TOPICS IN COMPOSITIONAL, SEASONAL AND SPATIAL-TEMPORAL TIME SERIES

BY KUN CHANG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics and Biostatistics

Written under the direction of

Professor Rong Chen

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2015

ABSTRACT OF THE DISSERTATION

Topics in compositional, seasonal and spatial-temporal time series

by Kun Chang

Dissertation Director: Professor Rong Chen

This dissertation studies several topics in time series modeling. The discussion on seasonal time series, compositional time series and spatial-temporal time series brings new insight to the existing methods. Innovative methodologies are developed for modeling and forecasting purposes. These topics are not isolated but to naturally support each other under rigorous discussions. A variety of real examples are presented from economics, social science and geoscience areas.

The second chapter introduces a new class of seasonal time series models, treating the seasonality as a stable composition through time. 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 distribution is formulated under the compositional time series framework. The probabilistic properties, the estimation and prediction, and the forecasting performance of the model are studied and demonstrated with simulation and real examples.

The third chapter focuses on the discussion of compositional time series theories in multivariate models. It provides an idea to the modeling procedure of the multivariate time series that has sum constraints at each time. It also proposes a joint MLE method for threshold vector-error correction models. This chapter interprets the methodologies

with an real example of the U.S. household consumption expenditures data. Threshold cointegration effects are analyzed on the U.S. real GDP growth rate. The estimation of TVECM is compared by the current two-step estimation method and the proposed joint MLE approach.

Sensor allocation problem is studied in Chapter 4 under spatial-temporal models in Gaussian random fields. Given observations from existing sensors, the problem is solved by minimizing the integrated conditional variance based on different forecasting purposes. In this chapter, the time series are measured at different locations with both spatial and time series correlations. The spatial-temporal covariance structure is extensively discussed under both separable and nonseparable cases. The model is finally applied to the ozone level measurements in Harris County, Texas.

Acknowledgements

First I would like to express my deepest gratitude to my thesis advisor, Professor Rong Chen for his continuous guidance, encouragement and support. It is him who inspires my motivation to start a journey in statistical science. His insight and enthusiasm of statistical research has made my journey enjoyable and fruitful. He is such a glamorous person whose natural talent, diligence and positive attitude beautifully compliment each other, making him a role model for many people including me. It is my privilege and my most exciting experience to work with him. His invaluable mentorship and constant inspiration will be my lifetime treasure.

I wish to give my unlimited gratitude to Professor Minge Xie, Professor Han Xiao and Professor Ying Hung for their brilliant ideas and generous advice on my research. I am deeply impressed by Professor Xie not only for his broad knowledge and great instructions, but also his rigorous mind and kindness to encourage the students. I feel very lucky to have numerous benefits from the time series mini-seminars with Professor Xiao at the early stage of my research. It is a fantastic experience to attend his lectures and having discussions with him. I have also received precious suggestions from Professor Hung on my research. Her top expertise in the field has provided deep insights to the problems I have.

I would like to thank Professor Thomas Fomby for his important contribution to our collaborated work. I am also grateful to Professor Fred Roberts for the precious research opportunity to learn from talented researchers in CCICADA. I feel gratitude to Professor Minge Xie, Professor Han Xiao and Professor Jerry Cheng for their efforts to be my committee member and their valuable comments on my dissertation.

Sincere thanks to Professor John Kolassa for his continuous advice and help throughout my graduate study, and thanks to the department for the financial support. I need

also express my appreciation to my peer colleagues for the enlightening discussions and encouragement. Thank all of them to make my PhD life memorable.

Finally, my deepest love to my family for their support and spirit that keeps me going.

Dedication

To my beloved parents for their constant source of love, inspiration and support.

Table of Contents

Chapter 1

Overview

1.1 Linear Time Series Models

We first consider a class of linear time series models that has been popular for several decades in time series analysis. The era of time series modeling began with linear models in Yule (1927) where the autoregressive (AR) model was first introduced in the study of sunspot numbers. After that, there are three main approaches to time series analysis, the exponential smoothing model, the linear stochastic model by Box and Jenkins (1994), and the spectral analysis, which is the time series analysis in the frequency domain (Priestley (1983) and Brockwell and Davis (1991)).

Among all models for scalar time series, the well-known autoregressive and moving average (ARMA) model by Box and Jenkins (1994) is a landmark of the mature development of statistical theory and methods for linear time series analysis. An ARMA(p, q) model is defined as

$$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

where x_t is a stationary time series and ε_t is a white noise with mean zero and variance σ_ε^2. The model parameters $(\phi_0, \phi_1, \ldots, \phi_p; \theta_1, \theta_2, \ldots, \theta_q; \sigma_\varepsilon)$ can be estimated by the maximum likelihood estimation or least squared procedures together with the Durbin-Levinson Algorithm and Innovations Algorithm. The estimation procedure and model properties are extensively introduced in Brockwell and Davis (1991).

For the scalar nonstationary time series that has unit roots, the autoregressive integrated moving average (ARIMA) model builds an ARMA structure on the differenced time series to achieve stationarity. Moreover, seasonal random walks and seasonalities can be included in the model as well. The seasonal ARIMA model has a general form

of

$$\Phi(B^s)\Delta_D^s x_t = \Theta(B^s)\varepsilon_t.$$

The seasonal ARIMA model is well-accepted in real applications. Moreover, it has been using as the baseline model for some innovative time series forecasting models. For example, Chen and Fomby (1999) proposed the stable seasonal pattern model and compared it with the seasonal ARIMA model on forecasting of the monthly tourists visiting Hawaii. More recently, Liu et al. (2015) compares the prediction power of the proposed functional-coefficient seasonal time series model with the seasonal ARIMA model on the Hawaii tourism forecasting problem as well.

The linear models are still popular since they are simple yet easy to explain. In addition, The models always have good approximation to the mean structure of real time series data. Various examples can be found in Shumway and Stoffer (2006) and Tsay (2010). Furthermore, this theory can been extended to the multivariate vector ARMA models, unit root and cointegration models. More details are provided in the subsequent sections.

1.2 Vector Time Series Models

1.2.1 Linear Vector Time Series Models

Multivariate time series analysis plays an substantial role in econometrics and finance recently. As extension of the univariate time series models, multivariate time series models investigate interrelations among different time series. It reflects the dynamic change of the full system and has great applications in topics such as stock return prediction, term structure models and a series of econometric analysis.

The vector autoregressive (VAR) model is a well-developed and the most often used model for interpreting the dynamic structure in multivariate time series analysis. It is a natural extension of the univariate autoregressive models but incorporates the casual relationship between the variables. For a stationary k-dimensional random vector $\boldsymbol{y}_t$, it follows a VAR(p) model if it satisfies

$$\boldsymbol{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \;\; p > 0, \tag{1.1}$$

where $\boldsymbol{\phi}_j, j = 1, \ldots, p$ are the $k \times k$ coefficient matrices, $\boldsymbol{\phi}_0$ is the mean vector. The error process $\boldsymbol{\varepsilon}_t$ is a k-dimensional white noise with zero mean and covariance matrix $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^{'}) = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$.

The model has a companion matrix $\boldsymbol{\phi}^*$ defined as

$$\boldsymbol{\phi}^* = \begin{pmatrix} \mathbf{0} & I & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & I \\ \boldsymbol{\phi_p} & \boldsymbol{\phi_{p-1}} & \boldsymbol{\phi_{p-2}} & \boldsymbol{\phi_{p-3}} & \cdots & \boldsymbol{\phi_1} \end{pmatrix}. \tag{1.2}$$

The weak stationary condition of the VAR(p) model in (1.1) is to ensure that all eigenvalues of the companion matrix in (1.2) are bounded by 1 in modulus.

The model building process starts from a cross-correlation analysis between the random vectors and confirms the necessity of VAR models. Then the order p is identified by model selection criterion such as AIC, BIC and HQ (Hannan and Quinn (1979)), or by chi-squared tests on models with consecutive orders. By gaussian assumption on the error process $\boldsymbol{\varepsilon}_t$, the model can be estimated using least square estimation or conditional maximum likelihood estimation, which are asymptotically equivalent. Multivariate Portmanteau tests are used for model checking of VAR models. The VAR model is natural for forecasting purposes if properly fitted. The application of VAR models are quite popular and universal. People can find more details and application examples in Reinsel (2003), Tsay (2010), Lütkepohl (2009) and Tsay (2013).

The vector autoregressive moving average (VARMA) models are generalized from ARMA models to multivariate cases. VARMA appear to be preferable from a theoretical viewpoint, however, they are scarcely used due to estimation difficulties. In fact, VAR models are much more widely employed because they are easier to implement by least squares methods, while VARMA models typically require nonlinear methods. Lütkepohl (2005) introduces the estimation methods of VARMA in details.

As a summary, linear VAR models will continue to be a powerful statistical model for multivariate time series. Nevertheless, many econometric and financial applications involve nonlinear dynamics, and linear time series models cannot capture the nonlinear

phenomena such as asymmetry, regime-shifts or jumps. Therefore, nonlinear time series models are of strong interests and will be introduced as one of the main topics in this dissertation.

1.2.2 Threshold VAR Models

As noted above, nonlinear features are more and more notable in many applications. Over the last three decades, nonlinear time series analysis has been advancing rapidly. Threshold model is a common type of nonlinear time series models. In univariate cases, threshold autoregressive model describes the regime switching behavior of the series. The current state of the variable depends on the past value of some weak stationary threshold variable. If the threshold variable is the series itself, the model is referred as self-exiting threshold autoregressive model (SETAR) by Tong and Lim (1980), Tong (1990) and Chan and Tong (1990). The threshold variable can be an exogenous variable and it is referred to the threshold autoregressive (TAR) model (Tong (2011)).

Similarly, the VAR models can also be generalized to capture the regime switching pattern to some threshold variables. For a k-dimensional time series $\boldsymbol{y}_t$, the TVAR(p) model with threshold variable γ_t and delay d is specified as,

$$\boldsymbol{y}_t = \boldsymbol{\mu}^{(j)} + \boldsymbol{\phi}_1^{(j)}\boldsymbol{y}_{t-1} + \boldsymbol{\phi}_2^{(j)}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\phi}_p^{(j)}\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t^{(j)}, \quad \text{when } \gamma_{j-1} < \gamma_{t-d} \leq \gamma_j, \tag{1.3}$$

where j is the number of regimes. This model is widely used in econometrics, for example, Fazzari et al. (2011) discussed the asymmetric fiscal policy shocks on the U.S. economic activities using threshold vector models.

Threshold tests are preliminary before the modeling process. The arranged autoregression method by Tsay (1998) is one of the most widely test for TVAR model. It is an extension of the univariate threshold test by Tsay (1989). With a pre-specified VAR order p and a given threshold variable γ_t, the arranged regression can detect the threshold pattern among possible values of γ_{t-d} by an F-test in univariate case and chi-squared test in multivariate case. The delay d corresponding to the most significant test statistic. The threshold value is located by concentrated LS method. It is picked to minimize the overall SSE, AIC or log-likelihood values by grid search among sorted

threshold values. The t-ratio plots of the recursive regression coefficients versus the ordered threshold values are informative in detecting threshold pattern and threshold location as well. Once the threshold value and regimes are determined, the other parameters can be estimated by OLS within each regime. Moreover, some extreme values in the sorted threshold variable are trimmed out to ensure the number of observations in each regime. This approach is both powerful and easy to understood, and it will be extensively used in the remaining of the dissertation in both univariate and multivariate cases.

Beyond Tsay's method, Hansen (1997) used an alternative testing approach for two-regime threshold models. It gives the threshold value and the model estimation simultaneously with the test result and allows a test on more than two regimes. The distribution properties of the threshold estimator are discussed in Hansen (1999) and Hansen (2000). There are other threshold tests such as the portmanteau test in Petruccelli and Davies (1986), the quasi-likelihood ratio test in Chan (1990), Chan and Tong (1990) and Ling and Tong (2005).

1.3 Cointegration Models

1.3.1 Linear Cointegration and Tests

Traditionally VAR models are suitable for stationary time series and the trends can be included as deterministic terms. Later in 1980s people discovered the stochastic trends and cointegration effects between the variables, indicating that new models are desired to separate the long-run relations in the traditional VAR model.

Cointegration is a statistical property that reflects the long-run relations between the multivariate time series variables. First of all, people need stationarity as the fundamental property in the modeling of time series. However, many series are non-stationary with trends or unit-roots. The unit-roots type non-stationarity is the one of particular interests, and people call the series is integrated. In case of unit-roots, we need to take differences on the original series to avoid the so-called spurious regression, which gives fake significance and therefore false regression results.

While it is common to achieve stationarity by taking differences on the original series for unit-root non-stationary series, the cointegration is the phenomenon that the linear combination of two or more non-stationary series is stationary. The concept of cointegration was first introduced by Engle and Granger (1987), describing the long-run equilibrium between the multivariate time series variables in economics. With presence of cointegration, there is a common trend between the series and taking differences on the original levels will result in a overdifferenced model. The existence of cointegration determines the necessity of the vector error correction model (VECM). The VECM was stated by the Granger representation theorem. For a general k-dimensional VARMA(p, q) model with m $(m < k)$ cointegrating vectors, the error correction representation is

$$\boldsymbol{\Delta y_t = \Gamma D_t + \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_t + \varepsilon_t - \sum_{j=1}^{q} \Theta_j^* \varepsilon_{t-j},} \tag{1.4}$$

where $t = 1, 2, \ldots, T$, and $\boldsymbol{\varepsilon_t}$ are independent k-dimensional Gaussian variable with mean zero and given variance structure. Here $\boldsymbol{\beta' y_{t-1}}$ is the error correction term that denotes the deviation from the equilibrium. The deterministic effects are represented by $\boldsymbol{D}_t$. The error correction term $\boldsymbol{\beta' y_{t-1}}$ is unit-root stationary because of cointegration; and the vector $\boldsymbol{\alpha}$ is the adjustment coefficient that measures how fast the deviation move to the long-run equilibrium. The matrices $\boldsymbol{\Phi}_i^*$ and $\boldsymbol{\Theta}_j^*$ are the adjusted coefficients of the VARMA(p, q) model. The system of VECM is stationary and the typical modeling procedure of the vector time series models are similar to apply.

Cointegration is a natural concept in economics. The purchasing power parity and the expected hypothesis of term structures are typical examples of cointegration. The former has cointegration between the nominal exchange rate and foreign and domestic prices, and the latter describes the common trend between the nominal interest rates at different maturities. In financial area, cointegration arises between the prices of the same stock but trading on different markets. Such an example is discussed on Tsay (2010). In high frequency trading area, cointegration effect is extensively discussed for statistical arbitrage opportunities. This dissertation focus on the application of cointegration in economics, where it is considered as a long-run equilibrium relationship.

Johansen's Cointegration Test

There are two main methods for testing of cointegration and estimation of cointegrating vectors. The Engle-Granger two-step method (1987) is straightforward as it assumes the cointegration relationship is known between two unit-root nonstationary time series, and the cointegrating vector can be estimated by OLS between the two. The cointegrating residual from the OLS is formed in the first step, and the null hypothesis of no cointegration can be tested by the unit root test on the residual series. In the same time, due to the spurious regression under the null hypothesis of no cointegration, the distributions for the cointegrating residual based unit-root tests are not standard Dickey-Fuller distributions. Phillips and Ouliaris (1990) gave the critical values of the distribution by simulation.

Engle-Granger's two-step method has the major restriction that it only allows the estimation of one cointegrating vector and assumes the cointegration between specific variables in advance. The test inJohansen (1991) is the maximum likelihood approach that allows the testing and estimation of multiple cointegrating vectors in the system. It is the major tool we use in this dissertation for cointegration analysis.

Johansen's test applies on the matrix $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and the cointegration relations are determined by the matrix $\mathbf{\Pi}$ in the following schemes,

1) $\mathbf{\Pi} = 0$, which means there is no cointegration, and the model is built on the differenced levels of $\Delta \boldsymbol{y}_t$.

2) $rank(\mathbf{\Pi}) = k$. In this case equation 3.1 can be arranged into the VARMA model on original levels, which means that the components of $\boldsymbol{y}_t$ are all stationary and no difference is needed.

3) $rank(\mathbf{\Pi}) = m < k$. Now we have decomposition that both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are k by m full rank matrices, representing that there are m cointegration relations in the system.

With this formulation, the existence of cointegration can be tested by analyzing the rank of $\mathbf{\Pi}$. Moreover, the specific number of cointegration relations in multivariate models can be detected by sequentially using the nested tests

$$H(0) \subset H(1) \subset \cdots \subset H(m) \subset \cdots \subset H(k)$$

where m is an integer between 0 and $k-1$. The likelihood ratio test results in two types of rank tests including the trace test and the maximum eigenvalue test. It is proved that the two test statistics are not chi-squared distributed due to the existence of unit roots. The limiting distribution of the test statistics are functions of Brownian motion and the critical values can be derived by simulation. The details of the testing methodology and procedure are described in Appendix.

1. *Trace Test*

It corresponding to the hypothesis testing

$$H_0 : rank(\mathbf{\Pi}) = m \quad vs \quad H_1 : rank(\mathbf{\Pi}) > m$$

where m is an integer between 0 and $k-1$. The trace test statistic is defined as

$$LR = -T \sum_{i=m+1}^{k} \log(1-\lambda_i).$$

2. *Maximum eigenvalue Test*

It corresponding to the hypothesis testing

$$H_0 : rank(\mathbf{\Pi}) = r \quad vs \quad H_1 : rank(\mathbf{\Pi}) = r+1$$

where m is an integer between 0 and $k-1$. The test statistic is defined as

$$LR = -T \log(1-\lambda_{m+1}).$$

It is understood that in Johansen's trace test, if the rank is m, then the eigenvalues of $\lambda_{m+1}, \cdots, \lambda_k$ should be zero, and the value of the test statistic should be small; otherwise, one rejects the null. The same rationale applies to the maximum eigenvalue test.

1.3.2 Threshold Cointegration

Balke and Fomby (1997) proposed the concept of threshold cointegration. They are motivated that with cointegration, the movement toward the long-run equilibrium may not occur in every period. In many economic situations, the system will move back to equilibrium only when the deviations from the equilibrium exceed some threshold. For

example in the financial markets, the asset returns are free to diverge within a band due to the transaction cost, and result in arbitrage possibilities. In economics, the government will not make interventions in the foreign exchange markets until the exchange rate fluctuates out of the band. In other words, the adjustment of the deviations from equilibrium is discrete and it can be modeled in terms of threshold cointegration. Mathematically, it means that the error correction term in the VECM model follows the SETAR model. The three regime threshold cointegration model is ,

$$z_t = \boldsymbol{\beta}^{'} \boldsymbol{y}_t = \begin{cases} \rho_0^{(1)} + \rho^{(1)}(L) z_{t-1}, & \text{when } z_{t-d} < \gamma_1, \\ \rho_0^{(2)} + \rho^{(2)}(L) z_{t-1}, & \text{when } \gamma_1 \leq z_{t-d} < \gamma_2. \\ \rho_0^{(3)} + \rho^{(3)}(L) z_{t-1}. & \text{when } \gamma_2 \leq z_{t-d} < \gamma_3. \end{cases} \tag{1.5}$$

where $\rho^{(2)}(L)$ are lag polynomials, d is the delay in the error correction process and $u_t^{(j)}$ is the innovation process with mean zero and standard deviations $\sigma^{(j)}, j = 1, 2, 3$. The threshold variable is the error correction term itself, though, it can be generalized to be an exogenous weak stationary variable γ_t. The stationarity conditions for the univariate threshold model when lag order is 1 can be found in Chan et al. (1985).

As suggested by Balke and Fomby, the testing of threshold cointegration involves two parts. The testing of cointegration and threshold effects are performed separately. We can first test the cointegration, which is called the global behavior. The available cointegration tests such as the Engle-Granger procedure and Johansen's procedure are ready to use. Once the cointegration relationship is confirmed, the error correction term is estimated and assumed to be known for the threshold test on it, which is referred as the local behavior. The arranged autoregression approach by Tsay (1989) or Sup-Wald test by Hansen (1999) can be applied. The latter allows testing for more than one threshold values and hence is more general than Tsay's method.

The application of threshold cointegration has been discussed extensively in economics. The empirical study of purchasing power parity has been one of the hot topics. Economists had great discussions of the nonlinearities such as threshold effect in the PPP theory, and proceeded to discover the threshold cointegration feature. For example, Heimonen (2006) found the evidence of threshold cointegration in the purchasing

power parity based on the asymmetry due to the sign of deviation, and the dependency of the asymmetry to the exchange rate regime. Similarly, the threshold cointegration among effective nominal exchange rates and import prices are studied in Al-Abri and Goodwin (2009) and Nakagawa (2010).

1.4 Compositional Time Series

Compositional data are data where the elements of the composition are non-negative and sum to unity. They usually arise from non-negative data such as counts, area, weights and expenditures that scaled by the total, describing the parts of the whole quantitatively. Compositional time series are the compositional data that changes by time. For a specific time t, we have a composition $\boldsymbol{x}_t$ consists of D elements and satisfies

$$\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Dt}), \quad x_{1t} > 0, \ldots, x_{Dt} > 0, \quad x_{1t}+x_{2t}+\ldots+x_{Dt} = 1, \quad t = 1, \ldots, T.$$

Notice that the composition is completely specified by the components of a $(D-1)$-dimension sub vector by the sum constrain. We could define $d = D - 1$ for simplicity.

The compositional data usually comes in the original measurements such as counts and lengths. Mathematically we call this original measurements as a basis z. It is a $D \times 1$ vector of positive components $(z_{1t}, \ldots, z_{Dt})$ that all recorded on the same measurement scale. And the components in a composition are determined by

$$x_{it} = z_{it}/(z_{1t} + \cdots + z_{Dt}), \quad i = 1, 2, \ldots, D, \quad t = 1, \ldots, T. \tag{1.6}$$

Aitchison (1986a) has developed simple statistical methodologies for compositional data analysis. He provided detailed information on the statistical properties, the estimation procedures and the applications of some compositional models. Normality is achieved by several transformation methods on original compositional data. Logratio transformations are the most popular ones that include the isometric logratio transformation (ILR), the centred logratio transformation (CLR) and the additive logratio transformation (ALR).

The isometric logratio transformation introduced by Egozcue et al. (2003) is a transformation from simplex $\mathcal{S}^{\mathcal{D}}$ into the real space $\mathcal{R}^{\lceil}$ but with more complex vectors and hence is hard to interpret.

The centred logratio transformation is defined as

$$\boldsymbol{y}_t = clr(\boldsymbol{x}_t) = \left(\log\left(\frac{x_{1t}}{g(\boldsymbol{x}_t)}\right), \log\left(\frac{x_{2t}}{g(\boldsymbol{x}_t)}\right), \ldots, \log\left(\frac{x_{D,t}}{g(\boldsymbol{x}_t)}\right)\right); \tag{1.7}$$

where $g(\boldsymbol{x}_t)$ is the geometric mean. Quintana and West (1988) was the first to use the CLR transformation to analyze the compositional time series data. Because of its retained singularity problem, people incorporate kinds of methods to solve the singularity under the CLR transformation such as Dirichilet approach.

The additive logratio transformation is the most widely used one because of its simplicity. For a composition $\boldsymbol{x}_t = (x_{1t}, x_{2t,} \ldots, x_{Dt})$, the additive logratio transformation is defined as

$$\boldsymbol{y}_t = alr(\boldsymbol{x}_t) = \left(\log\left(\frac{x_{1t}}{x_{Dt}}\right), \log\left(\frac{x_{2t}}{x_{Dt}}\right), \ldots, \log\left(\frac{x_{D-1,t}}{x_{Dt}}\right)\right). \tag{1.8}$$

This is a transformation from the data sample space in the simplex $\mathcal{S}^{\mathcal{D}}$ into the real space $\mathcal{R}^{\lceil}$ by assign a base component x_{Dt}. More important, the choice of base component x_{Dt} is flexible due to the invariance property of the transformation. Finally, $\boldsymbol{y}_t$ is shown to be multivariate normally distributed in Aitchison (1986a), and standard multivariate models such as VARIMA can be directly applied. This approach is widely used in compositional time series analysis. For example, Brunsdon and Smith (1998) used the log-normal based ARIMA model on the additive logratio transformation of UK poll data on vote intentions. A regression model with VARMA errors was fitted to illustrate the compositional time series of mortality events in Los Angeles in Ravishanker et al. (2001). Mills (2010) tried to predict UK GDP components through VAR model under compositional transformations. In this paper we apply this approach to analyze the compositions from US household expenditures data.

1.5 Spatial-Temporal Models

Spatial-temporal models arise when data are collected across time as well as space. It has been widely used in environmental sciences since environmental data usually have both spatial and temporal components. The history of spatial-temporal analysis dates back to 1980s when Bilonick (1985) started to build the model for the acid rainfall data

in in a series of his papers. Another example would be the monitoring sensor network where data are collected at equally spaced time intervals and fixed sensor locations. Then data analysis has to take into account of spatial dependence among the monitors as well as the temporal correlations. Until recently, researchers in diverse areas such as climatology, ecology, catastrophe insurance institutes and real estate marketing are developing increasing interests in analyzing the spatial-temporal models.

A spatial-temporal process $Y(s,t)$ is defined as

$$\{Y(\boldsymbol{s},t) : \boldsymbol{s} \in \Omega, t \in \mathbb{R}\},$$

where $\Omega \subset \mathbb{R}^d$ and t is discretized in $\mathbb{R}$. The key probabilistic properties of the process are stationarity and separability.

Weak stationarity

The process $Y(\boldsymbol{s},t)$ is weak stationary if it satisfies

- $\mathrm{E}(Y(\boldsymbol{s},t)) = \mu$, μ is a constant;
- $\mathrm{Cov}(Y(\boldsymbol{s},t), Y(\boldsymbol{s}+\boldsymbol{h}, t+u)) = C(\boldsymbol{h},u)$, for $\boldsymbol{h} \in \mathbb{R}^d, u \in \mathbb{R}$.

The weak-stationarity guarantees the shift invariance of the process. A stronger form is the strong stationarity but implies the weak stationarity in Gaussian assumption.

Separability

The process is separable if

$$\mathrm{Cov}(Y(\boldsymbol{s},t), Y(\boldsymbol{s}+\boldsymbol{h}, t+u)) = C_s(\boldsymbol{h})C_t(u), \quad \boldsymbol{h} \in \mathbb{R}^d, u \in \mathbb{R}.$$

It indicates that the covariance structure is the multiplication of the pure spatial and pure time covariance. Separability gives lots of advantage to the model such as the Kronecker product representation of the covariance structure, and it can greatly reduce the computational complexity if $Y(\boldsymbol{s},t)$ is high-dimensional. Moreover, the conditional independence property in separable covariance structures can reduce the spatial-temporal process to a pure spatial case. The latter will be discussed in details in Chapter 4.

The separable covariance structure is convenient, however, real data are usually not separable and more complex nonseparable structures are embedded. Actually, most of

the spatial-temporal covariance functions have a interaction parameter that controls the separability of the model. Therefore, separability tests are desirable in the procedure. Fuentes (2006) designs a separability test by spectral analysis, while Mitchell et al. (2005) proposes the likelihood ratio test on the admissible covariance families by Cressie and Huang (1999) and Gneiting (2002). The latter is discussed further in Genton (2007) and Li et al. (2008).

With measurements from locations $\boldsymbol{s}$ and time t, $Y(\boldsymbol{s}, t)$ has a typical model structure as,

$$Y(\boldsymbol{s}, t) = \mu(\boldsymbol{s}, t) + \varepsilon(\boldsymbol{s}, t), \tag{1.9}$$

where $\mu(\boldsymbol{s}, t)$ is the mean structure and $\varepsilon(\boldsymbol{s}, t)$ is the error process. The error process is usually assumed to be Gaussian $\varepsilon(\boldsymbol{s}, t) \sim N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{s}, t))$. The random vector $Y(\boldsymbol{s}, t)$ is multivariate normally distributed so that a lot of statistical theories such as maximum likelihood estimation and likelihood ratio test are available for use.

As can be seen, the key issue for spatial-temporal modeling will be the modeling of the covariance structure in the error process. Difficulties happen in using an admissible covariance function to model the interactions between spatial and time correlations. An admissible covariance function should be positive definite and stationary. People have been researched in this topic for decades and make great contributions in building admissible covariance function families.

In the early stage, literatures concentrate on spatial statistics in Matern (1986) and Cressie (1993). Cressie and Huang (1999) proposed a generic approach to develop parametric models for spatial-temporal processes. The method relies heavily on spectral representations for the theoretical space-time covariance structure as generalization results of Matern (1986) for pure spatial processes. This approach is powerful but depends much on Fourier transform yet not trivial for applications. Gneiting (2002) provides a general class of valid spatial-temporal covariance models based on the approach of Cressie and Huang (1999). The popularity of this topic keeps inspiring more advances in Mitchell et al. (2005) and De Luna and Genton (2005). Recent contributions to admissible covariance families are reviewed in Gneiting et al. (2007).

Among the valid covariance functions, Gneiting (2002)'s model worths more attention since it is more general and more practical to apply. In his model, the completely monotone function $\phi(t)$ and positive functions $\psi(t)$ with a completely monotone derivative can be modeled as the spatial and temporal covariance structure separately. Then a valid spatial-temporal covariance function is constructed as

$$C(\boldsymbol{h}; u) = \frac{\sigma^2}{\phi(|u|^2)^{d/2}} \psi\left(\frac{\|h\|^2}{\phi(|u|^2}\right), (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{1.10}$$

where σ^2 is the variance of the random process, $a, c > 0$ and the smoothness parameters α and γ are both restricted on $(0, 1]$. Furthermore, Gneiting (2002) provides lists of valid examples for $\phi(t)$ and $\psi(t)$. Generally, the model has many favored features for modeling the real data and it will be employed in this dissertation as well.

Chapter 2

Prediction-Based Adaptive Compositional Model For Seasonal Time Series Analysis

2.1 Background

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 ARIMA models (Box and Jenkins, 1994), the trend-and-seasonal models (Franzini and Harvey, 1983) and the stable seasonal pattern models (Oliver, 1987; Chen and Fomby, 1999). All these models provide different perspectives in dealing with seasonality. In particular, standard seasonal ARIMA models are in a multiplicative form while trend-and-seasonal models are in an additive form. There is a vast literature on seasonal time series analysis and seasonal adjustment, for instance, Cleveland and Tiao (1976); Zellner (1978); Balchin (1995); Bell and Hillmer (1984); Findley et al. (1998); Ghysels and Osborn (2001) and Box and Jenkins (1994).

On the other hand, compositional models of Aitchison (1986b) 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 models 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 type of rock specimens, as shown in a series of research by Thomas and Aitchison (2006). Another example is the study of budget pattern of a household reflected by the proportions of total expenditures allocated to several commodity groups. Aitchison (1986b) had analyzed such an example on five commodity groups.

The 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 twelve months can be seen as twelve components of the year (the composition), and the seasonality can be seen as certain systematic distributive pattern of the measurements among twelve 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 while 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 chapter we introduce a class of seasonal time series model 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 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 chapter we consider the objective of forecasting the next ℓ-season's total for different ℓ in a seasonal time series. Such forecasting tasks are often encountered in many applications. For example, for certain industry, an accurate prediction of the total quantity (e.g., sales, production) of next several months is important for better inventory management and marketing strategy. See, for example, Kilger and Wagner (2010).

The rest of the chapter is organized as follows. In section 2.2 we introduce a class of compositional seasonal time series model based on the theory of compositional analysis. Section 2.3 discusses the estimation, model checking and prediction procedures of the model. Section 2.4 contains two simulation studies and section 2.5 for three real data examples, including the forecasting comparison with standard seasonal ARIMA models. Finally, a predictive distribution approach is utilized to provide prediction with certainty in any confidence levels.

2.2 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, one complete season is formed by the previous $d-\ell$ measurements including X_t 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 (1986b), we assume that the ratio $Y_{2,t}/Y_{1,t}$ follows the lognormal distribution. That is,

$$Z_t = \log\left(\frac{Y_{2,t}}{Y_{1,t}}\right) \sim N\left(\mu_t + \boldsymbol{\beta}' \boldsymbol{r_t},\ \sigma_t^2\right), \tag{2.1}$$

where $\boldsymbol{r_t} = (r_{1t}, r_{2t}, \ldots, r_{mt})'$ is a set of exogenous variables and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_m)'$ is the coefficient vector. Let $\varepsilon_t = Z_t - \mu_t - \boldsymbol{\beta}' \boldsymbol{r_t}$ and $\varepsilon_t = \sigma_t e_t$. Model (3.4) 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. It 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 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

$$Z_t = \mu_{s(t)} + \boldsymbol{\beta}' \boldsymbol{r_t} + \sigma_{s(t)} e_t, \quad s(t) = t \bmod d. \tag{2.2}$$

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, which results 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,

$$e_t = \frac{Z_t - \mu_{s(t)} - \boldsymbol{\beta}' \boldsymbol{r_t}}{\sigma_{s(t)}} = \frac{\theta(B)}{\phi(B)} a_t, \quad a_t \sim N(0, \sigma_a^2). \tag{2.3}$$

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

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

Jointly (2.2) and (2.3) are referred to as a compositional seasonal component time series model with $(d-\ell, \ell)$ partition, denoted as $CSC(\ell)$ 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. Moreover, the observations in X_t can be inferred from Z_t once the latter is realized. This fact introduces the simulation of $CSC(\ell)$ in subsequent sections.

Remark 2: There are several ways of using 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 $CSC(\ell)$ model. Or 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 $CSC(\ell)$.

Remark 3: By combining (2.2) and (2.3), we have

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

It 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: To ensure uniqueness, we impose the constraint $\sigma_{s(t)} = 1$ when $s(t) = 0$. This constraint is slightly easier to meet than the more natural constraint of $\mathrm{Var}(a_t) = 1$.

Figure 2.1 shows a simulated series from $CSC(1)$ model with $d = 12$. Details of this series are given in section 2.4.1. A strong seasonality is seen together with certain nonstationarity. Figure 2.2 shows the sample ACF and PACF of the series. It is seen that those features can be easily misspecified as a seasonal ARIMA model.

2.3 Estimation, Model Checking and Prediction

2.3.1 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 (2.2) 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.

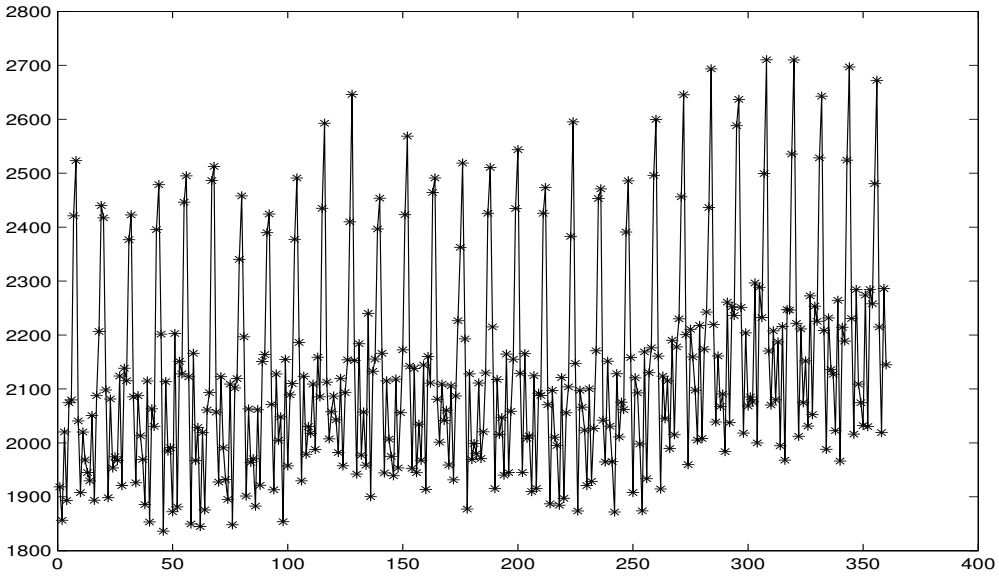


Figure 2.1: Original data plot of simulated series (I) from model $CSC(1)$.

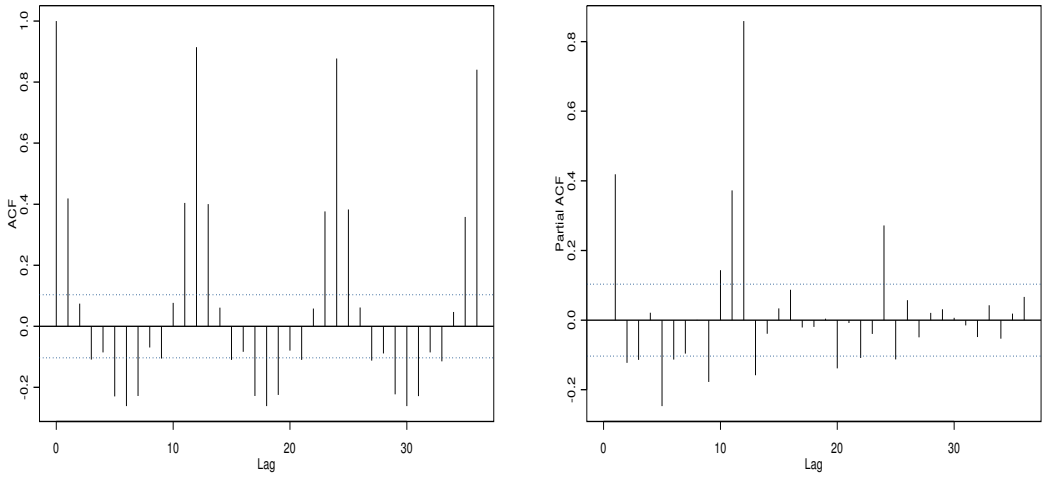


Figure 2.2: ACF and PACF of the simulated series (I).

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 $\boldsymbol{\Theta} = \left(\boldsymbol{\beta}, \boldsymbol{\mu}, \sigma_{s(t)}, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_a^2\right)$. Specifically, the likelihood function can be written as,

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

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

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

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 $(\boldsymbol{\beta}, \boldsymbol{\mu})$ can be estimated with standard regression estimators with known error covariance matrix. In the first iteration, we can assume e_t are $i.i.d.$

2. Form residuals from the first step with estimated coefficient $\hat{\epsilon}_t = Z_t - \hat{\boldsymbol{\beta}}^{'}\boldsymbol{r_t} - \hat{\boldsymbol{\mu}}^{\prime}\boldsymbol{\delta_t}$. 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 $\{e_t : e_t = \hat{\epsilon}_t / \hat{\sigma}_{s(t)}\}$.

2.3.2 Model Checking

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

1. Residual analysis

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

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

be the estimated residual series for $CSC(\ell)$ 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.

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, 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 2.3.3.

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

$$Q_\ell = \frac{1}{T-\ell-K+1}\sum_{j=K+1}^{T-\ell+1}(Y_{2,j} - \hat{Y}_{2,j})^2, \tag{2.4}$$

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

2.3.3 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.

Prediction under $CSC(\ell)$ model

The $CSC(\ell)$ model is relatively straightforward in forecasting of the next ℓ-observations total. Suppose we currently have the observations of X_t up to time t, the prediction of 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 (2.2) and (2.3). 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 $Z_{t-\ell+1}, \ldots, Z_t$.

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

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

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 (2.3). Hence

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

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

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

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

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 $CSC(\ell)$ model and yields more accurate predictions in many cases. In this setting, the prediction of the next ℓ-observations total can be done by the 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}. In the $CSC(1)$ model,

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

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

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

where $\hat{e}_{t+1|t}$ and $\sigma^2_{t+1|t}$ are the one-step prediction and its prediction variance under model (2.3). Here we assume $r_{t+1}, \ldots, r_{t+\ell}$ are observable at time t.

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

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

which is based on the two-step 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 ℓ-th step forecast $\hat{X}_t(\ell)$,

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

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

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

and the ℓ-th step prediction is

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

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

2.4 Numerical Examples and Forecasting Performance Comparison

Here we present two simulated examples and three real examples to demonstrate the prediction power of CSC models. We focus on out-of-sample forecasting performance comparisons between $CSC(\ell)$, $CSC(1)$ and seasonal ARIMA models.

2.4.1 Simulation Examples

Simulation Example (I)

As described in section 2.2, 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 2.1 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.

The parameters are set as $\boldsymbol{\sigma} = (\sigma_0, \ldots, \sigma_{d-1})'$ with $\sigma_0 = \cdots = \sigma_{d-1} = 1$ and

$$\boldsymbol{\mu} = (\mu_0, \ldots, \mu_{d-1})' = (-2.45 \quad -2.50 \quad -2.38 \quad -2.50 \quad -2.40 \quad -2.38$$
$$-2.25 \quad -2.20 \quad -2.40 \quad -2.50 \quad -2.40 \quad -2.45)'.$$

Based on various selection criteria, the seasonal ARIMA model is selected as

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

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 (2.4) for different prediction horizons, ℓ. The values of Q_ℓ are plotted in Figure 2.3. It can be seen that, the ARIMA model performs worse than CSC models for larger ℓ. As the true model, $CSC(1)$ performs the best among the three models.

Simulation Example (II)

The second simulated series is generated from a $CSC(1)$ model with AR(1)($\phi = 0.8$) errors. The mean vector is the same as the previous example but its variance is larger. Figure 2.4 shows the time series plot. It shows certain non-stationarity and strong seasonality in the series. The ACF and PACF plots (Figure 2.5) are very similar to what

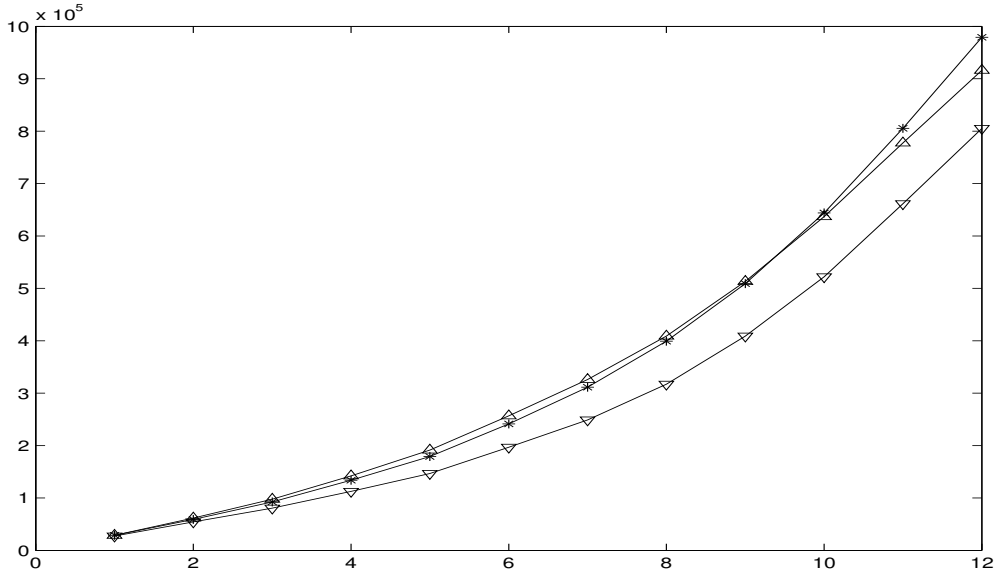


Figure 2.3: $'*'$ denotes $Q_{\ell,ARIMA}$; $'\nabla'$ denotes $Q_{\ell,CSC(1)}$; $'\triangle'$ denotes $Q_{\ell,CSC(\ell)}$.

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 boxplots shown in Figure 2.6.

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 $CSC(\ell)+ARMA(p_\ell, q_\ell)$ models are listed in table 2.1 and the rolling forecast starts at point $K = 168$. The $CSC(\ell) + ARMA(p_\ell, q_\ell)$ 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 time $\ell = 1, \ldots, 12$ (table 2.2). The $CSC(1)+AR(1)$ model gives better predictions than seasonal ARIMA for moderate prediction horizons.

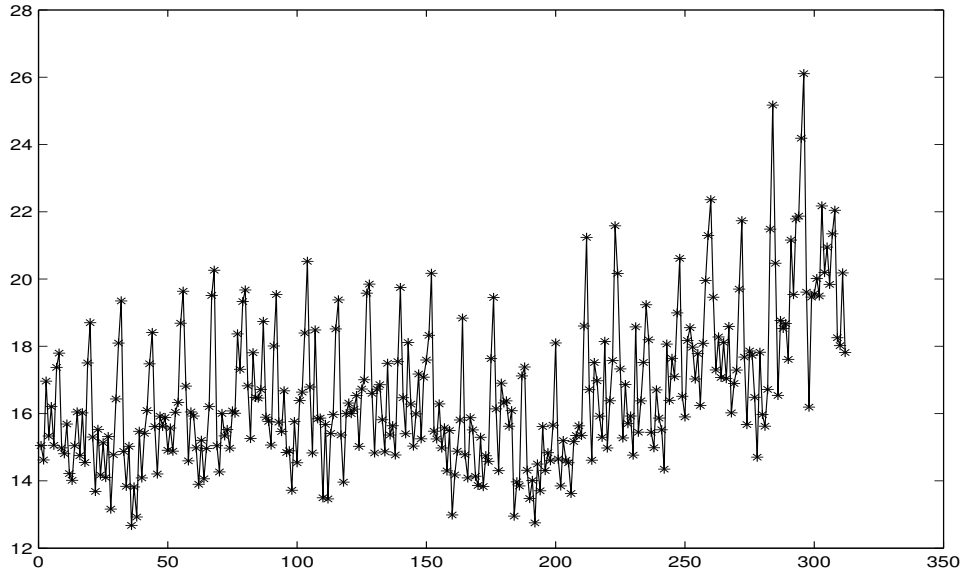


Figure 2.4: Original data plot of simulated series (II) from model $CSC(1) + AR(1)$.

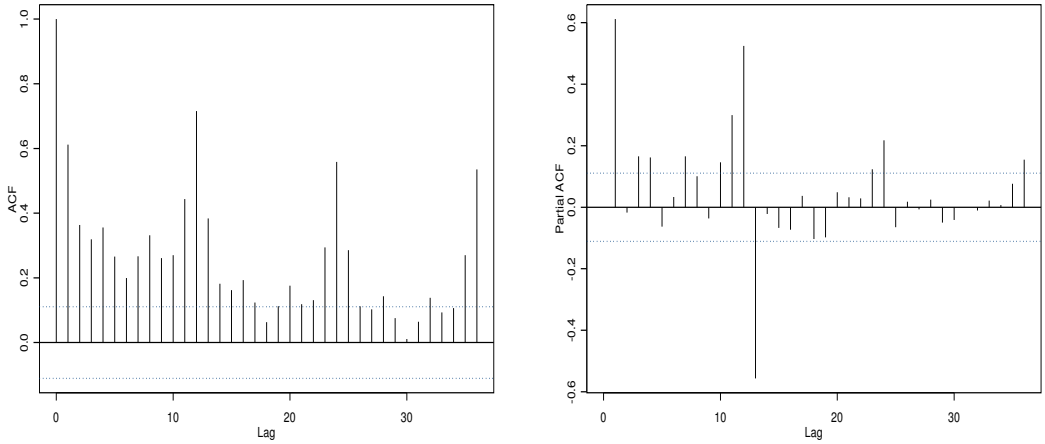


Figure 2.5: ACF and PACF of the simulated series (II).

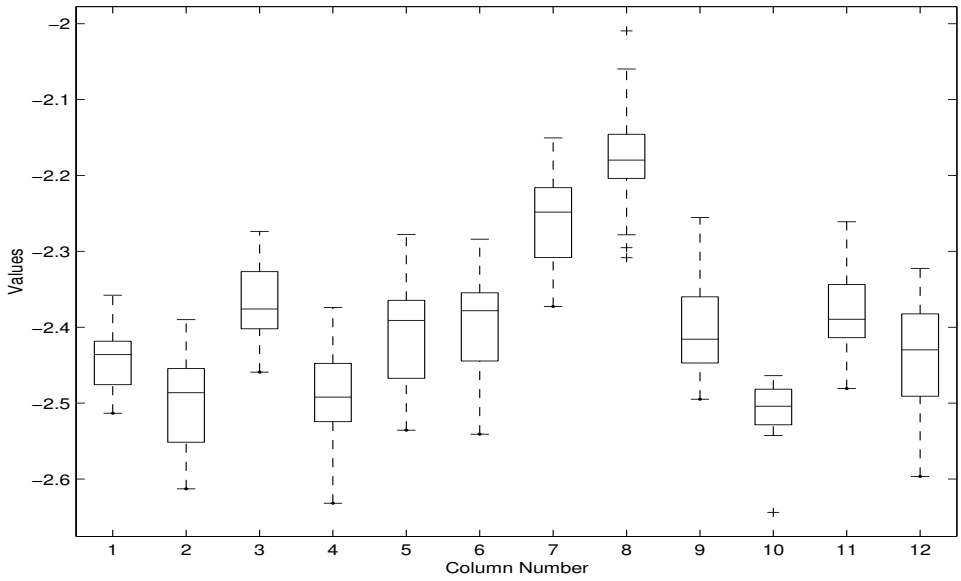


Figure 2.6: Boxplots of logratio for simulated series (II).

Table 2.1: Comparison of forecasting performance: $Q_{\ell,A}$ for seasonal ARIMA model, $Q_{\ell,C(1)}$ for $CSC(1) + AR(1)$ model and $Q_{\ell,C(\ell)}$ for $CSC(\ell) + ARMA(p_\ell, q_\ell)$ model for simulated series (II).

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

Table 2.2: Comparison of yearly-total forecasting performance for simulated series (II): M_A for seasonal ARIMA, $M_{C(1)}$ for $CSC(1) + AR(1)$, and 'Chge' denotes percent of change(%) between the two.

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

2.4.2 Real Examples

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 and the analyzed series is transformed in such a way that the models used are not affected. 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 2.7. The data is analyzed by the seasonal ARIMA, $CSC(\ell)$ and $CSC(1)$ model, respectively.

The series has a strong seasonality as observed in the ACF and PACF plots (Figure 2.8). Beyond this, the application numbers for another type of benefit have strong linear effect on the target series. We include the exogenous variable as r_t in the seasonal ARIMA model. 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 $CSC(\ell)$, the components are constructed by the value of ℓ that varies from 1 to 12. Figure 2.9 gives the boxplots of logratio series Z_t for $\ell = 3$ and $\ell = 6$ respectively. The patterns in the boxplots show that the logratio series captures the seasonality in the original series by the seasonal-dependent mean $\mu_{s(t)}$ and variance $\sigma^2_{s(t)}$. We use period $d = 12$ 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 2.4 summarizes the coefficients and standard errors for the AR estimation of the error process.

In 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 2.3.3. The $CSC(1) + AR(1)$ model is fitted based on the evident AR(1) structure of the error process.

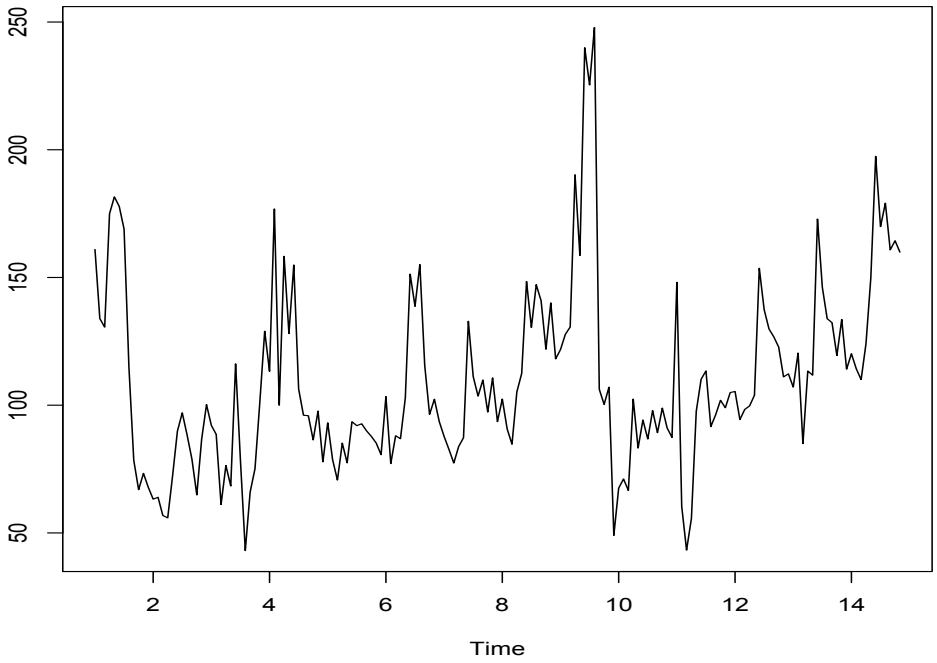


Figure 2.7: The time series plot of application volume for government benefit.

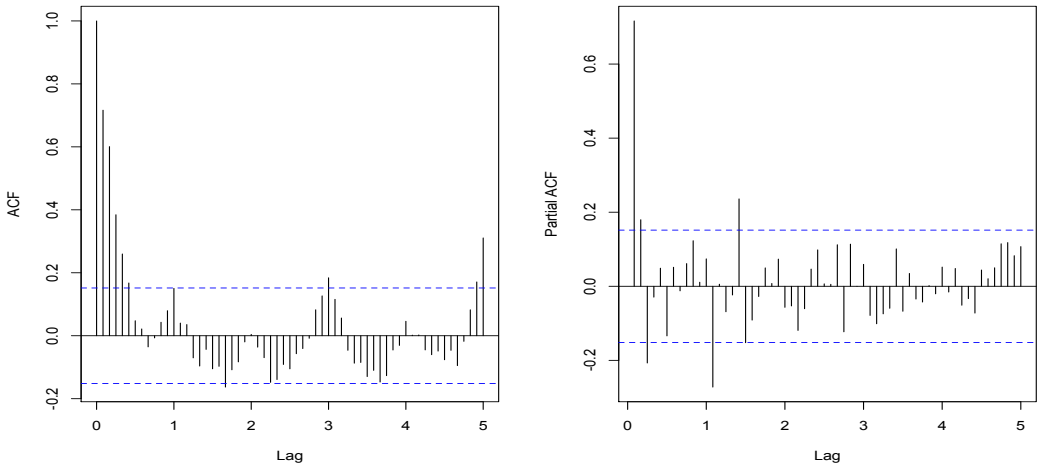


Figure 2.8: ACF and PACF of the application volume series.

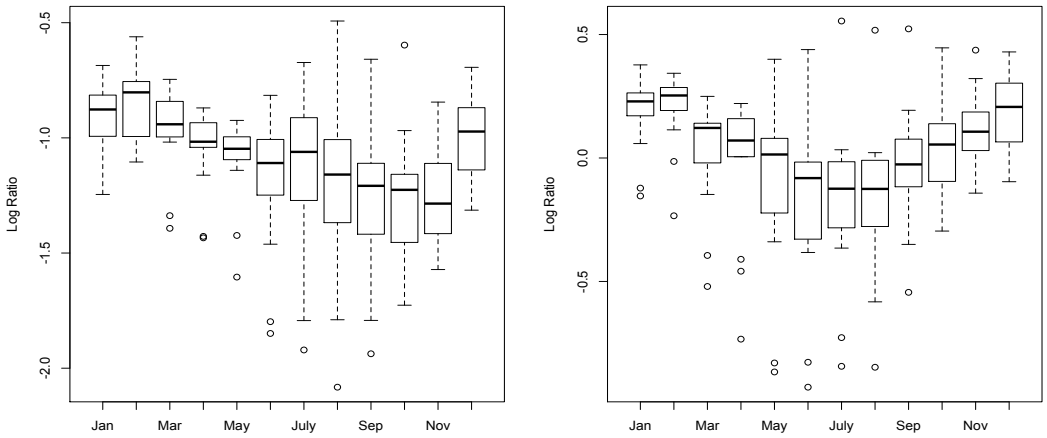


Figure 2.9: Boxplots of logratio for application volume. Left: $\ell = 3$, Right: $\ell = 6$.

Table 2.3: Model estimation for $CSC(1)$ (upper panel) and $CSC(3)$ (lower panel): $\mu_{s(t)}$ is the seasonal mean and $\sigma_{s(t)}$ is the seasonal standard deviation.

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

Table 2.4: Model estimation for AR structrue of e_t. $\hat{\phi}_j$: AR coefficients, $se(\hat{\phi}_j)$: Standard errors. Notation '–' for unavailable results due to different AR structrues.

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

The performance of the seasonal ARIMA, $CSC(\ell)$ 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 2.5 shows the square root of forecasting measure Q_ℓ of the models. From table 2.5, it is seen that both $CSC(\ell)$ 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 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 2.6 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 boxplots of the prediction errors (%) from the rolling forecast by ARIMA and $CSC(\ell)$ model in Figure 2.10. It provides clear evidence that the $CSC(\ell)$ model performs better than the ARIMA model with moderate prediction horizons.

Real Example (II)

In this example we analyze the U.S. industrial production index of energy (1997=100)

Table 2.5: Comparison of forecasting performance for benefit application: $Q_{\ell,A}(\times 10^2)$ for seasonal ARIMA model, $Q_{\ell,C(\ell)}(\times 10^2)$ for $CSC(\ell) + AR(p_\ell)$ model, $Q_{\ell,C(1)}(\times 10^2)$ for $CSC(1) + AR(1)$ model and 'Chge' denotes percent of change(%) between the two.

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

Table 2.6: Comparison of forecasting performance for benefit application: $\text{MAFE}_{\ell,A}$ for seasonal ARIMA model, $\text{MAFE}_{\ell,C(\ell)}$ for $CSC(\ell) + AR(p_\ell)$ model, $\text{MAFE}_{\ell,C(1)}$ for $CSC(1) + AR(1)$ model and 'Chge' denotes percent of change(%) between the two.

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

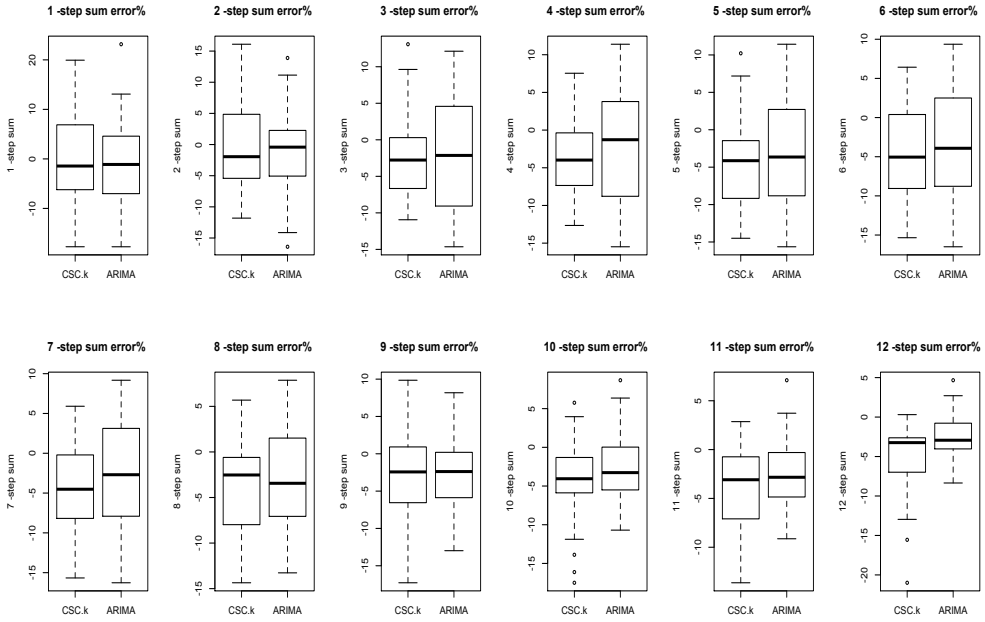


Figure 2.10: Boxplots of forecasting errors for ℓ-months total, $\ell = 1, 2, \ldots, 12$.

from January 1977 to December 2002, with 312 observations, from the website www.economagic.com.

The time series plot (Figure 2.11) shows strong seasonality with an upward trend. Based on ACF and PACF plots (Figure 2.12) and a model selection procedure, the seasonal ARIMA model

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

is used for comparison purpose.

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 2.7 presents detailed forecasting performance comparison of the two models. The values of Q_ℓ and the percentage change of Q_ℓ between the two models suggest that $CSC(1) + AR(3)$ performs better for longer prediction horizons such as $\ell = 6, \ldots, 12$.

Table 2.8 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 $CSC(1) + AR(3)$

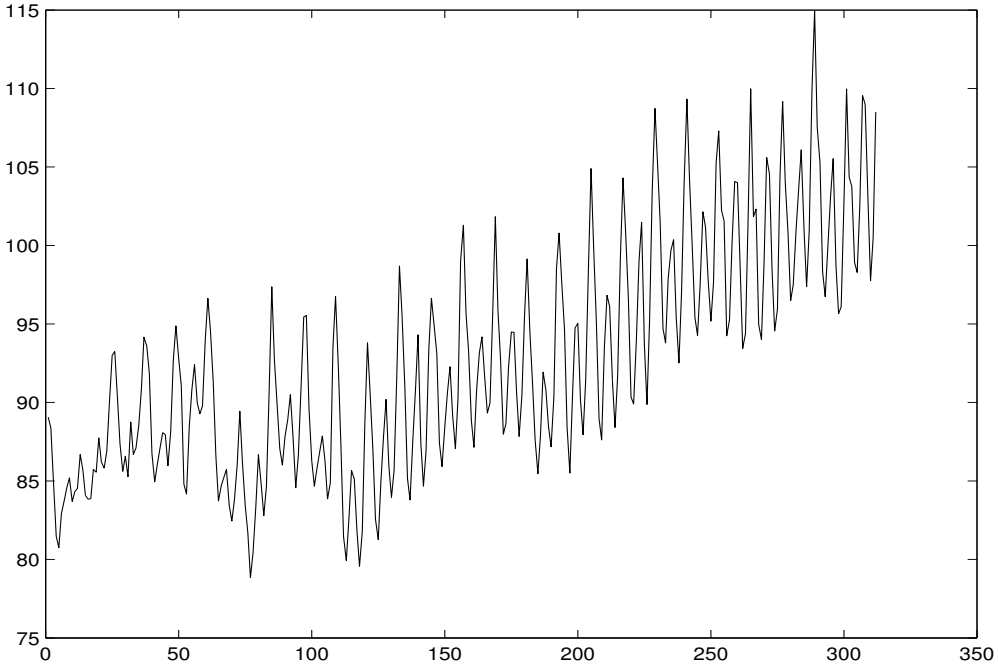


Figure 2.11: The total industrial production index of energy (1997=100) (US) (1977-2002).

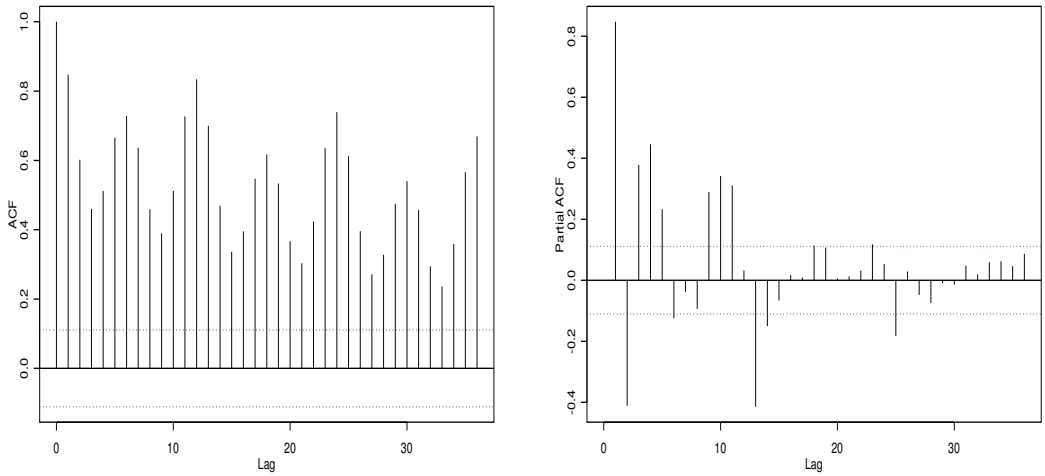


Figure 2.12: ACF and PACF of US industrial production index of energy.

Table 2.7: Comparison of forecasting performance for US industrial production index of energy: $Q_{\ell,A}$ for seasonal ARIMA model, $Q_{\ell,C(1)}$ for $CSC(1)+AR(3)$, and 'Chge' denotes percent of change(%) between the two.

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

Table 2.8: Comparison of yearly-total forecasting performance for US industrial production index of energy: MSE_A for seasonal ARIMA, $\text{MSE}_{C(1)}$ for $CSC(1)+AR(3)$, and 'Chge' denotes percent of change(%) between the two.

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

outperforms seasonal ARIMA in long horizon forecasting.

Real Example (III)

In this example we analyze the U.S. retail inventories/sales ratio for furniture, home furnishing, electronic and appliance, with totally 132 observations 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 2.13.

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 forecast starts from $K=72$. The comparison of forecasting performance is shown in table 2.9. It demonstrates that $CSC(1)+MA(3)$ outperforms seasonal ARIMA for almost all prediction horizons except for very short

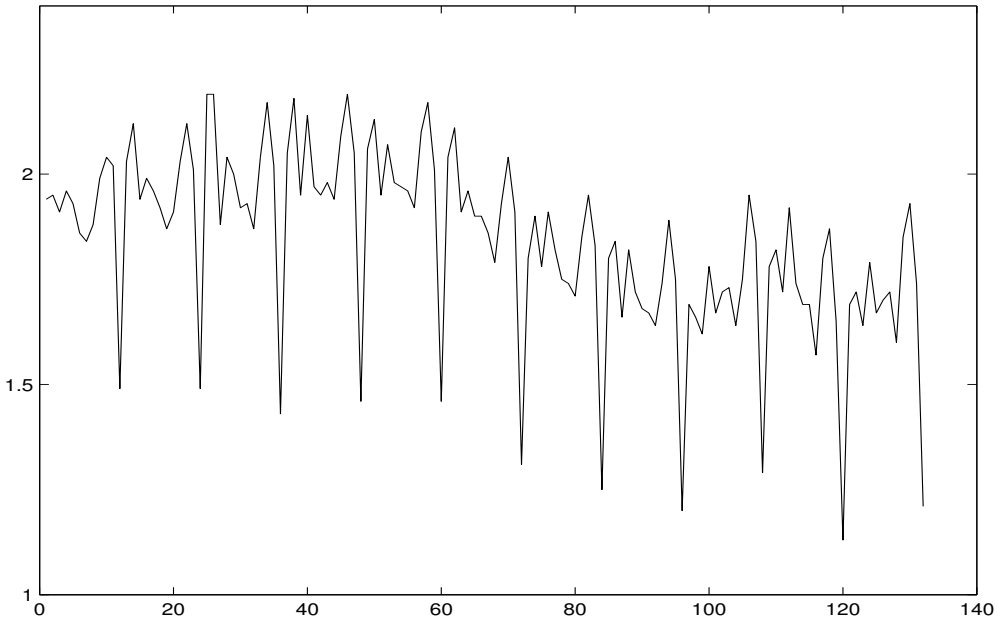


Figure 2.13: US retail inventories/sales ratio for furniture etc. (1992-2002).

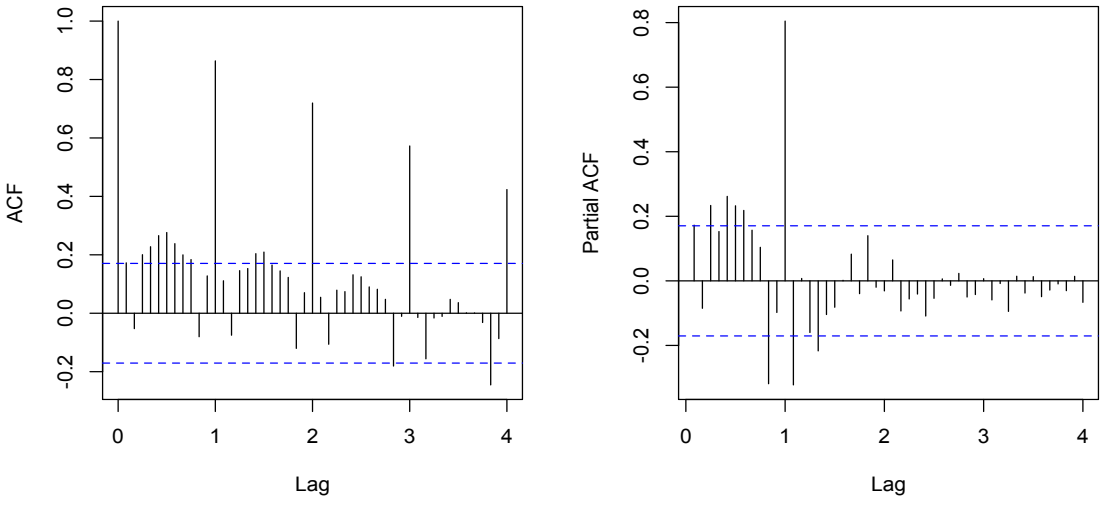


Figure 2.14: ACF and PACF of US retail inventories/sales ratio for furniture, etc.

Table 2.9: Comparison of forecasting performance for US retail inventories/sales ratio for furniture etc: $Q_{\ell,A}(\times 10^2)$ for seasonal ARIMA, $Q_{\ell,C(1)}(\times 10^2)$ for $CSC(1)+MA(3)$, and 'Chge' denotes percent of change(%) between the two.

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

terms when $\ell = 1$ and 2. In addition, the improvement of prediction is more significant for longer forecasting horizons.

2.4.3 Predictive Distribution

The CSC model we proposed here is a prediction-based model, and it shows the superior forecasting power in previous sections. In this section, we are going to explore the predictive distribution on the CSC model to give more insight of the prediction with confidence levels. The method comes from the theory of confidence distribution in a series of eminent work in Singh et al. (2007). Traditionally, people construct confidence intervals or regions at some given confidence level based on some specific distribution assumptions. By the predictive distribution approach, we are able to analyze the forecasting certainty at any confidence levels. For complex prediction models such as the CSC model, it is attractive to apply this approach to get more certainty of the predictions.

Note that we are able to construct prediction distribution for any prediction length ℓ and for any starting time t. The algorithm for deriving the predictive distribution is described as follows,

1. Construct the series of Z_t based on the CSC model, and get estimation of $\hat{\mu}_{s(t)}$ and $\hat{\sigma}_{s(t)}$ respectively.

2. Construct $\hat{e}_t$ from the previous step as,

$$\hat{e}_t = \frac{Z_t - \hat{\mu}_{s(t)}}{\hat{\sigma}_{s(t)}}.$$

Then perform standard ARMA procedure on $\hat{e}_t$ to get estimations

$$\left(\hat{\phi}_1, \cdots, \hat{\phi}_p; \hat{\theta}_1, \cdots, \hat{\theta}_q\right),$$

and their standard errors

$$\left(se(\hat{\phi}_1), \cdots, se(\hat{\phi}_p); se(\hat{\theta}_1), \cdots, se(\hat{\theta}_q)\right).$$

3. Simulate one observation from $N\left(\hat{\phi}_i, se(\hat{\phi}_i)\right)$ for each $i = 1, \cdots, p$, denoted as $\hat{\phi}_i^*$. Similarly, simulate one observation from $N\left(\hat{\theta}_j, se(\hat{\theta}_j)\right)$ for each $j = 1, \cdots, q$, denoted as $\hat{\theta}_j^*$.

4. Iteratively get ℓ-step prediction of $\hat{e}_t$,

$$\hat{e}_t^*(\ell) = \hat{\phi}_1^* \hat{e}_t^*(\ell - 1) + \cdots + \hat{\phi}_p^* \hat{e}_t^*(\ell - p).$$

5. Derive the predicted value of $Y_{2,t}$ as

$$\hat{Y}_{2,t}^{(1)} = Y_{1,t} \exp\left(\hat{\mu}_{s(t)} + \hat{\sigma}_{s(t)} \hat{e}_t^*(\ell)\right).$$

6. Perform Step 3 to Step 5 for $n = 1000$ times, get a series of predictive values,

$$\hat{\boldsymbol{Y}}_{2,t}^* = \left(\hat{Y}_{2,t}^{(1)}, \hat{Y}_{2,t}^{(2)}, \ldots, \hat{Y}_{2,t}^{(n)}\right),$$

then distribution of $\hat{\boldsymbol{Y}}_{2,t}^*$ is referred as the predictive distribution of the predicted value for the next ℓ-months total starting at time t.

We apply this method to the first empirical example in section 2.4.2. The algorithm is performed for three months total application volume prediction with $\ell = 3$. The histograms of $\hat{\boldsymbol{Y}}_{2,t}^*$ are showed in Figure 2.15 for predictions that start from $t = 154, 155, 156$ and 157 respectively.

The method of predictive distribution provides us the flexibility to have insight on the prediction certainties. Take the simulation for $t = 156$ for example, we are able to get answers for questions such as,

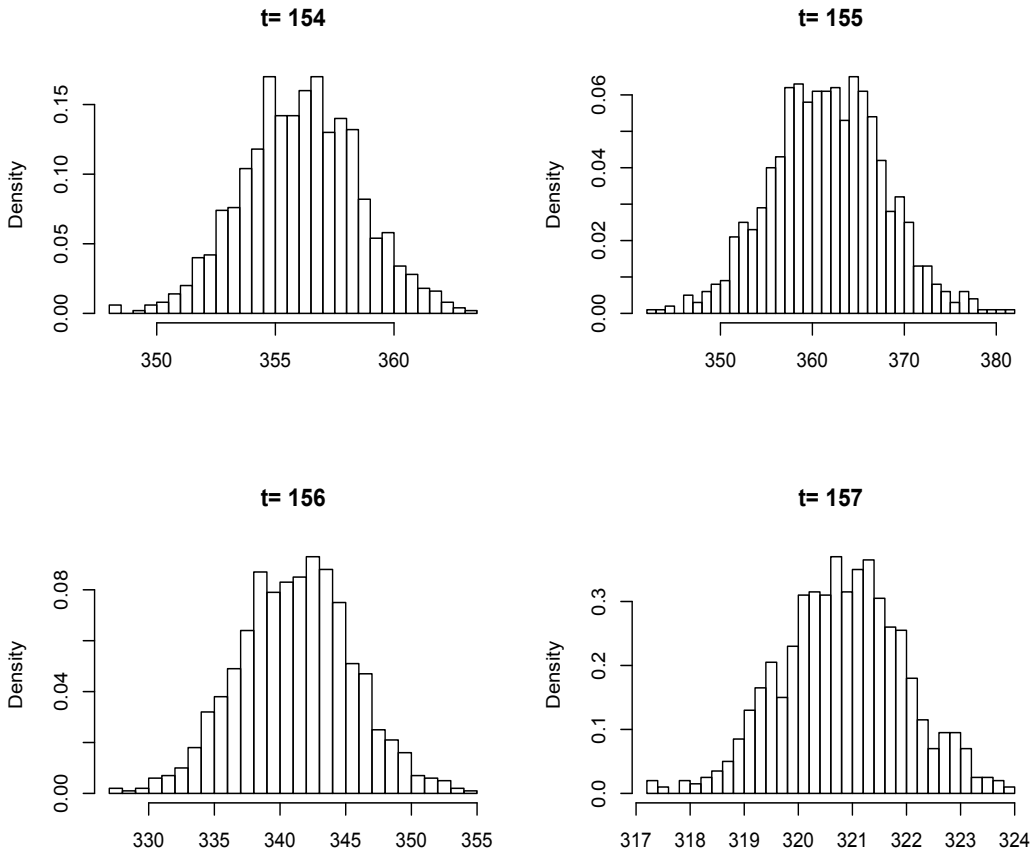


Figure 2.15: Predictive distribution of CSC model on real example (I) when $\ell = 3$. Prediction starts from $t = 154, 155, 156$ and 157 respectively.

1. What is the predictive interval at 90% confidence level?

   Answer: The predictive interval is (334, 348).

2. How confident that the total application volume from $t = 156$ to $t = 158$ will be greater than 340, 345 and 350?

   Answer: The confidence is 60.5%, 25.6% and 2.1% respectively.

3. What is the lowest expectation of the application volume at 90% confidence level?

   Answer: There is 90% confidence that the application volume will exceed 336.

The sample questions above provide us more indications of what we can get from the prediction. With predictive distributions, we can stress the certainty of prediction at any confidence levels in need.

2.5 Summary

While seasonal time series has been extensively discussed in a lot of literatures, this chapter offers an innovative perspective of seasonality by combining the stable seasonal pattern and compositional approach. The well-constructed compositional seasonal component model is introduced in details about the model setting, estimation and prediction procedures.

Compared with the Box-Jenkins seasonal ARIMA model, we provide a new model with more general assumptions and better forecasting performance as well. Stationarity is always a desirable property for most of the time series models. The nonstationary patterns such as unit-roots and trends are usually modeled by differencing and covariate terms. Therefore, nonstationarity significantly increases the model complexity especially in seasonal time series models where higher-order seasonal terms are presented. The proposed CSC model has the substantial advantage that it does not require stationarity of the original seasonal time series so that the modeling process will be simplified in this sense. In the other aspect, the compositional approach gives the normally of the logratio series and allows us to model the non-seasonal error process by pure ARMA procedures, which is much easier to deal with.

As a prediction-based model, the application of CSC model is flexible as well. It is designed to predict the next ℓ-seasons total numbers while ℓ can vary in a feasible range. This feature is highly desirable in many applications such as inventory management, marketing planing or operation management. The superior prediction power of the model has been demonstrated by a series of empirical studies that covers government operation, retail and energy production industries.

Chapter 3

Compositional Time Series Analysis with Threshold Cointegration Patterns

3.1 Background

The household consumption expenditure is an important topic that researchers and economists are keeping heated interests in. As the major component of the Gross Domestic Product in the United States, the household consumption expenditure takes seventy percent of the total Gross Domestic Product. It significantly reflects the economic health condition of the nation in tax, fiscal policy, inflation and purchasing powers. Numerous literatures appear in the analysis of household consumption behaviors. Krueger and Perri (2006) discussed the effects of income inequality on household consumption behaviors. Stephens Jr. (2004) examines the link between expectations of future job losses and the subsequent impact on household consumption behavior. The growth rate of household expenditures is a key factor in the model of asset pricing kernel for evaluating the aggregate risk in Perri et al. (2008).

The household expenditures can be naturally categorized by different research purposes. For example, it can be differentiated as durable and nondurable goods by major type of products. It also can be recorded by different consumption functions such as foods, education, health or transportation. In recent years, researchers are more and more interested in applying quantitative methods to household expenditure analysis. Barigozzi et al. (2012) try to find the statistical distribution properties of both the aggregated and disaggregated household consumption data. They take attention to the distribution of each component, but fail to consider the sum constrain in the individual analysis. Del Boca and Flinn (2012) take into account the endogenous interactions between the components and improve the prediction of household budget allocation by

adding a list of participating constraints.

Among various categorical methods, the budget allocation by consumption functions is of particular interests as they are substantial to reflect people's daily life. This paper focus on the cross-correlations between different consumption functions. as well as the evolution of the household budget pattern on exogenous economic factors. We release the sum constraint on the consumption category proportions by compositional time series theories and build a VAR model on the logratio series. Then we find the cointegration effect which describes the long-run equilibrium between the transformed variables. Furthermore, threshold effects arise in the system when analyzing the model tendency by exogenous variables. Therefore, the threshold cointegration model by Balke and Fomby (1997) is studied on the cointegrating vector from the household consumption data, and the threshold vector error-correction model is explored as well.

The discussion is organized as follows. The preliminary study on the household data and the compositional transformation are introduced in section 3.2. The vector error-correction model is also introduced based on the cointegration pattern in the VAR model. The threshold effects are discussed extensively in this section. We also propose a joint maximum likelihood estimation to the threshold VECM model as an extension of the existing estimation method in section 3.3. Section 3.4 shows the model estimation on the household data under joint MLE, and the model comparison results between the linear VECM model and the threshold VECM models. The rolling forecast performance are compared in section as well.

3.2 Preliminary Study on Household Consumption Data

3.2.1 The Data

The data are downloaded from the U.S. Bureau of Economic Analysis website. It is annually data from 1947 to 2012 and measured in billions of real dollars. The data is recorded for 13 consumption functions respectively in each year: Food and beverages purchased for off-premises consumption; Clothing, footwear, and related services; Housing, utilities, and fuels; Furnishings, household equipment, and routine household

maintenance; Health; Transportation; Communication; Recreation; Education; Food services and accommodations; Financial services and insurance; Other goods and services; Net foreign travel and expenditures abroad by U.S. residents. We refer the consumption functions as sectors in the remaining discussion.

Figure 3.1 shows the time series patterns of the expenditures in each sector. The expenditures are strictly increasing in raw quantities, reflecting the overall improvement of the economic condition since 1950's. However, the expenditures for each sector have different behaviors in percentages. In contrary to the substantial drop in food, clothing and furnishing, the proportions of health, commutation, recreation and education have readily increased over the past sixty years. The fact reveals that people have been allocating more and more budget in their demand for cultural life. The sharp drop for the transportation sector after 1980 is resulted from the rapid burgeon of computers and internet. As can be seen, the patterns of proportions give us the intuition that the household budget pattern has gradually moved its focus away from the basic needs in foods and clothing, and gets into the higher level needs such as health, education and recreations.

The 13 sectors cover almost all aspects of people's life. However, it is infeasible and unnecessary to simultaneously model 13 sectors. In this case, we aggregate similar sectors together, trying to reduce the model into a reasonable dimension. Finally, five components are defined with abbreviated names for convenience,

- z_1: Aggregation of food and clothing, footwear, and related services. Note as food and clothing later.
- z_2: Aggregation of housing, utilities, fuels and furnishings. Note as housing and furnishing later.
- z_3: Aggregation of health, recreation and education,
- z_4: Aggregation of communication and transportation,
- z_5: Aggregation of all services and net foreign travels. Note as services later.

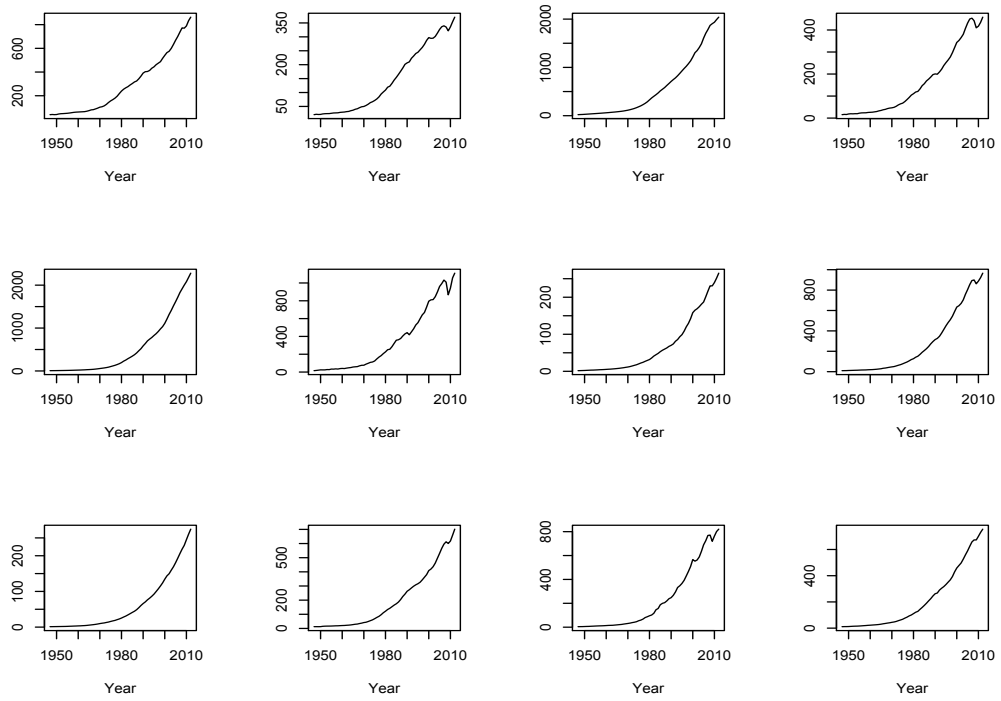


Figure 3.1: US household consumption expenditures by sector: real US dollars (1947-2012).

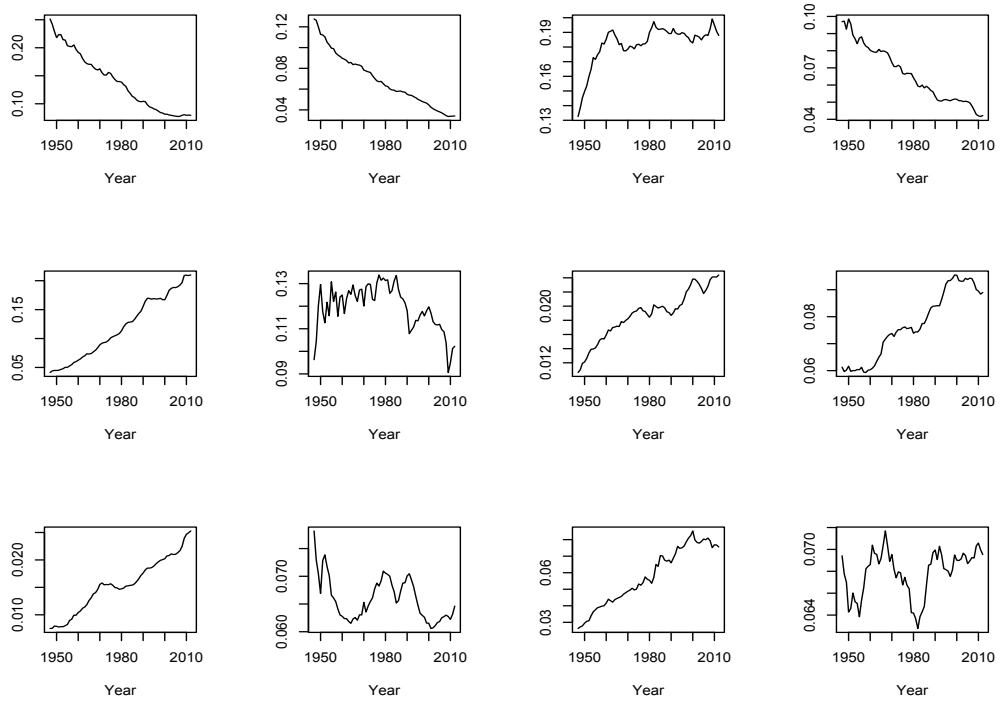


Figure 3.2: US household consumption expenditures by sector: percentages out of the total expenditure (1947-2012).

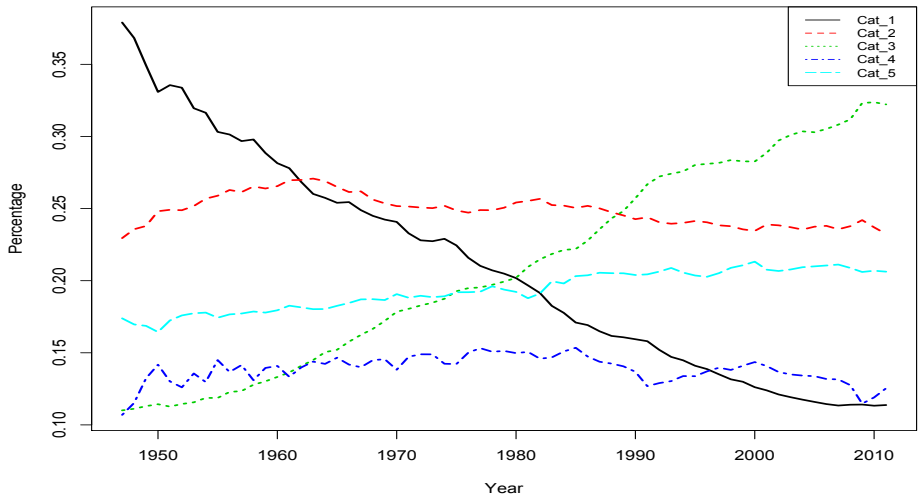


Figure 3.3: The compositions of US Household Consumption Expenditures by function(%) (1947-2012).

The observations in $\boldsymbol{z}_t$ are measured in US dollars, we can directly get the compositions $\boldsymbol{x}_t$ defined in Equation 1.6. Figure 3.3 shows the basic features of the compositions. By aggregation, it is clearer that the household budget allocation to food, clothing and related services keeps decreasing since 1947, and the consumption in health, education and recreation has a steady increase in recent years. The consumption on other categories are relatively stable with a slightly increase in services.

Based on Equation 1.8, we get the logratio series $\boldsymbol{y}_t$ by choosing x_1 as the base series,

$$y_{1t} = \log\left(\frac{x_{2t}}{x_{1t}}\right), \quad y_{2t} = \log\left(\frac{x_{3t}}{x_{1t}}\right), \quad y_{3t} = \log\left(\frac{x_{4t}}{x_{1t}}\right), \quad y_{4t} = \log\left(\frac{x_{5t}}{x_{1t}}\right).$$

Note that the choice of base series will not affect the model features due to the invariant property of additive logratio transformation in Aitchison (1986a).

3.2.2 Cointegration: Vector Error-Correction Model

The upward pattern in Figure 3.4 indicates that the logratio series are not stationarity. We further validate the existence of unit-roots by the exponentially decay patterns in Figure 3.5 and the unit-root test. Given the unit-roots in $\boldsymbol{y}_t$, we apply the rank tests by Johansen (1991) to find the cointegration effect and further determine the cointegrating

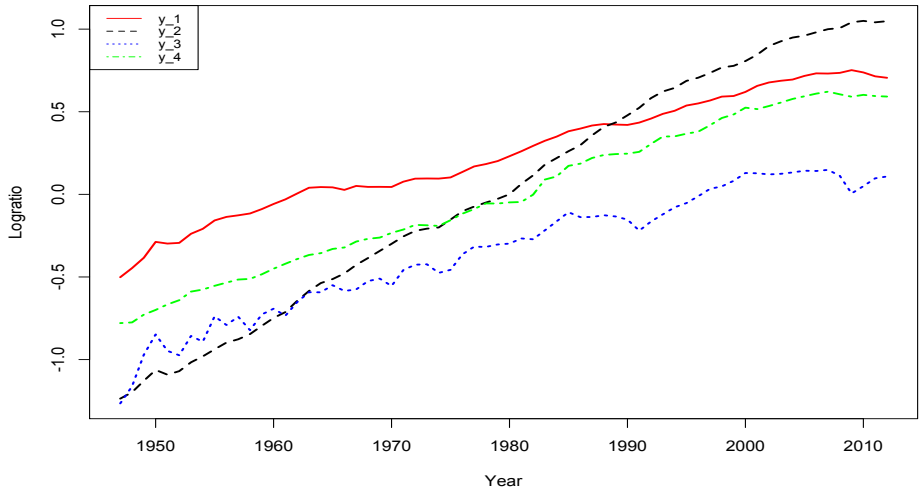


Figure 3.4: The time series after additive logratio transformation of US Household Consumption Expenditures components. The base series is z_1.

Table 3.1: Johansen's Trace Test Statistic and corresponding critical values

	Test Statistic	10% Level	5% Level	1% Level
$m \leq 3$	4.06	6.50	8.18	11.65
$m \leq 2$	11.46	15.66	17.95	23.52
$m \leq 1$	32.55*	28.71	31.52	37.22
$m = 0$	62.43	45.23	48.28	55.43

vectors. We assume a VAR(2) model on the differenced random vector $\Delta \boldsymbol{y}_t$ based on evidences from previous sections. Both the trace test and maximum eigenvalue test results are summarized in table 3.1 and table 3.2. They strongly support the fact that there is one cointegration relation between the four random variables in $\boldsymbol{y}_t$.

Cointegration is the phenomenon that the linear combination of two or more unit-root non-stationary series is stationary. With presence of cointegration, there is a common trend between the series and taking differences on the original levels will result in an over-differenced model. The linear vector error correction model (VECM) is introduced as,

$$\boldsymbol{\Delta y_t = \mu D_t + \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta y_t + \varepsilon_t - \sum_{j=1}^{q} \theta_j^* \varepsilon_{t-j},} \tag{3.1}$$

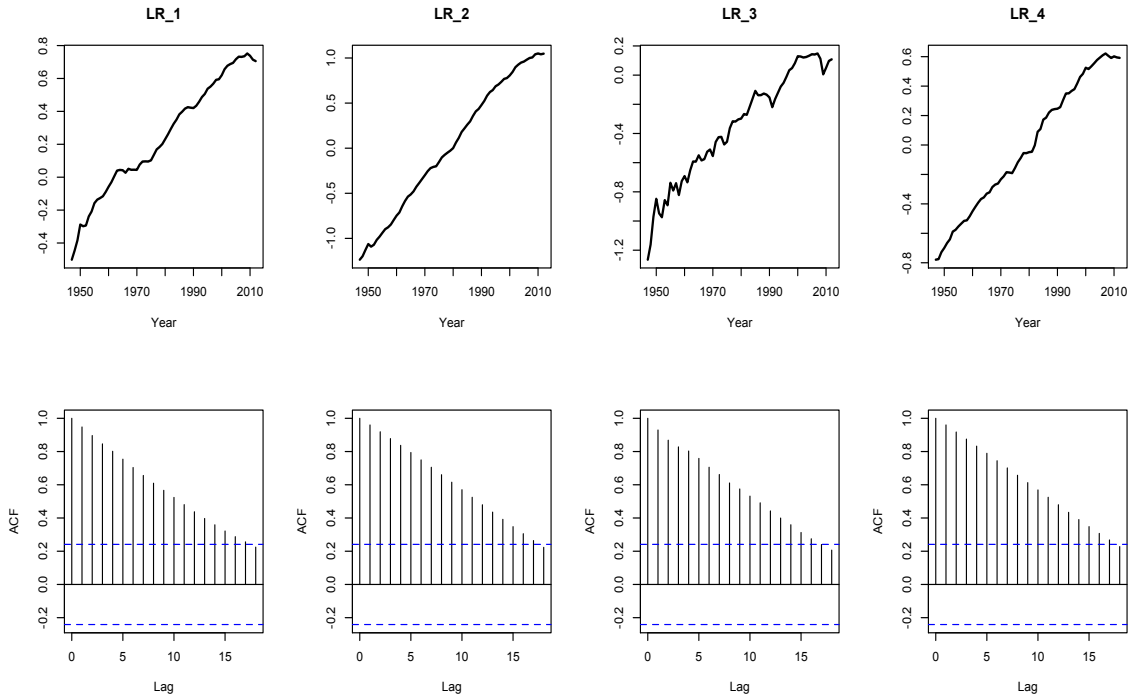


Figure 3.5: The logratio series and the corresponding ACF plot. Upper: Time series plots of the four logratio series from LR_1 to LR_4. Lower: ACF plots of the logratio series.

Table 3.2: Johansen's Maximum Eigenvalue Test Statistic and corresponding critical values

	Test Statistic	10% Level	5% Level	1% Level
$m = 3$	4.06	6.50	8.18	11.65
$m = 2$	7.40	12.91	14.90	19.19
$m = 1$	21.09*	18.90	21.07	25.75
$m = 0$	29.89	24.78	27.14	32.14

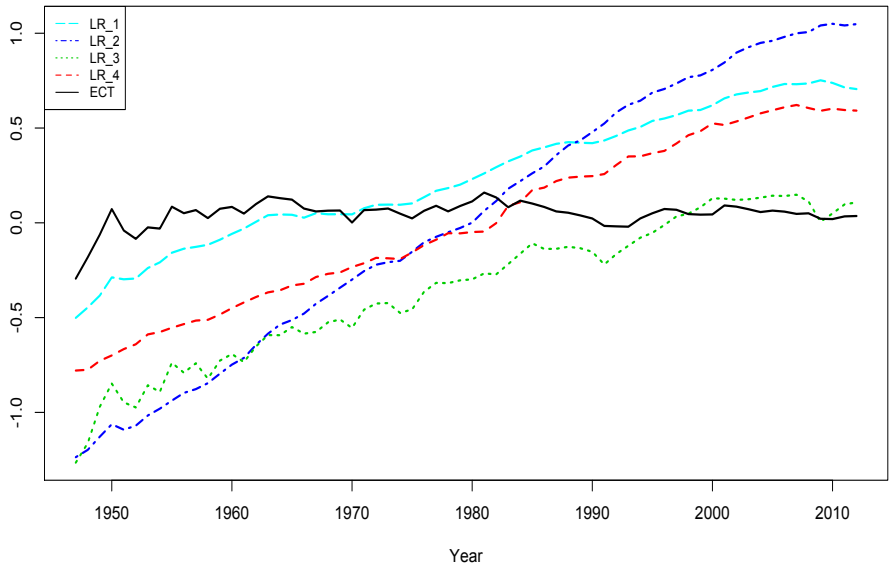


Figure 3.6: Pattern comparison of $y_{1t}, y_{2t}, y_{3t}, y_{4t}$ and the error correction term z_t (1948-2011).

where $t = 1, 2, \ldots, T$, $z_{t-1} = \boldsymbol{\beta}' \boldsymbol{y_{t-1}}$ is the the error correction term and $\boldsymbol{\varepsilon}_t$ is a k-dimensional Gaussian random variable such that $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. The term of $\boldsymbol{D}_t$ stands for the deterministic effects and the matrices $\boldsymbol{\phi}_i^*$ and $\boldsymbol{\theta}_j^*$ are the adjusted coefficients of the VARMA(p, q) model. The error correction term z_t is unit-root stationary based on cointegration feature. Combind with the stationarity of $\Delta \boldsymbol{y}_t$, Model 3.1 is stationary.

Applying Model 3.1 to the household data, the cointegration coefficient vector and the rate of cointegration are estimated by Johansen's MLE as,

$$\tilde{\boldsymbol{\alpha}}' = (-0.08, -0.04, 0.02, 0.26)', \quad \tilde{\boldsymbol{\beta}}' = (1, 0.31, 0.13, -1.38)'.$$

The estimated error correction term $\tilde{z}_t$ needs to be unit-root stationary as proposed, and we can validate this by a unit-root test with p-value of 0.01. Figure 3.6 shows the behavior of the original series in $\boldsymbol{y}_t$ and $\tilde{z}_t$. It is clearly seen that stationarity is achieved by linear combination of four unit-root pattern series. While $\tilde{z}_t$ fixed, the estimation of $\boldsymbol{\mu}_t$ and $\boldsymbol{\phi}_1^*$ in model 3.1 are shown in table 3.3 with significance indicated by standard errors .

Table 3.3: Linear VECM model estimation on US household expenditures data

	Δy_{1t}	Δy_{2t}	Δy_{3t}	Δy_{4t}
$\boldsymbol{\mu}_t$	0.013	0.024	0.013	0.010
	(0.005)	(0.005)	(0.014)	(0.005)
ECT_{t-1}	-0.150	-0.068	-0.540	-0.021
	(0.039)	(0.042)	(0.109)	(0.043)
$\Delta y_{1,t-1}$	0.227	-0.142	-0.202	0.200
	(0.152)	(0.164)	(0.428)	(0.169)
$\Delta y_{2,t-1}$	0.182	0.544	1.047	0.244
	(0.151)	(0.163)	(0.426)	(0.168)
$\Delta y_{3,t-1}$	0.024	-0.043	-0.159	-0.077
	(0.047)	(0.051)	(0.133)	(0.053)
$\Delta y_{4,t-1}$	0.054	-0.062	0.199	0.095
	(0.126)	(0.136)	(0.355)	(0.140)

3.2.3 Threshold Cointegration

The previous section validates the cointegration effect and gives the stationary part as

$$\tilde{z}_t = \tilde{\boldsymbol{\beta}}' \boldsymbol{y}_{t-1} = y_{1t} + 0.31 y_{2t} + 0.13 y_{3t} - 1.38 y_{4t}.$$

With stationarity, we can proceed with further research on $\tilde{z}_t$ for its economic behaviors. The household consumption is directly related to the household income level, the tax levels, and the national GDP condition. We explored the statistical connections between the long-run equilibrium $\tilde{z}_t$ and these key economic indicators, and found attractive pattern between $\tilde{z}_t$ and the U.S. GDP growth rate.

GDP is one of the most widely discussed economic indicators. It is the key criterion for policy makers to judge the economical condition of the nation and identify the impact of factors such as the fiscal and tax policies. As the largest component, personal consumption keeps contributing around 70% to the total US GDP. Therefore, we are interested in the influence of US real GDP change on the household consumption patterns. In particular, the threshold effect of US real GDP growth rate on the adjustment process $\tilde{z}_t$ is identified and will be discussed extensively in this section. The annual US real GDP growth rate (1947-2012) is shown in Figure 3.7. It is a stationary time series that is eligible to be a threshold variable in the subsequent models.

Given the estimated cointegrating vector, we try the arranged autoregression test

by Tsay (1989) for the threshold effect of GDP growth rate γ_t on $\tilde{z}_t$. An AR(2) model is imposed on $\tilde{z}_t$ and the threshold test is performed on lags 1 and 2 respectively. The threshold pattern in γ_{t-1} is stronger with p-value of 0.0045 hence it is picked as the threshold variable. The Sup-Wald test by Hansen (1999) is a widely-used threshold test as well. It allows testing for more than two regimes and is more general than Tsay's arranged autoregression model. Due to the length of our data, we do not expect more than two regimes in the threshold model and Tsay's test is straightforward in our case.

On the other hand, the threshold cointegration model by Balke and Fomby (1997) is introduced to describe the threshold effect in the long-run equilibrium in cointegration models. It identifies the threshold effect on itself but it can be generalized to use exogenous variable as the transition variable. Generally, a threshold cointegration model is defined as

$$z_t = \mu_j + \rho_{j,1} z_{t-1} + \rho_{j,2} z_{t-2} + \cdots + \rho_{j,p} z_{t-p} + e_t, \quad \text{when } \theta_{j-1} < r_{t-d} \leq \theta_j, \tag{3.2}$$

where $j = 1, 2, \ldots, m$ denotes the state of the model if there are m regimes. The coefficients in ρ_j depend on the state of r_t from d time lags ago. The transition variable r_t can be the error correction term itself, or a weak-stationary exogenous variable.

As suggested by Balke and Fomby, the testing of threshold cointegration can be done separately for cointegration and threshold effect. In previous discussions, we confirmed the cointegration pattern with Johansen's method, and discovered the threshold pattern in the estimated error correction term $\tilde{z}_t$ using Tsay's threshold test. In Tsay's test, we produced recursive least square estimates on arranged autoregression that is ordered on the transition variable γ_{t-1}. The threshold value is identified to be 4.1 by minimizing the overall AIC. The t-ratio plots in Figure 3.8 provide solid evidence for the threshold effect of γ_{t-1} on $\tilde{z}_t$. The discontinuities at $\gamma_{t-1} = 4.1$ in the t-ratio plots suggest that a linear model on the error correction term is inappropriate whereas the two-regime threshold model is strongly recommended.

The following equation gives the estimation of the threshold model on $\tilde{z}_t$ with US real GDP growth rate at time lag 1 as threshold variable. We assume an AR(2) model for each regime since all the coefficients are significant. The roots of the characteristic

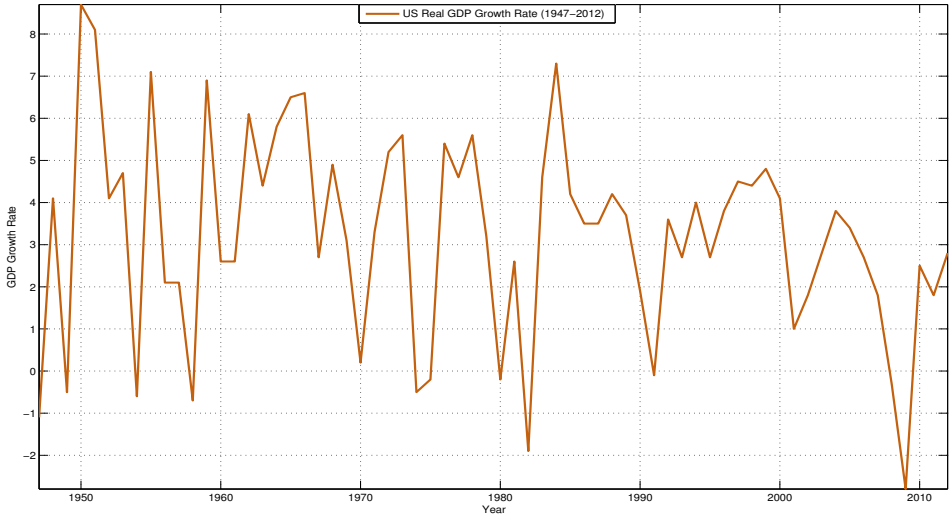


Figure 3.7: US real GDP growth rate in percentage (1947-2012).

Table 3.4: Comparison of the linear and threshold model on ECT

Criteria	SSE	AIC
Linear Model	0.074	-411.38
Threshold Model	0.056	-422.76

function for the outer regime are 1.147 and 5.25 and they ensure the stationarity of the threshold model. A comparison of SSE and AIC between the linear model and the threshold model on $\tilde{z}_t$ is summarized in table 3.4. It fits the process better than the linear model with smaller SSE and AIC values.

$$\tilde{z}_t = \begin{cases} 0.147 + 0.874\tilde{z}_{t-1} - 0.526\tilde{z}_{t-2} + u_t, & \text{if } \gamma_{t-1} \leq 4.1, \\ 0.030 + 0.682\tilde{z}_{t-1} + 0.166\tilde{z}_{t-2} + u_t. & \text{if } \gamma_{t-1} > 4.1. \end{cases}$$

3.3 Joint Maximum Likelihood Estimation for TVECM

The classic threshold cointegration model by Balke and Fomby concentrates on the threshold pattern of the cointegrating vector only, but the threshold effect may also exist in the multivariate structure since the cointegrating vector is a linear combination of

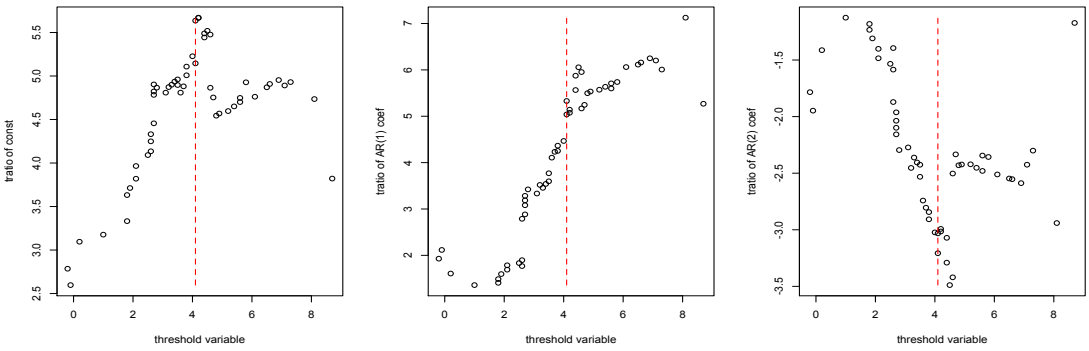


Figure 3.8: The t-ratio of arranged regression coefficients versus the ordered US real GDP growth rate at time lag 1. From left: the tratio of intercept, AR(1) coefficient and AR(2) coefficient.

them. Hansen and Seo (2002) provides the maximum likelihood estimation for the two-regime threshold vector error-correction model (TVECM). The model with unrestricted mean is defined as,

$$\Delta \boldsymbol{y}_t = \begin{cases} \boldsymbol{\mu}_t^{(1)} + \boldsymbol{\alpha}^{(1)}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \Phi_1^{(1)}\Delta \boldsymbol{y}_{t-1} + \cdots + \Phi_{p-1}^{(1)}\Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon_t}, & \text{if } \gamma_{t-d} \leq \gamma_0, \\ \boldsymbol{\mu}_t^{(2)} + \boldsymbol{\alpha}^{(2)}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \Phi_1^{(2)}\Delta \boldsymbol{y}_{t-1} + \cdots + \Phi_{p-1}^{(2)}\Delta \boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon_t}. & \text{if } \gamma_{t-d} > \gamma_0. \end{cases} \tag{3.3}$$

The parameters of the model are $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \Sigma_1, \boldsymbol{\beta}, \gamma)$ where $\boldsymbol{\theta}_j = \left(\boldsymbol{\mu}_t^{(j)}, \boldsymbol{\alpha}^{(j)}, \Phi_1^{(j)}, \ldots, \Phi_{p-1}^{(j)}\right)$, with $j = 1, 2$.

It worths attention that the linear estimation of the cointegrating vector is likely to be inaccurate due to the embedded nonlinearities. Hansen and Seo (2002) suggest updating the linear cointegration by maximum likelihood estimation of the threshold vector error-correction model, and the threshold value will be located simultaneously. The threshold variable is defined as the error correction term itself, but it can be an exogenous variable as long as it is stationary.

In the estimation of TVECM, Hansen's approach concentrates on the threshold structure in the vector error-correction model but ignores the threshold pattern in the cointegrating vector. While most of the literatures focus on the empirical analysis of two-dimensional $\boldsymbol{y}_t$ where only one cointegrating vectors might appear, multiple cointegration relations are common in higher dimensional cases. Hansen's method will lose information if there are multiple cointegrating vectors in the model. In this section,

we propose the joint maximum likelihood estimation of the threshold vector error-correction model and the threshold cointegration model on $\boldsymbol{z}_t$ together.

Also, the two-regime threshold model for the error correction term z_t is stated as

$$\boldsymbol{z}_t = \boldsymbol{\beta}^{'}\boldsymbol{y}_{t-1} = \begin{cases} \boldsymbol{\rho}_0^{(1)} + \boldsymbol{\rho}_1^{(1)}\boldsymbol{z}_{t-1} + \cdots + \boldsymbol{\rho}_\ell^{(1)}\boldsymbol{z}_{t-\ell} + \boldsymbol{u}_t, & \text{when } \gamma_{t-d} \leq \gamma_0, \\ \boldsymbol{\rho}_0^{(2)} + \boldsymbol{\rho}_1^{(2)} z_{t-1} + \cdots + \boldsymbol{\rho}_\ell^{(2)}\boldsymbol{z}_{t-\ell} + \boldsymbol{u}_t. & \text{when } \gamma_{t-d} > \gamma_0. \end{cases} \tag{3.4}$$

Assume that the errors $\boldsymbol{\varepsilon}_t$ and u_t are both *i.i.d* Gaussian random variable, and $\boldsymbol{\varepsilon}_t$ and u_t are independent. With this assumption, we propose the estimation of model (3.3) and model (3.4) simultaneously by joint maximum likelihood estimation. The joint log-likelihood is

$$\begin{aligned} & L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2; \Sigma_1, \Sigma_2, \boldsymbol{\beta}, \gamma) \\ = \; & -\frac{n-p}{2}\log|\Sigma_1| - \frac{1}{2}\sum_{t=p+1}^{n} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\beta}, \gamma)^{'}\Sigma_1^{-1}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\beta}, \gamma) \\ & -\frac{n^*}{2}\log|\Sigma_2| - \frac{1}{2}\sum_{t=p+\ell}^{n} \boldsymbol{u}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\beta}, \gamma)^{'}\Sigma_2^{-1}\boldsymbol{u}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\beta}, \gamma), \end{aligned} \tag{3.5}$$

where $n^* = n - p - \ell + 1$, $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\beta}, \gamma)$ is the error of model (3.3) with embedded parameters $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\beta}, \gamma)$, and $\boldsymbol{u}_t(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \boldsymbol{\beta}, \gamma)$ is the error of model (3.4). The joint maximum likelihood estimation of (3.3) and (3.4) is realized by maximizing the objective equation in (3.5).

Note that for threshold models defined above, we could first fix $(\boldsymbol{\beta}, \gamma)$ and get least square estimation for the linear coefficients in $(\hat{\boldsymbol{\theta}_1}, \hat{\boldsymbol{\theta}_2}, \hat{\boldsymbol{\rho}_1}, \hat{\boldsymbol{\rho}_2})$. Then the estimated covariance matrices can be easily obtained as

$$\hat{\Sigma}_1(\boldsymbol{\beta}, \gamma) = \frac{1}{n-p}\boldsymbol{\varepsilon}_t^{'}(\boldsymbol{\beta}, \gamma)\boldsymbol{\varepsilon}_t(\boldsymbol{\beta}, \gamma), \quad \hat{\Sigma}_2(\boldsymbol{\beta}, \gamma) = \frac{1}{n^*}\boldsymbol{u}_t^{'}(\boldsymbol{\beta}, \gamma)\boldsymbol{u}_t(\boldsymbol{\beta}, \gamma).$$

Thus the log-likelihood function in (3.5) reduces to

$$L(\boldsymbol{\beta}, \gamma) = -\frac{n-p}{2}\log|\hat{\Sigma}_1(\boldsymbol{\beta}, \gamma)| - \frac{n^*}{2}\log|\hat{\Sigma}_2(\boldsymbol{\beta}, \gamma)|. \tag{3.6}$$

Then we perform the following algorithm to get $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$,

1. Get initial value $\boldsymbol{\beta}_0$ from linear cointegration estimation, and derive its confidence region based on the discussion in Johansen (1991). Form a grid on $[\gamma_\ell, \gamma_u]$ such

that $P(\gamma_\ell \leq \gamma \leq \gamma_u)$ is constraint to a proper level. (Tsay (1998), Hansen and Seo (2002)).

2. For fixed $\boldsymbol{\beta}_0$, get $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \hat{\boldsymbol{\rho}}_1, \hat{\boldsymbol{\rho}}_2)$ and $\hat{\Sigma}_1(\boldsymbol{\beta}, \gamma), \hat{\Sigma}_2(\boldsymbol{\beta}, \gamma)$ for each grid value of $\gamma \in [\gamma_\ell, \gamma_u]$. Find $\hat{\gamma}$ that maximizes (3.6).

3. For fixed $\hat{\gamma}$ from step 2, perform nonlinear maximization on $\boldsymbol{\beta}$ within its confidence region and get optimal estimation $\hat{\boldsymbol{\beta}}$. Note that the initial value can be flexible but should be within the confidence region.

4. Fix $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$, record the corresponding estimation in $\left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \hat{\boldsymbol{\rho}}_1, \hat{\boldsymbol{\rho}}_2, \hat{\Sigma}_1, \hat{\Sigma}_2\right)$.

Then we get the estimation of all parameters for the joint MLE approach.

3.4 Real Example: Household Consumption Data

3.4.1 Model Estimation

In the preliminary study we had confirmed the threshold pattern in the error correction term, it is natural to check the threshold behavior of Model 3.1 on the same threshold variable as well. With the estimated cointegrating vector from linear model, we try the multivariate threshold test in Tsay (1998) to confirm the threshold effect of US real GDP growth rate to Model 3.1.

We first apply Hansen's approach on Model 3.3 to the household consumption data with US real GDP growth as the threshold variable. The modeling results are summarized in table 3.5. The proposed joint maximum likelihood estimation is then applied to the household consumption data as well. By the procedure in section 3.3 on the household consumption data, the updated cointegrating vector in threshold model is estimated as,

$$\hat{z}_t = \hat{\boldsymbol{\beta}}' \boldsymbol{y}_{t-1} = y_{1t} - 0.42 y_{2t} + 0.002 y_{3t} - 0.027 y_{4t}.$$

The updated threshold value is located as $\hat{\gamma} = 4.1$. As the linear combination of $\boldsymbol{y_t}$, $\hat{z}_t$

Table 3.5: TVECM estimation (Hansen's) with γ_{t-1} as threshold variable: lower regime when $\gamma_{t-1} \leq 4.1$ and higher regime when $\gamma_{t-1} > 4.1$.

Lower	Δy_{1t}	Δy_{2t}	Δy_{3t}	Δy_{4t}	Higher	Δy_{1t}	Δy_{2t}	Δy_{3t}	Δy_{4t}
$\boldsymbol{\mu}_t$	0.008	0.021	0.010	0.002	$\boldsymbol{\mu}_t$	0.005	0.022	-0.019	0.027
	(0.005)	(0.006)	(0.013)	(0.006)		(0.008)	(0.008)	(0.020)	(0.010)
ECT_{t-1}	-0.088	-0.062	-0.455	-0.076	ECT_{t-1}	0.004	0.049	0.139	0.081
	(0.026)	(0.028)	(0.066)	(0.031)		(0.064)	(0.059)	(0.151)	(0.075)
$\Delta y_{1,t-1}$	0.763	0.405	1.593	0.224	$\Delta y_{1,t-1}$	-0.236	-0.812	-1.181	0.263
	(0.194)	(0.204)	(0.489)	(0.227)		(0.245)	(0.226)	(0.578)	(0.287)
$\Delta y_{2,t-1}$	-0.217	0.279	-1.138	0.288	$\Delta y_{2,t-1}$	-0.032	0.250	0.874	-0.179
	(0.204)	(0.214)	(0.513)	(0.238)		(0.223)	(0.206)	(0.525)	(0.261)
$\Delta y_{3,t-1}$	0.003	-0.086	-0.389	-0.115	$\Delta y_{3,t-1}$	0.155	0.216	0.008	-0.022
	(0.053)	(0.056)	(0.134)	(0.062)		(0.113)	(0.104)	(0.266)	(0.132)
$\Delta y_{4,t-1}$	0.240	-0.018	1.239	0.228	$\Delta y_{4,t-1}$	0.263	0.212	0.530	-0.114
	(0.176)	(0.185)	(0.443)	(0.206)		(0.186)	(0.172)	(0.439)	(0.218)

is stationary by unit-root test. For the threshold cointegration model of $\hat{z}_t$, we have,

$$\hat{z}_t = \begin{cases} 0.042 + 1.590\hat{z}_{t-1} - 0.761\hat{z}_{t-2} + u_t, & \text{if } \gamma_{t-1} \leq 4.1, \\ 0.019 + 1.103\hat{z}_{t-1} - 0.179\hat{z}_{t-2} + u_t. & \text{if } \gamma_{t-1} > 4.1. \end{cases}$$

Evidences show that the threshold model on $\hat{z}_t$ is stationary as well. The roots of the characteristic function for the higher regime are 1.104 and 5.052, indicating the stationarity of the threshold model. The estimation of threshold vector error-correction model by joint MLE is summarized in table 3.6.

3.4.2 Model Comparison

In section 3.2 we discussed the cointegration effect in the multivariate compositional model of US Household consumption data with the linear vector error-correction model fitted. Then we identified the threshold effect on an exogenous variable and then apply the threshold vector error-correction model by Hansen's approach. Finally, we extended Hansen's approach to include the threshold effect of the error correction term on the same threshold variable. In this section, we compare the model performance of the linear VECM, the threshold VECM by Hansen's estimation and the threshold VECM by the joint MLE approach. The models are compared in two aspects. We first compare

Table 3.6: TVECM estimation from Joint MLE with γ_{t-1} (US real GDP growth rate) as threshold variable: lower regime when $\gamma_{t-1} \leq 4.1$ and higher regime when $\gamma_{t-1} > 4.1$.

Lower	Δy_{1t}	Δy_{2t}	Δy_{3t}	Δy_{4t}	Higher	Δy_{1t}	Δy_{2t}	Δy_{3t}	Δy_{4t}
$\boldsymbol{\mu}_t$	0.075	0.079	0.235	0.050	$\boldsymbol{\mu}_t$	0.003	-0.019	-0.074	-0.013
	(0.015)	(0.015)	(0.050)	(0.019)		(0.026)	(0.022)	(0.060)	(0.029)
ECT_{t-1}	-0.244	-0.212	-0.826	-0.172	ECT_{t-1}	0.010	0.180	0.225	0.170
	(0.050)	(0.052)	(0.167)	(0.065)		(0.111)	(0.095)	(0.262)	(0.128)
$\Delta y_{1,t-1}$	0.781	0.417	1.684	0.239	$\Delta y_{1,t-1}$	-0.232	-0.714	-1.176	0.303
	(0.173)	(0.180)	(0.577)	(0.225)		(0.249)	(0.213)	(0.588)	(0.286)
$\Delta y_{2,t-1}$	-0.326	0.162	-1.276	0.232	$\Delta y_{2,t-1}$	-0.038	0.115	0.776	-0.275
	(0.185)	(0.192)	(0.617)	(0.240)		(0.242)	(0.207)	(0.571)	(0.278)
$\Delta y_{3,t-1}$	-0.053	-0.141	-0.509	-0.148	$\Delta y_{3,t-1}$	0.156	0.217	0.030	-0.005
	(0.050)	(0.052)	(0.167)	(0.065)		(0.110)	(0.094)	(0.259)	(0.126)
$\Delta y_{4,t-1}$	0.253	0.004	1.156	0.225	$\Delta y_{4,t-1}$	0.264	0.224	0.593	-0.082
	(0.156)	(0.163)	(0.522)	(0.204)		(0.182)	(0.156)	(0.431)	(0.210)

the behavior of the estimated cointegrating vectors from the three models, and then compare the model adequacy of the fitted vector error-correction models.

The time series plots of the estimated error correction term are shown in Figure 3.10. The updated cointegrating vectors by threshold effect are more smooth than the linear estimation. The joint MLE estimation looks even more smooth than the other two plots. As discussed in section 4, all the three estimated cointegrating vector are free of unit-roots, and the univariate threshold models are stationary as well.

Next we compare the model adequacy of the vector error-correction models based on multiple criterion. The comparison is summarized in table 3.7. It suggests that the threshold vector error-correction model by the joint MLE estimation is the best fitting to the data. The model estimation results after significance selection are presented in table 3.8.

3.4.3 Prediction Performance

An important purpose of the threshold vector error-correction model is prediction. Based on the available household consumption data, we performed out-of-sample one-step rolling forecast for 10 years. We compare the forecasting performance of the linear

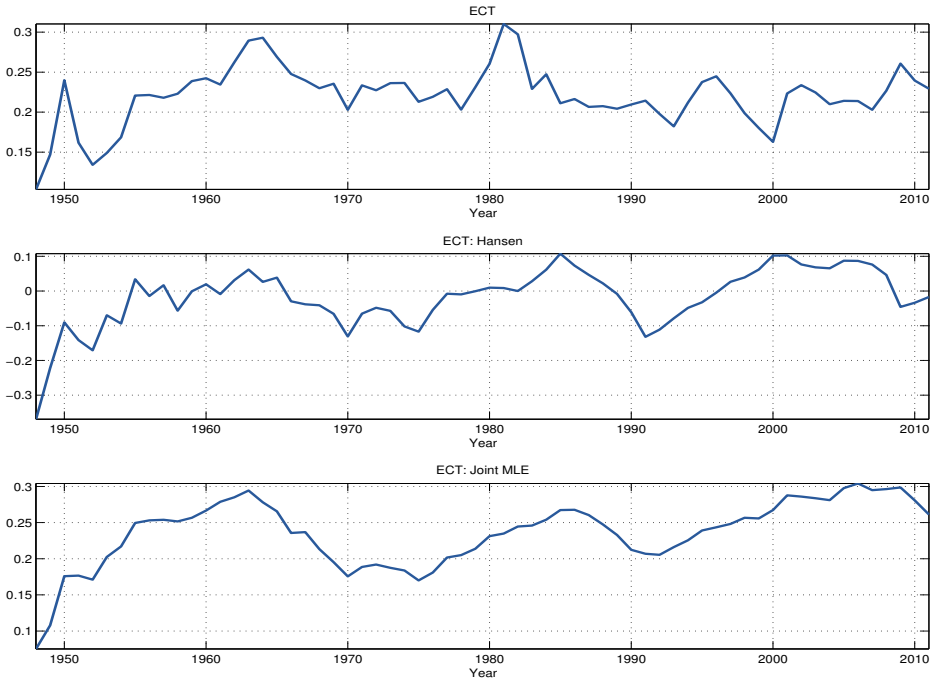


Figure 3.9: The time series plots of the estimated error correction terms.

Table 3.7: Comparison of the linear VECM, the TVECM by Hansen's and the TVECM by Joint MLE.

Model	Linear VECM	TVECM: Hansen	TVECM: Joint MLE
Unit-roots in ECT	No	No	No
Stationarity of $TAR(z_t)$	Yes	Yes	Yes
SSE of $TAR(z_t)$	0.028	0.089	0.008
SSE of $VECM$	0.206	0.109	0.121
Total SSE	0.234	0.298	0.129
Loglik of $TAR(z_t)$	151.92	115.09	187.82
Loglik of $VECM$	548.57	630.57	618.08
Total Loglik	700.49	745.66	805.90

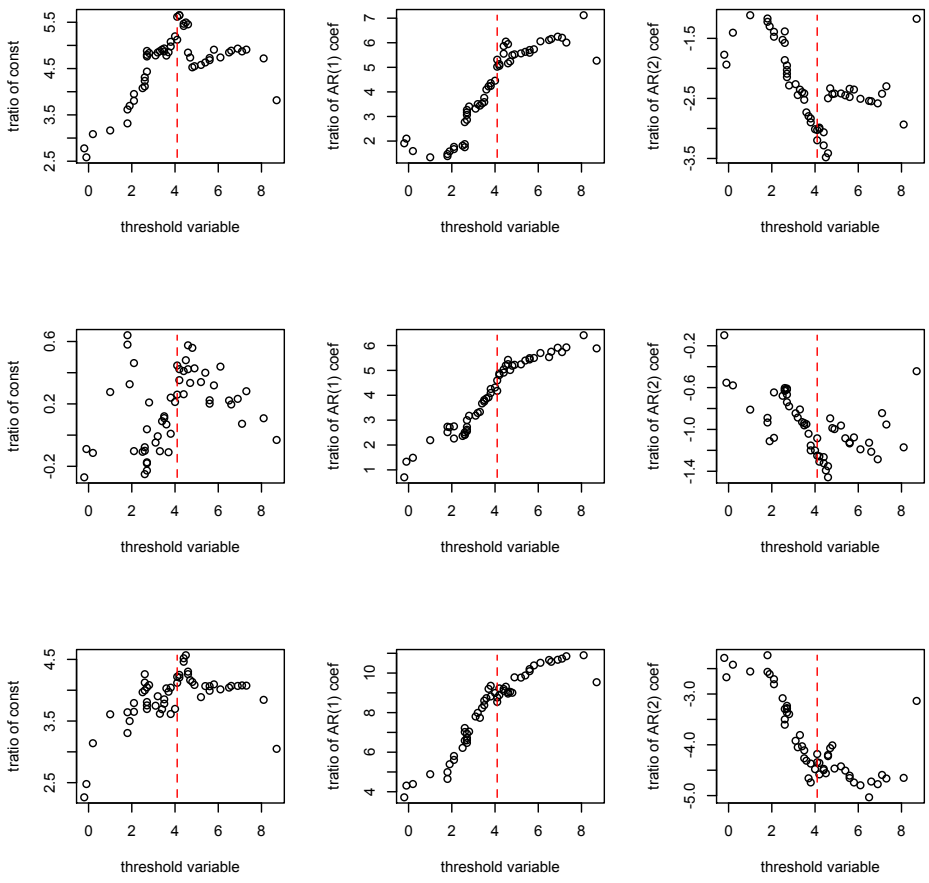


Figure 3.10: The t-ratio plots for the threshold AR(2) model of the estimated ECT.

Table 3.8: Reduced Threshold VECM estimation from joint maximum likelihood estimation with γ_{t-1} (US real GDP growth rate) as threshold variable: lower regime when $\gamma_{t-1} \leq 4.1$ and higher regime when $\gamma_{t-1} > 4.1$.

Lower	Δy_{1t}	Δy_{2t}	Δy_{3t}	Δy_{4t}	Higher	Δy_{1t}	Δy_{2t}	Δy_{3t}	Δy_{4t}
$\boldsymbol{\mu}_t$	0.063	0.079	0.222	0.055	$\boldsymbol{\mu}_t$	–	–	-0.025	–
	(0.013)	(0.012)	(0.045)	(0.015)		(–)	(–)	(0.018)	(–)
ECT_{t-1}	-0.208	-0.208	-0.813	-0.187	ECT_{t-1}	–	0.124	–	0.093
	(0.045)	(0.048)	(0.157)	(0.059)		(–)	(0.022)	(–)	(0.014)
$\Delta y_{1,t-1}$	0.728	0.529	1.727	0.361	$\Delta y_{1,t-1}$	–	-0.745	-1.316	–
	(0.163)	(0.145)	(0.565)	(0.185)		(–)	(0.188)	(0.395)	(–)
$\Delta y_{2,t-1}$	-0.273	–	-1.342	–	$\Delta y_{2,t-1}$	–	–	1.053	–
	(0.179)	(–)	(0.607)	(–)		(–)	(–)	(0.438)	(–)
$\Delta y_{3,t-1}$	–	-0.145	-0.519	-0.169	$\Delta y_{3,t-1}$	0.079	0.256	–	–
	(–)	(0.048)	(0.164)	(0.061)		(0.056)	(0.083)	(–)	(–)
$\Delta y_{4,t-1}$	0.210	–	1.218	0.328	$\Delta y_{4,t-1}$	0.290	0.212	0.604	–
	(0.149)	(–)	(0.513)	(0.177)		(0.108)	(0.152)	(0.411)	(–)

VECM, threshold VECM by Hansen's approach and threshold VECM by joint MLE approach. The performance criterion is the root of mean squared forecasting error that defined as

$$RMSE_j = \sqrt{\sum_{i=1}^{10} \left(y_{j,i}^{pred} - y_{j,i}^{true} \right)^2}, \quad j = 1, \ldots, k.$$

The rolling forecast starts from 2003 to 2012. The prediction behavior of the four logratio series are compared separately due to their different measurements scales. Table 3.9 summarizes the RMSE values from three different models respectively for $\Delta y_j, j = 1, 2, 3, 4$. From the values we concludes that the joint MLE approach has the best rolling forecasting performance for $\Delta y_j, j = 1, 2, 4$ and Hansen's method gives smaller RMSE for Δy_{3t}. Both the threshold models outperform the linear model, validating the nonlinearities in the vector error-correction model for household consumption data.

Figure 3.11 visually compares the true values of $\boldsymbol{y}_t$ and the predicted values by rolling forecast. In general, the prediction for Δy_{1t} and Δy_{2t} are more accurate than the other two. In fact, the model estimation results indicate that Δy_{1t} and Δy_{2t} are more significance than the other two random variables in the multivariate model. In other words, the categories corresponding to housing and utilities (y_{1t}), health, recreation and

Table 3.9: Comparison of RMSE by 1-step rolling forecast.

Models	Δy_1	Δy_2	Δy_3	Δy_4
Linear VECM	0.015	0.023	0.049	0.017
TVECM: Hansen	0.013	0.019	0.063	0.013
TVECM: Joint MLE	0.012	0.017	0.073	0.012

education (y_{2t}) play more important roles in modeling and forecasting the household consumption data.

So far we have been working in the compositional framework, the modeling process is built on the transformed data by additive logratio transformation. With the prediction of the logratio series, we derive the predicted numbers of the original percentages for each category in table 3.10.

3.5 Summary

In this chapter we aim to analyze the U.S. household consumption patterns between different consumption functions and how they change based on historical US real GDP change. The functions are categorized with similar functions and finally five major consumption category are formed for further discussion. With the fixed total amount of budget at each year, the percentages allocated to each category form a compositional time series. The transformed data have unit-roots pattern, though, we find cointegration effect between the vectors, and stationarity can be achieved by linear combination of the lograio vectors. The linear vector error-correction model is fitted to the data as an initial exploration.

Then we are more interested in the household consumption pattern change with exogenous economic factors. We find that the U.S. real GDP growth rate affects the consumption pattern through a threshold model. Tsay's threshold test confirms the phenomenon. With annually data, the level of real GDP change from the previous year determines the transition pattern of household budget allocation in the current year. To incorporate the threshold effect, we fit a threshold vector error-correction model based on Hansen's method. Furthermore, we believe that Hansen's method ignores the

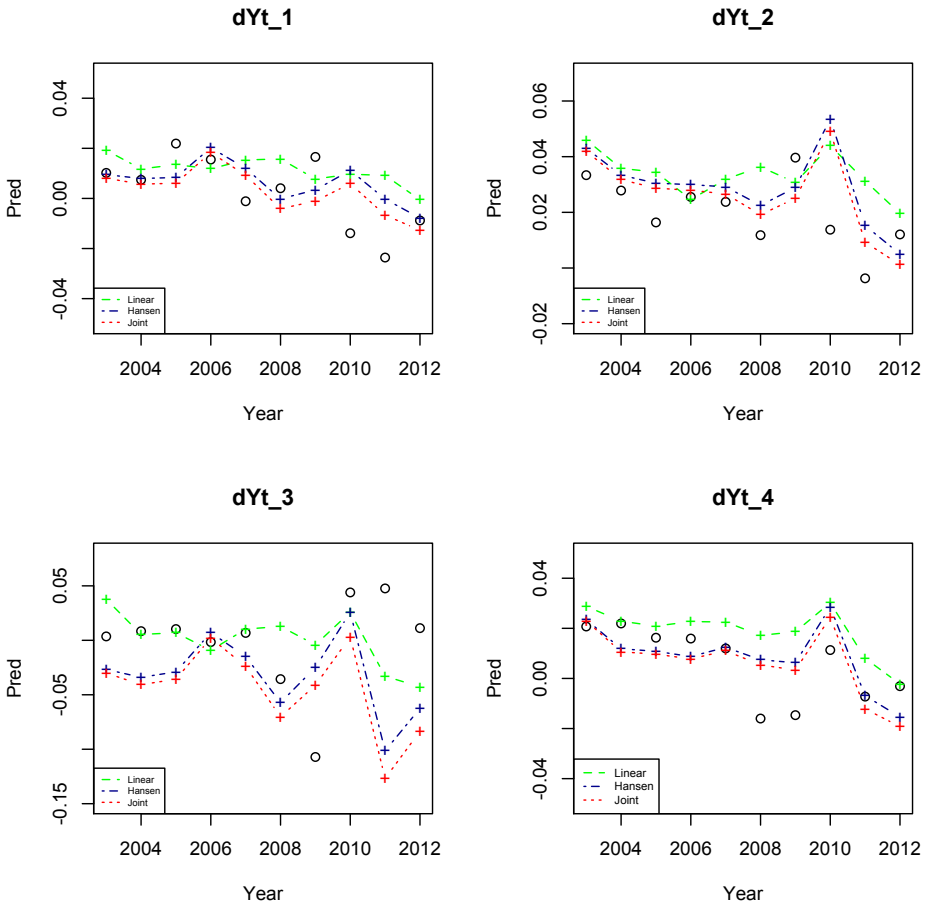


Figure 3.11: Comparison of rolling forecast performance from 2003 to 2012.

Table 3.10: Prediction in percentages by three models

Model	Functions	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
True	Food	12.11	11.92	11.75	11.59	11.44	11.34	11.40	11.41	11.33	11.38
	Housing	23.84	23.71	23.54	23.74	23.80	23.56	23.78	24.20	23.70	23.24
	Health	29.72	30.10	30.35	30.29	30.52	30.82	31.18	32.32	32.37	32.22
	Commu	13.66	13.50	13.42	13.38	13.18	13.15	12.76	11.47	11.90	12.54
	Service	20.67	20.78	20.93	20.99	21.05	21.12	20.88	20.60	20.69	20.63
VECM	Food	11.94	11.59	11.56	11.33	11.51	11.10	11.34	11.16	11.34	11.15
	Housing	23.96	24.13	24.05	23.88	24.22	23.91	23.78	24.22	24.04	23.57
	Health	29.50	29.87	30.22	30.33	30.42	30.87	31.49	32.64	32.74	32.29
	Commu	13.69	13.61	13.37	13.30	13.36	13.28	12.63	11.43	12.03	12.83
	Service	20.91	20.81	20.79	21.16	20.49	20.84	20.76	20.55	19.84	20.16
TVECM (I)	Food	12.12	11.66	11.84	11.38	11.63	11.20	11.50	11.25	11.49	11.21
	Housing	24.09	24.20	23.10	23.85	24.38	24.05	23.19	24.14	24.24	23.61
	Health	28.90	29.75	31.21	30.63	31.23	30.69	31.85	32.81	32.35	32.15
	Commu	13.67	13.50	12.77	13.23	13.44	13.37	12.56	11.38	12.21	13.03
	Service	21.23	20.90	21.09	20.92	19.33	20.70	20.91	20.42	19.71	20.01
TVECM (II)	Food	12.21	11.68	11.89	11.40	11.72	11.22	11.56	11.28	11.59	11.23
	Housing	24.23	24.22	23.12	23.89	24.51	24.06	23.16	24.17	24.38	23.62
	Health	28.92	29.77	31.28	30.64	31.31	30.73	31.93	32.80	32.32	32.17
	Commu	13.72	13.48	12.66	13.23	13.48	13.34	12.42	11.38	12.25	13.00
	Service	20.91	20.85	21.04	20.84	18.99	20.64	20.92	20.38	19.46	19.98

threshold structure in the error correction term. Information will be lost for situations that are more than two dimensions. We thus extend the method by joint maximum likelihood estimation on both threshold structures. The estimation results by the joint MLE approach provides smaller SSE and larger log-likelihood values than the other two models. In addition, the joint MLE approach gives better rolling forecast performance on the household consumption data as expected.

Chapter 4

Sensor Allocation under Separable and Nonseparable Models

4.1 Background

4.1.1 Spatial-Temporal Models

Spatial-temporal models are extensively discussed in the past decades. It was proposed to analyze the geographical patterns that usually have both spatial and temporal components. The applications of spatial-temporal models in environmental sciences are quite popular and conventional. The history of spatial-temporal analysis dates back to 1980s when Bilonick (1985) started to build the model for the acid rainfall data in New York. Recent examples can be found in Huerta et al. (2004) who applies the model for analyzing the ozone levels in Mexico city, and in Reich et al. (2011) who developed a covariate-dependent spatial-temporal model for the daily ozone levels in Southeast United States. Another typical example is the Irish wind data that was initially introduced by Haslett and Raftery (1989) and then being referred frequently in Gneiting (2002), Stein (2005) and De Luna and Genton (2005) for demonstration purposes.

One interesting application of the spatial-temporal model is the monitoring sensor network where data are collected at regular time and fixed sensor locations. The data possesses spatial dependence among the monitors as well as the temporal correlations. Until recently, researchers in diverse areas such as climatology, ecology, catastrophe insurance institutes and real estate marketing are raising increasing interests in this topic.

In the same time, there is a rapid development in sensor allocation problems. Sensor network has a wide range of applications including environmental monitoring, industrial

control and security problems. The goal is to efficiently allocate the sensors in a random field and in the same time best utilize the limit resources. People are interested in allocating the sensors one by one instead of allocating them simultaneously. Such examples exist in the environmental sciences where people construct sensor stations to measure the pollution level, wind speed or to manage the wild life and agriculture. There are several well-developed approaches for the sensor allocation problem that concentrates on different aspects of the problem. In the computer experiment area, the eminent work by Sacks et al. (1989) and Williams (2000) propose a sequential design by optimizing the expected improvement from a bayesian view. This area has been nourished greatly by recent work such as Hung et al. (2009) and Hung et al. (2010). The sensor allocation problem also worths attention from a machine learning approach, for example, Krause et al. (2008) and Das and Kempe (2008). Compared with these approaches that concentrate on higher dimension with pure spatial correlations, the spatial-temporal models provide a different perspective of view to the problem and will be a main focus in this chapter.

4.1.2 Sensor Allocation Problem

A lot of work has been done in proposing different forms of constraints to the sensor allocation problem. In our case, prediction is a primary topic hence a natural objective is to optimize the prediction performance of all sensors including the existing ones. Take the environmental monitoring for instance, a set of sensor stations are built over the region to measure the temperatures. They are not designed for temperature forecasting of their own locations only, but the temperatures at any other locations need to be efficiently predicted by nearby stations. In Gaussian random field, the predicted value has a conditional normal distribution and the correlation structure is well defined and convenient to model with. Moreover, since the observations are also recorded by time, time correlation plays a substantial role in forecasting as well. The criterion for new sensor allocation is naturally introduced by minimizing the overall prediction standard error, which is the integrated prediction variance.

To make the problem well-formulated, we first define a two-dimensional spatial region $\Omega \subset \mathbb{R}^2$, and the random process $Y(\boldsymbol{s};t)$ is generated over the region $(\boldsymbol{s};t) \in \Omega \times \mathbb{R}^+$. Suppose we have n different sensor locations in the region Ω, time series observations are recorded for each station. Both spatial correlations with respect to $\boldsymbol{s}$ and temporal correlation on t are considered in forecasting. The objective for allocating $\boldsymbol{s}_{new}$ is to minimize the expected overall prediction error in Ω, and we denote $\boldsymbol{s}_{new}$ by $\boldsymbol{s}_{k+1}$ for simplicity in the following discussions. Conditional on observations $Y(\boldsymbol{s}_1;t_1),\ldots,Y(\boldsymbol{s}_k;t_n)$ and the intended new sensor located at $\boldsymbol{s}_{new}$, the predicted value for any other locations $\boldsymbol{s} \in \Omega$ at the current time t is defined as

$$\hat{Y}(\boldsymbol{s};t) = \mathrm{E}\left(Y(\boldsymbol{s};t) \mid \vec{Y}(\boldsymbol{s}_1;t), \vec{Y}(\boldsymbol{s}_2;t), \ldots, \vec{Y}(\boldsymbol{s}_k;t); \vec{Y}(\boldsymbol{s}_{new};t)\right).$$

where $\vec{Y}(\boldsymbol{s}_j;t) = (Y(\boldsymbol{s}_j;t), Y(\boldsymbol{s}_j;t-2), \ldots, Y(\boldsymbol{s}_k;t-n)), j = 1,\ldots,k$. The prediction is based on the given k stations and the new station $\boldsymbol{s}_{new}$ that is to be located. Comparing with the predicted values, we are more interested in the expected prediction variance for sensor allocation purpose. For a fixed location $\boldsymbol{s}$ and time t, the expected prediction variance of $\hat{Y}(\boldsymbol{s};t)$ is defined as,

$$\mathrm{E}(\hat{Y}(\boldsymbol{s};t) - Y(\boldsymbol{s};t))^2.$$

As $\boldsymbol{s}$ varies in Ω, the integrated conditional variance over the entire region Ω should be minimized to efficiently allocating $\boldsymbol{s}_{new}$. Denote $\boldsymbol{s}_{new}$ as $\boldsymbol{s}_{k+1}$ for simplicity in the following discussions. The objective function is,

$$f\left(\boldsymbol{s}_{k+1};\boldsymbol{s}_1,\ldots,\boldsymbol{s}_k\right) = \int_\Omega \mathrm{E}(\hat{Y}(\boldsymbol{s};t) - Y(\boldsymbol{s};t))^2 d\boldsymbol{s}.$$

Specifically, it is,

$$\begin{aligned} & f\left(\boldsymbol{s}_{k+1};\boldsymbol{s}_1,\ldots,\boldsymbol{s}_k\right) \\ = & \int_\Omega \mathrm{Var}\left(Y(\boldsymbol{s};t) \mid \vec{Y}(\boldsymbol{s}_1;t), \vec{Y}(\boldsymbol{s}_2;t), \ldots, \vec{Y}(\boldsymbol{s}_k;t); \vec{Y}(\boldsymbol{s}_{k+1};t)\right) d\boldsymbol{s}, \end{aligned} \tag{4.1}$$

where $\vec{Y}(\boldsymbol{s}_j;t)$ are observed for $j = 1,2,\ldots,k$ and $t = 0,1,\ldots,n$. In the objective function, $Y(\boldsymbol{s}_{k+1};t)$ is unknown and $\boldsymbol{s}_{k+1}$ is the new sensor location that is to be determined by minimizing equation (4.1) on $\boldsymbol{s}_{k+1}$. Notice that the above objective

function is built under the goal of concurrent prediction; if the purpose concentrates more on future time predictions, the objective function can be generalized as,

$$\begin{aligned} & f_m\left(\boldsymbol{s}_{k+1}; \boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\right) \\ = & \int_\Omega \operatorname{Var}\left(Y(\boldsymbol{s}; t+m) \mid \vec{Y}(\boldsymbol{s}_1; t), \vec{Y}(\boldsymbol{s}_2; t), \ldots, \vec{Y}(\boldsymbol{s}_k; t); \vec{Y}(\boldsymbol{s}_{k+1}; t)\right) d\boldsymbol{s}, \end{aligned} \tag{4.2}$$

if the prediction horizon is $m \in \mathbb{R}^+$. Both situations will be discussed in this chapter.

This chapter is organized as follows. Section 4.1 introduces the motivation and the objective function for the sensor allocation problem. Section 4.2 discusses the objective function under separable spatial-temporal models. The objective function based on future time prediction is introduced as well. Section 4.3 introduces the full covariance structure under nonseparable models, which usually involves a spatial-temporal interaction parameter. A positive-definite nonseparable covariance family is discussed in details. We compare the difference between separable and nonseparable assumptions on the sensor allocation problem by several empirical examples in section 4.4. Finally, an real example of the ozone level measurement in Texas is extensively analyzed in section 4.5.

4.2 Separable Spatial-Temporal Models

In spatial-temporal models, we are given observations at different locations over a regular time period so that we have

$$\{Y(\boldsymbol{s}; t) : s \in \mathbb{R}^2; t \in \mathbb{R}^+\},$$

where t takes value in a regularly lattice, s is located irregularly in a region $\Omega \subset \mathbb{R}^2$. Given a stationary process $Y(\boldsymbol{s}; t)$, separability means that the spatial-temporal covariance can be the multiplication of the two. Specifically, it requires the assumptions that $Y(\boldsymbol{s}; t)$ needs to be weak stationary $w.r.t$ both space $\boldsymbol{s}$ and time t, and on top of this,

$$\operatorname{Cov}(Y(\boldsymbol{s}, t), Y(\boldsymbol{s}+\boldsymbol{h}, t+u)) = C_s(\boldsymbol{h}) C_t(u), \quad \boldsymbol{h} \in \mathbb{R}^d, u \in \mathbb{R}.$$

Separability is such a favored property that it is always assumed in real applications. It brings convenience to the modeling process from many aspects. In addition, We found the conditional independence property for separable spatial-temporal models (4.A.1).

Equation (4.17) gives the conclusion that, given all available information at location $\boldsymbol{s}_j$, the current observation at $\boldsymbol{s}_i$ is uncorrelated with the past observations at location $\boldsymbol{s}_j$ for $i \neq j$. This directly implies conditional independence when Gaussian. Therefore, the objective function in (4.1) can be simplified as,

$$\begin{aligned} & f(\boldsymbol{s}_{k+1}; \boldsymbol{s}_1, \ldots, \boldsymbol{s}_k) \\ = & \int_\Omega \mathrm{Var}(Y(\boldsymbol{s}; t) \mid Y(\boldsymbol{s}_1; t), \ldots, Y(\boldsymbol{s}_k; t); Y(\boldsymbol{s}_{k+1}; t)) d\boldsymbol{s}. \end{aligned}$$

4.2.1 Allocation for Concurrent Prediction

If the objective is concurrent prediction, the cost function above can be further simplified by ignoring the temporal correlations, and the new sensor can be located by a pure spatial model. With a pure spatial model, the cost function is formed as,

$$f(\boldsymbol{s}_{k+1}; s_1, \ldots, \boldsymbol{s}_k) = \int_\Omega \mathrm{Var}(Y(\boldsymbol{s}) \mid Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_k), Y(\boldsymbol{s}_{k+1})) d\boldsymbol{s}. \tag{4.3}$$

Given the simple structure in spatial only models, we can try to get the explicit form for $f(\boldsymbol{s}_{k+1}; \boldsymbol{s}_1, \ldots, \boldsymbol{s}_k)$ to simplify the optimization procedure and possibly derive an analytical solution. From the multivariate normal distribution in Gaussian, the conditional distribution of $Y(\boldsymbol{s})$ given $Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_k)$ and $Y(\boldsymbol{s}_{k+1})$ is,

$$Y(\boldsymbol{s}) \mid (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_k), Y(\boldsymbol{s}_{k+1})) \sim N(\mu, \boldsymbol{\sigma}^2),$$

where

$$\begin{aligned} \boldsymbol{\sigma}^2 = & C(\boldsymbol{s}, \boldsymbol{s}) - (C(\boldsymbol{s}, \boldsymbol{s}_1), C(\boldsymbol{s}, \boldsymbol{s}_2), \ldots, C(\boldsymbol{s}, \boldsymbol{s}_k) C(\boldsymbol{s}, \boldsymbol{s}_{k+1})) \\ & \times \begin{pmatrix} C(\boldsymbol{s}_1, s_1) & C(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots & C(\boldsymbol{s}_1, \boldsymbol{s}_{k+1}) \\ C(\boldsymbol{s}_2, s_1) & C(\boldsymbol{s}_2, \boldsymbol{s}_2) & \cdots & C(\boldsymbol{s}_2, \boldsymbol{s}_{k+1}) \\ \vdots & \vdots & & \vdots \\ C(\boldsymbol{s}_{k+1}, \boldsymbol{s}_1) & C(\boldsymbol{s}_{k+1}, s_2) & \cdots & C(\boldsymbol{s}_{k+1}, \boldsymbol{s}_{k+1}) \end{pmatrix}^{-1} \begin{pmatrix} C(\boldsymbol{s}, \boldsymbol{s}_1) \\ C(\boldsymbol{s}, \boldsymbol{s}_2) \\ \vdots \\ C(\boldsymbol{s}, \boldsymbol{s}_k) \\ C(\boldsymbol{s}, \boldsymbol{s}_{k+1}) \end{pmatrix}, \end{aligned} \tag{4.4}$$

and $C(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is the spatial covariance function imposed in the random field. For example, the exponential distance correlation $\exp(-\|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$ is frequently used.

Remark When $\Omega \in \mathbb{R}$, the correlation is $\rho(s_i, s_j) = \exp(-|s_i - s_j|)$. We have shown through calculations that in one-dimensional case,

1. If the new sensor is to be located between existing ones, it needs to be located in the middle of the largest interval by given sensors.

2. If the new sensor is located on an interval that no sensor is given at one of the endpoints, it will be closer to the endpoint without given sensors.

The conclusions in 1-D are practical that they provide an inspiration of the effectiveness of model (4.1) for more complex discussions later.

4.2.2 Allocation for Future Prediction

Suppose now we are more concerning about the prediction at time $t+m$ instead of t, the new sensor will be located differently since the model needs to take into account of the time correlation between $Y(\boldsymbol{s}, t+m)$ and the current observations at time t, thus it cannot be reduced to a spatial-only model as in section 4.2.1. However, the Markov property still exists in this situation. The joint covariance structure will be changed, and hence the conditional variance and the cost function to be considered.

The conditional distribution of $Y(\boldsymbol{s}; t+m)$ is given as,

$$Y(\boldsymbol{s}; t+m) \mid (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_k), Y(\boldsymbol{s}_{k+1})) \sim N(\mu, \boldsymbol{\sigma}^2),$$

and the corresponding covariance matrix is,

$$\begin{aligned}\boldsymbol{\sigma}^2 =& C(\boldsymbol{s}, \boldsymbol{s}; m) - (C(\boldsymbol{s}, \boldsymbol{s}_1; m), C(\boldsymbol{s}, \boldsymbol{s}_2; m), \ldots, C(\boldsymbol{s}, \boldsymbol{s}_k; m) C(\boldsymbol{s}, \boldsymbol{s}_{k+1}; m)) \\ &\times \begin{pmatrix} C(\boldsymbol{s}_1, \boldsymbol{s}_1; m) & C(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots & C(\boldsymbol{s}_1, \boldsymbol{s}_{k+1}; m) \\ C(\boldsymbol{s}_2, s_1; m) & C(\boldsymbol{s}_2, \boldsymbol{s}_2; m) & \cdots & C(\boldsymbol{s}_2, \boldsymbol{s}_{k+1}; m) \\ \vdots & \vdots & & \vdots \\ C(\boldsymbol{s}_{k+1}, \boldsymbol{s}_1; m) & C(\boldsymbol{s}_{k+1}, s_2; m) & \cdots & C(\boldsymbol{s}_{k+1}, \boldsymbol{s}_{k+1}; m) \end{pmatrix}^{-1} \begin{pmatrix} C(\boldsymbol{s}, \boldsymbol{s}_1; m) \\ C(\boldsymbol{s}, \boldsymbol{s}_2; m) \\ \vdots \\ C(\boldsymbol{s}, \boldsymbol{s}_k; m) \\ C(\boldsymbol{s}, \boldsymbol{s}_{k+1}; m) \end{pmatrix}, \end{aligned} \tag{4.5}$$

and $C(\boldsymbol{s}_i, \boldsymbol{s}_j; m)$ is the spatial-temporal covariance function. More details about $C(\boldsymbol{s}_i, \boldsymbol{s}_j; m)$ will be discussed in the subsequent sections.

4.3 Nonseparable Spatial-Temporal Models

Separable spatial-temporal models are commonly assumed in applications; however, many real examples present obvious nonseparable patterns. Nonseparable spatial-temporal models are highly demanded to better fit the empirical covariance structure. The conclusions in the previous section states that under separable covariance structures, the problem of minimizing the integrated conditional variance reduces to a pure spatial case. If the assumption of separability fails, the cost function has have the general form as in (4.1) for concurrent time prediction and in (4.2) for future time prediction purposes. They have different covariance structures, though, both structures need to take the full spatial-temporal model into consideration, and the former can be treated as a special case of the latter by setting $m = 0$. Let's revisit the general form of the objective function in(4.2),

$$\begin{aligned} & f\left(\boldsymbol{s}_{k+1}; \boldsymbol{s}_1, \ldots, \boldsymbol{s}_k, m\right) \\ = & \int_{\Omega} \operatorname{Var}\left(Y(\boldsymbol{s}; t+m) \mid \vec{Y}(\boldsymbol{s}_1; t), \vec{Y}(\boldsymbol{s}_2; t), \ldots, \vec{Y}(\boldsymbol{s}_k; t); \vec{Y}(\boldsymbol{s}_{k+1}; t)\right) d\boldsymbol{s}, \end{aligned}$$

where $\vec{Y}(\boldsymbol{s}_j; t) = \left(Y(\boldsymbol{s}_j; t), Y(\boldsymbol{s}_j; t-2), \ldots, Y(\boldsymbol{s}_k; t-n)\right), j = 1, \ldots, k.$

In this general equation, the time series observations are measured for the whole time horizon from 1 to n; however, the time lag can be restricted to u under proper constraints on time series correlations. Therefore, the objective function is a function of locations $\boldsymbol{s}_j$, time lag u and forecasting horizon m,

$$\begin{aligned} & f\left(\boldsymbol{s}_{k+1}; \boldsymbol{s}_1, \ldots, \boldsymbol{s}_k, m, u\right) \\ = & \int_{\Omega} \operatorname{Var}\left(Y(\boldsymbol{s}; t+m) \mid \vec{Y}(\boldsymbol{s}_1; t), \vec{Y}(\boldsymbol{s}_2; t), \ldots, \vec{Y}(\boldsymbol{s}_k; t); \vec{Y}(\boldsymbol{s}_{k+1}; t)\right) d\boldsymbol{s}, \end{aligned} \tag{4.6}$$

and $\vec{Y}(\boldsymbol{s}_j; t) = \left(Y(\boldsymbol{s}_j; t), Y(\boldsymbol{s}_j; t-2), \ldots, Y(\boldsymbol{s}_k; t-u)\right), j = 1, \ldots, k.$ Note that

$$\operatorname{Var}\left(Y(\boldsymbol{s}; t+m) \mid \vec{Y}(\boldsymbol{s}_1; t), \vec{Y}(\boldsymbol{s}_2; t), \ldots, \vec{Y}(\boldsymbol{s}_k; t); \vec{Y}(\boldsymbol{s}_{k+1}; t)\right) = 1 - \boldsymbol{b}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}^T;$$

where

$$\begin{aligned} \boldsymbol{b} = & \big(C(\boldsymbol{s}, \boldsymbol{s}_1, m), C(\boldsymbol{s}, \boldsymbol{s}_1, m+1), \ldots, C(\boldsymbol{s}, \boldsymbol{s}_1, m+u); \ldots; \\ & C(\boldsymbol{s}, \boldsymbol{s}_{k+1}, m), \ldots, C(\boldsymbol{s}, \boldsymbol{s}_{k+1}, m+u)\big), \end{aligned}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} [\boldsymbol{A}(\boldsymbol{s}_i, \boldsymbol{s}_j)]_{i,j=1}^{k} & [\boldsymbol{A}(\boldsymbol{s}_i, \boldsymbol{s}_{k+1})]_{i=1}^{k} \\ [\boldsymbol{A}(\boldsymbol{s}_i, \boldsymbol{s}_{k+1})]_{i=1}^{k'} & A(\boldsymbol{s}_{k+1}, \boldsymbol{s}_{k+1}) \end{pmatrix}, \tag{4.7}$$

and

$$\boldsymbol{A}(\boldsymbol{s}_i, \boldsymbol{s}_j) = \begin{pmatrix} C(\boldsymbol{s}_i, \boldsymbol{s}_j; m) & C(\boldsymbol{s}_i, \boldsymbol{s}_j; m+1) & \cdots & C(\boldsymbol{s}_i, \boldsymbol{s}_j; m+u) \\ C(\boldsymbol{s}_i, \boldsymbol{s}_j; m+1) & C(\boldsymbol{s}_i, \boldsymbol{s}_j; m+2) & \cdots & C(\boldsymbol{s}_i, \boldsymbol{s}_j; m+u-1) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ C(\boldsymbol{s}_i, \boldsymbol{s}_j; m+u) & C(\boldsymbol{s}_i, \boldsymbol{s}_j; m+u-1) & \cdots & C(\boldsymbol{s}_i, \boldsymbol{s}_j; m) \end{pmatrix}$$

where $C(\boldsymbol{s}_i, \boldsymbol{s}_j; m+u)$ is the stationary covariance function of $Y(\boldsymbol{s}_i; t)$ and $Y(\boldsymbol{s}_j; t - m - u)$.

4.3.1 Positive-Definite Covariance Functions

It worth notice that the choice of covariance structure $C(\boldsymbol{h}_{ij}, t)$ is not arbitrary, it needs to be a permissible function that guarantees the positive-definiteness of $\boldsymbol{A}(s_i, s_j)$ and $\boldsymbol{\Sigma}(s_i, s_j)$. It is always hard to verify one particular covariance function is valid; however, a more conventional approach is to apply the well-developed classes of positive-definite covariance functions.

There has been a lot of research in nonseparable spatial-temporal covariance models. Evidently, Cressie and Huang (1999) provided an important family of positive-definite space-time covariance functions by spectral decomposition approach. Gneiting (2002) proposed a larger class of such models. He gave several examples of completely monotone function $\phi(t)$ and a list of positive functions $\psi(t)$ with completely monotone derivative, both for $t \geq 0$. Moreover, $\phi(t)$ and $\psi(t)$ can be separately modeled as the spatial and temporal covariance structures, which provides great convenience in practice. Then a valid spatial-temporal covariance function is given as

$$C(\boldsymbol{h}; u) = \frac{\sigma^2}{\phi(|u|^2)^{d/2}} \psi\left(\frac{\|\boldsymbol{h}\|^2}{\phi(|u|)^2}\right), (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

In particular, among the lists of $\phi(t)$ and $\psi(t)$ in table 1 and table 2 in Gneiting (2002), if take $\psi(t) = \exp(-ct^\gamma)$ and $\phi(t) = (at^\alpha + 1)^\beta$, a valid space-time covariance family

will be specified as

$$C(\boldsymbol{h}; u|\beta) = \frac{\sigma^2}{(a|u|^{2\alpha}+1)^{d/2}} \exp\left(-\frac{c\|\boldsymbol{h}\|^{2\gamma}}{(a|u|^{2\alpha}+1)^{\beta\gamma}}\right), (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{4.8}$$

where σ^2 is the variance of the random process, $a, c > 0$ and the smoothness parameters α and γ are both restricted on $(0, 1]$.

Gneiting's model is structured by two parts. The first part preserves the temporal correlation and the second part describes the spatial correlation. The interaction between spatial and temporal correlations are modeled by β, which is confined between 0 to 1. A larger β indicates a stronger spatial-temporal interaction. The extreme case when $\beta = 0$ provides a separable spatial-temporal model with a lot of appealing properties, for which reason, the separability test is always necessary in modeling.

4.3.2 Numerical Form for Optimization

As can be seen, the objective function in (4.6) is highly nonlinear. It involves the inverse of a complicate matrix (4.7). The matrix has an unknown variable $\boldsymbol{s}_{k+1}$, making it impossible to be decomposed for numerical convenience under nonseparable case. In addition, each entry of the matrix, $C(\boldsymbol{h}_{ij}, t)$, is highly nonlinear itself. Furthermore, we need to take integration of the nonlinear term over the region $\Omega \in \mathbb{R}^2$. All the facts indicate that there is no explicit form or fast calculation methods for the objective function, hence numerical approximation is a practical method for evaluating equation (4.6).

Let's consider equation (4.7), the sub covariance matrix $[\boldsymbol{A}(\boldsymbol{s}_i, \boldsymbol{s}_j)]_{i,j=1}^{k}$ is completely known given $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k$ and

$$[\boldsymbol{A}(\boldsymbol{s}_i, \boldsymbol{s}_{k+1})]_{i=1}^{k'} = \big(C(\boldsymbol{s}_1, \boldsymbol{s}_{k+1}), C(\boldsymbol{s}_2, \boldsymbol{s}_{k+1}), \cdots, C(\boldsymbol{s}_k, \boldsymbol{s}_{k+1})\big),$$

with the unknown variable $\boldsymbol{s}_{k+1}$ to be located. When $\Omega \in \mathbb{R}^2$, the integrand in (4.6) is a function of unknown location $\boldsymbol{s}_{k+1} = (s_1, s_2)$ with all other parameters known as either pre-fixed or pre-estimated. Therefore, the integrand in (4.6) can be denoted as,

$$g\left(\boldsymbol{s} \mid \boldsymbol{s}_{k+1}\right) = \mathrm{Var}\left(Y(\boldsymbol{s}; t+m) \mid \vec{Y}(\boldsymbol{s}_1; t), \vec{Y}(\boldsymbol{s}_2; t), \ldots, \vec{Y}(\boldsymbol{s}_k; t); \vec{Y}(\boldsymbol{s}_{k+1}; t)\right).$$

For evaluating the integral, we can have some regularized grids on Ω to get a numerical approximation. Take $\mathbb{R}^2$ for example, assume a rectangle region,

$$\Omega = [0, a] \times [0, b],$$

for some $a, b \in \mathbb{R}$. We write

$$[0, a] = \bigcup_{i=0}^{N-1} [a_i, a_{i+1}], \text{ with } a_i = \frac{ia}{N}, i = 0, \ldots, N,$$

and

$$[0, b] = \bigcup_{j=0}^{M-1} [b_j, b_{j+1}], \text{ with } b_j = \frac{jb}{N}, j = 0, \ldots, M.$$

Then we can write

$$\Omega = \bigcup_{\substack{i=0,\ldots,N-1 \\ j=0,\ldots,M-1}} [a_i, a_{i+1}] \times [b_j, b_{j+1}].$$

By taking this partition of the domain Ω, for any integrable continuous function $\varphi(x)$, we take the "central point approximation". That is,

$$\int_\Omega \varphi(x)dx \approx \frac{ab}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} \varphi\left(\frac{(2n-1)a}{2N}, \frac{(2m-1)b}{2M}\right),$$

where $\left(\frac{(2n-1)a}{2N}, \frac{(2m-1)b}{2M}\right)$ is the center of the rectangle $\left[\frac{(n-1)a}{N}, \frac{na}{N}\right] \times \left[\frac{(m-1)b}{M}, \frac{mb}{M}\right]$. Therefore, equation (4.6) becomes

$$f\left(\boldsymbol{s}_{k+1}\right) \approx \frac{ab}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} g\left(\frac{(2n-1)a}{2N}, \frac{(2m-1)b}{2M} \mid \boldsymbol{s}_{k+1}\right), \tag{4.9}$$

and this is the numerical form that is to be implemented in the subsequent examples.

4.4 Empirical Examples

Two empirical examples are shown for demonstration purposes. In particular, the first example is presented to show the effectiveness of the objective function we proposed. The second example is repeatedly attempted in order to show the difference in sensor allocation under different assumptions. Given k different nodes $\boldsymbol{s}_i \in \mathbb{R}^2, i = 1, \ldots, k$, the new location s_{k+1} is derived by minimizing (4.9) using grid search over the region. The grid lattice for numerical integration and the lattice for grid search are imposed

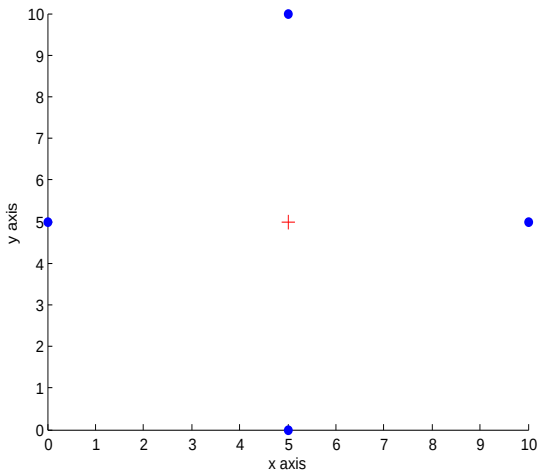


Figure 4.1: New sensor given four symmetric sensors.

differently on purpose. In the figures, the blue dots represent the given sensor locations and the new location s_{k+1} is labeled by red plus signs.

Example 1

The first example is given in the spatial-only case when no time correlation is preserved. The exponential correlation is imposed. In $\mathbb{R}^2$, assume $\{\Omega = [0, 10] \times [0, 10]\}$ and four sensors are symmetrically located around $(5, 5)$: $\{\boldsymbol{s}_1 = (0, 5), \boldsymbol{s}_2 = (10, 5), \boldsymbol{s}_3 = (5, 10), \boldsymbol{s}_4 = (5, 0)\}$. The new sensor is then ideally located at $s_{k+1} = (5, 5)$, and the corresponding integrated conditional variance is $f(\boldsymbol{s}_{k+1}; s_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \boldsymbol{s}_4) = 95.0046$. See Figure 4.1.

Example 2

Given another region $\{\Omega = [0, 40] \times [0, 60]\}$ and five locations randomly distributed as $\{\boldsymbol{s}_1 = (10, 20); \boldsymbol{s}_2 = (30, 19); \boldsymbol{s}_3 = (28, 25); \boldsymbol{s}_4 = (25, 10); \boldsymbol{s}_5 = (40, 40)\}$. We assume a spatial-temporal covariance structure in (4.8) for the random field we considered. Specifically, we consider the following spatial-temporal covariance,

$$C(\boldsymbol{h}; u|\beta) = \frac{1}{0.2|u|^{1.5} + 1} \exp\left(-\frac{0.4\|\boldsymbol{h}\|}{(0.2|u|^{1.5} + 1)^{\beta/2}}\right), (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

here $\boldsymbol{h}$ is given by the existing nodes and u is taken as 2 for temporal correlations.

Under this structure, we further impose distinctive assumptions on the model to see

the potentially different sensor allocation results. Four situations are considered:

1. Separable for concurrent prediction: $\beta = 0$ and $m = 0$;

2. Separable for future prediction: $\beta = 0$ and $m = 6$;

3. Nonseparable for concurrent prediction: $\beta = 1$ and $m = 0$;

4. Nonseparable for future prediction: $\beta = 1$ and $m = 6$.

The corresponding sensor allocation results are: 1). $\boldsymbol{s}_{k+1} = (13.60, 41.53)$ with objective function minimized at $f_{min}(\boldsymbol{s}_{k+1}) = 2345.2$; 2). $\boldsymbol{s}_{k+1} = (13.60, 41.53)$ with $f_{min}(\boldsymbol{s}_{k+1}) = 2395.1$; 3). $\boldsymbol{s}_{k+1} = (10.47, 40.53)$ with $f_{min}(\boldsymbol{s}_{k+1}) = 2339.7$ and 4). $\boldsymbol{s}_{k+1} = (12.00, 42.60)$ with $f_{min}(\boldsymbol{s}_{k+1}) = 2380.3$. See Figure 4.2.

The results have great instructive purposes. The sensor allocation for case 1 and case 2 are identical, indicating that the sensor allocation problem is indifferent with the prediction horizons under separable spatial-temporal models. Actually, the conditional independence property under separability guarantees that all future predictions are based on the current observations. The objective function values have an overall increasing as greater forecasting horizons, though, they always achieve the same global minimum for each m.

In contrast, the sensor allocation problem is sensitive to the length of prediction horizons under nonseparable models. Case 3 and case 4 have different new sensor locations, and a longer prediction horizon yields greater prediction variance as expected. It also makes sense that nonseparable models always give smaller prediction variances than separable models since in this case, the interaction between spatial and temporal correlations helps the overall correlation to reduce less rapidly than in the separable case. By intuition, more information is utilized from earlier observations and hence the smaller prediction variance in the nonseparable case.

4.5 Real Example: Ozone Levels in Harris County, Texas

In this section we apply the method in section 4.3 on the ozone level measurements in Harris county. Harris county is the most populous county in the U.S. state of Texas. Its

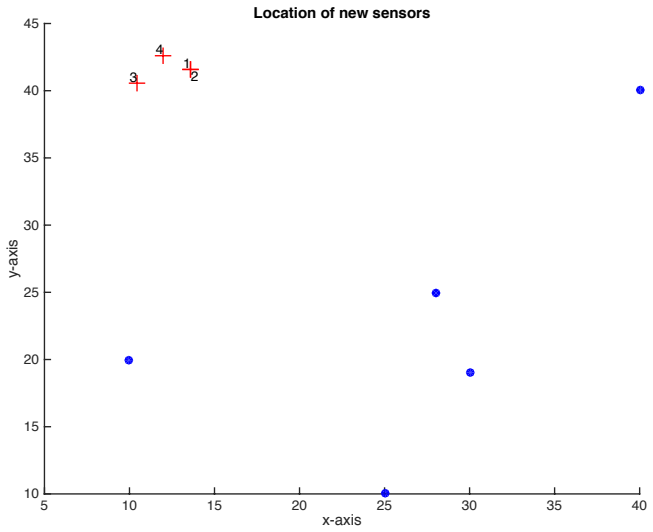


Figure 4.2: New sensor given five sensors under different assumptions.

county seat is Houston, the largest city in Texas and fourth-largest city in the United States. Hourly measurements of the ozone levels in Harris County were recorded at the 11 monitoring stations in 1993. The data was initially used in Carroll et al. (1997) for analyzing the pattern of population density to ozone exposures. They built a model on the spatial-temporal structure to predict the ozone levels in locations other than the sensor stations. Figure 4.3 shows the geographical distribution of the existing monitoring stations, and the predicted ozone levels over Harris county.

In this project, our purpose is to build a new monitoring station in addition to the 11 existing ones. Instead of the full dataset, we utilize the hourly data from 10 stations since the majority observations of station 4 are missing in year 1993. It is time series data with length 8760 for 10 different stations. Following Carroll (1997), square root transformation of the original ozone data is implemented to stabilize the variance. In addition, the trend part is taken out to make the residual process $\varepsilon(t, s)$ a real-valued stationary Gaussian random field with mean zero. Figure 4.4 shows the behavior of de-trend hourly ozone level measured by station 3 and station 6.

Before the spatial-temporal covariance model is fitted, a strong seasonal pattern exhibits from the ACF plot of residual series after de-trend process. Take station 3 and station 8 for example, the ACF of de-trend residual series are presented in the Figure

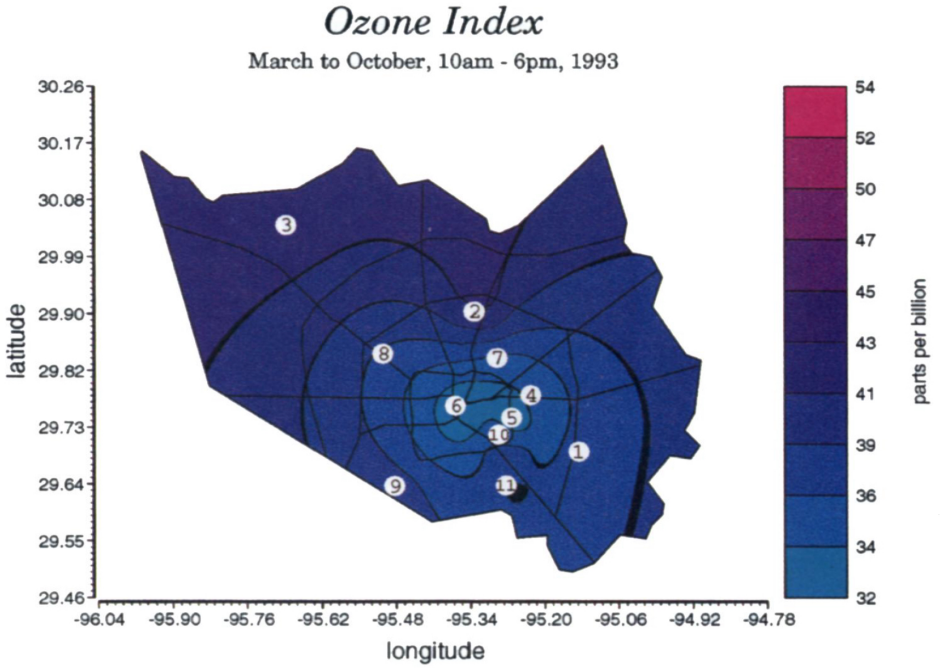


Figure 4.3: Locations of existing monitoring sensors in Harris County, Texas.

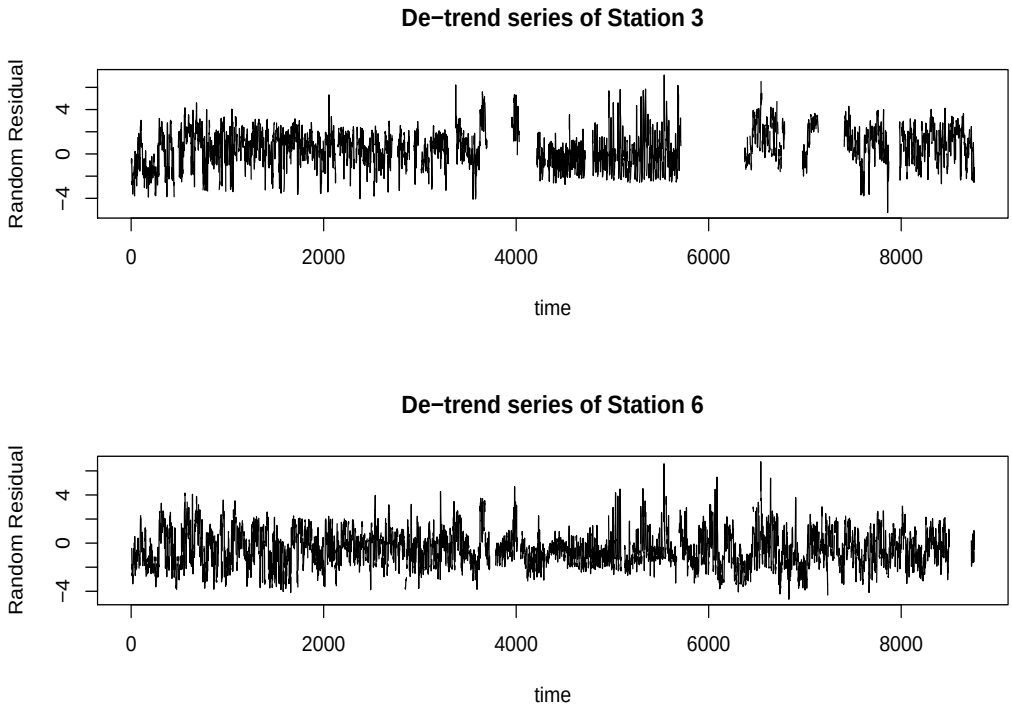


Figure 4.4: De-trend time series plots of the square root of hourly ozone level measured by station 3 and station 6.

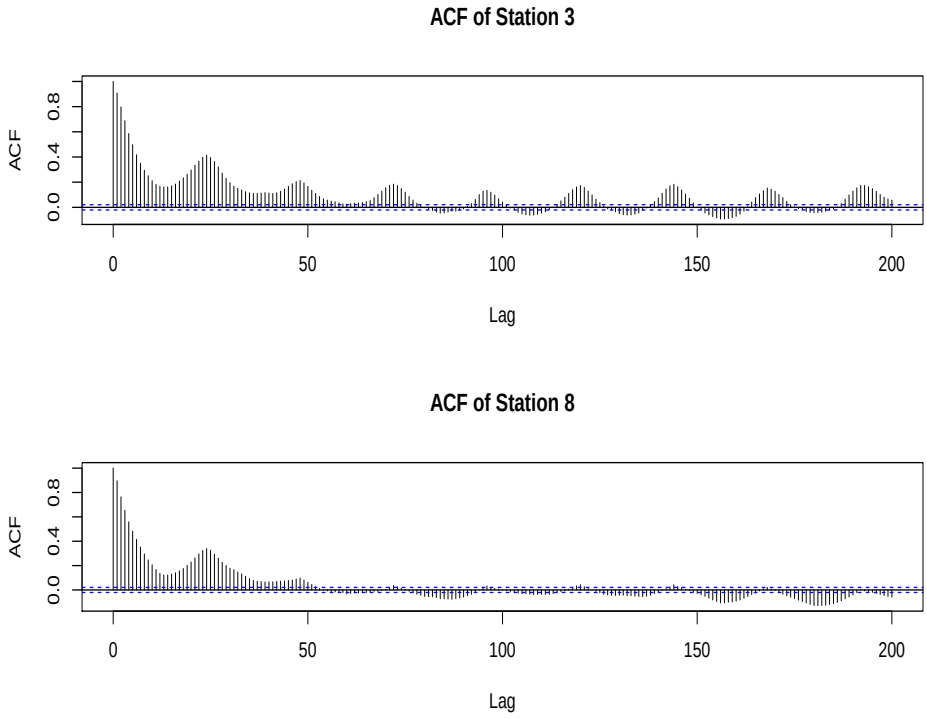


Figure 4.5: ACF of detrend time series for hourly ozone level measured by station 3 and station 8.

4.5.

Figure 4.5 shows a seasonal period of 24 hours, which coincides with the real condition of hourly data. Based on the autocorrelation pattern, $SAR(1)$ is fitted in order to remove seasonality. Figure 4.6 shows the autocorrelation pattern of the residual from $SAR(1)$. The seasonal patterns are alleviated significantly with small fluctuations for large time lags. Although higher order seasonal model is available, they fail to refine the small fluctuations better than $SAR(1)$. In other words, models such as $SAR(2)$ or $SAR(3)$ doesn't perform better than $SAR(1)$. On the other hand, the spatial-temporal analysis concentrates on small time lag only. Most of the spatial-temporal models claim that the time lags should be small to ensure the accuracy of model fitting since the information from large time lags cannot differentiate between separable and non-separable models efficiently. Therefore, $SAR(1)$ is sufficient to remove the seasonality for further analysis.

Take $SAR(1)$ for further discussion, the time correlations of hourly ozone level fall

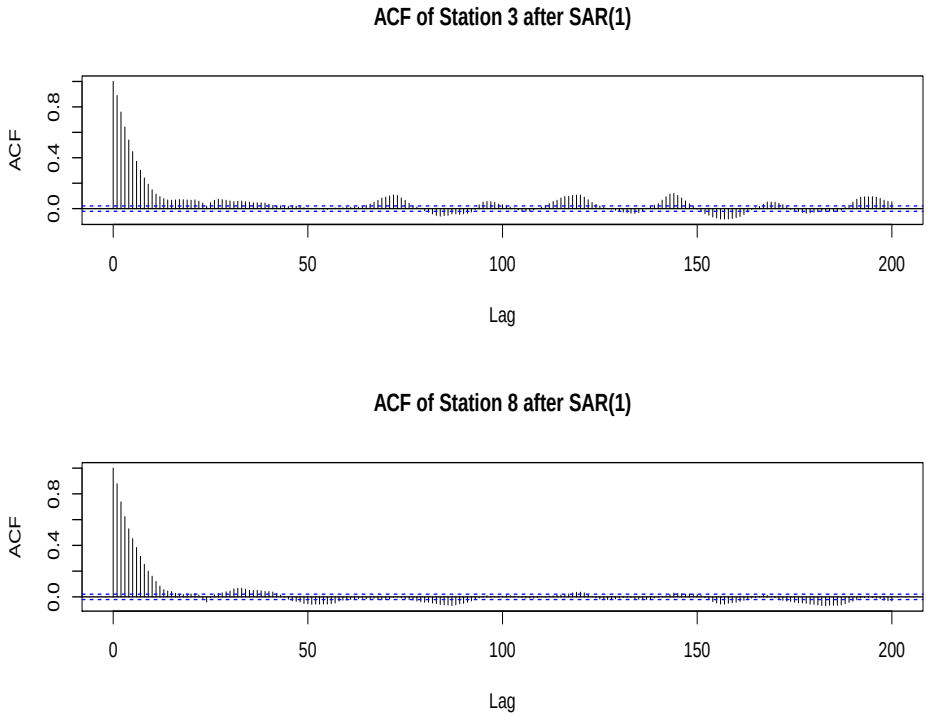


Figure 4.6: ACF of $SAR(1)$ residuals for hourly ozone level measured by station 3 and station 8.

off rapidly and are getting extremely small for time lags larger than 12. The residual series after de-seasonal are taken as $Y(\boldsymbol{s};t)$ for discussion of the spatial-temporal model. Note that t varies from 1 to 8736 and $\boldsymbol{s}$ varies among the given 10 monitoring sensor locations.

4.5.1 Spatial-Temporal Structure

As introduced in section 4.4, it is necessary to restrict attention to the well-accepted classes of positive-definite functions in $\mathbb{R}^2 \times \mathbb{R}$ for the ozone data. The covariance function family by Gneiting (2002) is utilized here. In particular, model (4.8) is a very popular one that fits well for the ozone data.

First, the purely spatial correlations are fitted for 45 pairs of stations. As discussed in Carroll (1997), a strong linear pattern is observed in the log spatial correlations for fixed time. In Figure 4.7, the log correlations versus distance for time lags fixed at 0, 2, 4, 6 are shown with a linear fitted line in each time lag. The scatter plot of spatial

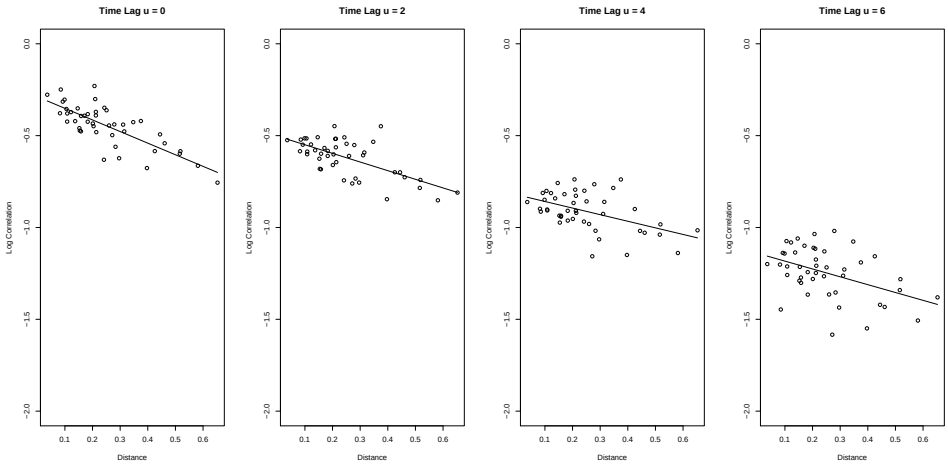


Figure 4.7: Spatial correlations: log-correlation versus distance for time lags fixed at 0, 2, 4, 6.

correlations shows a solid linear pattern for each time lag, and moves down as distance increases. Thus an exponential correlation model is strongly indicated and fitted as

$$C(\boldsymbol{h};0) = \begin{cases} 1, & \boldsymbol{h} = 0, \\ 0.751 \exp\left(-0.627\|\boldsymbol{h}\|\right), & \text{otherwise}, \end{cases} \tag{4.10}$$

with $\boldsymbol{h}$ ranges in $[0.036, 0.652]$.

Then we fit the pure temporal correlations by a Cauchy model. Since the distinction between separable and nonseparable covariance structures may not be feasible at higher time lags, Gneiting (2002) chooses to use time lags up to 3 days. Here we choose to fit time lags up to 12 hours since our samples consists of hourly measurements without any aggregation of them. In the other aspect, the time correlations for lags higher than 12 become very small with slight fluctuations afterwards. The fluctuations may due to variations of other environmental factors that cannot be modeled by usual seasonal factors. Therefore, time lags of 12 hours is reasonable to model the pure temporal correlation.

The pure temporal model below fits the empirical temporal correlations in the Ozone data well,

$$\rho(\mathbf{0};u) = \frac{1}{0.05|u|^{1.875}+1}, \tag{4.11}$$

with $u = 0, 1, 2, \ldots, 12$. Figure 4.8 gives examples of the time correlations between

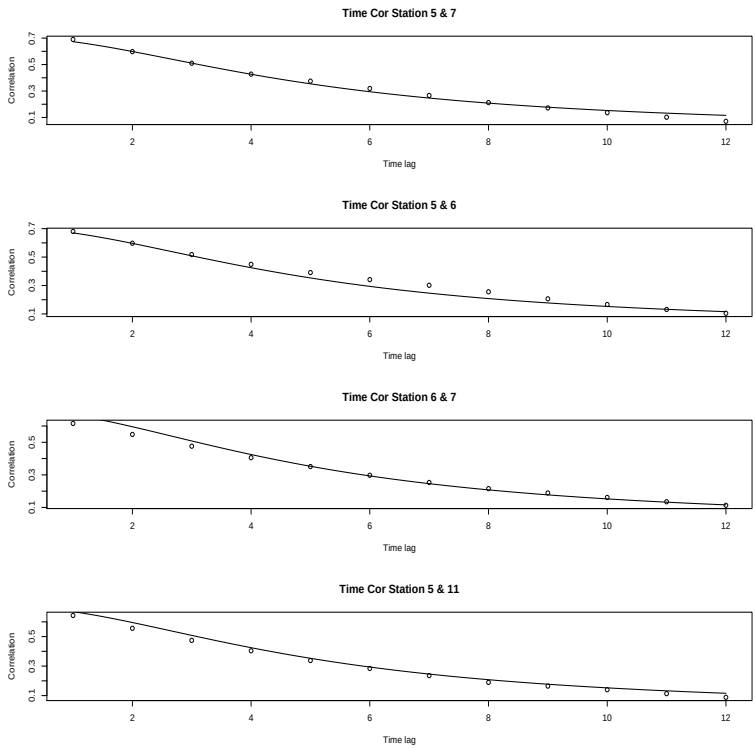


Figure 4.8: Time series correlations: log correlations versus time lag for fixed distances.

several pairs of stations, with fitted line added.

The pure spatial covariance and temporal covariance consist of the permissible parametric model,

$$C(\boldsymbol{h}; u|\beta) = \begin{cases} (0.05|u|^{1.875}+1)^{-1}, & \boldsymbol{h} = 0, \\ 0.751(0.05|u|^{1.875}+1)^{-1} \exp\left(-\frac{0.627\|\boldsymbol{h}\|}{(0.05|u|^{1.875}+1)^{\beta/2}}\right), & \text{otherwise}, \end{cases} \tag{4.12}$$

where $\beta \in [0,1]$ is the interaction parameter. The spatial-temporal interaction is controlled by β. The separable model corresponds to $\beta = 0$, and $\beta = 1$ indicates an extreme nonseparable case. As β increases, the overall correlation falls less rapidly than the separable model for nonzero temporal lags.

For the estimation of β, we use maximum likelihood estimation in Gaussian. The log-likelihood function from multivariate normal distribution is

$$\log L(\beta|\boldsymbol{h}; u) = -\frac{n|\boldsymbol{\Sigma}|}{2} - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}).$$

Here the covariance matrix $\boldsymbol{\Sigma}$ is positive-definite and defined in (4.7). The sample vector

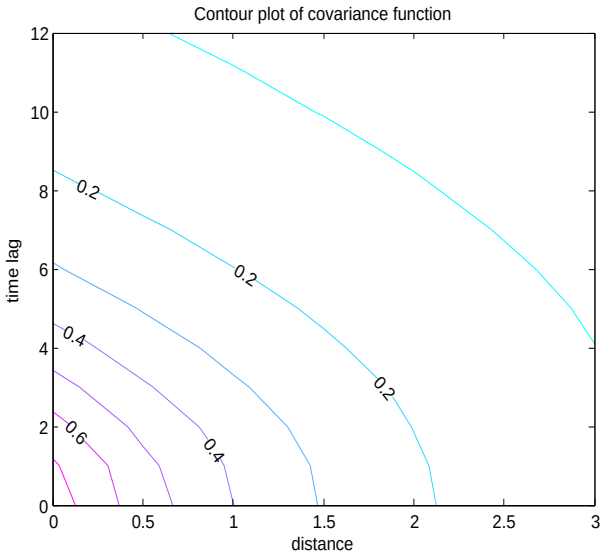


Figure 4.9: Contour plot of $C(\boldsymbol{h}; u|\beta)$ in (4.12) versus $\boldsymbol{h}$ and u. The x-axis is the spatial lag, y-axis is the temporal lag.

$\boldsymbol{x}_i$ is the i_{th} row in data matrix $\boldsymbol{X}$, where $\boldsymbol{X}$ is constructed to match the structure of the covariance matrix $\boldsymbol{\Sigma}$. That is

$$\boldsymbol{X} = (\boldsymbol{x}_{1,0}, \ldots, \boldsymbol{x}_{1,u}; \ldots; \boldsymbol{x}_{i,j}; \ldots; \boldsymbol{x}_{k,0}, \ldots, \boldsymbol{x}_{k,u}),$$

where $i = 1, \ldots, k$ and k is the number of given sensor locations, $j = 1, \ldots, u$ and u is the maximum time lag in the model. Here $\boldsymbol{x}_{i,j}$ is the time series measurements from $1 + j$ to $8724 + j$ at station i. Since the trend part has been removed before modeling, the mean vector $\boldsymbol{\mu}$ for the random process $\boldsymbol{X}_i$'s are assumed to be zero. Based on the real observations, the maximum likelihood estimation of β is $\hat{\beta} = 0.89$. The contour plot of the correlation function in (4.12) with $\beta = 0.89$ is shown in Figure 4.9.

4.5.2 Likelihood-Ratio Test for Separability

Separable covariance models are always preferable because of their advantages in computation, especially in the prediction problem regarding a spatial-temporal process. In prediction problem with k given sensor locations and u time lags, we need the inverse of a $(k+1)(u+1)$ by $(k+1)(u+1)$ covariance matrix under a nonseparable covariance

structure; however, this can be reduced to the inverse of a $(k+1)$ by $(k+1)$ matrix if separability can be assumed. We has shown the conditional independence property that can simplify the computational work by reducing the spatial-temporal model into a spatial one. Although we got a nonzero estimation for β in this example, the separability test is always necessary to exclude the situation that the curve of likelihood function on $\hat{\beta}$ is too flat to support the opposite.

There isn't an universal test for separability. Actually, the separability test is model-dependent as introduced in Mitchell et al. (2005). The covariance is separable in one model doesn't necessarily mean that it is separable in a different one. Therefore, the likelihood ratio test is a reasonable approach here to test the separability of the specific model we use.

Specifically, the test aims at the value of the interaction parameter β in Gneiting's model on the Ozone data. Recall the model with $d = 2$ and $\gamma = 1/2$:

$$C(\boldsymbol{h}; u|\beta) = \frac{\sigma^2}{(a|u|^{2\alpha}+1)} \exp\left(-\frac{c\|\boldsymbol{h}\|}{(a|u|^{2\alpha}+1)^{\beta/2}}\right), \quad (\boldsymbol{h}, u) \in \mathbb{R}^2 \times \mathbb{R}, \tag{4.13}$$

where $a \geq 0$, $c \geq 0$, $0 < \alpha \leq 1$ and $0 \leq \beta \leq 1$. The separable model corresponds to $\beta = 0$, and the extreme nonseparable case occurs when $\beta = 1$. The hypothesis is naturally constructed as

$$\boldsymbol{H}_0 : \beta = 0; \quad \boldsymbol{H}_1 : \beta > 0.$$

As the data frame constructed in the estimation procedure, there is a group of repeated measures of the spatial-temporal process $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_n)$, which are $i.i.d$ normal random vectors with the same covariance matrix defined in (4.7). Actually $\boldsymbol{x}_i$ is the i_{th} row in data frame $\boldsymbol{X}$, which is constructed to match the covariance structure of (4.7):

$$\boldsymbol{X} = (\boldsymbol{x}_{1,0}, \ldots, \boldsymbol{x}_{1,u}; \ldots; \boldsymbol{x}_{i,j}; \ldots; \boldsymbol{x}_{k,0}, \ldots, \boldsymbol{x}_{k,u}),$$

with k and u being the number of given spatial locations and time lags respectively. The log-likelihood function is

$$\log L(\beta \mid \boldsymbol{X}) = -\frac{n|\boldsymbol{\Sigma}|}{2} - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}),$$

and the test statistic from likelihood-ratio test is in the following form,

$$\Lambda(\boldsymbol{X}) = \frac{\sup_{H_0} L(\beta \mid \boldsymbol{X})}{\sup_{H_0 \cup H_1} L(\beta \mid \boldsymbol{X})}. \tag{4.14}$$

In our case, $L(\beta|\boldsymbol{X})$ cannot be analytically determined since the model construction is complicate; however, it can be calculated numerically as it is a completely defined function on β when $\boldsymbol{X}$ are observed. Therefore, the test statistic $\Lambda(\boldsymbol{x})$ can be obtained for observed $\boldsymbol{X} = \boldsymbol{x}$.

Although the sampling distribution of $\Lambda(\boldsymbol{X})$ is hard to derive here, by the asymptotic distribution of likelihood-ratio tests, the random variable $-2\log\Lambda(\boldsymbol{X})$ follows χ^2_1 when n is large. This asymptotic property is appropriate in this test since the sample size is large as $n = 8724$. Thus, the test statistic $\tilde{\Lambda}(\boldsymbol{X}) = -2\log\Lambda(\boldsymbol{X})$ is used to compare with the Chi-squared critical value. We have

$$\tilde{\Lambda}(\boldsymbol{X}) = 2\log L(\hat{\beta}|\boldsymbol{X}) - 2\log L(\hat{\beta}_0|\boldsymbol{X}).$$

The restricted estimation of β under H_0 is $\hat{\beta}_0 = 0$ and the unrestricted maximum likelihood estimation of β is given in the previous section as $\hat{\beta} = 0.89$, with log-likelihood value of -16968 and -16910 respectively.

Based on the asymptotic property, H_0 is rejected at level α if $\tilde{\Lambda}(\boldsymbol{x}) > \chi^2_{1,\alpha}$. The value of the test statistic is 116 comparing with the critical value 3.84 when $\alpha = 0.05$. This indicates that we will reject the null hypothesis $\beta = 0$. In other words, the covariance structure in (4.14) is nonseparable for the ozone data, and hence the estimation for β is acceptable.

4.5.3 New Station Allocation

Allocation by concurrent prediction

In the previous section, the spatial-temporal structure for ozone data is modeled as,

$$\rho(\boldsymbol{h}; u) = \begin{cases} (0.05|u|^{1.875} + 1)^{-1}, & \boldsymbol{h} = 0, \\ 0.751(0.05|u|^{1.875} + 1)^{-1} \exp\left(-\frac{0.627\|\boldsymbol{h}\|}{(0.05|u|^{1.875}+1)^{0.89/2}}\right), & \text{otherwise.} \end{cases}$$

To minimize the cost function (4.6), we use the grid search method to have the new location s_{k+1} walking over the region to get the one that corresponds to the global

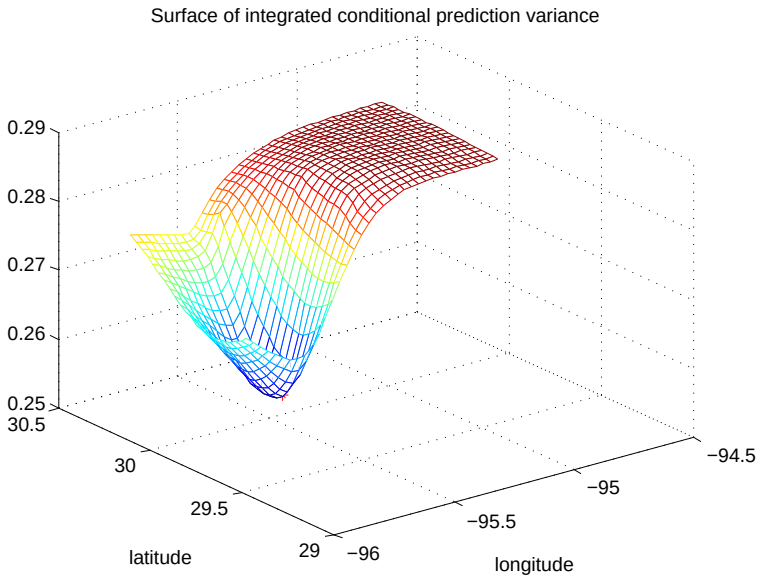


Figure 4.10: Surface of the integrated conditional variance over Harris County, Texas.

minimum of the objective function. For different locations of s_{k+1}, the surface for the objective function is shown in Figure 4.10. The contour plot of the objective function over this region is shown in Figure 4.11.

Observe from Figure 4.10 and 4.11 that the surface is convex with a global minimum at $[29.63, -95.72]$, the negative sign for longitude means it degrees to the west. The location of the new station is shown on the two-dimensional region in Harris County in Figure 4.12. The values of the objective function are high at locations in the south-eastern part since there has been a lot of sensor stations given in that area. Therefore, building a new station in the south-eastern area is meaningless since that the overall prediction variance could be larger than building it in the area with fewer given sensors. Figure 4.13 shows the final location of the new sensor together with the existing ones in the real map of the region. Apparently, the new sensor is located out of the boundary of Harris county since we take the whole rectangle region for numerical calculations. A more rigorous calculation could be done by simply imposing a non-regular lattice that fits perfectly around the county if needed.

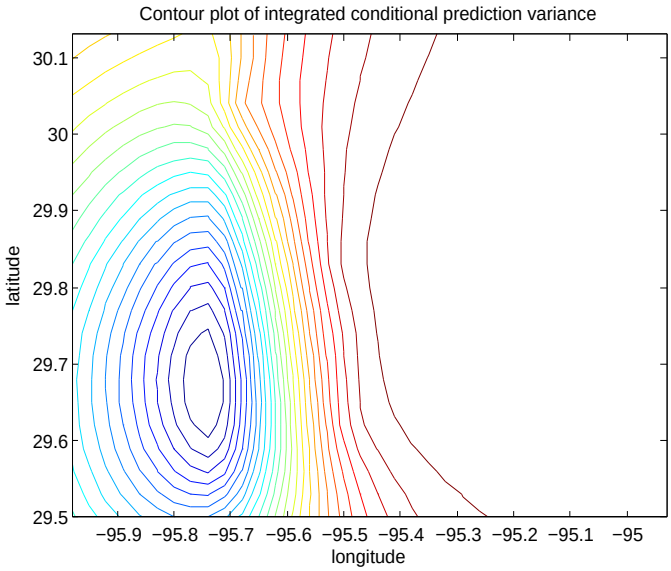


Figure 4.11: Contour plot of integrated conditional variance over Harris County.

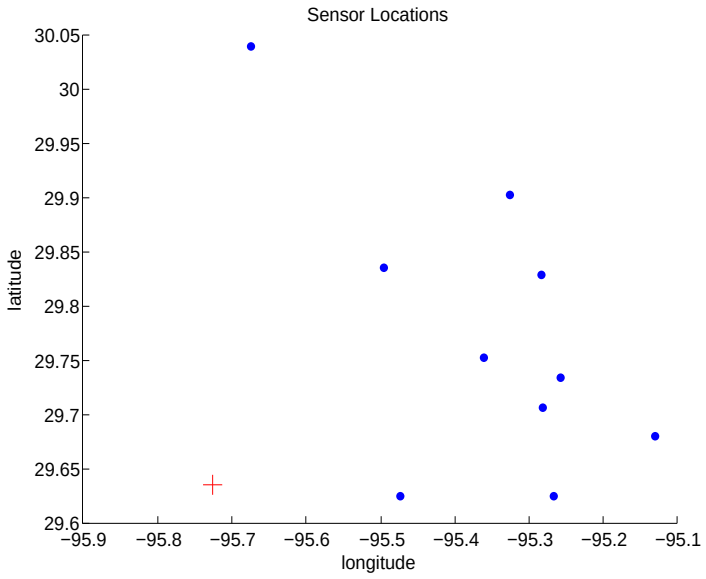


Figure 4.12: New sensor location in addition to existing ones in Harris County, Texas.

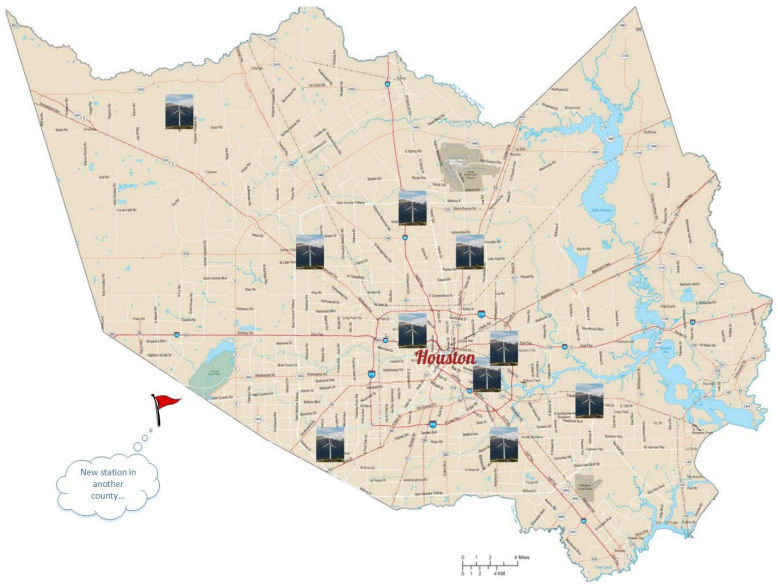
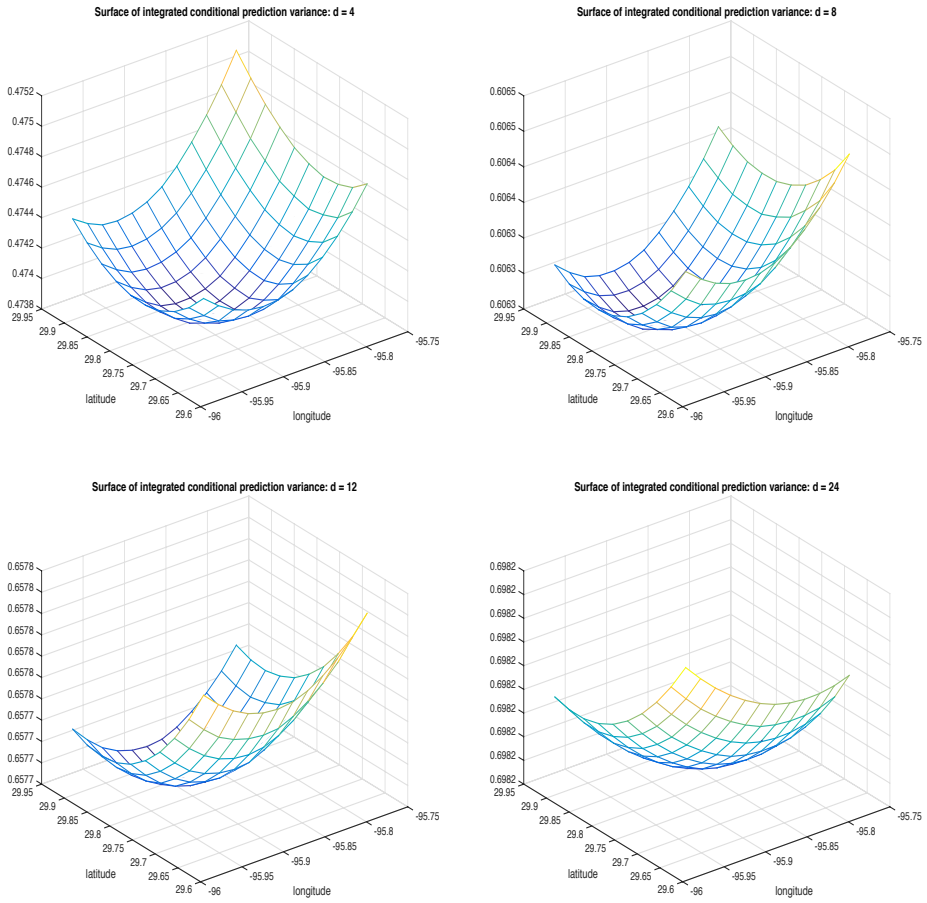


Figure 4.13: Sensor locations in real map of Harris County, Texas.

Allocation by future prediction

The previous section gives the new sensor location under the concurrent prediction purpose, but we can also explore the problem by the objective of future time predictions. The same procedure can be repeated on the objective function (4.6) for different forecasting horizons $m > 0$. The local behavior of the objective function are compared for $m = 4, 8, 12, 24$ hours respectively. See Figure 4.5.3. The local region is specified as $(-95.98, -95.76)$ by $(29.63, 29.95)$ for longitude and latitude.

As expected, the new sensor will not be allocated to the same location under non-separable models, especially for different prediction horizons. Different local behaviors of the objective function in Figure 4.5.3 strongly support this conclusion again. On the other hand, the allocation results are different though, they are all located closely to each other at the south-western area, including the new sensor location for concurrent prediction in section 4.5.3. It might be hard to differentiate if putting them all together into the full map. This is also the reason to present the local behaviors instead of the whole region.



4.6 Summary and Future Work

Spatial-temporal models are widely applied in sensor allocation problem. In contrary to other models, we focus on the minimization of overall prediction error in the spatial region. In Gaussian random field, the overall prediction error is defined as an integral of conditional variance given observations from existing sensors. The conditional variance structure relies on both spatial and temporal correlations.

Under separable assumptions, the sensor allocation problem can be solved by the observations from the current time. The conditional independence property says that the spatial-temporal process will reduce to a spatial-only case if the covariance is separable. If the covariance function is not separable, both spatial and temporal correlations will be included in the covariance structure by an interaction parameter. An admissible spatial-temporal covariance function is in need to guarantee the positive-definiteness

of the covariance matrix. In the ozone data example, one valid nonseparable spatial-temporal covariance model from Gneiting (2002) is well fitted. Further likelihood ratio test on the interaction parameter in the model is implemented to convince the nonseparability. The optimization of the integrated conditional variance gives the new station location in Harris County, Texas. The new station can thus be used to monitor and forecast environmental measures such as temperature, wind speed and ozone level for any other locations in the county.

As an inspiration of future work, the discussion concentrates on the sensor allocation problem for different prediction perspectives. It is demonstrated that the model gives different results given different prediction horizons. However, the distinct results may not be favored in practical use. People always need one final decision that can solve the problem simultaneously for multiple prediction purposes. Given this concern, future work can be done for an universal solution for the sensor allocation problem. One potential method is to generalize the objective function as the sum of those for different forecasting horizons. Or alternatively, when new locations are derived for each horizon separately, the final decision can be located as the geographical center of them. Such approaches are practical and anticipated to be explored further.

4.A Appendix

4.A.1 Appendix I

Conditional Independence Property

This section tries to show the conditional independence property of Gaussian random field under separable spatial-temporal models. This property can be proved by solving the conditional covariance of $Y(\boldsymbol{s}_i;t)$ and $Y(\boldsymbol{s}_j;t-u)$ given $Y(\boldsymbol{s}_j;t)$. Start to notice that

$$\begin{aligned}&\operatorname{Cov}\left(Y(\boldsymbol{s}_i;t), Y(\boldsymbol{s}_j;t-u) \mid Y(\boldsymbol{s}_j;t)\right)\\ =&E\left\{\left[Y(\boldsymbol{s}_i;t) - E(Y(\boldsymbol{s}_i;t) \mid Y(\boldsymbol{s}_j;t))\right]\left[Y(\boldsymbol{s}_j;t-u) - E(Y(\boldsymbol{s}_j;t-u) \mid Y(\boldsymbol{s}_j;t))\right] \mid Y(\boldsymbol{s}_j;t)\right\},\end{aligned} \tag{4.15}$$

with conditional expectation

$$E\left(Y(\boldsymbol{s}_i;t) \mid Y(\boldsymbol{s}_j;t)\right) = C_{\boldsymbol{h}}Y(\boldsymbol{s}_j;t), \mathrm{E}\left(Y(\boldsymbol{s}_j;t-u) \mid Y(\boldsymbol{s}_j;t)\right) = C_u Y(\boldsymbol{s}_j;t),$$

so that we can reduce (4.15) to

$$\mathrm{Cov}\left(Y(\boldsymbol{s}_i;t), Y(\boldsymbol{s}_j;t-u) \mid Y(\boldsymbol{s}_j;t)\right) = E\left\{Y(\boldsymbol{s}_i;t)Y(\boldsymbol{s}_j;t-u) \mid Y(\boldsymbol{s}_j;t)\right\} - C_{\boldsymbol{h}}C_u Y(\boldsymbol{s}_j;t)^2.$$

Next, for solving $E\left\{Y(\boldsymbol{s}_i;t)Y(\boldsymbol{s}_j;t-u)|Y(\boldsymbol{s}_j;t)\right\}$, we can perform the following decomposition on $Y(\boldsymbol{s}_i;t)$ and $Y(\boldsymbol{s}_j;t-u)$:

$$\begin{aligned} Y(\boldsymbol{s}_i;t) &= C_{\boldsymbol{h}}Y(\boldsymbol{s}_j;t) + \left(Y(\boldsymbol{s}_i;t) - C_{\boldsymbol{h}}Y(\boldsymbol{s}_j;t)\right) \\ Y(\boldsymbol{s}_j;t-u) &= C_u Y(\boldsymbol{s}_j;t) + \left(Y(\boldsymbol{s}_j;t-u) - C_u Y(\boldsymbol{s}_j;t)\right). \end{aligned} \tag{4.16}$$

In this way, by the definition of $C_{\boldsymbol{h}}$ and C_u, the random variable $Y(\boldsymbol{s}_j;t)$ and $Y(\boldsymbol{s}_i;t) - C_{\boldsymbol{h}}Y(\boldsymbol{s}_j;t)$ are independent given $Y(\boldsymbol{s}_j;t)$. Similarly, $Y(\boldsymbol{s}_j;t-u) - C_u Y(\boldsymbol{s}_j;t)$ and $Y(\boldsymbol{s}_j;t)$ are conditionally independent given $Y(\boldsymbol{s}_j;t)$, so the term of $\mathrm{E}\left[Y(\boldsymbol{s}_i;t)Y(\boldsymbol{s}_j;t-u)|Y(\boldsymbol{s}_j;t)\right]$ reduces to

$$\begin{aligned} &\mathrm{E}\left[Y(\boldsymbol{s}_i;t)Y(\boldsymbol{s}_j;t-u) \mid Y(\boldsymbol{s}_j;t)\right] \\ =&C_{\boldsymbol{h}}C_u Y(\boldsymbol{s}_j;t)^2 + \mathrm{E}[(Y(\boldsymbol{s}_i;t) - C_{\boldsymbol{h}}Y(\boldsymbol{s}_j;t))(Y(\boldsymbol{s}_j;t-u) - C_u Y(\boldsymbol{s}_j;t)) \mid Y(\boldsymbol{s}_j;t)]. \end{aligned}$$

With assumption of separability we have

$$\mathrm{Cov}\left[Y(\boldsymbol{s}_i;t)Y(\boldsymbol{s}_j;t-u)\right] = C_{\boldsymbol{h}}C_u,$$

A simple calculation leads to

$$\mathrm{E}\left[Y(\boldsymbol{s}_i;t)Y(\boldsymbol{s}_j;t-u) \mid Y(\boldsymbol{s}_j;t)\right] = C_{\boldsymbol{h}}C_u \mathrm{E}[Y(\boldsymbol{s}_j;t)^2 \mid Y(\boldsymbol{s}_j;t)].$$

Finally we get:

$$\mathrm{Cov}\left(Y(\boldsymbol{s}_i;t), Y(\boldsymbol{s}_j;t-u) \mid Y(\boldsymbol{s}_j;t)\right) = 0. \tag{4.17}$$

In Gaussian random filed, this directly implies conditional independence and hence the Markov property.

4.A.2 Appendix II

This section tries to show some theoretical results for the sensor allocation problem in the 1-D spatial-only case. Moreover, we assume the spatial correlation to be exponential: $\mathrm{Corr}(\mathrm{Y}(\mathrm{s_i}), \mathrm{Y}(\mathrm{s_j})) = \exp(-|\mathrm{s_i} - \mathrm{s_j}|)$. Suppose there are k different sensors given in a bounded line $l \subset \mathbb{R}$ with order $s_1, s_2, \ldots, s_k$. These k points construct $k-1$ intervals. To determine the location of new sensor s_{k+1}, it is reasonable to determine the location of s_{k+1} inside each of the $k-1$ intervals between s_i and s_j, $i, j = 1, 2, \ldots, k$. Next we can decide which interval should s_{k+1} be located in order to minimize the integrated conditional variance defined in (4.3).

First we can discuss the location of s_{k+1} inside the interval $[s_i, s_j]$, with $i, j = 1, \ldots, k$. Suppose the new sensor s_{k+1} is located between s_i and s_j, $s_i < s_{k+1} < s_j$, then by the Markov property, the observations in $[s_i, s_{k+1}]$ is independent with observations within $[s_{k+1}, s_j]$, the conditional variance of any other locations given s_i, s_{s+1} and s_j is

$$\mathrm{Var}[Y(s) \mid Y(s_i), Y(s_{k+1}), Y(s_j)] = \mathrm{Var}[Y(s) \mid Y(s_j), Y(s_{k+1})] + \mathrm{Var}[Y(s_{k+1}) \mid Y(s_j)].$$

By the definition of objective function (4.3), we also have

$$f(s_{k+1}; s_i, s_j) = f(s_{k+1}; s_i) + f(s_{k+1}; s_j).$$

The new sensor location s_{k+1} is derived by minimizing the above function to set

$$\frac{\partial f(s_{k+1}; s_i, s_j)}{\partial s_{k+1}} = \frac{\partial f(s_{k+1}; s_i)}{\partial s_{k+1}} + \frac{\partial f(s_{k+1}; s_j)}{\partial s_{k+1}}.$$

It implies that

$$\frac{1 - e^{-4\alpha(s_{k+1} - s_i)} - 4\alpha(s_{k+1} - s_i)e^{-2\alpha(s_{k+1} - s_i)}}{\left(1 - e^{-2\alpha(s_{k+1} - s_i)}\right)^2} = \frac{1 - e^{-4\alpha(s_j - s_{k+1})} - 4\alpha(s_j - s_{k+1})e^{-2\alpha(s_j - s_{k+1})}}{\left(1 - e^{-2\alpha(s_j - s_{k+1})}\right)^2},$$

it proceeds to imply that $s_{k+1} - s_i = s_j - s_{k+1}$, which results in the solution $s_{k+1} = \frac{s_i + s_j}{2}$ to this equation. The result concludes that in the one dimensional case, the new sensor should be located in the middle of two given sensors within each interval.

Next we can compare the integrated conditional variance for different intervals to decide which interval should s_{k+1} be located. Given two different intervals $[s_i, s_j]$ and

$[s_m, s_n]$, with $i, j, m, n = 1, \ldots, k$. Since s_{k+1} is to be located in one of these two intervals with all other intervals unchanged, it is essentially to compare the objective function $f(s_{k+1}; s_i, s_j, s_m, s_n)$ with $s_{k+1} \in [s_i, s_j]$ and $s_{k+1} \in [s_m, s_n]$ respectively. When s_{k+1} is located in $[s_i, s_j]$, the objective function (4.3) is calculated as

$$\begin{aligned}
&f(s_{k+1}; s_i, s_j, s_m, s_n) \\
&= \int_{s_i}^{s_{k+1}} \mathrm{Var}[Y(s) \mid Y(s_i), Y(s_{k+1})] + \int_{s_{k+1}}^{s_j} \mathrm{Var}[Y(s) \mid Y(s_j), Y(s_{k+1})] \\
&+ \int_{s_m}^{s_n} \mathrm{Var}[Y(s) \mid Y(s_m), Y(s_n)] \\
&= d_2 \left(1 + \frac{2e^{-\alpha d_2}}{1 - e^{-\alpha d_2}}\right) - d_1 \left(1 + \frac{2e^{-\alpha d_1}}{1 - e^{-\alpha d_1}}\right),
\end{aligned}$$

where $d_1 = s_j - s_i$ and $d_2 = s_n - s_m$.

A similar equation holds when s_{k+1} is located in $[s_m, s_n]$ with d_1 and d_2 exchanged in place in the above equation. Compared these two equations, $f(s_{k+1}; s_i, s_j, s_m, s_n)$ would be smaller in $[s_i, s_j]$ if $d_1 > d_2$ and similarly, it will be smaller in $[s_m, s_n]$ if $d_2 > d_1$. This concludes that the new sensor should be located in the larger interval.

There is another situation that the sensor would fall beyond the intervals confined by $s_1, \ldots, s_k$. In this case, one similar formula is derived to minimize the objective function (4.3). The solution here is not necessarily to be in the middle of this interval. For example, if we want to put a new sensor in the interval $[0, 10]$, and there is one given sensor at point 10, the new sensor will be located at point 2.9. If the interval becomes $[0, 5]$ instead of $[0, 10]$, the location of s_{k+1} falls at point 1.36. This fact reveals that the new sensor tends to be close to the end point 0 where no sensors are given.

Bibliography

Aitchison, J. (1986a). *Review of Economics of the Household*. Chapman and Hall, New York.

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

Al-Abri, A. S. and Goodwin, B. K. (2009). Re-examining the exchange rate pass-through into import prices using non-linear estimation techniques: Threshold cointegration. *International Review of Economics and Finance*, 18:142161.

Balchin, S. (1995). A description of the seasonal adjustment methods X11, X11ARIMA, X12ARIMA, GLAS, STL, SEATS, STAMP and microCAPTAIN. Central Statistical Office.

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration. *International Economic Review*, 38(3):627–645.

Barigozzi, M., Alessi, L., Capasso, M., and Fagiolo, G. (2012). The distribution of household consumption-expenditure budget shares. *Structural Change and Economic Dynamics*, 23:69–91.

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

Bilonick, R. (1985). The space-time distribution of sulfate deposition in the northeastern united states. *Atmospheric Environment*, 19:1829–1845.

Box, G. and Jenkins, G. (1994). *Time Series Analysis: Forecasting and Control (3rd Edition)*. Prentice Hall, New Jersey.

Brockwell, P. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer.

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

Carroll, R. J., Chen, R., George, E. I., Li, T. H., Newton, H. J., Schmiediche, H., and Wang, N. (1997). Ozone exposure and population density in harris county, texas. *Journal of the American Statistical Association*, 92:392–415.

Chan, K. S. (1990). Testing for threshold autoregression. *Ann. Statist.*, 18(4):1886–1894.

Chan, K. S., Petruccelli, J. D., Tong, H., and Woolford, S. W. (1985). A multiple-threshold $ar(1)$ model. *Journal of Applied Probability*, 22(2).

Chan, K. S. and Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society*, 52:469–476.

Chen, R. and Fomby, T. (1999). Forecasting with stable seasonal pattern models with an application of hawaiian tourist data. *Journal of Business & Economic Statistics*, 17:497–504.

Cleveland, W. P. and 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.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, second edition.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatiotemporal stationary covariance functions. *Journal of the American Statistical Association*, 94(4):392–415.

Das, A. and Kempe, D. (2008). Sensor selection for minimizing worst-case prediction error. In *Proceedings of the 7th International Conference on Information Processing in Sensor Networks, IPSN 2008, St. Louis, Missouri, USA, April 22-24, 2008*, pages 97–108.

De Luna, X. and Genton, M. G. (2005). Predictive spatio-temporal models for spatially sparse environmental data. *Statistica Sinica*, 15:547568.

Del Boca, D. and Flinn, C. (2012). Endogenous household interaction. *Journal of Econometrics*, 166:49–65.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barcel-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35:279–300.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.

Fazzari, S., Morley, J., and Panovska, I. (2011). Fiscal policy asymmetries: A threshold vector autoregression approach.

Findley, D., Monsell, B., Bell, W., Otto, M., and 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. and Harvey, A. (1983). Testing for deterministic trend and seasonal components in time series model. *Biometrika*, 70(3):673–682.

Fuentes, M. (2006). Testing for separability of spatialtemporal covariance functions. *Journal of Statistical Planning and Inference*, 136:447–466.

Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics*, 18:681–695.

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

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97:590–600.

Gneiting, T., Genton, M. G., and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In *Statistical Methods for Spatio-Temporal Systems*. Chapman and Hall.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):190–195.

Hansen, B. E. (1997). Inference in tar models. *Studies in Nonlinear Dynamics and Econometrics*, 2:1558–3708.

Hansen, B. E. (1999). Testing for linearity. *Journal of Economic Surveys*, 13:551–576.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68:575–603.

Hansen, B. E. and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics*, 110:293 318.

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: assessing irelands wind-power resource (with discussion). *Applied Statistics*, 38:1–50.

Heimonen, K. (2006). Nonlinear adjustment in pppevidence from threshold cointegration. *Empirical Economics*, 31(2).

Huerta, G., Sans, B., and Stroud, J. R. (2004). A spatiotemporal model for mexico city ozone levels. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 53(2):231–248.

Hung, Y., Amemiya, Y., and Wu, C. F. J. (2010). Probability-based latin hypercube designs. *Biometrika*, 97:961–968.

Hung, Y., Joseph, V. R., and Melkote, S. N. (2009). Design and analysis of computer experiments with branching and nested factors. *Technometrics*, 51:354–365.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):15511580.

Kilger, C. and Wagner, M. (2010). Demand planning. In Stadtler, H. and Kilger, C., editors, *Supply Chain Management and Advanced Planning: Concepts, Models, Software, and Case Studies (4th Edition)*, chapter 7, pages 133–160. Springer.

Krause, A., Singh, A., and Guestrin, C. (2008). Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.*, 9:235–284.

Krueger, D. and Perri, F. (2006). Does income inequality lead to consumption equality? evidence and theory. *Review of Economic Studies*.

Li, B., Genton, M. G., and Sherman, M. (2008). Testing the covariance structure of multivariate random fields. *Biometrika*, 95(4):813–829.

Ling, S. and Tong, H. (2005). Testing for a linear ma model against threshold ma models. *Ann. Statist.*, 33(6):2529–2552.

Liu, X., Cai, Z., and Chen, R. (2015). Functional coefficient seasonal time series models with an application of hawaii tourism data. *Computational Statistics*, page 1.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

Lütkepohl, H. (2009). *Handbook of Computational Econometrics*. John Wiley and Sons Inc, New Jersey.

Matern, B. (1986). *Spatial Variation, Lecture Notes in Statistics*. Number 36. Springer Varlag, New York.

Mills, T. C. (2010). Forecasting compositional time series. *Qual Quant*, 44:673690.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2005). A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis*, 97:1025–1043.

Nakagawa, H. (2010). Investigating nonlinearities in real exchange rate adjustment: Threshold cointegration and the dynamics of exchange rates and relative prices. *Journal of International Money and Finance*, 29(5):770–790.

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

Perri, F., Krueger, D., and Lustig, H. (2008). Evaluating asset pricing models with limited commitment using household consumption data. *Journal of the European Economic Association*.

Petruccelli, J. D. and Davies, N. (1986). A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika*, 73(3):687–694.

Phillips, P. C. B. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometric*, 58:165193.

Priestley, M. (1983). *Spectral Analysis and Time Series*. Wiley.

Quintana, J. M. and West, M. (1988). The time series analysis of compositional data. *Bayesian Statistics*, 3:747756.

Ravishanker, N., Dey, D. K., and Iyengar, M. (2001). Compositional time series analysis of mortality proportions. *Communications in Statistics - Theory and Methods*, 30(11):2281–2291.

Reich, B. J., Eidsvik, J., Guindani, M., Nail, A. J., and Schmidt, A. M. (2011). A class of covariate-dependent spatiotemporal covariance functions for the analysis of daily ozone concentration. *Ann. Appl. Stat.*, 5(4):2425–2447.

Reinsel, G. C. (2003). *Elements of multivariate time series analysis*. Springer.

Sacks, J., Welch, W. J., Mitchell, T. J., , and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423.

Shumway, R. and Stoffer, D. (2006). *Time Series Analysis and Its Applications: With R Examples*. Springer.

Singh, K., Xie, M., and Strawderman, W. E. (2007). *Confidence distribution (CD) – distribution estimator of a parameter*, volume Volume 54 of *Lecture Notes–Monograph Series*, pages 132–150. Institute of Mathematical Statistics, Beachwood, Ohio, USA.

Stein, M. L. (2005). Space time covariance functions. *Journal of the American Statistical Association*, 100(469):310–321.

Stephens Jr., M. (2004). Job loss expectations, realizations, and household consumption behavior. *Review of Economics and Statistics*.

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

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

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

Tong, H. (1990). *Nonlinear Time Series Analysis: A Dynamical System Approach*. London: Oxford University Press.

Tong, H. (1997). Some comments on nonlinear time series analysis. In Cutler, C. D. and Kaplan, D. T., editors, *Nonlinear Dynamics and Time Series*, volume Fields Institute Communications 11. American Mathematical Society, Rhode Island.

Tong, H. (2011). Threshold models in time-series analysis30 years on. *Statistics and Its Interface*, 4:123–127.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc., B*, 42:245–292.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84:231–240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, 93:1188–1202.

Tsay, R. S. (2010). *Analysis of Financial Time Series (Third Edition)*. John Wiley and Sons Inc, New Jersey.

Tsay, R. S. (2013). *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley, New Jersey.

Williams, B. J. (2000). *Sequential Design of Computer Experiments to Minimize Integrated Response Functions*. PhD thesis. AAI9983009.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *The Royal Society*, 226:267–298.

Zellner, A. (1978). Retrospect and prospect. In Zellner, A., editor, *Seasonal Analysis of Economic Time Series*. US Department of Commerce, Bureau of the Census, Washington, DC.