Received: 1 July 2016 | Revised: 29 January 2017 | Accepted: 21 March 2017

DOI: 10.1002/for.2474

RESEARCH ARTICLE

WILEY

Prediction-based adaptive compositional model for seasonal time series analysis

Kun Chang[1] | Rong Chen[1] | Thomas B. Fomby[2]

[1]Department of Statistics and Biostatistics, Rutgers, the State University of New Jersey, Piscataway, NJ, USA

[2]Department of Economics, Southern Methodist University, Dallas, TX, USA

Correspondence
Rong Chen, Department of Statistics and Biostatistics, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA.
Email: rongchen@stat.rutgers.edu

Funding information
NSF, Grant/Award Number: DMS-1540863, DMS-1209085; Department of Homeland Security, Grant/Award Number: 2009-ST-061-CCI002-05 and 2009-ST-061-CCI002-06

Abstract

In this paper we propose a new class of seasonal time series models, based on a stable seasonal composition assumption. With the objective of forecasting the sum of the next ℓ observations, the concept of rolling season is adopted and a structure of rolling conditional distributions is formulated. The probabilistic properties, estimation and prediction procedures, and the forecasting performance of the model are studied and demonstrated with simulations and real examples.

KEYWORDS

adaptive forecasting, prediction, stable seasonal pattern

1 | INTRODUCTION

Seasonal time series are encountered in a wide range of applications. Traditionally, there are three general classes of seasonal time series models, namely the seasonal autoregressive integrated moving average (ARIMA) models (Box & Jenkins, 1994), the trend-and-seasonal models (Franzini & Harvey, 1983), and the stable seasonal pattern models (Chen & Fomby, 1999; Oliver, 1987). All these models provide different perspectives in dealing with seasonality. In particular, standard seasonal ARIMA models are in a multiplicative form, whereas trend-and-seasonal models are in an additive form. There is a vast literature on seasonal time series analysis and seasonal adjustment (e.g., Bell & Hillmer, 1984; Box & Jenkins, 1994; Cleveland & Tiao, 1976; Findley, Monsell, Bell, Otto, & Chen, 1998; Ghysels & Osborn, 2001; Zellner, 1978).

On the other hand, compositional models of Aitchison (1986) concentrate on the proportion of each component relative to the whole. Compositional models, by modeling ratios of proportions, successfully release the unit sum constraint, which makes it possible to apply standard statistical methodologies. This type of model has been used in many applications. For example, statistical analysis of percentages by weight of major oxides in rock specimens can be used to identify new types of rock specimens, as shown in a series of research topics by Thomas and Aitchison (2006). Another example is the study of budget patterns of a household reflected by the proportions of total expenditures allocated to several commodity groups. Aitchison (1986) analyzed such an example on five commodity groups.

Seasonality in a time series can often be viewed as a certain type of regular composition of seasons over time. For example, for a monthly time series with an annual seasonality, the 12 months can be seen as 12 components of the year (the composition), and the seasonality can be seen as a certain systematic distributive pattern of the measurements among 12 months with respect to the total measurement of the year. In the sales industry, the percentage of sales amount in each quarter out of the year is often stable across different years, whereas the yearly total may vary. Chen and Fomby (1999) touched upon this observation and introduced a stable seasonal pattern model, by assuming that the proportion (composition) of each part in a period remains the same (probability-wise) across seasons.

In this paper we introduce a class of seasonal time series models using the compositional principle to deal with seasonality. This class of models has the flexibility in adapting to different forecasting objectives. Tiao and Xu (1993) first proposed the adaptive idea using different estimation criteria for different forecasting horizons (objectives). This is a powerful idea and has been used by Tiao and Tsay (1994), who proposed an adaptive scheme to approximate certain long-memory processes, and by Tong (1997), who gave further discussions on adaptive procedures. Here we adopt this idea to adaptively choose different models for different forecasting objectives. Specifically, in this paper we consider the objective of forecasting the next ℓ-observations total for different ℓ in a seasonal time series. Such forecasting tasks are often encountered in many applications. For example, for certain industries, an accurate prediction of the total quantity (e.g., sales, production) for the next several months is important for better inventory management and marketing strategy (see, e.g., Kilger & Wagner, 2010).

The rest of the paper is organized as follows. In Section 2 we introduce a class of compositional seasonal time series models based on the theory of compositional analysis. Section 3 discusses the estimation, model checking and prediction procedures of the model. Section 4 contains several simulation and real data examples, including the forecasting comparison with standard seasonal ARIMA models.

2 | THE MODEL

Seasonal time series, in some sense, is a form of compositional data. Suppose we have a seasonal time series

$$X_1, X_2, \ldots, X_t, \ldots$$

with period d. Each seasonal cycle can be viewed as a basis. More specifically, the observations of $\{X_{t+1}, X_{t+2}, \ldots, X_{t+d}\}$ comprise the basis of a d-parts composition. This feature can be used to model the seasonal behavior of the time series.

The use of compositional analysis can be flexible. For example, in the seasonal time series analysis there are several different ways to construct the seasonal components. Given monthly observations, the annual total can also be viewed as the sum of a four-part composition of four quarters total, or the sum of a two-part composition of two semi-annual totals. In this paper, we concentrate on the objective of forecasting the next ℓ-observations with ℓ varying from 1 to d.

Under this objective, we are motivated to partition the seasonal total into a two-part composition that consisting of the sum of the first $d-\ell$ measurements and the remaining ℓ measurements within one cycle. The seasonal time series of X_t can be viewed as

$$\ldots; |\underbrace{X_{t-d+\ell+1}, \ldots, X_{t-1}, X_t}_{Y_{1,t}}, \underbrace{X_{t+1}, \ldots, X_{t+\ell}}_{Y_{2,t}}; |\underbrace{X_{t+\ell+1}, \ldots, X_{t+d-1}, X_{t+d}}_{Y_{1,t+d}}, \underbrace{X_{t+d+1}, \ldots, X_{t+d+\ell}}_{Y_{2,t+d}}; | \ldots.$$

Under this setting, at each time t a complete season is formed by the previous $d-\ell$ measurements inclusive, and the next ℓ measurements in a rolling basis. Then we can construct a rolling two-component $(d-\ell, \ell)$ partition:

$$Y_{1,t} = \sum_{i=0}^{d-\ell-1} X_{t-i} \quad \text{and} \quad Y_{2,t} = \sum_{i=1}^{\ell} X_{t+i}.$$

Following Aitchison (1986), we assume that the ratio $Y_{2,t}/Y_{1,t}$, conditioning on a set of exogenous variables $\boldsymbol{r_t} = (r_{1t}, r_{2t}, \ldots, r_{mt})'$ observed at time t, follows the log-normal distribution. That is,

$$Z_t|\boldsymbol{r_t} = \log\left(\frac{Y_{2,t}}{Y_{1,t}}\right)|\boldsymbol{r_t} \sim N\left(\mu_t + \boldsymbol{\beta}'\boldsymbol{r_t},\ \sigma_t^2\right), \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_m)'$ is the coefficient vector and μ_t is a series of unknown time-varying coefficients. Let $\varepsilon_t = Z_t - \mu_t - \boldsymbol{\beta}'\boldsymbol{r_t}$ and $\varepsilon_t = \sigma_t e_t$, where $e_t \sim N(0,1)$. Model 1 can be written as

$$Z_t = \mu_t + \boldsymbol{\beta}'\boldsymbol{r_t} + \sigma_t e_t.$$

Compared with the traditional d-parts compositional model, this $(d-\ell, \ell)$ partition avoids the excessive estimation of the high-dimensional parameters that are not essential in dealing with seasonality. Moreover, it is designed for the objective of forecasting the next ℓ-observations total. This provides a much simpler forecasting scheme following the objective-based adaptive model selection principle.

As $Z_t, Z_{t+d}, Z_{t+2d}, \ldots$ are constructed with the same partition of non-overlapping periods (i.e., $\sum_{i=0}^{d-\ell-1} X_{t-i+kd}$, $\sum_{i=1}^{\ell} X_{t+i+kd}$), it is reasonable to assume that this subseries is stationary with the same mean and variance. On the other hand, Z_t and Z_{t+1} are from different partitions and hence would have different mean and variance.

As a result, with period d, we assume

$$Z_t = \mu_{s(t)} + \boldsymbol{\beta}'\boldsymbol{r_t} + \sigma_{s(t)} e_t, \quad s(t) = t \bmod d. \tag{2}$$

The intercept $\mu_{s(t)}$ and error variance $\sigma_{s(t)}$ reflect the variation of the proportions in the season. Note that the time series $Y_{1,t}$ and $Y_{2,t}$ both consist of partial sums of overlapping windows, resulting in strong autocorrelations in Z_t. In addition, it is natural for a time series to possess serial correlations. To accommodate serial correlation beyond the seasonal components, we introduce an ARMA(p,q) structure to the standardized error in the compositional model. That is,

$$e_t = \frac{Z_t - \mu_{s(t)} - \boldsymbol{\beta}'\boldsymbol{r_t}}{\sigma_{s(t)}} = \frac{\theta(B)}{\phi(B)}a_t, \quad a_t \sim N\left(0, \sigma_a^2\right). \qquad (3)$$

Here B is the back-shift operator $BX_t = X_{t-1}$, and θ and ϕ are MA and AR polynomials:

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q, \ \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p.$$

Jointly, Equations 2 and 3 are referred to as a compositional seasonal component (CSC) time series model with $(d - \ell, \ell)$ partition, denoted as the CSC(ℓ) model.

Remark 1. *The major advantage of the CSC model is its flexibility in the assumptions of the process X_t. This model is designed to analyze the proportions of seasonal components instead of individual observations, allowing nonstationarity in the process X_t. Given the observed sequence X_t, the series Z_t can be obtained through $Y_{1,t}$ and $Y_{2,t}$ in our construction. Moreover, given a series of Z_t and a set of initial values of $X_1, \ldots, X_{d+\ell}$, the series of X_t can be reconstructed iteratively. The data-generating process (Equations 2 and 3) of Z_t also specifies the data-generating process of X_t, which is used in our simulation study.*

Remark 2. *There are several ways of using the CSC model for the prediction of the next ℓ-observations total. The next ℓ-observations total can be predicted by a $(d-\ell, \ell)$ partition in the CSC(ℓ) model. Alternatively, it can be done by the summation of the one to ℓ-step-ahead predictions from a CSC(1) model, a special class of CSC(ℓ).*

Remark 3. *By combining Equations 2 and 3, we have*

$$\phi(B)\log Y_{2,t} = \phi(B)\left(\log Y_{1,t} + \mu_{s(t)} + \boldsymbol{\beta}'\boldsymbol{r_t}\right) + \theta(B)\sigma_{s(t)}e_t.$$

This shows that this model is a special case of a transfer function model, with the sum of the preceding seasons as an input variable. For transfer function modeling, see, for example, Box and Jenkins (1994).

Remark 4. *In model 2, we have assumed that $var(e_t) = 1$. Such an assumption imposes complex constraints on the parameters in model 3. Instead, we will put a constraint on one of the $\sigma_{s(t)}$'s. Specifically, we assume $\sigma_0 = 1$. Such a reparametrization is equivalent to the original setting.*

Figure 1 shows a simulated series from the CSC(1) model with $d = 12$. Details of this series are given in Section 4.1. A strong seasonality is seen together with certain nonstationarity. Figure 2 shows the sample ACF and PACF of the series. It is seen that those features can be easily misspecified as a seasonal ARIMA model.

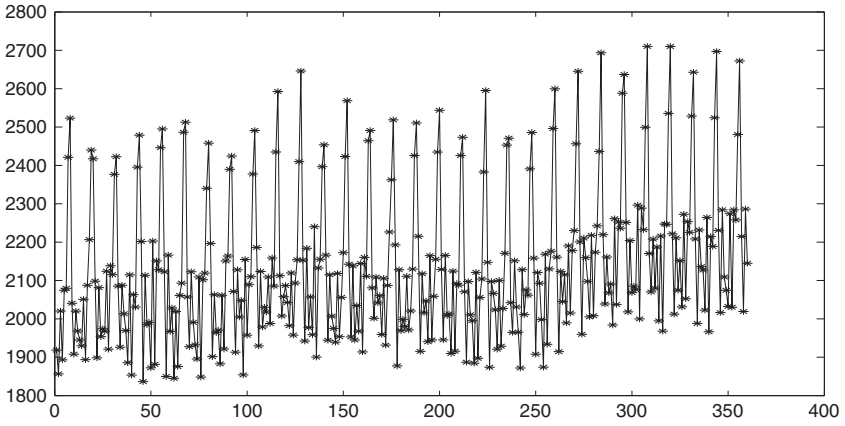

FIGURE 1 Original data plot of simulated series (I) from model CSC(1)

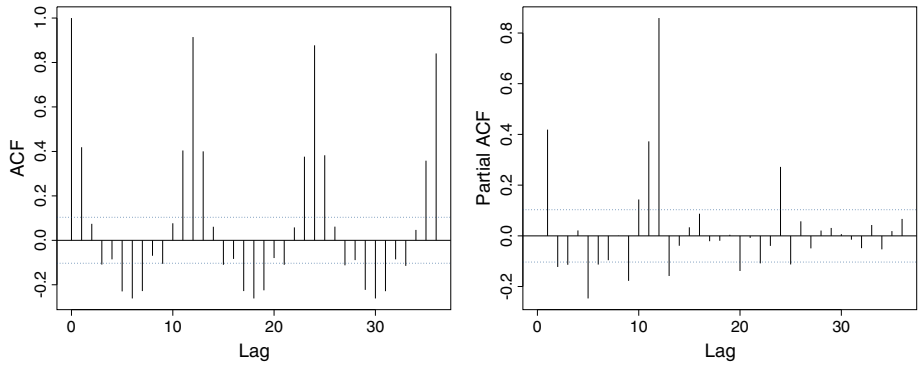

FIGURE 2 ACF and PACF of the simulated series (I) [Colour figure can be viewed at wileyonlinelibrary.com]

3 | ESTIMATION, MODEL CHECKING, AND PREDICTION

In this section, we discuss estimation, model checking, and prediction procedures for CSC models. The time series X_t is observed, and the deduced process Z_t in (1) is constructed through rolling partition as discussed in Section 2. We then estimate Model 3. Model checking is based on the residual analysis of Model 3 and out-of-sample performance of predicting the sum of the next ℓ X_t's.

3.1 | Estimation

Define a set of indicator variables $\delta_{j,t}$, $\delta_{j,t} = 1$ if $j = s(t)$ and $\delta_{j,t} = 0$ otherwise. Equation 2 can be rewritten as

$$Z_t = \sum_{i=1}^{m} \beta_i r_{i,t} + \sum_{j=0}^{d-1} \mu_j \delta_{j,t} + \sigma_{s(t)} e_t,$$

if there are m exogenous variables.

This is essentially a regression problem with time series errors that can be formulated as

$$Z_t = \boldsymbol{\beta}' \boldsymbol{r_t} + \boldsymbol{\mu}' \boldsymbol{\delta_t} + \sigma_{s(t)} e_t \quad \text{and} \quad e_t = \frac{\theta(B)}{\phi(B)} a_t,$$

with $a_t \sim N(0, \sigma_a^2)$ and $\sigma_0 = 1$. Conditional maximum likelihood estimation procedures can be used to estimate the parameters $\Theta = \left(\boldsymbol{\beta}, \boldsymbol{\mu}, \sigma_{s(t)}, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_a^2\right)$. Specifically, the likelihood function can be written as

$$\boldsymbol{L}(\Theta) = \left(2\pi\sigma_a^2\right)^{-\frac{T-p-d+1}{2}} \exp\left(- \sum_{t=p+d-\ell}^{T-\ell} \frac{\left(e_t - \phi_1 e_{t-1} - \cdots - \phi_p e_{t-p} - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}\right)^2}{2\sigma_a^2} \right),$$

where $\boldsymbol{e_t} = \frac{Z_t - \boldsymbol{\mu}'\boldsymbol{\delta_t} - \boldsymbol{\beta}'\boldsymbol{r_t}}{\sigma_{s(t)}}$ and a_t can be iteratively calculated by

$$\begin{aligned} a_t =& e_t - \phi_1 e_{t-1} - \ldots \phi_p e_{t-p} - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \\ t =& p + d - \ell, \ldots, T - \ell, \end{aligned}$$

conditional on the assumptions that $a_{p+d-\ell-1} = \cdots = a_{p+d-\ell-q} = 0$.

To find good initial values for MLE, we perform several iterations of the following steps:

1. Given variance σ_a^2 and ARMA coefficients for e_t, parameters in $(\boldsymbol{\beta}, \mu)$ can be estimated with standard regression estimators with known error covariance matrix. In the first iteration, we can assume that e_t are i.i.d.
2. Form residuals from the first step with estimated coefficient $\hat{\epsilon}_t = Z_t - \hat{\boldsymbol{\beta}}' \boldsymbol{r_t} - \hat{\boldsymbol{\mu}}' \boldsymbol{\delta_t}$. The seasonal residual variance can be estimated as

$$s_j^2 = \frac{1}{\lfloor T/d \rfloor} \sum_{k=1}^{\lfloor T/d \rfloor} \hat{\epsilon}_{j+kd}^2, \quad j = s(t) = 0, 1, \ldots, d-1,$$

where $\lfloor T/d \rfloor$ is the floor function and $\hat{\sigma}_j^2 = \frac{s_j^2}{s_1^2}$, $\hat{\sigma}_0^2 = 1$.

3. The ARMA coefficients $\hat{\phi}(B)$ and $\hat{\theta}(B)$ are estimated using the standardized residual time series $\{e_t : e_t = \hat{\epsilon}_t / \hat{\sigma}_{s(t)}\}$.

3.2 | Model checking

We focus on the following aspects of model checking and validation.

3.2.1 | Residual analysis

For model-building procedures that involve time series analysis, the residual autocorrelation analysis is an important step. Specifically, let

$$\hat{\mathrm{a}}_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} \hat{\mathrm{e}}_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} \frac{\hat{\epsilon}_t}{\hat{\sigma}_{s(t)}} = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} \frac{Z_t - \hat{\mu}_{s(t)} - \hat{\boldsymbol{\beta}}' \boldsymbol{r_t}}{\hat{\sigma}_{s(t)}}$$

be the estimated residual series for the CSC(ℓ) model. A standard white noise test such as the Box–Ljung test can be used. Normality tests such as marginal univariate distribution test, bivariate angle distribution test, and radius test can be used as well.

3.2.2 | Out-of-sample forecasting performance

With the objective of making predictions, the prediction procedure and performance measure are determined.

The out-of-sample rolling forecast is implemented for predicting the next ℓ-observations total $Y_{2,T} = \sum_{j=1}^{\ell} X_{T+j}$ for given observations $X_1, \ldots, X_T$. In the rolling forecast procedure, we define the starting point of the rolling forecast as K, and the value of $Y_{2,t}$ is predicted for each time t between $K+1$ and $T - \ell + 1$. The prediction can be done with least square criterion or least absolute deviation criterion, described in Section 3.3.

Denote the predicted value of $Y_{2,t}$ as $\hat{Y}_{2,t}$. We use the mean squared forecasting error (MSFE) as the performance measure, denoted as Q_ℓ:

$$Q_\ell = \frac{1}{T - \ell - K + 1} \sum_{j=K+1}^{T-\ell+1} (Y_{2,j} - \hat{Y}_{2,j})^2, \tag{4}$$

where ℓ varies from 1 to d. The criterion of Q_ℓ is measured for both the seasonal ARIMA model and the CSC models in Section 4.

3.3 | Prediction

Here we discuss the prediction procedure for the next ℓ-observations total $\sum_{j=1}^{\ell} X_{t+j}$, under both the least square criterion and the least absolute deviation criterion.

3.3.1 | Prediction under CSC(ℓ)model

The CSC(ℓ) model is relatively straightforward in forecasting the next ℓ-observations total. Suppose we currently have the observations of X_t up to time t, the prediction of the next ℓ-observations total $Y_{2,t} = \sum_{j=1}^{\ell} X_{t+j}$ can be realized by noting that $Y_{2,t} = Y_{1,t}\exp\{Z_t\}$, where $Y_{1,t}$ is observed at time t and Z_t can be predicted using the joint model (Equations 2 and 3). Note that at time t the time series Z_s is observed only up to time $t-\ell$. In other words, we will need to predict $Z_{t-\ell+1}, \dots, Z_t$.

Let $\mathcal{F}_t$ be the sigma field generated by $\{e_i, i = 1, \dots, t\}$. The least square prediction is the conditional mean

$$\begin{aligned}\hat{Y}_{2,t} &= \mathrm{E}\left[Y_{2,t}|Y_{1,t}, \mathcal{F}_{t-\ell}\right] = Y_{1t}\mathrm{E}\left[\exp\{Z_t\}|\mathcal{F}_{t-\ell}\right] \\ &= Y_{1,t}\mathrm{E}\left[\exp\{\mu_{s(t)} + \boldsymbol{\beta}'\boldsymbol{r_t} + \sigma_{s(t)}e_t\}|\mathcal{F}_{t-\ell}\right].\end{aligned}$$

Here $Y_{1,t}$ is completely known as of time t, and e_t is the random part that follows a stationary ARMA process with normal errors in Equation 3. Hence

$$\hat{Y}_{2,t} = Y_{1,t}\exp\left\{\mu_{s(t)} + \boldsymbol{\beta}'\boldsymbol{r_t} + \sigma_{s(t)}\hat{\mathrm{e}}_{t|t-\ell} + 0.5\sigma^2_{s(t)}\sigma^2_{t|t-\ell}\right\},$$

where $\hat{\mathrm{e}}_{t|t-\ell} = \mathrm{E}[e_t|\mathcal{F}_{t-\ell}]$ is the ℓ-step-ahead forecast from the ARMA process of e_t, and $\sigma^2_{t|t-\ell} = \mathrm{var}[e_t|\mathcal{F}_{t-\ell}]$ is the prediction variance.

On the other hand, the prediction under the absolute deviation criterion is the conditional median. Since the exponential function is a monotone function and, for normal distribution, the median equals the mean, we have

$$\tilde{Y}_{2,t} = \mathrm{median}\left[Y_{2,t}|Y_{1,t}, \mathcal{F}_{t-\ell}\right] = Y_{1,t}\exp\{\mu_{s(t)} + \boldsymbol{\beta}'\boldsymbol{r_t} + \sigma_{s(t)}\hat{\mathrm{e}}_{t|t-\ell}\},$$

where $\hat{\mathrm{e}}_{t|t-\ell}$ is the same ℓ-step-ahead prediction as above.

3.3.2 | Prediction under CSC(1)model

Alternatively, the CSC(1) model can be used for the prediction of next ℓ-observations total, and in certain cases it is more convenient than the CSC(ℓ) model and yields more accurate predictions in many cases. In this setting, the prediction of the next ℓ-observations total can be done by thez summation of each individual season that is predicted by the CSC(1) model. Specifically, the prediction of $Y_{2,t} = \sum_{j=1}^{\ell} X_{t+j}$ involves the prediction for each single X_{t+j}. Like all prediction models that involve exogenous variables, we need the future $r_{t+1}, \dots, r_{t+\ell}$ available at time t. There is no problem if $\boldsymbol{r_t}$ are deterministic such as a time trend or seasonal dummies; or $\boldsymbol{r_t}$ can be ℓ-lag observations of a process so $r_{t+1}, \dots, r_{t+\ell}$ are available at time t.

In the CSC(1) model,

$$X_{t+1} = Y_{1,t}\exp\left\{\mu_{s(t+1)} + \boldsymbol{\beta}'\boldsymbol{r_{t+1}} + \sigma_{s(t+1)}e_{t+1}\right\},$$

where $Y_{1,t} = X_{t-d+2} + \cdots + X_t$ is known up to time t. Under the least square criterion, the one-step forecast of X_{t+1} is the conditional expectation

$$\begin{aligned}\hat{X}_t(1) = \mathrm{E}\left[X_{t+1}|Y_{1,t}, \mathcal{F}_{t-1}\right] = Y_{1,t}\exp\Big\{&\mu_{s(t+1)} + \boldsymbol{\beta}'\boldsymbol{r_{t+1}} \\ &+ \sigma_{s(t+1)}\hat{\mathrm{e}}_{t+1|t} + 0.5\sigma^2_{s(t+1)}\sigma^2_{t+1|t}\Big\},\end{aligned}$$

where $\hat{\mathrm{e}}_{t+1|t}$ and $\sigma^2_{t+1|t}$ are the one-step prediction and its prediction variance under Model 3.

The two-step-ahead least square forecast is also the conditional expectation:

$$\begin{aligned}\hat{X}_t(2) &= \mathrm{E}\left[X_{t+2}|Y_{1,t}, \mathcal{F}_t\right] \\ &= \mathrm{E}\left[(X_{t-d+3} + \cdots + X_{t+1})\exp\left\{\mu_{s(t+2)}\right.\right. \\ &\quad \left.\left.+\boldsymbol{\beta}'\boldsymbol{r_{t+2}} + \sigma_{s(t+2)}\hat{\mathrm{e}}_{t+2}\right\}|\mathcal{F}_t\right] \\ &= \left(X_{t-d+3} + \cdots + \hat{X}_t(1)\right)\exp\left\{\mu_{s(t+2)}\right. \\ &\quad \left.+ \boldsymbol{\beta}'\boldsymbol{r_{t+2}} + \sigma_{s(t+2)}\hat{\mathrm{e}}_{t+2|t} + 0.5\sigma^2_{s(t+2)}\sigma^2_{t+2|t}\right\},\end{aligned}$$

which is based on the two-steo ahead forecast of e_{t+2}, as well as the one-step forecast of X_{t+1} from the previous step. Similarly, we can get the ℓth step forecast $\hat{X}_t(\ell)$:

$$\begin{aligned}\hat{X}_t(\ell) = &\left(X_{t-d+\ell+1} + \cdots + X_t + \hat{X}_t(1) + \cdots + \hat{X}_t(\ell-1)\right) \\ &\exp\left\{\mu_{s(t+\ell)} + \boldsymbol{\beta}'\boldsymbol{r_{t+\ell}} + \sigma_{s(t+\ell)}\hat{\mathrm{e}}_{t+\ell|t} + 0.5\sigma^2_{s(t+\ell)}\sigma^2_{t+\ell|t}\right\}.\end{aligned}$$

Then the prediction of $Y_{2,t}$ is $\hat{Y}_{2,t} = \hat{X}_t(1) + \cdots + \hat{X}_t(\ell)$.

Prediction is simpler under the absolute deviation criterion. The one-step prediction is

$$\begin{aligned}\tilde{X}_t(1) &= \mathrm{median}\left[X_{t+1}|Y_{1,t}, \mathcal{F}_t\right] \\ &= Y_{1,t}\exp\left\{\mu_{s(t+1)} + \boldsymbol{\beta}'\boldsymbol{r_{t+1}} + \sigma_{s(t+1)}\hat{\mathrm{e}}_{t+1|t}\right\},\end{aligned}$$

and the ℓth step prediction is

$$\begin{aligned}\tilde{X}_t(\ell) &= \mathrm{median}\left[X_{t+\ell}|Y_{1,t}, \mathcal{F}_t\right] \\ &= \left(X_{t-d+\ell+1} + \cdots + X_t + \tilde{X}_t(1) + \cdots + \tilde{X}_t(l-1)\right) \\ &\quad \exp\left\{\mu_{s(t+\ell)} + \boldsymbol{\beta}'\boldsymbol{r_{t+\ell}} + \sigma_{s(t+\ell)}\hat{\mathrm{e}}_{t+\ell|t}\right\},\end{aligned}$$

where $\hat{e}_{t+\ell|t}$ is the ℓ-step-ahead least square forecast of $e_{t+\ell}$ from the ARMA process.

4 | NUMERICAL EXAMPLES AND FORECASTING PERFORMANCE COMPARISON

Here we present two simulated examples and three real examples to demonstrate the predictive power of CSC models. We focus on out-of-sample forecasting performance comparisons between CSC(ℓ), CSC(1) and seasonal ARIMA models. In all our examples, the ARIMA models used for comparison are selected based on a detailed analysis, combining residual analysis, Akaike information criterion and out-of-sample rolling forecasting performance.

4.1 | Simulated example (I)

As described in Section 1, the data-generating process produces Z_t based on model 1, and the observations in X_t can be inferred from Z_t. The first simulated series in Figure 1 is generated by a CSC(1) process with white noise errors $e_t \sim N(0, 0.02^2)$. No exogenous variables are assumed in this simulation example.

Set $d = 12$, $\sigma_0 = \cdots = \sigma_{d11} = 1$ and

$$\boldsymbol{\mu} = (\mu_0, \ldots, \mu_{11})' = (-2.45 - 2.50 - 2.38 - 2.50 - 2.40 - 2.38 - 2.25 - 2.20 - 2.40 - 2.50 - 2.40 - 2.45)'.$$

The seasonal ARIMA model selected for comparison is

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_3 B^{12})\varepsilon_t$$

and the standard rolling forecast procedure is applied.

We perform out-of-sample predictions using the three models respectively and obtained Q_ℓ defined in Equation 4 for different prediction horizons, ℓ. The values of Q_ℓ are plotted in Figure 3. It can be seen that the ARIMA model performs worse than the CSC models for larger ℓ. As the true model, CSC(1) performs the best among the three models.

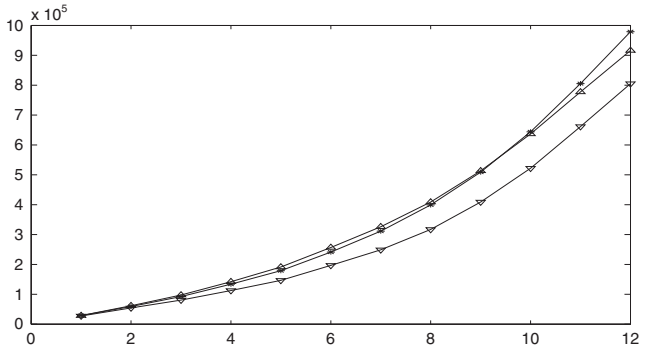

FIGURE 3 '*' denotes $Q_{\ell,\text{ARIMA}}$; '$\triangledown$' denotes $Q_{\ell,\text{CSC(1)}}$; '$\triangle$' denotes $Q_{\ell,\text{CSC}(\ell)}$

4.2 | Simulated example (II)

The second simulated series is generated from a CSC(1) model with AR(1)($\phi = 0.8$) errors. The simulation is based on the same seasonal mean and seasonal variance vectors as the previous example. We assume an AR(1) error process $e_t = 0.8e_{t-1} + a_t$ and the white noise process $a_t \sim N(0, 0.02^2)$. Figure 4 shows the time series plot. It shows certain nonstationarity and strong seasonality in the series. The ACF and PACF plots (Figure 5) are very similar to what we commonly see for seasonal time series, so that such series can be easily misspecified by seasonal ARIMA models. In addition, the seasonal pattern is clearly seen through the box plots shown in Figure 6.

According to various criteria including the rolling forecast performance, the best seasonal ARIMA model for the simulated series is

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \theta_2 B^{12})\varepsilon_t.$$

The values of Q_ℓ for seasonal ARIMA, CSC(1) + AR(1) and CSC(ℓ) + ARMA(p_ℓ, q_ℓ) models are listed in Table 1 and the rolling forecasts start at point $K = 168$. The CSC(ℓ) + ARMA(p_ℓ, q_ℓ) model gives a poor performance and the forecasting performance of the seasonal ARIMA is not as good as CSC(1) + AR(1).

We also compare the yearly-total prediction performance of seasonal ARIMA and CSC(1) + AR(1) by comparing the values of MSE at different lead times $\ell = 1, \ldots, 12$ (Table 2). Different from the rolling forecast of ℓ-observations total in Table 1, the yearly-total prediction is the sum of ℓ-forecasts, with the end-of-year forecast being the endpoint. For example, the yearly total is the total forecasts from July to December if $\ell = 6$.

The CSC(1) + AR(1) model gives better predictions than seasonal ARIMA for moderate prediction horizons.

4.3 | Real example (I)

In this real example, the monthly number of applications for a certain type of government benefit is analyzed. There are 167 observations in total. The series we show and analyze here is a transformed series for confidentiality reasons. It is transformed in such a way that the models used are not affected.

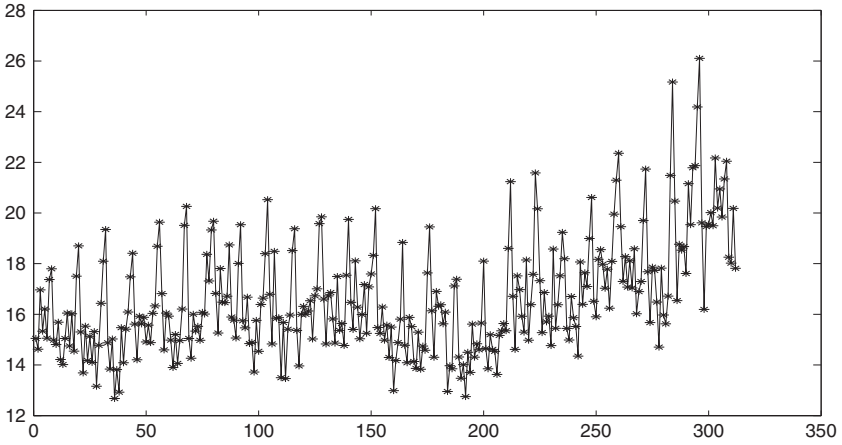

FIGURE 4 Original data plot of simulated series (II) from model CSC(1) + AR(1)

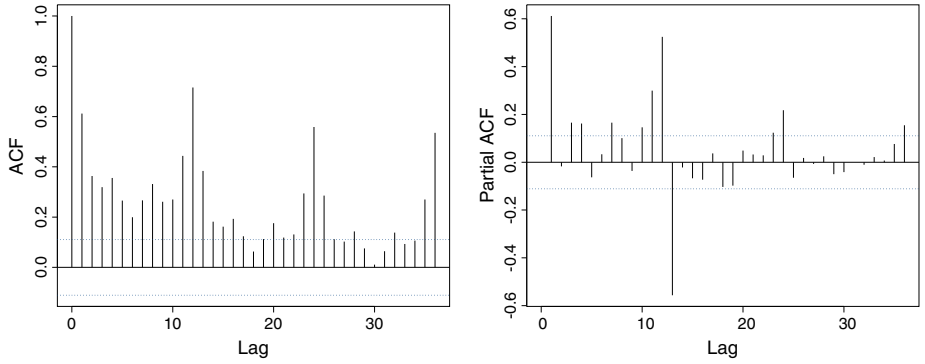

FIGURE 5 ACF and PACF of the simulated series (II) [Colour figure can be viewed at wileyonlinelibrary.com]

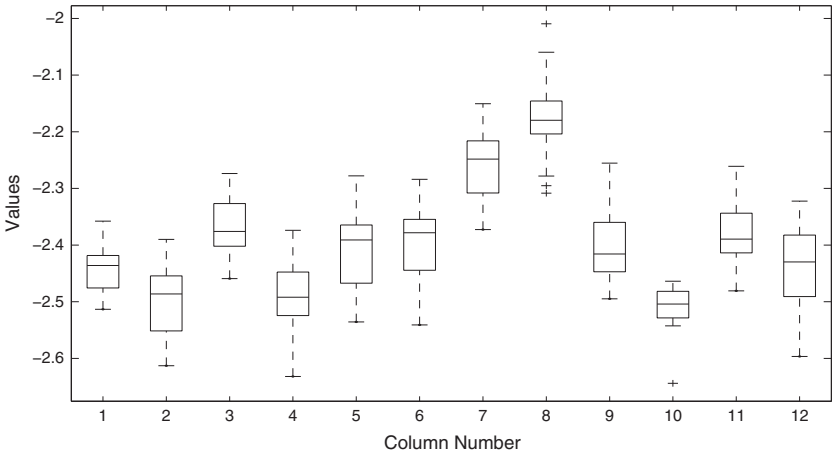

FIGURE 6 Box plots of log ratio for simulated series (II)

The prediction is performed for the total application volumes of the next ℓ months for $\ell = 1, 2, \ldots, 12$. The original series has known outliers at observations 105, 106, and 107. With the focus of seasonality in this paper, we smooth these three observations with historical means of the corresponding months in order to keep an objective discussion of the CSC model. The transformed series after outlier smoothing is shown in Figure 7. The data are analyzed by the seasonal ARIMA, CSC(ℓ) and CSC(1) model, respectively.

The series has a strong seasonality, as observed in the ACF and PACF plots (Figure 8). Beyond this, the application numbers for another type of benefit have a strong linear effect on the target series. We include the series as r_t in the seasonal ARIMA model. The events captured by r_t occurred many

TABLE 1 Comparison of forecasting performance

Lead ℓ	1	2	3	4	5	6	7	8	9	10	11	12
$Q_{\ell,A}$	0.89	3.88	8.67	15.34	24.45	34.81	45.69	57.27	71.66	91.07	115.45	145.84
$Q_{\ell,C(1)}$	0.84	3.56	7.82	13.78	22.27	32.29	43.24	55.29	70.32	89.79	114.15	144.09
$Q_{\ell,C(\ell)}$	0.88	4.08	9.44	18.05	28.97	39.65	48.80	51.44	70.30	102.06	130.05	187.99

$Q_{\ell,A}$, seasonal ARIMA model; $Q_{\ell,C(1)}$, CSC(1) + AR(1) model; $Q_{\ell,C(\ell)}$, CSC(ℓ) + ARMA(p_ℓ, q_ℓ) model for simulated series (II).

TABLE 2 Comparison of yearly-total forecasting performance for simulated series (II)

Lead ℓ	12	11	10	9	8	7	6	5	4	3	2	1
M_A	104.11	114.20	143.04	103.97	63.79	66.72	53.56	30.00	8.84	5.53	2.82	0.70
$M_{C(1)}$	107.26	117.47	138.62	104.49	59.34	60.24	49.86	29.04	7.78	5.60	3.11	0.66
Chge	3.03	2.87	−3.09	0.50	−6.97	−9.71	−6.91	−3.20	−11.99	1.34	10.16	−6.25

M_A, seasonal ARIMA; $M_{C(1)}$, CSC(1) + AR(1); 'Chge' denotes percentage change between the two.

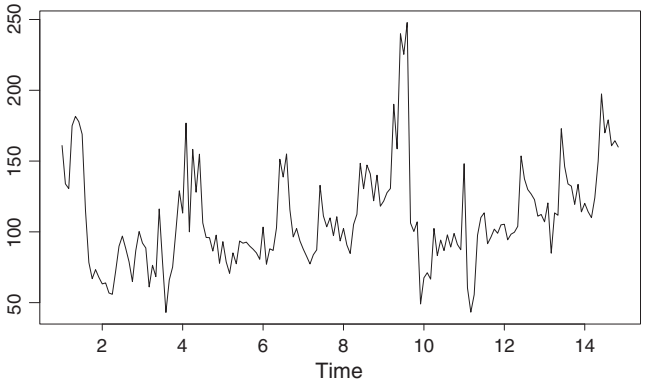

FIGURE 7 Time series plot of application volume for government benefit

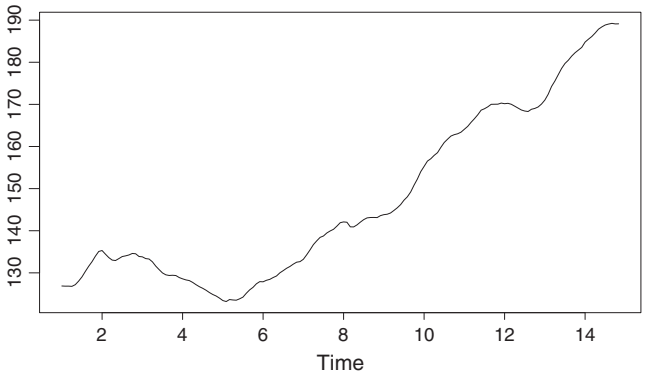

FIGURE 9 Time series plot of exogenous variable r_t

years ago, so we do have all the future r_t's available in our prediction exercise. Figure 9 shows the time series plot of r_t.

Based on model selection criteria, the following seasonal ARIMA model is analyzed for comparison:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^{12})(X_t + \beta r_t) = (1 + \Theta_1 B^{12})\varepsilon_t.$$

For CSC(ℓ), the components are constructed by the value of ℓ that varies from 1 to 12. Figure 10 gives the box plots of log-ratio series Z_t for $\ell = 3$ and $\ell = 6$, respectively. The patterns in the box plots show that the log-ratio series captures the seasonality in the original series by the seasonal-dependent mean $\mu_{s(t)}$ and variance $\sigma^2_{s(t)}$. Table 3 shows the estimation of the seasonal-dependent mean $\mu_{s(t)}$ and standard error $\sigma_{s(t)}$ for CSC(1) and CSC(3), respectively. The seasonality of the log-ratio series is clearly seen from the results, not only in the mean but also in the variance. In our prediction exercise, we use different d periods to construct

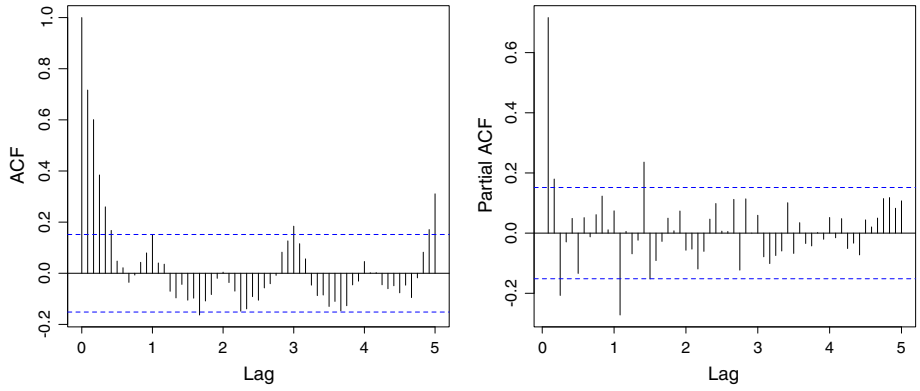

FIGURE 8 ACF and PACF of application volume [Colour figure can be viewed at wileyonlinelibrary.com]

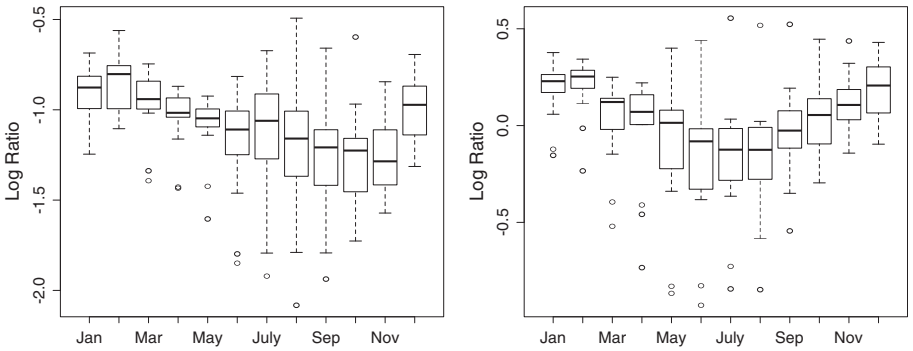

FIGURE 10 Box plots of log ratio for application volume: left, $\ell = 3$; right, $\ell = 6$

TABLE 3 Model estimation for CSC(1) (upper panel) and CSC(3) (lower panel)

$s(t)$	0	1	2	3	4	5	6	7	8	9	10	11
$\mu_{1,s(t)}$	−2.535	−2.449	−2.539	2.698	−2.477	−2.434	−2.095	−2.233	−2.289	−2.383	−2.443	−2.370
$\sigma_{1,s(t)}$	1	0.877	0.867	0.694	0.690	0.458	0.471	0.410	0.687	0.444	0.454	0.423
$\mu_{3,s(t)}$	−1.198	−1.169	−1.212	−1.270	−1.288	−1.245	−0.995	−0.900	−0.839	−0.968	−1.056	−1.100
$\sigma_{3,s(t)}$	1	1.134	1.255	1.031	0.997	0.732	0.615	0.500	0.544	0.610	0.578	0.620

$\mu_{s(t)}$ is the seasonal mean and $\sigma_{s(t)}$ is the seasonal standard deviation.

TABLE 4 Model estimation for AR structure of e_t

ℓ	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\phi}_1$	0.72	1.27	1.19	1.23	1.26	1.32	1.25	1.22	1.52	1.59	1.54	1.58
se($\hat{\phi}_1$)	0.056	0.079	0.079	0.077	0.077	0.078	0.077	0.077	0.067	0.062	0.066	0.063
$\hat{\phi}_2$	—	−0.63	−0.23	−0.15	−0.16	−0.30	−0.19	−0.18	−0.58	−0.65	−0.60	−0.64
$se(\hat{\phi}_2)$	—	0.119	0.123	0.125	0.127	0.130	0.127	0.124	0.067	0.062	0.066	0.063
$\hat{\phi}_3$	—	0.18	−0.17	−0.27	−0.29	−0.21	−0.26	−0.25	—	—	—	—
$se(\hat{\phi}_3)$	—	0.080	0.080	0.078	0.078	0.079	0.079	0.079	—	—	—	—

AR coefficients, $\hat{\phi}_j$; standard errors, se($\hat{\phi}_j$); '—' indicates that results are unavailabled{3,2} due to different AR structures.

TABLE 5 Comparison of forecasting performance for benefit application

Lead ℓ	1	2	3	4	5	6	7	8	9	10	11	12
$Q_{\ell,A}$	1.45	4.42	9.91	17.91	29.25	44.26	55.58	55.43	54.04	49.57	40.78	38.13
$Q_{\ell,C(\ell)}$	1.27	3.12	6.18	14.01	27.89	39.93	51.58	57.61	65.69	77.81	65.98	112.96
$Q_{\ell,C(1)}$	1.27	3.07	6.67	11.45	18.69	29.18	42.59	48.50	58.31	63.88	60.57	67.70
Chge$_{C(\ell)}$	−11.95	−29.40	−37.69	−21.77	−4.66	−9.78	−7.19	3.93	21.57	56.97	61.78	196.21
Chge$_{C(1)}$	−11.95	−30.56	−32.66	−36.06	−36.11	−34.07	−23.37	−12.50	7.90	28.87	48.51	77.54

$Q_{\ell,A}(\times 10^2)$, seasonal ARIMA model; $Q_{\ell,C(\ell)}(\times 10^2)$, CSC($\ell$) + AR($p_\ell$) model; $Q_{\ell,C(1)}(\times 10^2)$, CSC(1) + AR(1) model; 'Chge' denotes percentage change between the two.

the CSC model, so the base component $Y_{1,t}$ contains a sufficient number of observations. Specifically, we use $d = 12$ for $\ell = 1, \ldots, 8$ and $d = 24$ for $\ell = 9, \ldots, 12$. The standardized error process e_t follows strong autoregressive patterns, but the AR orders are not identical for different ℓ. Table 4 summarizes the coefficients and standard errors for the AR estimation of the error process.

At the same time, the prediction of the next ℓ-months total can be achieved by taking summation of ℓ-steps prediction from the CSC(1) model, as described in Section 3.3. The CSC(1) + AR(1) model is fitted based on the evident AR(1) structure of the error process.

The performance of the seasonal ARIMA, CSC(ℓ) and CSC(1) models are compared for the forecasting of the next ℓ-months total application volume. The rolling forecast starts from $K = 134$. Table 5 shows the square root of forecasting measure Q_ℓ of the models. From Table 5 it is seen that both CSC(ℓ) and CSC(1) outperform seasonal ARIMA significantly in relatively short-term prediction when $\ell = 1, 2, \ldots, 7$. They do not do as well as the ARIMA model in

TABLE 6 Comparison of forecasting performance for benefit application

Lead ℓ	1	2	3	4	5	6	7	8	9	10	11	12
$\text{MAFE}_{\ell,\mathcal{A}}$	7.40	5.78	6.42	6.33	6.41	6.63	6.39	5.73	4.87	4.35	3.83	3.57
$\text{MAFE}_{\ell,C_\ell}$	7.85	5.84	5.11	5.49	6.18	6.22	5.61	5.53	5.58	5.61	4.62	5.48
$\text{MAFE}_{\ell,C(1)}$	7.85	5.75	5.55	5.32	5.41	5.37	5.56	5.33	5.16	5.04	4.72	4.67
$\text{Chge}_{C(\ell)}$	6.06	1.06	−20.33	−13.28	−3.60	−6.07	−12.28	−3.47	14.48	29.01	20.68	53.67
$\text{Chge}_{C(1)}$	6.06	−0.62	−13.47	−15.90	−15.62	−18.88	−13.01	−7.11	5.81	15.81	23.19	30.76

$\text{MAFE}_{\ell,\mathcal{A}}$, seasonal ARIMA model; $\text{MAFE}_{\ell,C(\ell)}$, CSC($\ell$) + AR($p_\ell$) model; $\text{MAFE}_{\ell,C(1)}$, CSC(1) + AR(1) model; 'Chge' denotes percentage change between the two.

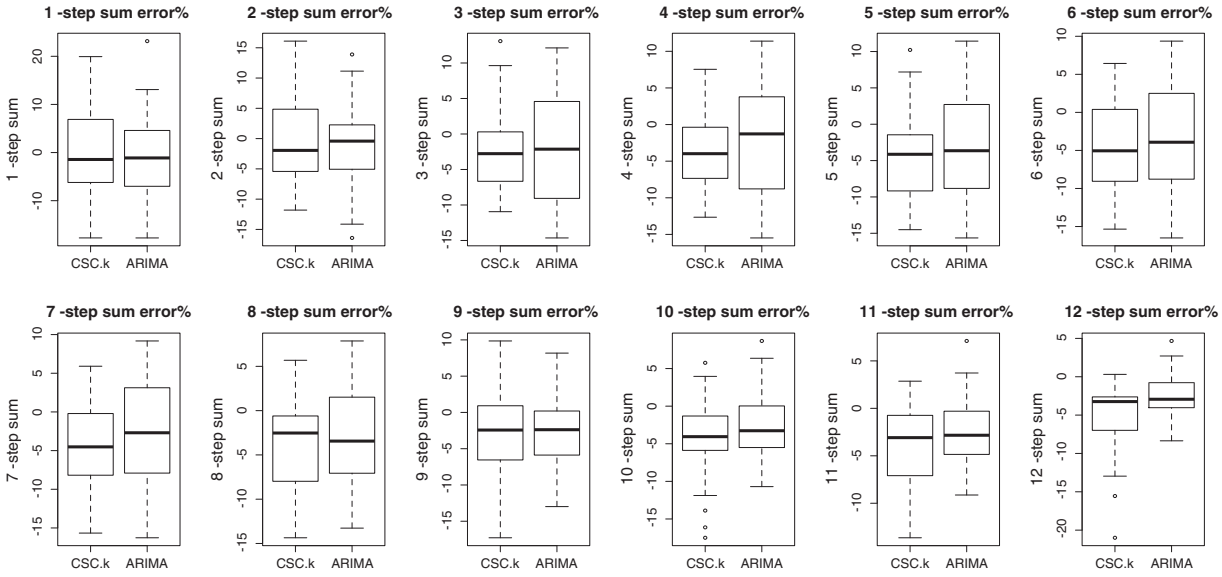

FIGURE 11 Box plots of forecasting errors for ℓ-months total, $\ell = 1, 2, \ldots, 12$

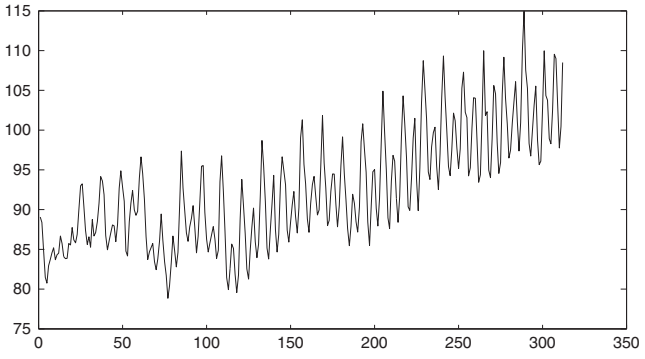

FIGURE 12 Total industrial production index of energy (1997 = 100) (US) (1977–2002)

longer-term predictions when $\ell = 9, 10, 11, 12$. Between the two CSC models, CSC(1) + AR(1) gives even better performance. As an additional forecasting measure, the mean absolute forecasting errors (MAFE) in Table 6 validate the better performance of CSC models in prediction lengths between 2 and 8.

To avoid the misleading conclusion from potential large prediction outliers, we show the box plots of the prediction errors (%) from the rolling forecast by ARIMA and CSC(ℓ) model in Figure 11. This provides clear evidence that the CSC(ℓ) model performs better than the ARIMA model with moderate prediction horizons.

4.4 | Real example (II)

In this example we analyze the monthly US industrial production index of energy from January 1977 to December 2002, with 312 observations, using 1997 as the base year (with value set to 100). The values of all other years are relative to year 1997. The data are from the website www.economagic.com.

The time series plot (Figure 12) shows strong seasonality with an upward trend. Based on ACF and PACF plots (Figures 13 and 14) and a model selection procedure, the seasonal ARIMA model

$$(1 - \phi B)(1 - B^{12})\log(X_t) = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_3 B^{12})\epsilon_t,$$

is used for comparison purposes.

Using a model selection procedure and detailed residual analysis, CSC(1) + AR(3) is selected to model the series. Rolling forecast starts from $K = 168$. Table 7 presents a detailed forecasting performance comparison of the two models. The values of Q_ℓ and the percentage change of Q_ℓ

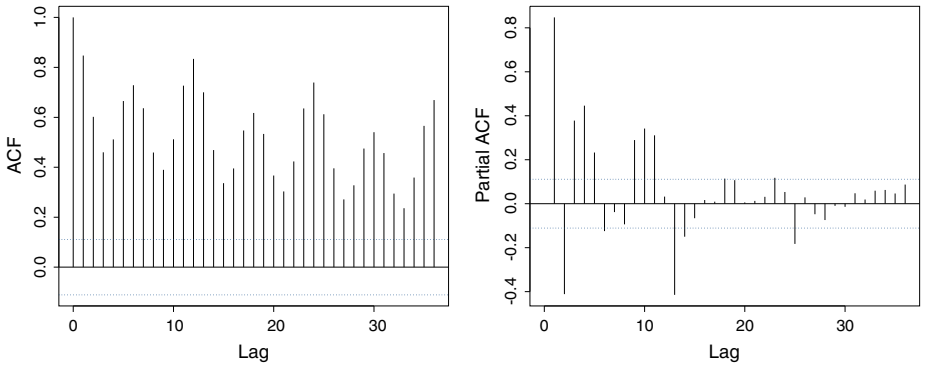

FIGURE 13 ACF and PACF of US industrial production index of energy [Colour figure can be viewed at wileyonlinelibrary.com]

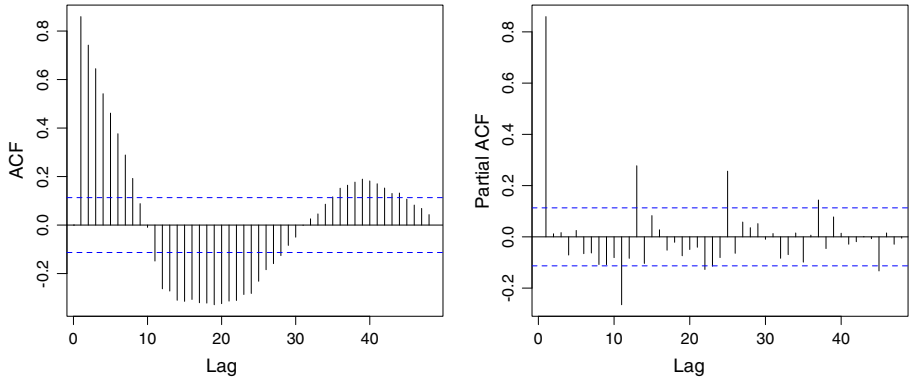

FIGURE 14 ACF and PACF of US industrial production index of energy (logged, seasonal differenced at $d = 12$) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 7 Comparison of forecasting performance for US industrial production index of energy

Lead ℓ	1	2	3	4	5	6	7	8	9	10	11	12
$Q_{\ell,A}$	2.15	8.62	20.61	39.27	64.64	97.44	138.60	187.35	243.86	301.48	360.56	427.87
$Q_{\ell,C}$	2.84	11.95	27.92	47.53	69.88	95.43	129.40	171.72	221.49	267.72	302.87	342.13
Chge	32.11	38.61	35.46	21.02	8.11	−2.06	−6.64	−8.35	−9.17	−11.20	−16.00	-20.04

$Q_{\ell,A}$, seasonal ARIMA model; $Q_{\ell,C(1)}$, CSC(1) + AR(3); 'Chge' denotes percentage change between the two.

TABLE 8 Comparison of yearly-total forecasting performance for US industrial production index of energy

Lead ℓ	12	11	10	9	8	7	6	5	4	3	2	1
MSE_A	633.22	433.95	409.85	290.68	189.02	93.81	72.94	49.55	43.67	23.74	13.40	1.33
$MSE_{C(1)}$	347.28	276.26	300.51	223.44	209.64	118.51	78.16	86.24	62.49	27.99	15.45	1.64
Chge	−45.16	−36.34	−26.68	−23.13	10.91	26.34	7.16	74.03	43.08	17.93	15.32	23.60

MSE_A, seasonal ARIMA; $MSE_{C(1)}$, CSC(1) + AR(3); 'Chge' denotes percentage change between the two.

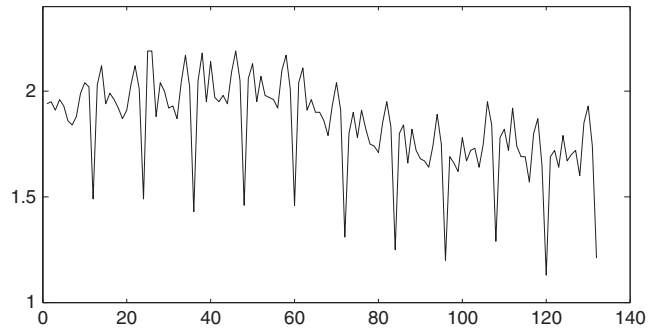

FIGURE 15 US retail inventories/sales ratio for furniture etc. (1992–2002)

between the two models suggest that CSC(1) + AR(3) performs better for longer prediction horizons such as $\ell = 6, \ldots, 12$.

Table 8 contains the values of MSE from the forecasting of yearly total by both seasonal ARIMA and CSC(1) + AR(3) models. It further shows that CSC(1) + AR(3) outperforms seasonal ARIMA in long-horizon forecasting.

4.3 | Real example (III)

In this example we analyze the US retail inventories/sales ratio for furniture, home furnishing, electronics, and

TABLE 9 Comparison of forecasting performance for US retail inventories/sales ratio for furniture etc.

Lead ℓ	1	2	3	4	5	6	7	8	9	10	11	12
$Q_{\ell,A}$	0.29	1.21	2.62	4.61	7.54	11.89	17.25	22.76	28.41	35.04	41.99	49.28
$Q_{\ell,C(1)}$	0.32	1.23	2.52	4.48	6.88	10.11	14.21	18.75	23.81	29.21	34.49	40.98
Chge	9.90	1.72	−3.85	−2.72	−8.83	−14.97	−17.67	−17.62	−16.19	−16.63	−17.86	−16.85

$Q_{\ell,A}(\times 10^2)$, seasonal ARIMA; $Q_{\ell,C(1)}(\times 10^2)$, CSC(1) + MA(3); 'Chge' denotes percentage change between the two.

appliances, with 132 observations in total from January 1992 to December 2002, obtained from the website www.economagic.com. It is a nonstationary time series with strong seasonality, as shown in Figure 15.

The seasonal ARIMA model $(1-\phi B)(1-B^{12})\log(X_t) = \varepsilon_t$, and CSC(1)+MA(3) are selected to model the series. Rolling forecasts start from $K = 72$. A comparison of forecasting performance is shown in Table 9. This demonstrates that CSC(1)+MA(3) outperforms seasonal ARIMA for almost all prediction horizons except for very short terms when $\ell = 1$ and 2. In addition, the improvement of prediction is more significant for longer forecasting horizons.

ACKNOWLEDGMENTS

Rong Chen's research is sponsored in part by NSF grants DMS-1540863 and DMS-1209085. The authors Kun Chang and Rong Chen thank the Department of Homeland Security for its support under grants 2009-ST-061-CCI002-05 and 2009-ST-061-CCI002-06 to Rutgers University. We also thank an AE and two anonymous referees for their insightful comments, which led to significant improvement of the paper.

REFERENCES

Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman & Hall: New York, NY.

Bell, W., & Hillmer, S. (1984). Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economic Statistics*, *2*, 291–320.

Box, G., & Jenkins, G. (1994). *Time series analysis: Forecasting and control* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Chen, R., & Fomby, T. B. (1999). Forecasting with stable seasonal pattern models with an application to Hawaiian tourism data. *Journal of Business and Economic Statistics*, *17*(4), 497–504.

Cleveland, W. P., & Tiao, G. C. (1976). Decomposition of seasonal time series: A model for the X-11 program. *Journal of the American Statistical Association*, *71*, 581–587.

Findley, D., Monsell, B., Bell, W., Otto, M., & Chen, S. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics*, *16*, 127–177.

Franzini, L., & Harvey, A. (1983). Testing for deterministic trend and seasonal components in time series model. *Biometrika*, *70*(3), 673–682.

Ghysels, E., & Osborn, D. R. (2001). *The econometric analysis of seasonal time series*. Cambridge, UK: Cambridge University Press.

Kilger, C., & Wagner, M. (2010). Demand planning, (4th ed.). In Stadtler, H., & Kilger, C. C. (Eds.), *Supply chain management and advanced planning: Concepts, models, software, and case studies*. Berlin, Germany: Springer, pp. 133–160.

Oliver, R. (1987). Bayesian forecasting with stable seasonal patterns. *Journal of Business and Economic Statistics*, *5*, 77–85.

Thomas, C., & Aitchison, J. (2006). *Log-ratios and geochemical discrimination of Scottish Dalradian limestones: A case study*. Special Publication. London, UK: Geological Society.

Tiao, G. C., & Tsay, R. S. (1994). Some advances in non-linear and adaptive modeling in time series. *Journal of Forecasting*, *13*, 109–131.

Tiao, G. C., & Xu, D. (1993). Robustness of maximum likelihood estimates for multi-step predictions: The exponential smoothing case. *Biometrika*, *80*(3), 623–641.

Tong, H. (1997). Some comments on nonlinear time series analysis, *Nonlinear dynamics and time series: Fields Institute Communications*, Vol. *11*. Providence, RI: American Mathematical Society, pp. 17–27.

Zellner, A. (1978). Retrospect and prospect. In Zellner, A. (Ed.), *Seasonal analysis of economic time series*. Washington, DC: US Department of Commerce, Bureau of the Census.

How to cite this article: Chang K, Chen R, Fomby TB. Prediction-based adaptive compositional model for seasonal time series analysis. *Journal of Forecasting*. 2017;36:842–853. https://doi.org/10.1002/for.2474