, and we denote by $\mathbf{D}_0$ the values of the externally estimated parameters and all relevant information available at time $t=0$. Equation (2.11) has the property that the distribution of $(\boldsymbol{\theta}_{t+1} | \mathbf{D}_t)$ has mode unchanged from that of $(\boldsymbol{\theta}_t | \mathbf{D}_t)$ but has greater dispersion.

Knowledge about the state $\boldsymbol{\theta}_t$ given data is specified by the standard recursive updating scheme, which follows from Bayes's theorem and yields

$$(\boldsymbol{\theta}_s | \mathbf{D}_t) \sim \mathrm{DC}(\sigma_{s|t}, \boldsymbol{\kappa}_{s|t}, \tau_t) \tag{2.12}$$

for $s=t$ or $s=t+1$, provided that the initial distribution of $(\boldsymbol{\theta}_1 | \mathbf{D}_0)$ is also DC. The recursion starts with $p(\boldsymbol{\theta}_t | \mathbf{D}_t)$ and consists of two steps. The first step, called the *prediction step*, consists of obtaining $p(\boldsymbol{\theta}_{t+1} | \mathbf{D}_t)$ from expressions (2.11) and (2.12). The second step, called the *updating step*, consists of obtaining $p(\boldsymbol{\theta}_{t+1} | \mathbf{D}_{t+1})$ by using expression (2.10) and Bayes's theorem.

The prediction step reduces to

$$\sigma_{t+1|t} = \gamma \sigma_{t|t}, \tag{2.13}$$

$$\boldsymbol{\kappa}_{t+1|t} = \boldsymbol{\kappa}_{t|t}, \tag{2.14}$$

where the notation is defined by distribution (2.12). The updating step is

$$\sigma_{t+1|t+1} = \sigma_{t+1|t} + 1, \tag{2.15}$$

$$\boldsymbol{\kappa}_{t+1|t+1} = (1-g_{t+1})\boldsymbol{\kappa}_{t+1|t} + g_{t+1}\mathbf{z}_{t+1}, \tag{2.16}$$

where $\mathbf{z}_t = \mathrm{slr}(\mathbf{y}_t)$ as defined in Section 2.1, and $g_{t+1} = 1/\sigma_{t+1|t+1}$ is analogous to the gain in the usual Kalman filter. In the absence of specific prior information, the recursions may be initialized by setting $\sigma_{1|0}=0$ (and then $\boldsymbol{\kappa}_{1|0}$ is ignored), which specifies a uniform prior distribution on the simplex for $\boldsymbol{\theta}_1$.

It remains to specify τ_{t+1}. This is done by specifying the average conditional variance of the components of $\mathbf{z}_{t+1}$, conditional on the predicted state $\hat{\boldsymbol{\theta}}_{t+1}$, to be constant over time. This is analogous to the Gaussian Kalman filter, in which the variance of the observation distribution, conditional on the state, is assumed to be constant over time. Define a new model parameter ξ as follows. The average conditional variance when $\tau_t = \xi$ and $\boldsymbol{\theta}_t = \mathbf{u}/(d+1)$ (the centre of the simplex) is, from equation (2.4),

$$\frac{d}{(d+1)^2}\psi'\left(\frac{\xi}{d+1}\right).$$

Let $\hat{\boldsymbol{\theta}}_{t+1}$ be the mode of $p(\boldsymbol{\theta}_{t+1} | \mathbf{D}_t)$. Then the constant variance assumption requires that, by equations (2.4) and (2.6), τ_{t+1} is the solution of the equation

$$\psi'(\tau_{t+1}\hat{\boldsymbol{\theta}}_{t+1})^{\mathrm{T}}\mathbf{u} = (d+1)\,\psi'\{\xi/(d+1)\}. \tag{2.17}$$

This one-dimensional equation is readily solved by Newton–Raphson iteration.

2.4. *Covariates, Trends, Seasonality and Interventions*

We incorporate independent variables by changing the location of the predictive state distribution $p(\boldsymbol{\theta}_{t+1} | \mathbf{D}_t)$ to take account of such information at time $t+1$. Let $\hat{\boldsymbol{\theta}}_t$ be the mode of $p(\boldsymbol{\theta}_t | \mathbf{D}_t)$. Then we define a new mode $\hat{\boldsymbol{\theta}}^*_{t+1}$ for the predictive state distribution by

$$g(\hat{\boldsymbol{\theta}}^*_{t+1}) = f_{\mathbf{B}}(\hat{\boldsymbol{\theta}}_t, \mathbf{x}_{t+1}), \tag{2.18}$$

where $\mathbf{x}_{t+1}$ is an r-vector of independent variables at time $t+1$ and $\mathbf{B}$ is the matrix of regression parameters. f and g are functions; g is similar to the link function in generalized linear models (McCullagh and Nelder, 1983).

Here we consider only a subset of the class of models defined by equation (2.18). This consists of models which work with $\hat{\boldsymbol{\theta}}_{t+1}$ on the symmetric log-ratio scale and treat the covariates linearly, namely

$$\mathrm{slr}(\hat{\boldsymbol{\theta}}^*_{t+1}) = \mathrm{slr}(\hat{\boldsymbol{\theta}}_t) + \mathbf{B}\mathbf{x}_{t+1}, \tag{2.19}$$

where the symmetric log-ratio is defined as in Section 2.1. The matrix $\mathbf{B}$ of regression parameters is $(d+1)\times r$, and each column must lie in the space $\mathbf{H}^d$ to ensure that $\mathbf{B}\mathbf{x}_{t+1} \in \mathbf{H}^d$ also. The recursion is still given by equations (2.13)–(2.17) with the exception that, using equation (2.6), $\boldsymbol{\kappa}_{t+1|t}$ is now specified by

$$\boldsymbol{\kappa}_{t+1|t} = \psi(\tau_t\hat{\boldsymbol{\theta}}^*_{t+1}) - \psi(\tau_t)\mathbf{u}, \tag{2.20}$$

so that equation (2.14) is replaced by equation (2.20). There is some similarity between this approach and the 'guide relations' used by West *et al.* (1985), though their approach is strictly univariate.

For interpretation in terms of the original proportions, this prediction at time $t+1$ can be written in terms of relative odds for two categories j and k as

$$\frac{\hat{\theta}^*_{t+1|t,j}}{\hat{\theta}^*_{t+1|t,k}} = \exp\{(\mathbf{B}\mathbf{x}_{t+1})_j - (\mathbf{B}\mathbf{x}_{t+1})_k\}\,\frac{\hat{\theta}^*_{t|t,j}}{\hat{\theta}^*_{t|t,k}}. \tag{2.21}$$

As will be seen in Section 3, this gives an easily described interpretation of the effects of independent variables.

Model (2.19) can represent trends, seasonality and interventions as well as covariates. For a constant linear trend (on the symmetric log-ratio scale), $\mathbf{x}_t = 1$ ($t=1, \ldots, T$). Seasonal effects may be represented by a set of dummy variables, one for each season, or by a set of deterministic periodic functions such as sinusoids. Given an estimated seasonal effect, $\mathbf{s}_t$, at time t, a time series of continuous proportions may be deseasonalized, for example, by forming the quantities $\mathrm{slr}^{-1}\{\mathrm{slr}(\mathbf{y}_t) - \mathbf{s}_t\}$. An intervention may be represented by a dummy variable (Box and Tiao, 1975).

2.5. *Forecasting, Estimation, Model Checking and Model Comparison*

The predictive distribution of $\mathbf{z}_{t+1}$ given the past is

$$p(\mathbf{z}_{t+1} | \mathbf{D}_t) = \int p(\mathbf{z}_{t+1} | \boldsymbol{\theta}_{t+1}) p(\boldsymbol{\theta}_{t+1} | \mathbf{D}_t) \, \mathrm{d}\boldsymbol{\theta}_{t+1}. \tag{2.22}$$

This density is the basis for forecasting since it no longer conditions on the unobservable state $\boldsymbol{\theta}_{t+1}$. Although no analytical expression for the density in equation (2.22) seems available, we have had good results with the approximation of Tierney and Kadane (1986), section 3, which is both fast and accurate in this case. The forecast mean is, from equation (2.7), $\boldsymbol{\kappa}_{t+1|t}$. (As an example of the accuracy of the approximations, the Tierney and Kadane approximation reproduced the theoretically calculated forecast means to at least three significant figures in the example of Section 3.)

The external parameters, or hyperparameters, in the model are γ, ξ and, if there are independent variables, $\mathbf{B}$. Given data at times $t=1, \ldots, T$, these may be estimated by maximum likelihood. The log-likelihood is

$$L(\gamma, \xi, \mathbf{B}) = \sum_{t=2}^{T} \log p(\mathbf{z}_t | \mathbf{D}_{t-1}; \gamma, \xi, \mathbf{B}),$$

and this can be maximized numerically. The log-likelihood is a smooth function of γ, ξ and $\mathbf{B}$ provided that $\mathbf{B}$ is written in terms of the rd independent parameters that it contains, since each column is constrained to sum to 0. Thus, standard arguments and similar results for the hyperparameters of other, Gaussian, linear models (e.g. Pagan (1980) and Los (1985)) suggest the maximum likelihood estimator to be asymptotically normal with the usual limiting distribution. The sum beginning at 2 gives a log-likelihood conditional on the first observation, as suggested by Harvey and Fernandes (1989) in the non-Gaussian setting. Confidence intervals for estimates of parameters in $\mathbf{B}$ are obtained by using the usual large sample approximations.

To compare models involving different covariates, we prefer to use an approximation to the posterior odds as a measure of evidence. We do not use the alternative approach of significance testing because the models are often non-nested and multiple comparisons are involved. Suppose that we have models M_i with covariates $\mathbf{x}_t^{(i)}$ of dimensions r_i ($i=0, 1$). Then, given that the maximum likelihood estimators of the hyperparameters have the usual limiting distribution, the arguments of Schwarz (1978) show that

$$-2 \log B_{01} \overset{P}{\sim} -2(L_1 - L_0) - d(r_1 - r_0) \log(Td),$$

where B_{01} is the posterior odds for M_0 against M_1, L_i is the maximized log-likelihood for M_i ($i=0, 1$), and $\overset{P}{\sim}$ denotes asymptotic equivalence in probability. If we are comparing several models, we thus prefer the model for which

$$\mathrm{BIC} = -2L + rd \log\{(T-1)d\}$$

is smallest. The rules of thumb of Jeffreys (1961) suggest that such a preference should not be decisive unless the smallest value of BIC is exceeded by the next smallest by at least $2 \log_e 100 = 9.2$.

The model can be checked by examining the standardized residuals

$$\begin{aligned} \mathbf{R}_t &= (\mathbf{z}_t - E[\mathbf{z}_t | \mathbf{D}_{t-1}]) / \mathrm{var}[\mathbf{z}_t | \mathbf{D}_{t-1}] \\ &= (\mathbf{z}_t - \boldsymbol{\kappa}_{t|t-1}) / \mathrm{var}[\mathbf{z}_t | \mathbf{D}_{t-1}]. \end{aligned} \tag{2.23}$$

The equality follows from equation (2.7), but a similar theoretical expression for $\mathrm{var}[\mathbf{z}_t | \mathbf{D}_{t-1}]$ does not seem to exist. We have again used the Tierney and Kadane (1986) approximation of this variance for standardization. Visual analysis of the residuals is important, and the use of dynamic interactive graphics is helpful (e.g. the spin command in S-Plus (Statistical Science, 1988), or the DataViewer (Hurley and Buja, 1988)). We shall give an example of residual analysis in the next section.

3. EXAMPLE

We now return to the world motor vehicle data described in Section 1. The results of fitting several of the models discussed in Section 2 to these data are shown in Table 2. For the steady model of Section 2.3, the maximum likelihood estimators are $\hat{\gamma} \approx 0$ with $\hat{\gamma}\hat{\xi} = 122$. The likelihood surface is a ridge aligned roughly along the latter curve. In the analogous normal Kalman filter steady model (random walk with observation error), $\gamma\xi$ is related to the limiting reciprocal forecast variance, and here this is well estimated. The parameter γ itself is not easily interpretable because of its relation to ξ through τ_t.

The standardized residuals of equation (2.23) show that the steady model does not fit well. For instance, the residuals for the component for Japan are nearly all positive because of the strong trend.

The model incorporating a constant time trend to account for this does better, with the smaller BIC indicating a significant improvement. Using equation (2.21), the quantitative information in the parameter estimates can be described in terms of the odds as follows. The ratio of the Japanese to US shares of production has increased, on average, by a factor of about $\exp(\hat{B}_{\mathrm{Japan},1} - \hat{B}_{\mathrm{US},1}) = 1.129$, or 12.9%, per year. A 95% confidence statement can be made by forming an interval

$$\hat{B}_{\mathrm{Japan},1} - \hat{B}_{\mathrm{US},1} \pm 1.96\,\mathrm{SE}(\hat{B}_{\mathrm{Japan},1} - \hat{B}_{\mathrm{US},1}),$$

with the standard error found as usual from $\mathrm{cov}(\hat{\mathbf{B}})$, and exponentiating these limits. The resulting lower and upper limits for the factor increase are 1.078 and 1.181, showing this effect to be very significant.

In addition to the underlying trend, Fig. 1 indicates that the general state of the

TABLE 2
Maximum likelihood estimates and BIC for three models

Parameter		*Model*	
	Steady	*Trend*	*Covariate*
γ	0	0.16	0.001
ξ	∞	1010	213000
$\gamma\xi$	122	162	213
$\beta_{\mathrm{Japan},1}$		+0.053	+0.052
$\beta_{\mathrm{USA},1}$		−0.068	−0.067
$\beta_{\mathrm{Other},1}$		+0.015	+0.015
$\beta_{\mathrm{Japan},2}$			−0.013
$\beta_{\mathrm{USA},2}$			+0.018
$\beta_{\mathrm{Other},2}$			−0.005
BIC	−383.5	−394.0	−407.5

US economy is also a factor in accounting for the relative components of production. We let G_t denote the percentage change in the US GNP in year t ($G_t = 100(\mathrm{GNP}_t - \mathrm{GNP}_{t-1})/\mathrm{GNP}_{t-1}$). Plots of the standardized residuals from the trend model against the first difference, $\nabla G_t = G_t - G_{t-1}$, of G_t show a roughly linear relation for all three components.

The inclusion of ∇G_t as a covariate again gives a significant improvement in the model. Quantitatively, a 1% increase in the growth rate from the previous year is associated with a (significant) change in the ratio of the Japanese to US production shares by a factor of about 0.969, with lower and upper limits of 0.955 and 0.984.

Table 3 shows factors for other ratios for the full covariate model. The trend effects

TABLE 3
Change factors for the covariate model

Ratio	*Factor change per year*	*Lower limit*	*Upper limit*	*Factor change per 1% GNP*	*Lower limit*	*Upper limit*
Japan to USA	1.126	1.075	1.181	0.969	0.955	0.984
Other to USA	1.085	1.059	1.113	0.977	0.971	0.983
Japan to other	1.038	0.992	1.086	0.992	0.978	1.006

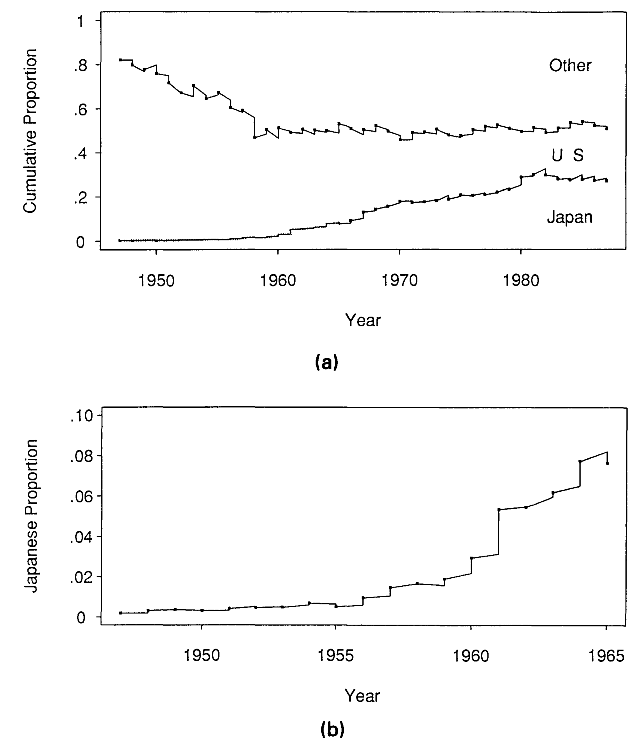

Fig. 3. (a) Proportions of world motor vehicle production with prior and posterior means; (b) Japanese proportion of world motor vehicle production with prior and posterior means

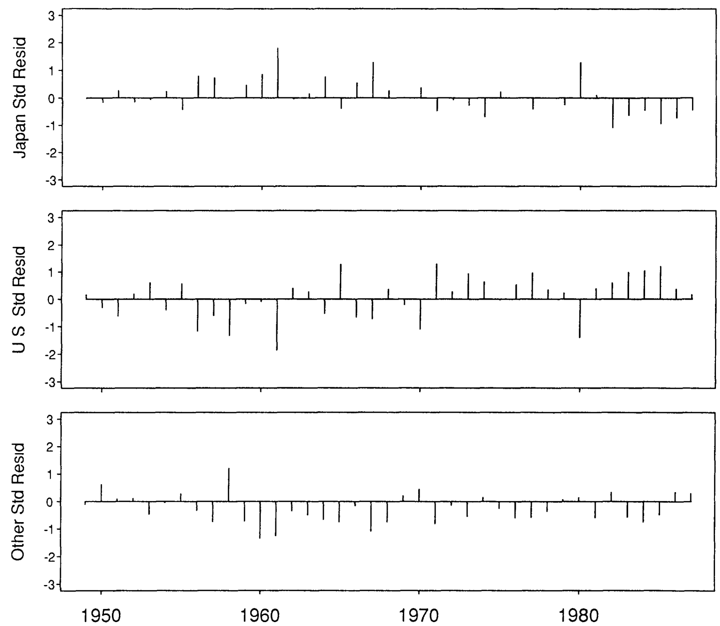

Fig. 4. Standardized residuals for the covariate model

change only slightly compared with the previous model because the trend and covariate terms are nearly orthogonal. The overall picture is that, on average, the USA has lost production share to the other countries and has lost even faster to Japan each year, but a good US economy helps the US against both.

Fig. 3 shows the prior and posterior state modes for the full model. The effect of GNP on the predictions is obvious and usually in the correct direction. Several poor predictions are also evident, and these can be studied on a more appropriate scale through the standardized residuals shown in Fig. 4. The largest residuals, in 1961 and in 1980, are not quite significant. Both Fig. 3 and Fig. 4 indicate that the model is performing well throughout a very wide range of the simplex.

The residual plots in Fig. 4 show some remaining patterns, indicating that gains from further modelling might be possible. Care is needed in interpreting the standardized residuals. In particular, they sum nearly to 0 at a given time, and hence some negative correlation is induced. Still, there is one huge correlation—at lag 0 the correlation between the residuals for Japan and the USA is -0.8. This negative correlation is also evident in Fig. 4. Thus, although the main competition in the trend and covariate model components was found to be between Japan and the USA, there is still some remaining competition in the unexplained variation. There is also a period of overprediction of the Japanese share and underprediction of the US share during the 1980s, due to a change in the trend effect. An extension of the present methods might thus include a dynamic trend.

ACKNOWLEDGEMENTS

The research of Grunwald was supported by National Science Foundation grant SES-8615541 and by a grant from the Weyerhaeuser Company. The research of

Raftery was supported by the same National Science Foundation grant and by Office of Naval Research contracts N-00014-88-K-0265 and N-00014-91-J-1074. The research of Guttorp was partially supported by a co-operative agreement between the United States Environmental Protection Agency and the Societal Institute for Mathematical Sciences. The authors are grateful to Michael J. Kahn, the Editor and two referees for helpful comments.

REFERENCES

Aitchison, J. (1982) The statistical analysis of compositional data (with discussion). *J. R. Statist. Soc.* B, **44**, 139–177.

——(1986) *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

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

Azzalini, A. (1984) A Markov process with Beta marginal distribution. *Statistica*, **44**, 241–243.

Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*, 2nd edn. San Francisco: Holden-Day.

Box, G. E. P. and Tiao, G. C. (1975) Intervention analysis with applications to economic and environmental problems. *J. Am. Statist. Ass.*, **70**, 70–79.

Brunsdon, T. M. (1986) Time series of compositional data. *PhD Thesis*. University of Southampton.

Chayes, F. (1971) *Ratio Correlation*. Chicago: University of Chicago Press.

Diaconis, P. and Ylvisaker, D. (1979) Conjugate priors for exponential families. *Ann. Statist.*, **7**, 269–281.

Grunwald, G. K. (1987) Time series models for continuous proportions. *PhD Dissertation*. Department of Statistics, University of Washington, Seattle.

Guttorp, P. (1990) Statistical analysis of biological monitoring data. *SIMS Technical Report 147*. Department of Statistics, University of Washington, Seattle.

Harrison, P. J. and Stevens, C. F. (1976) Bayesian forecasting (with discussion). *J. R. Statist. Soc.* B, **38**, 205–247.

Harvey, A. C. and Fernandes, C. (1989) Time series models for count or qualitative observations. *J. Bus. Econ. Statist.*, **7**, 407–422.

Hurley, C. and Buja, A. (1988) Analyzing high-dimensional data with motion graphics. *Technical Report STAT-88-03*. Department of Statistics and Actuarial Science, University of Waterloo.

Jeffreys, H. (1961) *Theory of Probability*, 3rd edn, appendix B. Oxford: Clarendon.

Kalman, R. E. (1960) A new approach to linear filtering and prediction problems. *ASME J. Bas. Engng* D, **82**, 33–45.

Key, P. B. and Godolphin, E. J. (1981) On the Bayesian steady forecasting model. *J. R. Statist. Soc.* B, **43**, 92–96.

Los, C. A. (1985) Comment. *J. Am. Statist. Ass.*, **80**, 92–93.

McCullagh, P. and Nelder, J. A. (1983) *Generalized Linear Models*. London: Chapman and Hall.

McKenzie, E. (1985) An autoregressive process for Beta random variables. *Mangmnt Sci.*, **31**, 988–997.

Motor Vehicle Manufacturers Association of the U.S. (1988) *World Motor Vehicle Facts and Figures*. Detroit: Motor Vehicle Manufacturers Association of the U.S.

Pagan, A. (1980) Some identification and estimation results for regression models with stochastically varying coefficients. *J. Econometr.*, **12**, 341–363.

Quintana, J. M. and West, M. (1988) Time series analysis of compositional data. In *Bayesian Statistics 3* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith). Oxford: Oxford University Press.

Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.

Smith, J. Q. (1979) A generalization of the Bayesian steady forecasting model. *J. R. Statist. Soc.* B, **41**, 375–387.

——(1981) The multiparameter steady model. *J. R. Statist. Soc.* B, **43**, 256–260.

——(1988) A comparison of the characteristics of some Bayesian forecasting models. *Technical Report 140*. Department of Statistics, University of Warwick, Coventry.

——— (1990) Non linear state space models with partially specified distributions on states. *J. Forecast.*, **9**, 137–150.
Smith, R. L. and Miller, J. E. (1986) A non-Gaussian state space model and application to prediction of records. *J. R. Statist. Soc.* B, **48**, 79–88.
Smith, T. F. M. and Brundson, T. M. (1986) Time series methods for small areas. *Technical Report*. University of Southampton.
Statistical Science (1988) *S-Plus Reference Manual*. Seattle: Statistical Science.
Tiao, G. C. and Box, G. E. P. (1981) Modeling multiple time series with applications. *J. Am. Statist. Ass.*, **76**, 802–816.
Tierney, L. and Kadane, J. B. (1986) Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Ass.*, **81**, 82–86.
US Government (1989) *Economic Report of the President to Congress*. Washington DC: US Government Printing Office.
Wallis, K. F. (1987) Time series analysis of bounded economic variables. *J. Time Ser. Anal.*, **8**, 115–123.
West, M., Harrison, P. J. and Migon, H. S. (1985) Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Am. Statist. Ass.*, **80**, 73–97.