RICE UNIVERSITY

An Examination of Some Open Problems in Time Series Analysis

by

Ginger Michelle Davis

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Dr. Katherine B. Ensrud
Chairman and Advisor
Professor and Chair of Statistics
Rice University

Dr. James R. Thompson
Noah Harding Professor of Statistics
Rice University

Dr. Shannon W. Anderson
Associate Professor of Management
Rice University
Principal Fellow
University of Melbourne
Dept. of Economics and Commerce

HOUSTON, TEXAS

MAY, 2005
UMI Number: 3216690

Copyright 2005 by
Davis, Ginger Michelle

All rights reserved.

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform 3216690

Copyright 2006 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346
Copyright
Ginger Michelle Davis
2005
Abstract

An Examination of Some Open Problems in Time Series Analysis

by

Ginger Michelle Davis

We investigate two open problems in the area of time series analysis. The first is developing a methodology for multivariate time series analysis when our time series has components that are both continuous and categorical. Our specific contribution is a logistic smooth transition regression (LSTR) model whose transition variable is related to a categorical variable. This methodology is necessary for series that exhibit nonlinear behavior dependent on a categorical variable. The estimation procedure is investigated both with simulation and an economic example. The second contribution to time series analysis is examining the evolving structure in multivariate time series. The application area we concentrate on is financial time series. Many models exist for the joint analysis of several financial instruments such as securities due to the fact that they are not independent. These models often assume some type of constant behavior between the instruments over the time period of analysis. Instead
of imposing this assumption, we are interested in understanding the dynamic covariance structure in our multivariate financial time series, which will provide us with an understanding of changing market conditions. In order to achieve this understanding, we first develop a multivariate model for the conditional covariance and then examine that estimate for changing structure using multivariate techniques. Specifically, we simultaneously model individual stock data that belong to one of three market sectors and examine the behavior of the market as a whole as well as the behavior of the sectors. Our aims are detecting and forecasting unusual changes in the system, such as market collapses and outliers, and understanding the issue of portfolio diversification in multivariate financial series from different industry sectors. The motivation for this research concerns portfolio diversification. The false assumption that investment in different industry sectors is uncorrelated is not made. Instead, we assume that the comovement of stocks within and between sectors changes with market conditions. Some of these market conditions include market crashes or collapses and common external influences.
Acknowledgments

I was first introduced to Time Series Analysis by Dr. Katherine Ensor in STAT 421 at Rice University. My interest was piqued by the concept that the internal structure in data could be captured with an intuitively sound model. My interest in this challenging field has continued to strengthen, and I hope to use my knowledge in this area to contribute to the many applications of time series analysis, including computational finance, environmental applications, econometrics, and many others. I would like to thank Dr. Ensor who served as advisor and chair of my committee. She has been a mentor to me academically, intellectually, professionally, and personally. Thank you to Dr. Anderson who has allowed me to broaden my horizons and conduct research with her in the business arena. Thanks as well to Dr. Thompson for introducing me to computational finance and providing me with the background to start research in computational finance.

Many other people have helped me in this research. Dr. Scott Baggett has helped me with many tasks over the past 4 years not only due to the fact that he supervised my Statistical Consulting Lab work but also that he is a very knowledgeable and helpful person. I would like to thank Dr. William Wojciechowski for his contributions during the early stages of the evolutionary principal component research. I am indebted to Dr. Denis Pelletier as well for his Ox computer programs and his answers to my questions regarding them.
I would like to thank the statistics students at Rice University for providing a fun and intellectual environment in which I could succeed. Some of these people include Mike Lecocke, Chris Rudnicki, Jason Gershman, Alena Scott, Rick Ott, Jenny Zhang, Li Deng, Kalatu Davies, Chad Bhatti, Blair Christian, Talithia Daniel, Hector Flores, Garrett Fox, Lada Kyj, Jong Soo Lee, and Matthias Mathaes.

This thesis is dedicated to my husband Brian whose encouragement, sacrifice, interest, understanding, and love has proven him indispensable in my success. Thanks also to my roommate and sister Kelly Holt who has cheered me on for the past two years. The rest of my family (Holts and Davies alike) have provided me with greatly appreciated support. Jodi Diamond was my confidante and motivator. Dan and Diane Parisian have also provided me with special encouragement. Finally, I would like to thank the rest of my family and friends for all of their support over the past 5 years. I could not have done it without you.

I would like to acknowledge the financial support for this research from the Statistical Consulting Lab at Rice University, the National Science Foundation DMS - 0240588, and the Army Research Office ARO-1-0354 support for my advisor Dr. Katherine Enser.
Contents

Abstract ................................................................. ii

Acknowledgments ......................................................... iv

List of Figures ............................................................. viii

List of Tables ............................................................. xii

1 Background .............................................................. 1

1.1 Overview .................................................................. 1

1.2 Multivariate Time Series Analysis with Categorical and Continuous
   Variables in an LSTR model ........................................... 4

1.3 Evolving Structure in Multivariate Time Series with Application to
   Stock Sector Data ....................................................... 13

2 Multivariate Time Series Analysis with Categorical and Continuous
   Variables in an LSTR model ........................................... 15

2.1 Introduction .............................................................. 15
2.2 Model Definition .............................................. 18
2.3 Estimation ..................................................... 19
2.4 Simulation Results .......................................... 22
2.5 Application of Methodology .............................. 25

3 Evolving Structure in Multivariate Time Series with Application to
Stock Sector Data .......................................................... 42
3.1 Introduction ....................................................... 42
3.2 Data Description ................................................ 43
3.3 Motivation ......................................................... 45
3.4 Methodology ....................................................... 46
3.5 Simulation Study ................................................ 51
3.6 Results ............................................................. 52

4 Conclusions and Further Work .................................. 78
4.1 Conclusions ....................................................... 78
4.2 Further Work ..................................................... 80

A Multivariate Generalized Autoregressive Conditional
Heteroscedasticity (MGARCH) model development ......... 82

B Multiple step-ahead Forecasting in LSTR model ............ 88

Bibliography ............................................................. 94
List of Figures

2.1 Latent Process $X(t)$ and its Estimators. $X.t$ denotes the true process $X_t$ as in Equation 2.2. $X.fit1$ denotes the first estimate $\hat{X}_t^{(1)}$ as in Equation 2.9. $X.fit3$ denotes the iterative estimate $\hat{X}_t^{(3)}$ as in Equation 2.15. .................................................. 32

2.2 Categorical Time Series $Y(t)$ and its Estimators. $Y.t$ denotes the categorical time series $Y_t$ as in Equation 2.3. $Y.fit1$ denotes the first estimate $\hat{Y}_t^{(1)}$ as in Equation 2.10. $Y.fit3$ denotes the iterative estimate $\hat{Y}_t^{(3)}$ as in Equation 2.16. .................................................. 33

2.3 LSTR process $y(t)$ and its Estimators. $y.t$ denotes the LSTR process $y_t$ as in Equation 2.6. $y.fit1$ denotes the first estimate $\hat{y}_t^{(1)}$ as in Equation 2.12. $y.fit2$ denotes the iterative estimate $\hat{y}_t^{(2)}$ as in Equation 2.18. ... 34

2.4 Monthly seasonally unadjusted US employment rate per 100, males aged 20 and above, January 1968-December 2004 ......................... 35

2.5 State of the Economy (Recession = 1, Contraction = 2, Expansion = 3) 36
2.6 State of the Economy (Recession = 1, Contraction = 2, Expansion = 3) with Fitted Values ............................................. 37
2.7 Monthly seasonally unadjusted US employment rate, males aged 20 and above, January 1968-December 1989 with Fitted Values ...... 38
2.8 Monthly seasonally unadjusted US employment rate, males aged 20 and above, January 1968-December 1989 with 95% Confidence Intervals 39
2.9 One step ahead forecasts for LSTR model for December 1989 - November 1999 compared to actual values ......................... 40
2.10 One step ahead forecasts for LSTR model for December 1989 - November 1999 with 95% Confidence Intervals compared to actual values 41
3.2 Daily (shifted) Stock returns ............................................ 61
3.3 Principal Component Scores by Year, where “f” denotes Financial companies, “e” denotes Energy companies, and “t” denotes Technology companies. The x- and y-axis are scores for the first and second principal component, respectively ................................. 62
3.4 Simulated Daily (shifted) stock returns ................................ 63
3.5 Estimated Probability / Indicator of Higher Correlation Regime. The points (circles) are an indicator function for the higher correlation regime (0 if we are in the lower correlation regime and 1 if we are in the higher correlation regime). The points (crosses) are the estimated probabilities of being in the higher correlation regime, and the path of the estimates is connected with the line.

3.6 Indicator of Correctly Predicted Regime. The points (circles) are an indicator function for whether the estimated probability of the true regime is > 0.5 (0 if the estimated probability of the true regime is < 0.5 and 1 if the estimated probability of the true regime is > 0.5).

3.7 One step ahead forecasts for average correlations

3.8 One step ahead forecasts for average correlations

3.9 Daily (shifted) returns and Estimated probability of being in higher correlated regime

3.10 Daily (shifted) returns and Estimated probability of being in higher correlated regime

3.11 Number of necessary principal components and Estimated probability of being in lower correlated regime

3.12 Biplots on 10-01-2001 and 10-02-2001

3.13 Biplot on 10-04-2001

3.14 Biplot on 06-02-2000

3.16 Biplot on 01-11-2000 .................................................. 76

Right panel: Daily (shifted) stock returns for WDC for 1/7/2000 - 1/21/2000. ................................. 77

B.1 Two steps ahead forecasts for LSTR model for December 1989 - October 1999 compared to actual values ...................... 90

B.2 Two steps ahead forecasts for LSTR model for December 1989 - October 1999 with 95% Confidence Intervals compared to actual values .... 91

B.3 Three steps ahead forecasts for LSTR model for December 1989 - September 1999 compared to actual values .................. 92

B.4 Two steps ahead forecasts for LSTR model for December 1989 - September 1999 with 95% Confidence Intervals compared to actual values .... 93
List of Tables

2.1 Simulation Study of Parameter Estimates based on 100 replications of series of length 500 ......................................................... 24

2.2 Forecast Evaluation for Monthly US Unemployment Rate .............. 30

3.1 Correlation and Transition Probability Estimates in Simulation Study based on 25 realizations ....................................................... 52

3.2 Univariate Volatility Estimates in Simulation Study based on 25 realizations ................................................................. 53

3.3 Correlation and Transition Probability Estimates ......................... 54

3.4 Univariate Volatility Estimates .................................................. 55

3.5 Outliers detected from Biplots .................................................... 73

B.1 Forecast Evaluation for Monthly US Unemployment Rate ............. 89
Chapter 1

Background

1.1 Overview

Many problems of interest involve the analysis of data collected over time. Such problems arise in many subject areas, and we will address particular problems in computational finance and econometrics. Some terminology is required in order to understand time series analysis. A time series is a collection of observations indexed by the time of the observation. We denote a time series by a vector y:

\[ y = (y_1, y_2, \ldots, y_T), \]  \hspace{1cm} (1.1)

where our observations are collected from time \( t = 1, \ldots, T \). We denote the mean of our time series as \( \mu_t \). Autocovariance is the covariance of our random variable \( Y_t \) with its own lagged value:

\[ \gamma_{jt} = E[(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})]. \]  \hspace{1cm} (1.2)
A covariance-stationary, or weakly stationary process has a mean and autocovariance function that is not dependent on the date of observation. That is,

\[ E [Y_t] = \mu, \]  
(1.3)

\[ E [(Y_t - \mu_t) (Y_{t-j} - \mu_{t-j})] = \gamma_j. \]  
(1.4)

This implies that the covariance between \(Y_t\) and \(Y_{t-j}\) depends only on \(j\), the length of time separating the observations, and not on the date of the observation. Strict stationarity means that for any values of \(j_1, j_2, \ldots, j_n\), the joint distribution of \((Y_t, Y_{t+j_1}, \ldots, Y_{t+j_n})\) depends only on the intervals separating the dates and not on the date itself. A Gaussian process is one whose joint density \(f_{Y_t, Y_{t+j_1}, \ldots, Y_{t+j_n}}(y_t, y_{t+j_1}, \ldots, y_{t+j_n})\) is Gaussian for any \(j_1, j_2, \ldots, j_n\). There are many processes which can describe the dynamics of a time series. The Autoregressive Moving Average (ARMA) process of order \((p, q)\) is a common one which has been well-developed in the literature. The process is defined as

\[ Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}. \]  
(1.5)

If the time series of interest is not covariance-stationary, then the assumptions for using the ARMA model are not met. A series that has variance which is not constant over time can be described by an Autoregressive Conditional Heteroscedastic (ARCH) process of some order \(m\):

\[ u_t^2 = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 + \omega_t. \]  
(1.6)
Many problems of interest involve more than one variable. Thus we are interested not only in the dynamics of one variable, but the dynamics of multiple variables and the relationships between them. This motivated the development of multivariate time series analysis. We define a multivariate time series as follows: Suppose we have a collection of univariate time series observed at equally spaced time intervals. Each of the \( k \) series is denoted as

\[
\{y_{1t}, \ldots, y_{kt}\}, \quad t = 0, 1, 2, \ldots, T. \tag{1.7}
\]

The \( k \)-dimensional vector time series will be denoted as

\[
y_t = (y_{1t}, \ldots, y_{kt})'. \tag{1.8}
\]

Multivariate time series analysis is a heavily researched area with many developed tools that parallel tools used in univariate time series analysis. There exist many alternative models as well as strategies for building them, estimating them, and performing diagnostic checks for the misspecification of them. Some of the literature encompassing these developments include Peña et al. (2001), Quenouille (1957), Whittle (1963), Hannan (1970), Zellner and Palm (1974), Brillinger (1975), Dunsmuir and Hannan (1976), Box and Haugh (1977), Parzen (1977), Wallis (1977), Chan and Wallis (1978), Deistler et al. (1978), Hallin (1978), Jenkins (1979), Hsiao (1979), Akaike (1980), Hannan (1980), Hannan et al. (1980), Quinn (1980), and Granger and Newbold (1986). Most of these contributions rely on the assumption that either 1) the joint distribution of the components in the vector time series is multivariate
normal, or 2) each component in the series has a continuous probability distribution function.

The problems of interest in this research involve one or more violations of the previous assumptions throughout this introduction. One problem involves a multivariate time series in which multiple data types (both continuous and categorical) are observed and the dynamics of the time series involve nonlinearities. Another problem involves a large scale multivariate time series in which the assumption of second-order stationarity is not valid.

1.2 Multivariate Time Series Analysis with Categorical and Continuous Variables in an LSTR model

When a multivariate time series consists of mixed data types, we cannot use most multivariate time series models since they assume a multivariate normal distribution or some other continuous distribution. However, work has been done which relaxes these assumptions. We review nine classes of models in this section.

There are approaches that allow analysis of multivariate time series in which every component has a discrete distribution. A Bayesian model for multivariate time series of count observations was introduced in Ord et al. (1993). Their model is described as follows: Suppose there are $m$ series of count data observations with $y_{it}$ denoting
the number in the $i$th series at time $t$ for $i = 1, \ldots, m$ and $t = 1, \ldots, N$. Let $y_t$ be the sum of all of the series at time $t$:

$$y_t = \sum_{i=1}^{m} y_{it}, \quad t = 1, \ldots, N.$$  

Ord et al. (1993) model $y_t$ with a Poisson-gamma model:

$$p(y_t | \mu_t) = \frac{\mu_t^y e^{-\mu_t}}{y_t!},$$

$$p(\mu | a, b) = \frac{e^{-b\mu} \mu^{a-1} b^a}{\Gamma(a)}, \quad a, b > 0.$$  

For a given value of $y_t$, the split into individual series is modeled by a multinomial-Dirichlet scheme where the $y_{it}$ given the parameters have a multinomial distribution. The conjugate prior for the multinomial distribution is a Dirichlet (multivariate beta) distribution. This approach can also be expanded to include explanatory variables in the Poisson model.

Another approach for analyzing multivariate categorical time series was given in van Buuren (1996). He adapts the classic linear autoregressive model to categorical data by means of optimal scaling. Recall an autoregressive model of order $p$ as

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \ldots + \phi_p u_{t-p} + \epsilon_t,$$

where $E \left[ (\sum_p \phi_p u_{t-p}) \epsilon_t \right] = 0$ and $\epsilon_t$ is a serially uncorrelated normal random variable with zero mean. If the data $u_t$ are transformed by a function $f()$ as $x_t = f(u_t)$ then we can model the transformed data with an autoregressive model:

$$f(u_t) = \phi_1 f(u_{t-1}) + \phi_2 f(u_{t-2}) + \ldots + \phi_p f(u_{t-p}) + \epsilon_t,$$
where $\epsilon_t$ is white noise. The class of transformations used is $x_t = g_t y$ where $g_t$ is a time varying binary row vector of length $m$ that indicates in which of the $m$ categories each $u_t$ falls, and $y = [y_1, \ldots, y_m]'$ is a column vector containing unknown scaling weights, or category quantifications. This model can be generalized to the multivariate case.

A methodology for time series analysis of non-Gaussian observations based on state space models is introduced in Durbin and Koopman (2000). They provide both classical and Bayesian approaches to inference. Their methodology is based on simulation using importance sampling and antithetic variables and uses the Kalman filter and smoother. They do not rely on Markov chains, but instead on independent samples, which allows them to avoid convergence problems and to obtain accurate variance estimates. Their model is as follows:

$$y_t = Z_t \alpha_t + \epsilon_t, \quad \epsilon_t \sim p(\epsilon_t) \quad (1.9)$$

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t, \quad \eta_t \sim p(\eta_t) \quad (1.10)$$

for $t = 1, \ldots, N$. Equation 1.9 is called the observation equation and equation 1.10 is called the state equation of the state space model. The vector of observations is $y_t$ and the unobserved $m \times 1$ state vector is $\alpha_t$. The method for classical inference uses common simulation methodology. Let $\alpha = (\alpha'_1, \ldots, \alpha'_N)'$ and $y = (y'_1, \ldots, y'_N)'$. Suppose we are interested in estimating the conditional mean of an arbitrary function $x(\alpha)$ given the observation vector $y$:

$$\bar{x} = E[x(\alpha)|y].$$
For example, we might be interested in 1) the mean of the state vector $\mathbf{x}_t$ given $\mathbf{y}$ and its conditional variance matrix, or 2) estimates of the conditional distribution function of $\mathbf{x}(\mathbf{\alpha})$. The estimation of $\mathbf{\bar{x}}$ is performed by simulation methods similar to those used in Shephard and Pitt (1997) and Durbin and Koopman (1997) for estimating the likelihood in non-Gaussian state space models.

A final approach by Jorgensen et al. (1999) consists of a state space model for multivariate longitudinal count data. Their model is driven by a latent gamma Markov process. In this setting, denote the $m$-dimensional vector of Poisson counts as $\mathbf{Y}_t$, and assume that all components of the vector reflect the same underlying tendency $\theta_t$. Their approach is one of log-linear regression based on a state space model for $\mathbf{Y}_t$ where $\theta_t$ is a latent gamma Markov model. That is,

$$Y_{it}|\theta_t \sim \text{Poisson} (a_{it}, \theta_t)$$

with $a_{it} = e^{\mathbf{x}_t^T \mathbf{\alpha}_t}$. Note that $\mathbf{x}_t$ is a vector of time-varying covariates and $\mathbf{\alpha}_t$ are the regression parameters that may vary among the $m$ categories. The latent process is defined by the transition distribution

$$\theta_t|\theta_{t-1} \sim \text{Gamma} \left( b_t \theta_{t-1}, \frac{\alpha}{\theta_{t-1}} \right).$$

Estimation of the model is performed based on the Kalman smoother.

The analysis of multivariate time series with mixed data types (both continuous and categorical data) has also received investigation. State-space models for multivariate longitudinal data of mixed types were introduced by Jorgensen et al. (1996).
The components of the response vector could follow any of the Tweedie exponential dispersion models, which include a variety of discrete and continuous distributions. Their model was very similar to a later paper by the same authors (Jorgensen et al. (1999)) which is described above. The latent process is assumed to be a Markov process, and the estimation is carried out using the Kalman filter. Another reference which presents similar material is Jorgensen and Tsao (1999).

Andrieu et al. (2003) investigates time series involving continuous and discrete variables by using a particle filtering method to perform optimal estimation in Jump Markov (Nonlinear) Systems (JMS). Instead of continuous state space hidden Markov models, JMS combine discrete and continuous state spaces in a hierarchical manner. Let \( \{ r_t \} \) be a stationary, finite, discrete, first order homogeneous Markov chain with values in a set \( S \) with transition probabilities

\[
\pi_{ij} \equiv \Pr\{ r_{t+1} = j | r_t = i \}, \quad (i, j \in S).
\]

If \( s \) is the finite number of elements in \( S \), then we consider a family of \( s^2 \) densities \( \{ f_{ij}(\mathbf{x}'|\mathbf{x}) \} \) where \( \mathbf{x} \in \mathbb{R}^{n_x} \) and \( \mathbf{x}' \in \mathbb{R}^{n_x'} \). We define the state transition conditional densities as

\[
p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, r_{1:t}) = f_{r_{t-1}r_t}(\mathbf{x}_t|\mathbf{x}_{t-1}).
\]

Note that for a set of variables \( l_\mathbf{t}, l_{ab} \equiv \{ l_a, l_{a+1}, \ldots, l_b \} \). We observe \( \{ y_t \} \) (not \( \{ r_t \} \) or \( \{ x_t \} \)) where

\[
p(\mathbf{y}_t|\mathbf{x}_{0:t}, r_{1:t}, y_{1:t-1}) = g_{r_t}(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \mathbf{x}_t)
\]
with \( y_t \in \mathbb{R}^{n_y} \). The class of processes known as Jump Markov Linear Models (JMLS) are defined as:

\[
x_t = A(r_t) x_{t-1} + B(r_t) v_t, \quad x_0 \sim N(m_0, P_0),
\]

\[
y_t = C(r_t) x_t + D(r_t) \epsilon_t
\]

where \( v_t \sim N(0, I_{n_v}) \) and \( \epsilon_t \sim N(0, I_{n_\epsilon}) \) are mutually independent sequences of independent and identically distributed Gaussian random variables. The class of processes studied in this paper is a generalization of the JMLS class where the above linearity and Gaussianity assumptions (which are unrealistic in many applications) are relaxed. Estimation is carried out through a combination of the Auxiliary Particle Filter (Pitt and Shephard (1999)) and the Unscented Kalman Filter (Julier and Uhlmann (1997)). One of the models in this class is the Time-Varying Autoregressive model with time-varying model order for which they motivate their study.

Linear time series models were considered adequate for most applications until computing power became more advanced and some applications encountered limitations with linear models (Peña et al. (2001)). Examples of parametric models developed in nonlinear time series analysis are bilinear models (Granger and Andersen (1978)), threshold autoregressive models (Tong (1978), Tong (1990)), exponential autoregressive models (Haggan and Ozaki (1981)), state-dependent models (Priestly (1980)), Markov switching models (Hamilton (1989)), and the ARCH model (Engle (1982)).
A specific class of nonlinear multivariate time series models of interest is the smooth transition regression (STR) model. The following description is taken from Granger and Terasvirta (1993). It can be written as:

\[ y_t = \alpha_0 + \alpha_1 x_{t-1} + (\beta_0 + \beta_1 x_{t-1}) F x_{t-1} - \mu + \epsilon_t \]  

(1.11)

where \(\epsilon_t\) is i.i.d., \(F(x)\) is a continuous function (even or odd), \(x_t\) is an explanatory variable, \(\alpha\) is the time delay, and \(x_{t-1}\) is the transition variable. Some examples of \(F(x)\) are the cumulative distribution function of a \(N(\mu, \sigma^2)\) variable and the density function of a \(N(\mu, \sigma^2)\) variable. If \(F\) is odd and monotonically increasing, then if \(|x_{t-1} - \mu|\) is large with \(x_{t-1} < \mu\), \(y_t\) is effectively generated by the following model:

\[ y_t = \alpha_0 + \alpha_1 x_{t-1} + \epsilon_t. \]

If \(|x_{t-1} - \mu|\) is large with \(x_{t-1} > \mu\) instead, \(y_t\) is effectively generated by the following model:

\[ y_t = (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1) x_{t-1} + \epsilon_t. \]

Values of \(x_{t-1}\) closer to \(\mu\) give mixtures of these values.

Special cases of this model are equivalent to previous nonlinear models. If \(\sigma^2 \to 0\), \(F\) becomes the Heaviside function: \(F = 0\) for \(x_{t-1} < \mu\) and \(F = 1\) for \(x_{t-1} \geq \mu\), and Equation 1.11 is a switching-regression model (Quandt (1983), Goldfeld and Quandt (1973)) with switching variable \(x_{t-1}\). If the switching variable is instead \(y_{t-1}\), then Equation 1.11 becomes a Smooth-Transition Autoregressive (STAR) model (Chan and Tong (1986), Luukkonen et al. (1988), Terasvirta (1990), Terasvirta et al. (1993)).
When $\sigma^2 \to 0$ and $y_{t-1}$ is the switching variable, then Equation 1.11 becomes the two-regime (single threshold) Threshold Autoregressive (TAR) model of Tong (1990). If one assumes

$$F(z) = 1 - e^{-z^2}$$

then Equation 1.11 is a multivariate generalization of the exponential autoregressive (EAR) model Haggan and Ozaki (1981). Note that we can obtain a more general model in Equation 1.11 if $F(x_{t-1})$ is replaced by $F(x_{t-d})$ for some $d > 0$.

An even more general model uses a vector $w_t$ of $v$ explanatory variables

$$w_t' = (y_{t-1}, \ldots, y_{t-p}, x_{1t}, \ldots, x_{kt}),$$

and a model

$$y_t = \alpha_0 + \alpha'w_t + (\beta_0 + \beta'w_t) F(z_t) + e_t \tag{1.12}$$

where

$$z_t = \gamma (\delta'w_t - c), \quad \delta = (\delta_1, \ldots, \delta_v)', \quad v = p + k$$

and $e_t$ is a zero mean sequence of independent variables. When we assume that $F$ has the form

$$F(z) = \frac{1}{1+e^{-z}}$$

then Equation 1.11 becomes the logistic smooth transition regression (LSTR) model. The coefficient $\gamma$ is the smoothness parameter and the $\delta_j$ are normalized so that their sum is one. Estimating Equation 1.12 becomes difficult when $p$ is not known
or if the number of parameters to be estimated is too large given the sample size. A constrained model which is used often in practice considers indicator functions of only one transition variable (Granger and Terasvirta (1993)),

$$z_t = \gamma (x_{t-d} - c),$$

(1.13)

or

$$z_t = \gamma (y_{t-d} - c).$$

(1.14)

Applying the LSTR model involves 1) testing linearity against STR 2) specifying the model 3) estimating the parameters, and 4) evaluating the model. These details are discussed in Granger and Terasvirta (1993).

The transition variables in Equations 1.13 and 1.14 and in Quandt (1983) and Goldfeld and Quandt (1973) are assumed to be observed. If the switching variable is assumed to be independent of other variables and unobservable, then the switching-regression model with two regimes becomes:

$$y_t = \alpha_0 + \alpha'w_t + (\beta_0 + \beta'w_t) s_t + e_t$$

where $s_t = 0$ or $1$, $E[e_t] = 0$, $\text{var}(e_t|s_t) = \sigma_0^2 + s_t\sigma_1^2 > 0$, and $s_t$ is generated by a two-state Markov chain with constant transition probabilities over time Goldfeld and Quandt (1973).

In Chapter 2, we explore an alternative model formulation for mixed continuous and categorical time series data.
1.3 Evolving Structure in Multivariate Time Series with Application to Stock Sector Data

Modeling the variance of an asset return is an important step in estimating how much risk a particular asset carries. The seminal paper of Engle (1982) which introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model allowed researchers to obtain such an estimate. Since then, many more developments in modeling heteroscedasticity have been made. It is widely accepted that financial volatilities are correlated across assets and markets, and many models have been developed which make use of this fact. The multivariate Generalized Autoregressive Conditional Heteroscedasticity (MGARCH) model is one for studying the relationships between the volatilities and co-volatilities of several markets. A good survey on MGARCH models is provided in Bauwens et al. (2004). Some important MGARCH models that have been developed include those introduced in Bollerslev et al. (1988), Engle and Kroner (1995), Engle et al. (1990), Vrontos et al. (2003), Kariya (1988), Alexander (1997), Bollerslev (1990), Christodoulakis and Satchett (2002), Engle (2002), Tse (2002), and Klaassen (1999). Appendix 1 provides further detail on these previously developed MGARCH models. An overview of the MGARCH model is provided in the following equations. Suppose we have a vector stochastic process \( \{y_t\} \) with dimension \( K \times 1 \). We model this process as:

\[
y_t = \mu_t(\theta) + \epsilon_t,
\]
where $\mu_t(\theta)$ is the mean vector conditional on the information up to time $t$ and

$$\epsilon_t = H_t^{1/2}(\theta) z_t,$$

where $H_t^{1/2}(\theta)$ is a $K \times K$ positive definite matrix. Also, $z_t$ has the following two properties:

$$E[z_t] = 0$$

$$\text{Var}[z_t] = I_K,$$

where $I_K$ is the identity matrix of dimension $K \times K$. The matrix $H_t^{1/2}$ is any $K \times K$ positive definite matrix such that $H_t$ is the covariance matrix of $y_t$ conditional on the information up to time $t$.

We define a new strategy for examining changing multivariate covariance structure in Chapter 3 of this thesis.
Chapter 2

Multivariate Time Series Analysis with Categorical and Continuous Variables in an LSTR model

2.1 Introduction

Our contribution involves the study of an LSTR model whose transition is related to a categorical variable. We make use of some recent work by Fokianos and Kedem (2003) on categorical time series. Suppose we have a categorical time series \( \{Y_t\}, t = 1, \ldots, N \) and let \( m \) be the number of categories. We can express the \( t \)th observation by the vector \( Y_t = (Y_{t1}, \ldots, Y_{tq}) \) of length \( q = m - 1 \) with elements
\[ Y_{tj} = \begin{cases} 1, & \text{if the } j\text{th category is observed at time } t, \\ 0, & \text{otherwise,} \end{cases} \]

for \( t = 1, \ldots, N \) and \( j = 1, \ldots, q \). Denote the vector of conditional probabilities given \( \mathcal{F}_{t-1} \) as \( \pi_t = (\pi_{t1}, \ldots, \pi_{tq})' \) where

\[ \pi_{tj} = E[Y_{tj}|\mathcal{F}_{t-1}] = P(Y_{tj} = 1|\mathcal{F}_{t-1}), \quad j = 1, \ldots, q. \]

The \( \sigma \)-field \( \mathcal{F}_{t-1} \) is generated by \( Y_{t-1}, Y_{t-2}, \ldots, Z_{t-1}, Z_{t-2}, \ldots \) where \( \{Z_{t-1}\}, \; t = 1, \ldots, N \) is a \( p \times q \) matrix that represents a covariate process. Using the theory of generalized linear models (McCullagh and Nelder (1989)) we assume that the vector of transition probabilities is linked to the covariate process through the equation

\[ \pi_t (\beta) = h (Z_{t-1}'\beta), \quad (2.1) \]

with \( \beta \) a \( p \)-dimensional vector of time-invariant parameters.

We assume our transition variable for the LSTR model is an ordinal categorical variable. For this type of variable, we use the cumulative odds model of Snell (1964) and McCullagh (1980). We assume the observed data results from the following:

\[ X_t = -\gamma'z_{t-1} + e_t \quad (2.2) \]

where \( e_t \) is a sequence of i.i.d. random variables with continuous c.d.f. \( F \), \( \gamma \) is a \( d \)-dimensional vector of parameters and \( z_{t-1} \) is a \( d \)-dimensional covariate vector. The latent process \( \{X_t\} \) may or may not be observed. The categorical time series that results is \( \{Y_t\}, \; t = 1, \ldots, N \) where

\[ Y_t = j \iff Y_{tj} = 1 \iff \theta_{j-1} \leq X_t < \theta_j \quad (2.3) \]
for $j = 1, \ldots, m$. The set of threshold parameters $\{\theta_0, \theta_1, \ldots, \theta_m\}$ satisfies

$$-\infty = \theta_0 < \theta_1 < \cdots < \theta_m = \infty.$$  

Then

$$\pi_{tj} = P(\theta_{j-1} \leq X_t < \theta_j | \mathcal{F}_{t-1})$$

$$= F(\theta_j + \gamma'z_{t-1}) - F(\theta_{j-1} + \gamma'z_{t-1})$$

for $j = 1, \ldots, m$, or

$$P(Y_t \leq j | \mathcal{F}_{t-1}) = F(\theta_j + \gamma'z_{t-1}), \quad j = 1, \ldots, m.$$  

We use the cumulative logistic or proportional odds model where $F$ is the logistic distribution function,

$$F_l(x) = \frac{1}{1+e^{-x}},$$

which makes

$$\log\left\{\frac{P(Y_t \leq j | \mathcal{F}_{t-1})}{P(Y_t > j | \mathcal{F}_{t-1})}\right\} = \theta_j + \gamma'z_{t-1}$$

for $j = 1, \ldots, q$.

Thus, we can define our model in Equation 2.1 more explicitly by stating the following definitions:

$$\beta = (\theta_1, \ldots, \theta_q, \gamma')',$$  \hspace{1cm} (2.4)

$Z_{t-1}$ as the $(q + d) \times q$ matrix
\[
Z_{t-1} = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 \\
z_{t-1} & z_{t-1} & \cdots & z_{t-1}
\end{bmatrix}
\]

\[
h = (h_1, \ldots, h_q)',
\]

\[
\eta_t = (\eta_{t1}, \ldots, \eta_{tq})' = Z_{t-1}' \beta,
\]

\[
\pi_{t1} (\beta) = h_{1} (\eta_t) = F(\eta_{t1}),
\]

\[
\pi_{tj} (\beta) = h_{j} (\eta_t) = F(\eta_{tj}) - F(\eta_{t(j-1)}), \quad j = 2, \ldots, q.
\]

### 2.2 Model Definition

We define a specific model from the more general class of models described above. Suppose we have a proportional odds model with a periodic component with 3 categories resulting from a latent process which also generates our LSTR model. That is,

\[
\log \left\{ \frac{P(Y_t \leq 1 | \chi_{t-1})}{P(Y_t > 1 | \chi_{t-1})} \right\} = \theta_1 + \gamma'z_{t-1} = \theta_1 + \gamma_1 \cos \left( \frac{2\pi t}{12} \right) + \gamma_2 Y_{(t-1)1} + \gamma_3 Y_{(t-1)2}
\]

and
\[
\log \left\{ \frac{P[Y_{t-1} \leq 2]}{P[Y_{t-1} > 2]} \right\} = \theta_2 + \gamma' \mathbf{z}_{t-1} = \theta_2 + \gamma_1 \cos \left( \frac{2\pi t}{12} \right) + \gamma_2 Y_{t(1)-1} + \gamma_3 Y_{t(2)-1}
\]

with

\[
X_t = \gamma' \mathbf{z}_{t-1} + \epsilon_t = \gamma_1 \cos \left( \frac{2\pi t}{12} \right) + \gamma_2 Y_{t(1)-1} + \gamma_3 Y_{t(2)-1} + \epsilon_t
\]  \tag{2.5}

and

\[
y_t = \alpha_0 + \alpha_1 y_{t-1} + (\beta_0 + \beta_1 y_{t-1}) \frac{1}{1 + e^{-X_{t-1}}} + \epsilon_t.
\]  \tag{2.6}

### 2.3 Estimation

The estimation procedure we propose for this model is as follows. We estimate the ordinal time series parameters $\mathbf{\beta}$ as in Equation 2.4 using the partial likelihood approach of Fokianos and Kedem (2003). The partial likelihood function relative to $\mathbf{\beta}$, $\mathcal{F}_t$, and the data is given by (Fahrmeir and Kaufmann (1987), Kaufmann (1987), Fokianos and Kedem (1998)):

\[
PL(\mathbf{\beta}) = \prod_{t=1}^{N} f \left( y_t; \mathbf{\beta} \middle| \mathcal{F}_{t-1} \right) = \prod_{t=1}^{N} \prod_{j=1}^{m} \pi_{tj}(\mathbf{\beta})^{y_{tj}},
\]

so that the partial log-likelihood is given by

\[
l(\mathbf{\beta}) = \text{log}PL(\mathbf{\beta}) = \sum_{t=1}^{N} \sum_{j=1}^{m} y_{tj} \log \pi_{tj}(\mathbf{\beta}).
\]  \tag{2.7}

We obtain the maximum partial likelihood estimator $\hat{\mathbf{\beta}}$ by maximizing the partial log-likelihood in Equation 2.7. If the maximum exists it is given as the solution of the partial score equations

\[
\nabla l(\mathbf{\beta}) = \nabla \text{log}PL(\mathbf{\beta}) = 0
\]
assuming differentiability. This solution is obtained by Fisher scoring. When we
differentiate Equation 2.7 we obtain the partial score

\[ S_N (\beta) = \nabla l (\beta) = \left( \frac{\partial l (\beta)}{\partial \beta_1}, \ldots, \frac{\partial l (\beta)}{\partial \beta_p} \right)' = \sum_{t=1}^{N} Z_{t-1} D_t (\beta) \Sigma_t^{-1} (\beta) (Y_t - \pi_t (\beta)), \]

(2.8)

where

\[ D_t (\beta) = \left[ \frac{\partial h(t)}{\partial m} \right] \]

and \( \Sigma_t (\beta) \) is the conditional covariance matrix of \( Y_t \) with elements

\[ \sigma_t^{(ij)} (\beta) = \begin{cases} -\pi_{ti} (\beta) \pi_{tj} (\beta), & \text{if } i \neq j, \\ \pi_{ti} (\beta) (1 - \pi_{ti} (\beta)), & \text{if } i = j, \end{cases} \]

for \( i, j = 1, \ldots, q \). Also, we have defined the dimension of \( \beta \) as \((p \times 1)\) where \( p = q + d \).

We solve Equation 2.8 using the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) optimization routine. This is a quasi-Newton optimization method (also known as a variable metric algorithm) that uses function values and gradients to build up a picture of the surface to be optimized. (Nocedal and Wright (1999), R Development Core Team (2003), Byrd et al. (1995)). Our first-stage parameter estimates for \( \beta \), denoted as \( \hat{\beta}^{(1)} = \left( \hat{\theta}_1^{(1)}, \ldots, \hat{\theta}_q^{(1)}, \hat{\gamma}^{(1)} \right) \), are used to compute estimates for Equations 2.2 and 2.3:

\[ \hat{X}_t^{(1)} = -\hat{\gamma}^{(1)}' z_{t-1} \]

(2.9)

\[ \hat{Y}_t^{(1)} = j \iff \hat{Y}_{ij}^{(1)} = 1 \iff \hat{\theta}_{j-1}^{(1)} \leq \hat{X}_t^{(1)} < \hat{\theta}_j^{(1)}. \]

(2.10)

Our parameter estimates for the LSTR model in Equation 2.6 are obtained using a nonlinear least squares approach. We minimize the following equation using the
BFGS quasi-Newton optimization method:

\[
\sum_{t=2}^{N} \left[ y_t - \left( \alpha_0 + \alpha_1 y_{t-1} + (\beta_0 + \beta_1 y_{t-1}) \frac{1}{1 + e^{-x_{t}^{(1)}}} \right) \right]^2.
\] (2.11)

We will denote these estimates as \( (\hat{\alpha}_0^{(1)}, \hat{\alpha}_1^{(1)}, \hat{\beta}_0^{(1)}, \hat{\beta}_1^{(1)}) \). We estimate \( y_t \) as follows:

\[
\hat{y}_t^{(1)} = \hat{\alpha}_0^{(1)} + \hat{\alpha}_1^{(1)} y_{t-1} + (\hat{\beta}_0^{(1)} + \hat{\beta}_1^{(1)} y_{t-1}) \frac{1}{1 + e^{-\tilde{x}_t^{(1)}}}.
\] (2.12)

Once these initial estimates are obtained, we begin the iterative estimation procedure. Our second-stage estimate for \( X_t \) is obtained using our observed \( y_t \) and the first-stage estimates: \( (\hat{\alpha}_0^{(1)}, \hat{\alpha}_1^{(1)}, \hat{\beta}_0^{(1)}, \hat{\beta}_1^{(1)}) \). That is, we rearrange Equation 2.6 and plug in our first-stage estimates:

\[
\hat{X}_{t-1}^{(2)} = -\log \left( \frac{\hat{\beta}_0^{(1)} + \hat{\beta}_1^{(1)} y_{t-1}}{y_t - \hat{\alpha}_0^{(1)} - \hat{\alpha}_1^{(1)} y_{t-1}} - 1 \right).
\] (2.13)

Using \( \hat{X}_{t-1}^{(2)} \) and our threshold parameter estimates, \( \{\hat{\theta}_0^{(1)}, \hat{\theta}_1^{(1)}, \ldots, \hat{\theta}_m^{(1)}\} \), we update our estimate of \( Y_t \) using Equation 2.3:

\[
\hat{Y}_t^{(2)} = j \iff \hat{Y}_{tj}^{(2)} = 1 \iff \hat{\theta}^{(1)}_{j-1} \leq \hat{X}_{t-1}^{(2)} < \hat{\theta}^{(1)}_j.
\] (2.14)

Subsequently, all of the terms involved in Equation 2.8 are computed using \( \hat{Y}_t^{(2)} \) instead of \( Y_t \) and the equation is solved again using the BFGS quasi-Newton optimization method. We will denote these estimates as \( \hat{\beta}^{(2)} = (\hat{\beta}_1^{(2)}, \ldots, \hat{\beta}_q^{(2)}, \hat{\gamma}^{(2)}) \). Next we update our estimate of \( X_t \) using Equation 2.5:

\[
\hat{X}_t^{(3)} = \hat{\gamma}(2)^{\top} z_{t-1} = \gamma_1^{(2)} \cos \left( \frac{2\pi t}{12} \right) + \gamma_2^{(2)} Y_{(t-1)1} + \gamma_3^{(2)} Y_{(t-1)2}.
\] (2.15)
With $\hat{X}_t^{(3)}$ and our threshold parameter estimates \{\hat{\theta}_0^{(2)}, \hat{\theta}_1^{(2)}, \ldots, \hat{\theta}_m^{(2)}\} we compute $\hat{Y}_t^{(3)}$ using Equation 2.3:

$$\hat{Y}_t^{(3)} = j \iff \hat{Y}_{tj}^{(3)} = 1 \iff \hat{\theta}_{j-1}^{(2)} \leq \hat{X}_t^{(3)} < \hat{\theta}_j^{(2)}.$$  (2.16)

Finally, we minimize Equation 2.11 using the BFGS quasi-Newton optimization method after updating our estimate of $X_t$:

$$\sum_{i=2}^{N} \left[ y_t - \left( \alpha_0 + \alpha_1 y_{t-1} + (\beta_0 + \beta_1 y_{t-1}) \frac{1}{1 + e^{-X_i^{(3)}}} \right) \right]^2.$$  (2.17)

These second-stage estimates are denoted as $\hat{\theta}_0^{(2)}, \hat{\theta}_1^{(2)}, \hat{\theta}_0^{(2)}, \hat{\theta}_1^{(2)}$. The updated estimate of $y_t$ is computed as:

$$\hat{y}_t^{(2)} = \hat{\alpha}_0^{(2)} + \hat{\alpha}_1^{(2)} y_{t-1} + (\hat{\beta}_0^{(2)} + \hat{\beta}_1^{(2)} y_{t-1}) \frac{1}{1 + e^{-X_i^{(3)}}}.$$  (2.18)

Equations 2.13 - 2.18 comprise the first iteration in this estimation procedure. Subsequent iterations will produce new parameter estimates. Iteration $i$ will produce $\left(\hat{\alpha}_0^{(i+1)}, \hat{\alpha}_1^{(i+1)}, \hat{\beta}_0^{(i+1)}, \hat{\beta}_1^{(i+1)}, \ldots, \hat{\theta}_q^{(i+1)}, \hat{\gamma}^{(i+1)}\right)$. We continue iterations of this procedure until convergence occurs, i.e. the maximum difference between parameter estimates in iterations $i$ and $i+1$ is small.

### 2.4 Simulation Results

A simulation study is conducted in order to assess the effectiveness of the new estimation procedure introduced in the Estimation section. To simulate a realization from an LSTR model with categorical transitional variable, we first simulate
the latent process \( \{X_t\} \) as in Equation 2.5 and generate the categorical time series \( \{Y_t\} \). We use a time series length of \( N = 500 \) and parameters \( \theta = (-0.2, 0.2) \) and \( \gamma = (0.2, -0.3, 0.5) \). The LSTR process \( y_t \) as in Equation 2.6 is simulated using the parameters \((\alpha_0, \alpha_1, \beta_0, \beta_1) = (0.2, 0.3, 0.25, 0.1)\). The error processes \( e_t \) and \( u_t \) in Equations 2.5 and 2.6 are generated from a Normal(0,0.1) distribution. Thus, our model can be defined as

\[
X_t = 0.2\cos\left(\frac{2\pi t}{12}\right) - 0.3Y_{(t-1)1} + 0.5Y_{(t-1)2} + e_t
\]

\[
Y_t = 1 \iff Y_{t1} = 1, Y_{t2} = 0 \iff -\infty \leq X_t < -0.2
\]

\[
Y_t = 2 \iff Y_{t1} = 0, Y_{t2} = 1 \iff -0.2 \leq X_t < 0.2
\]

\[
Y_t = 3 \iff Y_{t1} = 0, Y_{t2} = 0 \iff 0.2 \leq X_t < \infty
\]

\[
\log \left\{ \frac{P[Y_t \leq 1|X_{t-1}]}{P[Y_t > 1|X_{t-1}]} \right\} = -0.2 + 0.2\cos\left(\frac{2\pi t}{12}\right) - 0.3Y_{(t-1)1} + 0.5Y_{(t-1)2}
\]

\[
\log \left\{ \frac{P[Y_t \leq 2|X_{t-1}]}{P[Y_t > 2|X_{t-1}]} \right\} = 0.2 + 0.2\cos\left(\frac{2\pi t}{12}\right) - 0.3Y_{(t-1)1} + 0.5Y_{(t-1)2}
\]

\[
y_t = 0.2 + 0.3y_{t-1} + (0.25 + 0.1y_{t-1})\frac{1}{1+e^{-X_{t-1}}} + u_t
\]

We simulate 100 realizations from this process and perform the estimation procedure described in the Estimation section. Table 2.1 details the results from the simulation study.

We see that the parameter estimates from the iterative estimation are more accurate and precise than those from the first-stage estimation. This is expected since
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>First-stage Estimate</th>
<th>First-stage Std. Error</th>
<th>Final Estimate</th>
<th>Final Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>-0.2</td>
<td>-0.329</td>
<td>0.0749</td>
<td>-0.236</td>
<td>0.0293</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.2</td>
<td>0.387</td>
<td>0.109</td>
<td>0.245</td>
<td>0.0339</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.2</td>
<td>0.303</td>
<td>0.0643</td>
<td>0.188</td>
<td>0.0184</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.3</td>
<td>-0.246</td>
<td>0.102</td>
<td>-0.350</td>
<td>0.0766</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.5</td>
<td>0.270</td>
<td>0.106</td>
<td>0.409</td>
<td>0.0740</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.2</td>
<td>0.198</td>
<td>0.00223</td>
<td>0.199</td>
<td>0.00219</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.3</td>
<td>0.297</td>
<td>0.00295</td>
<td>0.298</td>
<td>0.00274</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.25</td>
<td>0.248</td>
<td>0.00258</td>
<td>0.248</td>
<td>0.00245</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.1</td>
<td>0.0994</td>
<td>0.00162</td>
<td>0.0996</td>
<td>0.00176</td>
</tr>
</tbody>
</table>

Table 2.1: Simulation Study of Parameter Estimates based on 100 replications of series of length 500

we are using both observed series $y_t$ and $Y_t$ in the estimation of $y_t$ for the iterative estimation. The average number of iterations until convergence (when the maximum difference between successive parameter estimates is less than 0.00001) is 2.47. Figure 2.1 demonstrates the improved accuracy of the iterative estimate over the initial estimate of our latent process $X_t$. Figure 2.2 further supports the increased accuracy of our estimator. Figure 2.3 illustrates the performance of the estimation for the LSTR process $y_t$. Note that we do not get much improvement from $\hat{y}_t^{(1)}$ to $\hat{y}_t^{(2)}$. It is impor-
tant to note that the major importance in this contribution is not the improvement in fit for $y_t$, but the ability of estimating the model at all when the transition involves a categorical variable.

2.5 Application of Methodology

Smooth transition autoregressive models have been successfully applied to describe the behavior of macroeconomic time series such as unemployment at different phases of the business cycle (van Dijk et al. (2002)). We propose a smooth transition regression model for unemployment that is not autoregressive but instead makes use of a categorical variable (state of the economy) and compare it with a smooth transition autoregressive model which was the final model selected in van Dijk et al. (2002).

The unemployment data used is the same as in van Dijk et al. (2002) and consists of the seasonally unadjusted unemployment rate among US males aged 20 and over, at the monthly frequency covering the period January 1968 until December 2004. We construct the series by computing the ratio of the unemployment level and civilian labor force of this population group, which are obtained from the Bureau of Labor Statistics. The series is shown in Figure 2.4, and we will denote this series as $y_t$.

As mentioned in van Dijk et al. (2002), the series contains three dominant features: business cycle asymmetry, large persistence, and a distinct seasonal pattern. A nonlinear model is useful for modeling this series since the unemployment rate steeply increases during recessions and declines at a slower rate during expansions.
In fact, many other contributions have provided evidence for nonlinear behavior in this series (Hansen (1997), Bianchi and Zoega (1998), Montgomery et al. (1998), Rothman (1998), Brannas and Ohlsson (1999), Koop and Potter (1999), Caner and Hansen (2001), Skalin and Terasvirta (2002)). The persistence of the series has also been researched. Two competing hypotheses have been developed concerning this persistence. The “natural rate” hypothesis states that the unemployment rate is mean-reverting while the “hysteresis” hypothesis states that the unemployment rate is non-stationary. We follow the viewpoint of van Dijk et al. (2002) along with Bianchi and Zoega (1998) and Skalin and Terasvirta (2002) which is that the unemployment rate is globally stationary but possibly nonlinear and locally nonstationary. The third feature of seasonality is easy to see when examining Figure 2.4. In general, unemployment is above average during the winter and below average during the late summer and fall. In order to account for the seasonality, we use monthly dummy variables denoted as $D_{s,t}$, $s = 1, \ldots, 11$, where $D_{s,t} = 1$ if observation $t$ corresponds to month $s$ and $D_{s,t} = 0$ otherwise.

The categorical transition variable we use is the state of the economy. The data are obtained from the National Bureau of Economic Research (NBER)\footnote{http://www.nber.org/cycles.html/} and are coded as follows:
\[ Y_t = \begin{cases} 
1, & \text{recession,} \\
2, & \text{contraction,} \\
3, & \text{expansion} 
\end{cases} \]

\[ Y_{tj} = \begin{cases} 
1, & \text{if the } j\text{th category is observed at time } t, \\
0, & \text{otherwise}, 
\end{cases} \]

for \( j=1,2 \). The data are displayed in Figure 2.5.

According to the NBER, the relevant indicators for the state of the economy include the growing or declining rates in Gross Domestic Product, change in real income, change in the amount of industrial production, and the unemployment level. We use the following model for \( Y_{tj} \):

\[ X_t = \gamma_1 \Delta GDP_t + \gamma_2 Y_{t-1,1} + \gamma_3 Y_{t-1,2} + \epsilon_t \quad (2.19) \]

where \( \Delta GDP_t \) is the change in Gross Domestic Product, and

\[ Y_t = j \iff Y_{tj} = 1 \iff \theta_{j-1} \leq X_t < \theta_j. \]

We aim to compare the model fit between an LSTAR model (which uses an autoregressive transitional variable) and an LSTR whose transition is related to a categorical variable. Thus, we compare our model to the final LSTAR model estimated in van
Dijk et al. (2002):

\[
\Delta y_t = 0.479 + 0.645D_{1,t} - 0.342D_{2,t} - 0.680D_{3,t} - 0.725D_{4,t} \\
-0.649D_{5,t} - 0.317D_{6,t} - 0.410D_{7,t} - 0.501D_{8,t} - 0.554D_{9,t} \\
-0.306D_{10,t} + [ -0.040y_{t-1} - 0.146\Delta y_{t-1} - 0.101\Delta y_{t-6} \\
+0.097\Delta y_{t-8} - 0.123\Delta y_{t-10} + 0.129\Delta y_{t-13} - 0.103\Delta y_{t-15} ] \\
\times [1 - G(\Delta_{12}y_{t-1}; \hat{\gamma}, \hat{\varphi})] + [-0.011y_{t-1} + 0.225\Delta y_{t-1} \\
+0.307\Delta y_{t-2} - 0.119\Delta y_{t-7} - 0.155\Delta y_{t-13} - 0.215\Delta y_{t-14} \\
-0.235\Delta y_{t-15}] \times G(\Delta_{12}y_{t-1}; \hat{\gamma}, \hat{\varphi}) + \hat{\epsilon}_t
\]  

(2.20)

where

\[
G(\Delta_{12}y_{t-1}; \hat{\gamma}, \hat{\varphi}) = \frac{1}{1 + \exp(-23.15(\Delta_{12}y_{t-1} - 0.27)/\hat{\sigma}_{\Delta_{12}y_{t-1}})}.
\]

Using the estimation procedure we described above, we estimate the same model using \(X_t\) as in Equation 2.19. Our model is estimated as:

\[
\Delta y_t = 0.473 + 0.659D_{1,t} - 0.270D_{2,t} - 0.763D_{3,t} - 0.814D_{4,t} \\
-0.621D_{5,t} - 0.249D_{6,t} - 0.403D_{7,t} - 0.555D_{8,t} - 0.527D_{9,t} \\
-0.304D_{10,t} + [-0.053y_{t-1} - 0.112\Delta y_{t-1} - 0.048\Delta y_{t-6} \\
+0.093\Delta y_{t-8} - 0.037\Delta y_{t-10} + 0.062\Delta y_{t-13} - 0.034\Delta y_{t-15} ] \\
\times [1 - G(\tilde{X}_{t-1}; \gamma)] + [-0.001y_{t-1} + 0.279\Delta y_{t-1} \\
+0.132\Delta y_{t-2} - 0.068\Delta y_{t-7} - 0.172\Delta y_{t-13} - 0.264\Delta y_{t-14} \\
-0.111\Delta y_{t-15}] \times G(\tilde{X}_{t-1}; \gamma) + \hat{\epsilon}_t
\]  

(2.21)

where \(\{\tilde{X}_t\}\) is a shifted (mean=0) version of \(\hat{X}_t\):
\[ \hat{X}_t = 180.259 \Delta GDP_t - 47.470 Y_{t-1,1} - 85.406 Y_{t-1,2} \]

and

\[ G \left( \hat{X}_{t-1}; \gamma \right) = \frac{1}{1 + \exp \left( \hat{X}_{t-1} \right)} \quad (2.22) \]

Note that since van Dijk et al. (2002) only used the time period of 1968-1989 for model fitting, we use the same time period for comparison purposes. Figures 2.6 - 2.7 display the actual and fitted values for the state of the economy and level of unemployment.

The residual standard deviation for the LSTAR model in Equation 2.20 is 0.185. The residual standard deviation for the LSTR model in Equation 2.21 is 0.210. The ratio of the residual standard deviations is 1.135. That is, the residual standard deviation of the LSTAR model is about 13.5% smaller than that of the fitted LSTR model. The fitted values for the series are very close to the actual values in both models, however. Figure 2.8 displays the actual values along with the 95% confidence intervals for the fitted values.

Our next aim is to evaluate the forecasting performance of the estimated model. We use the same out-of-sample time period as in van Dijk et al. (2002) in order to compare our results to theirs. We use 10 years of data, from January 1990 until December 1999. For each point from December 1989 up to September 1999, we compute 1 step ahead forecasts of the unemployment rate. Multiple step-ahead forecast results are shown in Appendix 2. The parameter estimates are not updated as new observations become available. The forecasts compared to actual values and confi-
<table>
<thead>
<tr>
<th>Forecast Description</th>
<th>LSTAR MPE</th>
<th>LSTAR MSPE</th>
<th>LSTR MPE</th>
<th>LSTR MSPE</th>
<th>Regime MPE</th>
<th>Regime MSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconditional</td>
<td>-0.02</td>
<td>0.05</td>
<td>0.02</td>
<td>0.05</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>Recession/contraction</td>
<td>-0.01</td>
<td>0.05</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>Expansion</td>
<td>-0.03</td>
<td>0.04</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 2.2: Forecast Evaluation for Monthly US Unemployment Rate

dence intervals are displayed in Figures 2.9 - 2.10. Table 2.2 contains two forecast evaluation criteria: mean prediction error (MPE) and mean squared prediction error (MSPE) for 3 different models. The first model (LSTAR) is Equation 2.20 of van Dijk et al. (2002). The second model (LSTR) is our model in Equation 2.21. The third is a regime switching model which uses average values of the transition function \( G(\hat{X}_{t-1}; \gamma) \) for each state of the economy. The forecast of the regime switching model is computed as in Equation 2.21 with the exception of \( G(\hat{X}_{t-1}; \gamma) \) taking on only 1 of 3 values (average value for \( G(\hat{X}_{t-1}; \gamma) \) from each state of the economy) instead of a value computed by Equation 2.22. Table 2.2 also contains forecasting results which are conditional upon the regime that is realized at the forecast origin.

The mean prediction error suggests that both the LSTAR and LSTR models produce biased forecasts in both regimes. During periods of expansions, when the unemployment is declining, the LSTAR model is overly pessimistic and predicts unemployment rates that are too high on the average while the LSTR model is overly
optimistic and predicts unemployment rates that are too low on the average. The regime switching model seems to produce unbiased estimates during expansions. During recessions when the unemployment rate is increasing, the LSTAR model again is overly pessimistic and predicts unemployment rates that are too high on the average. This is also true (but to a greater extent) in the regime switching model. The LSTR model (again) predicts unemployment rates that are too low on the average. While comparing the overall forecasting performance of the 3 models, the regime switching model appears to perform the best. However, when examining the conditional forecasts, we see poor predictive power during recessions and good predictive power during expansions. Due to the state of the economy during this 10-year period (few recessionary periods), the good predictive power during expansions masks the model's weakness. It is interesting to note that the overall biases of the LSTAR and LSTR model are of the same magnitude but of opposite sign. Perhaps it would be helpful to combine the two forecasts for an overall forecast.
Figure 2.1: Latent Process $X(t)$ and its Estimators. $X.t$ denotes the true process $X_t$ as in Equation 2.2. $X.fit1$ denotes the first estimate $\hat{X}_t^{(1)}$ as in Equation 2.9. $X.fit3$ denotes the iterative estimate $\hat{X}_t^{(3)}$ as in Equation 2.15.
Figure 2.2: Categorical Time Series Y(t) and its Estimators. Y.t denotes the categorical time series $Y_t$ as in Equation 2.3. Y.fit1 denotes the first estimate $\hat{Y}_t^{(1)}$ as in Equation 2.10. Y.fit3 denotes the iterative estimate $\hat{Y}_t^{(3)}$ as in Equation 2.16.
Figure 2.3: LSTR process $y(t)$ and its Estimators. $y.t$ denotes the LSTR process $y_t$ as in Equation 2.6. $y.fit1$ denotes the first estimate $\hat{y}_t^{(1)}$ as in Equation 2.12. $y.fit2$ denotes the iterative estimate $\hat{y}_t^{(2)}$ as in Equation 2.18.
Figure 2.4: Monthly seasonally unadjusted US employment rate per 100, males aged 20 and above, January 1968-December 2004
Figure 2.5: State of the Economy (Recession = 1, Contraction = 2, Expansion = 3)
Figure 2.6: State of the Economy (Recession = 1, Contraction = 2, Expansion = 3) with Fitted Values
Figure 2.7: Monthly seasonally unadjusted US employment rate, males aged 20 and above, January 1968-December 1989 with Fitted Values
Figure 2.8: Monthly seasonally unadjusted US employment rate, males aged 20 and above, January 1968-December 1989 with 95% Confidence Intervals
Figure 2.9: One step ahead forecasts for LSTR model for December 1989 - November 1999 compared to actual values
Figure 2.10: One step ahead forecasts for LSTR model for December 1989 - November 1999 with 95% Confidence Intervals compared to actual values.
Chapter 3

Evolving Structure in Multivariate
Time Series with Application to
Stock Sector Data

3.1 Introduction

Financial data lends itself to multivariate analysis due to its hierarchical structure (e.g. individual securities within sectors within markets). Many models exist for the joint analysis of several financial instruments such as securities due to the fact that they are not independent. These models often assume some type of constant behavior between the instruments over the time period of analysis. Instead of imposing this assumption, we are interested in understanding the dynamic covariance structure in
our multivariate financial time series, which will provide us with an understanding of changing market conditions. In order to achieve this understanding, we first develop a multivariate model for the conditional covariance and then examine that estimate for changing structure using multivariate techniques. Specifically, we simultaneously model individual stock data that belong to one of three market sectors and examine the behavior of the market as a whole as well as the behavior of the sectors. Our aims are detecting and forecasting unusual changes in the system, such as market collapses and outliers, and understanding the issue of portfolio diversification in multivariate financial series from different industry sectors.

The motivation for this research concerns portfolio diversification. The false assumption that investment in different industry sectors is uncorrelated is not made. Instead, we assume that the comovement of stocks within and between sectors changes with market conditions. Some of these market conditions include market crashes or collapses and common external influences.

3.2 Data Description

We collected daily stock returns from the CRSP database for three different stock sectors: Energy, Financial, and Technology for the time period January 2, 1998 - December 31, 2001. In order to obtain stocks from these specific sectors, we used the Standard Industrial Classification Code List published by the U.S. Securities and
Exchange Commission. Our aim was to use the twenty companies with the largest market capitalization (on January 6, 1998) in each sector. However, due to missing data issues and classification code errors, we considered up to the largest twenty-eight companies. For each sector, we used the largest twenty-one companies with the correct industry classification code and no missing data. Thus our system includes 63 series of daily stock returns for the time period January 2, 1998 to December 31, 2001. We define the daily stock return at time \( t \) as \( Y_t \) where

\[
Y_t = \frac{P_t - P_{t-1}}{P_{t-1}}.
\]

Figure 3.1 displays the data for a time period of six months. Each cluster includes the twenty-one time series plots for a particular sector during the six month period. The returns for the financial companies have been shifted by +25 %, and the returns for the technology companies have been shifted by -25%. This convention will be kept in all remaining similar plots (top series = financial, middle series = energy, bottom series = technology). Included in Figure 3.2 are the same data for a narrower timeframe. Each panel in Figure 3.2 includes six days. Each of the two panels demonstrates a different type of comovement behavior in the stocks. The left panel demonstrates a time period when the stock returns are following each other closely. The right panel demonstrates a time period with less correlation. We are interested in detecting this difference.

\footnotetext[1]{http://www.sec.gov/info/edgar/siccodes.htm}
3.3 Motivation

The models given in Appendix 1 are infeasible and/or inappropriate to apply to our problem. The BEKK and VEC models are highly parameterized and difficult to estimate for a large scale system. It is also difficult to guarantee the positivity of $H_t$ in the VEC model without imposing strong restrictions on the parameters. The factor models and principal component models are inappropriate since we cannot assume that the factor or principal component structure remains constant throughout the time period of interest. We can verify this fact using the singular value decomposition of our covariance matrix. Figure 3.3 displays the scores for the second principal component versus the scores for the first principal component. By examining these four plots simultaneously, we can see a difference between the time periods 1998-2000 and 2001. In 1998-2000 there is more separation between the three sectors and in 2001 the second principal component scores are virtually constant with the exception of one company. This provides evidence for the need of a model which allows for dynamic underlying structure in the principal components.

The conditional constant correlation model does not allow for time-varying correlation between the series, which is present in our problem. The DCC models seem more appropriate. However, allowing the correlation to change for every time period seems inappropriate for the type of behavior we are trying to detect. Instead, we develop a DCC model that incorporates our intuitive sense for the problem.
3.4 Methodology

The Regime Switching for Dynamic Correlations (RSDC) model of Pelletier (2004) allows time-varying correlation between the series by allowing the system to switch between regimes. The covariances in the system are decomposed into correlations and standard deviations and the correlation matrix follows a regime switching model, that is, the correlations are constant within regime but different across regimes. A Markov chain governs the transitions between the regimes. The model is described as follows. Assume our \(K\)-variate process \(Y_t\) has the form:

\[
Y_t = H_t^{1/2}U_t,
\]

where \(U_t\) is an i.i.d. \((0, I_k)\) process. That is, each element in \(U_t\) has zero mean and unit variance. Also, each series is uncorrelated with all remaining series. The time varying covariance matrix \(H_t\) can be decomposed as:

\[
H_t \equiv S_t \Gamma_t S_t
\]

(3.1)

where \(S_t\) is a diagonal matrix composed of the standard deviations \(s_{k,t}\), for \(k = 1, \ldots, K\) and the matrix \(\Gamma_t\) contains the correlations. Specifically, \(\Gamma_t\) follows a regime switching model:

\[
\Gamma_t = \sum_{r=1}^{R} I_{(\Delta_t = r)} \Gamma_r
\]

where \(\Delta_t\) is an unobserved Markov process independent of \(U_t\) which can take \(R\) possible values \((\Delta_t = 1, 2, \ldots, R)\) and \(I\) is the indicator function. The \(K \times K\) matrices
The probability law governing $\Delta_t$ is defined by its transition probability matrix $\Pi$. The probability of going from regime $i$ in period $t$ to regime $j$ in period $t + 1$ is denoted by $\pi_{i,j}$ and the limiting probability of being in regime $n$ is $\pi_n$. We assume that the Markov chain is ergodic and irreducible.

This model has several good properties that make it appropriate for our problem. First, we expect that our system will be in different regimes depending on the state of the market. Secondly, we can obtain estimates of the parameters, even when the number of time series is large. Our problem involves detecting changes in an overall market which involves many time series within multiple sectors. Another property of the RSDC model is that it is easy to impose that the variance matrices are positive semidefinite. Finally, when we model the standard deviations with the Autoregressive Moving Average Conditional Heteroscedasticity (ARMACH) model of Taylor (1986) we can compute multi-step ahead predictions for the covariance matrix.

In order to fit the model, we use a two-step estimation procedure. We will denote the vector of parameters in our model as $\theta$ and split them into two groups $\theta = (\theta_1, \theta_2)$ where $\theta_1$ contains our parameters estimated in the first step and $\theta_2$ contains our parameters estimated in the second step. In the first step, we estimate the univariate volatility models. We use the ARMACH model of Taylor (1986) in which the conditional standard deviation (rather than variance) is modeled as follows:

$$s_t = \omega + \sum_{i=1}^q a_i |y_{t-i}| + \sum_{j=1}^p \beta_j s_{t-j}$$
with $\bar{\alpha}_i = \alpha_i / E|\bar{u}_t|$.

Denote $QL_1$ as the log-likelihood where the correlation matrix is taken to be an identity matrix:

$$QL_1(\theta_1; Y) = -\frac{1}{2} \sum_{t=1}^{T} (K \log (2\pi) + 2 \log \left(|S_t|\right) + U_t^T U_t).$$

Since this is just the sum of $K$ univariate log-likelihoods, maximizing it is equivalent to maximizing (separately) each univariate log-likelihood.

For the second step of the estimation, we need to maximize $QL_2$ which is the log-likelihood given $\theta_1$:

$$QL_2(\theta_2; Y, \theta_1) = -\frac{1}{2} \sum_{t=1}^{T} (K \log (2\pi) + 2 \log \left(|\Gamma_t|\right) + U_t^T \Gamma_t^{-1} U_t).$$

We can maximize $QL_2$ using the EM algorithm. We make use of the results in Hamilton (1994). We are using Hamilton's filter because the Markov chain $\Delta_t$ is unobserved.

In order to make inference on the state of the Markov chain, we compute values for $\hat{\xi}_{it}$ and $\eta_i$ which are defined as follows: Let $\hat{\xi}_{it}$ be an $R \times 1$ vector containing the probability of being in each regime at time $t$ conditional on observations up to time $t$. Thus, $\hat{\xi}_{it}$ is an $R \times 1$ vector containing the elements $P\{\Delta_t = j|\hat{U}_t; \theta_2\}$. Let $\eta_i$ be the $R \times 1$ vector whose $j$th element is the density of $U_t$ conditional on past observations and being in the $j$th regime at time $t$. In order to obtain these estimates, we use the following steps:

1) Choose a starting value for $\hat{\xi}_{10}$ and $\theta_2 = (\Pi, \Gamma_1, \ldots, \Gamma_R)$.

We set each element in $\hat{\xi}_{10} = 1/R$. In order to obtain starting values for the correlation matrices in each regime, we compute an estimate of the correlation matrix
of the data using the numerically accurate corrected two-pass method described in Chan et al. (1983). We will denote the empirical estimator as $\hat{\Sigma}$. Next we set $\Gamma_1 = 0.8\hat{\Sigma}$ and $\Gamma_2 = 1.2\hat{\Sigma}$. We initially set the diagonal elements of $\Pi$ equal to 0.7 since we expect some persistence in our Markov chain. We tried several reasonable starting values, and obtained convergence to the same values. Thus, the starting values did not affect our results—only the computing time.

2) Compute $\eta_t$ by evaluating the multivariate normal density of $U_t$ conditional on past observations and being in each regime at time $t$. For high dimensional data, we found that working on the log-scale was necessary. That is,

$$\tilde{\eta}_{r,t} = -\frac{1}{2} \left( K \log(2\pi) - \log |\hat{\Gamma}_r| - U_r \hat{\Gamma}_r^{-1} U_r' \right), \quad r = 1, \ldots, R$$

$$\eta_{r,t} = \exp \tilde{\eta}_{r,t}.$$

Then compute iteratively for $t = 1, \ldots, T$

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t-1|t} \odot \eta_t)}{\sum(\hat{\xi}_{t-1|t} \odot \eta_t)}$$

$$\hat{\xi}_{t+1|t} = \Pi \cdot \hat{\xi}_{t|t}.$$

Note that $\odot$ denotes element-by-element multiplication. The $R \times 1$ vector $\hat{\xi}_{t+1|t}$ contains the probabilities of being in each regime at time $t + 1$ conditional on observations up to time $t$.

3) Compute the smoothed state variables starting from $t = T - 1, \ldots, 1$ as follows:

$$\hat{\xi}_{t|T} = \hat{\xi}_{tt} \odot \left\{ \Pi' \left[ \hat{\xi}_{t+1|T} \left( \div \right) \hat{\xi}_{t+1|t} \right] \right\}.$$
Note that (\divide) denotes element-by-element division.

4) Update $\theta_2$ as follows:

$$\hat{\Pi}_{i,j} = \frac{\sum_{t=2}^{T} F[\Delta_t=j, \Delta_{t-1}=i|U_T; \theta_2]}{\sum_{t=2}^{T} F[\Delta_{t-1}=i|U_T; \theta_2]}$$

$$\hat{\Gamma}_r = \frac{\sum_{t=1}^{T} \left(U_t U_t'\right) P[\Delta_t=r|U_T; \theta_2]}{\sum_{t=1}^{T} P[\Delta_t=r|U_T; \theta_2]}.$$

Note that we must scale the $\hat{\Gamma}_r$ matrices so that we have ones on the diagonal since this does not happen naturally. Let us denote $\hat{\Gamma}_r$ as our estimate before rescaling and $\hat{\Gamma}_r$ as our estimate after rescaling. We compute $\hat{\Gamma}_r$ as:

$$\hat{\Gamma}_t = D_r^{-1} \hat{\Gamma}_r D_r^{-1}$$

where $D_r$ is a diagonal matrix with $\sqrt{\hat{\Gamma}_{i,i,r}}$ on row $i$ and column $i$. This rescaling ensures that we have ones on the diagonal and off-diagonal elements between -1 and 1.

Now we will replace our starting value $\hat{\xi}_{1|0}$ with its smoothed estimate $\hat{\xi}_{1|T}$ (its estimate conditional on information from $t = 1, \ldots, T$). We do this since it has been shown in Hamilton (1994) that the MLE estimate for $\hat{\xi}_{1|0}$ is given by $\hat{\xi}_{1|T}$. This makes the choice of our starting value for $\hat{\xi}_{1|0}$ less important since we are replacing it by the MLE in the next iteration.

5) Repeat steps 2) through 4) until convergence occurs (maximum element of the difference in successive parameter estimates is small).

Pelletier (2004) showed that under the usual assumptions of quasi- maximum likelihood estimation, the two-step maximum likelihood estimates obtained from the
procedure detailed above are consistent and asymptotically normal.

3.5 Simulation Study

In order to investigate the estimation procedure, we simulate an RSDC model with two regimes and compare our estimates to the true values. We simulate this model 25 times and Tables 3.1 - 3.2 detail the comparison. We use the same parameters and parameter estimates from the stock sector example (detailed in the next section) for our simulated series. That is, $K = 63, T = 1004, R = 2$, and the correlation model parameters from the Results section. For the univariate volatility models, we simulate from a GARCH(1,1) model:

$$s_{k,t}^2 = 0.0005 + 0.6\epsilon_{k,t-1}^2 + 0.1s_{k,t-1}^2.$$ 

These GARCH(1,1) parameters were chosen in order to obtain the same level of volatility as in our stock sector data. Figure 3.4 displays the data for one of the simulations in the same format as Figure 3.1. Each cluster includes the twenty-one time series plots for a particular sector during a six month period. Again, the returns for the financial companies have been shifted by +25 %, and the returns for the technology companies have been shifted by -25% while the returns for the energy companies remain unshifted.

Figure 3.5 displays our model’s ability to estimate the regime at each time point for one of the simulations. We classify our predictions as correct if the true regime
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True R1</th>
<th>Est. R1</th>
<th>Std. Dev.</th>
<th>True R2</th>
<th>Est. R2</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy/Energy</td>
<td>0.3539</td>
<td>0.3364</td>
<td>0.0175</td>
<td>0.6437</td>
<td>0.5295</td>
<td>0.0157</td>
</tr>
<tr>
<td>Energy/Fin.</td>
<td>0.0926</td>
<td>0.0996</td>
<td>0.0156</td>
<td>0.1696</td>
<td>0.1287</td>
<td>0.0263</td>
</tr>
<tr>
<td>Energy/Tech.</td>
<td>0.0227</td>
<td>0.0453</td>
<td>0.0144</td>
<td>0.1424</td>
<td>0.0980</td>
<td>0.0282</td>
</tr>
<tr>
<td>Fin./Fin.</td>
<td>0.4802</td>
<td>0.4256</td>
<td>0.0167</td>
<td>0.5248</td>
<td>0.4398</td>
<td>0.0220</td>
</tr>
<tr>
<td>Fin./Tech.</td>
<td>0.1186</td>
<td>0.1407</td>
<td>0.0124</td>
<td>0.3613</td>
<td>0.2803</td>
<td>0.0237</td>
</tr>
<tr>
<td>Tech./Tech.</td>
<td>0.1489</td>
<td>0.1961</td>
<td>0.0205</td>
<td>0.6736</td>
<td>0.5460</td>
<td>0.0158</td>
</tr>
<tr>
<td>Transition</td>
<td>0.8340</td>
<td>0.7979</td>
<td>0.0165</td>
<td>0.8305</td>
<td>0.7930</td>
<td>0.0214</td>
</tr>
</tbody>
</table>

Table 3.1: Correlation and Transition Probability Estimates in Simulation Study based on 25 realizations

has an estimated probability > 0.5. We can see that we are more often than not predicting the correct regime. Figure 3.6 details which timepoints are correct and incorrect for Figure 3.5. For the entire time period, we predict the correct regime 76% of the time.

### 3.6 Results

We fit the RSDC model to our stock sector data (3 sectors consisting of 21 different companies) using two regimes, namely higher correlation and lower correlation. We
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True</th>
<th>Estimated</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0.0005</td>
<td>0.0005072</td>
<td>0.00006383</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.6</td>
<td>0.5927</td>
<td>0.06920</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.1</td>
<td>0.09518</td>
<td>0.05318</td>
</tr>
</tbody>
</table>

Table 3.2: Univariate Volatility Estimates in Simulation Study based on 25 realizations

used Ox software to fit this model.\(^2\) We model the correlations within each sector and between the sectors. Tables 3.3-3.4 include parameter estimates for this model. Several different starting values which were far apart in the parameter space were used with convergence to the same values.

Of additional interest is the relative magnitude of correlations within and between the three sectors. Figure 3.7 displays the one-step ahead forecasts for the average correlation between and within sectors for a nine-month time period in 1999-2000. Additionally, Figure 3.8 displays the same information for a shorter time period. Note that although the magnitudes of the within and between sector correlations differ, the pattern in the forecasts is identical because this is driven by the regimes.

Let us revisit the the issue of differing comovement behavior types. In Figure 3.2 we displayed two panels: one in which the stock returns followed each other closely and the other in which the stock returns exhibited less correlation. Our goal is to detect this difference. Figures 3.9 - 3.10 display the returns and the estimated probability

\(^2\)http://www.doornik.com/index.html
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Regime 1</th>
<th>Regime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy/Energy</td>
<td>0.3539</td>
<td>0.6437</td>
</tr>
<tr>
<td>Energy/Financial</td>
<td>0.0926</td>
<td>0.1696</td>
</tr>
<tr>
<td>Energy/Technology</td>
<td>0.0227</td>
<td>0.1424</td>
</tr>
<tr>
<td>Financial/Financial</td>
<td>0.4802</td>
<td>0.5248</td>
</tr>
<tr>
<td>Financial/Technology</td>
<td>0.1186</td>
<td>0.3613</td>
</tr>
<tr>
<td>Technology/Technology</td>
<td>0.1489</td>
<td>0.6736</td>
</tr>
<tr>
<td>Transition Probabilities</td>
<td>0.8340</td>
<td>0.8305</td>
</tr>
</tbody>
</table>

Table 3.3: Correlation and Transition Probability Estimates

of being in the higher correlation regime for a typical six day period. Our model is detecting these differing comovement behaviors by predicting with high probability being in the higher correlation regime when the stocks are following each other closely and predicting with high probability being in the lower correlation regime when the stocks are following a more random pattern.

With the RSDC model, we can estimate the covariance matrix at time $t$, $H_t$, by combining our estimates for the standard deviations for each asset and the correlation matrix as in Equation 3.1. In order to detect changes in the system, we use the singular value decomposition of our covariance matrix and examine the pattern over time. The use of principal components allows us to examine market structure in lower dimensional space (instead of examining the $K \times K$ covariance matrix). This reduc-
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.661</td>
</tr>
<tr>
<td>$\bar{\alpha}_1$</td>
<td>0.0465</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.317</td>
</tr>
</tbody>
</table>

Table 3.4: Univariate Volatility Estimates

Reduction in dimensionality is achieved by finding a set of standardized linear combinations which are orthogonal and taken together explain all the variance of the original data. The principal components are defined as follows (Mardia et al. (1979)):

If $x$ is a random vector with mean $\mu$ and covariance matrix $\Sigma$ then the principal component transformation is the transformation

$$ x \rightarrow y = \Phi' (x - \mu), $$

where $\Phi$ is orthogonal, $\Phi' \Sigma \Phi = \Lambda$ is diagonal, and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. The $i$th principal component of $x$ is defined as the $i$th element of the vector $y$, namely, as

$$ y_i = \phi'_i (x - \mu). $$

Here $\phi_i$ is the $i$th column of $\Phi$ and is called the $i$th vector of principal component loadings.

By studying the evolution of the principal components, we examine the interactive structure of financial instruments which directs us to unusual events such as a market crash or another extreme event. Also, the evolution of such structures identify
favorable market conditions for trading strategies, ultimately leading to performance improvements.

Of particular interest is how many dimensions in which we can reduce the complexity of our multivariate data by transforming the data into the principal components space. The criteria we use in order to decide how many principal components to retain is to include just enough components to explain 90% of the total variance of the processes. We restrict our graphics to shorter time periods for clarity. Figure 3.11 displays the number of principal components needed to explain 90% of the total variation as a function of time from October 12 - November 16, 2001 along with the estimated probability of being in the lower correlated regime. The reason this time period was chosen will become clear when we discuss further results concerning outlier detection. We can compare the behavior of these two plots contemporaneously in order to get a sense of market conditions. Between October 12 and October 19, we see very similar patterns in the two plots. This demonstrates the notion that when we need a small number of principal components to explain most of the variance, we have a high estimated probability of being in the highly correlated regime. Between November 2 and 8, however, these two plots have an opposite pattern. The number of principal components necessary to explain 90% of the variance is small, and there is a high estimated probability of being in the lower correlated regime. This time period is actually one of low variance for the entire market. Comparing these plots gives us a sense of market behavior.
In order to gain a comprehensive view of both the principal components and the original data, we can examine the biplot introduced in Gabriel (1971). Its interpretation is explained in Insightful (2001). By examining the transformed observations, one can interpret the original data in terms of the principal components. By examining the original variables, one can graphically see the relationships between the original variables and the principal components. The x-axis and y-axis represent scores for the first and second principal component, respectively. The original variables are represented by arrows which indicate the proportion of the original variance explained by the first two principal components. Figure 3.12 includes biplots for 2 consecutive days in October 2001. This demonstrates one instance of the evolution of the principal components of our system.

Examination of biplots for subsequent dates demonstrate the ability of our model to detect outliers in our data which may help us predict future market behavior. A well-known outlier in this data is the Enron collapse. A time line of the Enron collapse is well-documented due to the litigation between the U.S. Securities and Exchange Commission and various Enron executives. On October 16, 2001, Enron announced allegedly “nonrecurring” losses of approximately $1 billion. When examining the daily biplots for our model, we first detect Enron as a well-identified outlier on October 4, 2001. Figure 3.13 includes the biplot for this date. Subsequent date’s biplots look very similar and Enron’s effects are still detected until the end of our sample (December 31, 2001). The events of interest during this time frame take place on the
following time line. On October 29 and November 1, 2001, the two leading credit rating agencies downgraded Enron’s credit rating. On November 8, 2001, Enron announced its intention to restate its financial statements for 1997 through 2000 and the first and second quarters of 2001 to reduce previously reported net income by an aggregate of $586 million. On November 21, 2001, Enron’s credit rating was downgraded to “junk” status. On December 2, 2001, Enron filed for bankruptcy, making its stock, which less than a year earlier had been trading at over $80 per share, virtually worthless. We compared our model’s outlier detection ability to a model using a weighted estimate of the covariance matrix using a moving window of 70 trading days, with weights based on the minimum volume ellipsoid estimator. The biplots for this estimator only detect Enron as an outlier on November 28, 2001, well after this detection could be of any use to the investor.

There were other (albeit less well-known) outliers which were detected in the biplots. Figure 3.14 displays the biplot for June 2, 2000. The company with the unusual pattern is Silicon Graphics, Inc. (SGI). The pattern in the biplot is present from 6/1/2000 - 6/8/2000. Figure 3.15 displays price information for all of the stocks in our study alongside price information for SGI. Upon further investigation of the company’s filings with the U. S. Securities and Exchange Commission, we learned that there was an 8-K filing on 6/9/2000 concerning a spin-off of the company. 8-K filings are defined as “current report filings” and are required when a company has

specific "trigger" events transpire. These reports must be filed within a few days of the event, and the report must include details about the event, how it affects the company, and its impact on shareholders. The broad categories of the 8-K filing are 1) registrant's business and operations, 2) financial information, 3) securities and trading markets 4) matters related to accountants and financial statements, 5) corporate governance and management, 6) regulation FD disclosure 7) other events, and 8) financial statements and exhibits.

Another example includes events which occurred with Western Digital Corporation (WDC). Figure 3.16 displays the biplot for January 10, 2000. The pattern in the biplot is present from 1/10/2000 - 1/20/2000. Figure 3.17 displays price information for all of the stocks in our study alongside price information for WDC. Upon further investigation of the company's filings with the U. S. SEC, we learned of several filings within that time period. On 1/13/2000, WDC issued a press release announcing the appointment of a new President and C.E.O. On 1/19, WDC issued a press release announcing it would shift its strategic focus and resources. Additionally, on 1/20, WDC issued a press release announcing a net loss for the last quarter of 1999.

Again, these detections could be useful to the investor as they were detected before these filings with the SEC. Table 3.5 lists these and additional outliers found in our study period. There were 17 total outliers: 14 of which we were able to attribute possible explanations and 3 of which we were not able to attribute possible explanations.
Figure 3.1: Daily (shifted) stock returns for 1/1998 - 5/1998
Figure 3.2: Daily (shifted) Stock returns
Figure 3.3: Principal Component Scores by Year, where “f” denotes Financial companies, “e” denotes Energy companies, and “t” denotes Technology companies. The x- and y-axis are scores for the first and second principal component, respectively.
Figure 3.4: Simulated Daily (shifted) stock returns
Figure 3.5: Estimated Probability / Indicator of Higher Correlation Regime. The points (circles) are an indicator function for the higher correlation regime (0 if we are in the lower correlation regime and 1 if we are in the higher correlation regime). The points (crosses) are the estimated probabilities of being in the higher correlation regime, and the path of the estimates is connected with the line.
Figure 3.6: Indicator of Correctly Predicted Regime. The points (circles) are an indicator function for whether the estimated probability of the true regime is > 0.5 (0 if the estimated probability of the true regime is < 0.5 and 1 if the estimated probability of the true regime is > 0.5.)
Figure 3.7: One step ahead forecasts for average correlations
Figure 3.8: One step ahead forecasts for average correlations
Figure 3.9: Daily (shifted) returns and Estimated probability of being in higher correlated regime
Figure 3.10: Daily (shifted) returns and Estimated probability of being in higher correlated regime
Figure 3.11: Number of necessary principal components and Estimated probability of being in lower correlated regime.
Figure 3.12: Biplots on 10-01-2001 and 10-02-2001
Figure 3.13: Biplot on 10-04-2001
<table>
<thead>
<tr>
<th>Company</th>
<th>Dates</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enron</td>
<td>10/4-12/31/2001</td>
<td>10/16/2001 announcement of losses of approximately $1 billion&lt;br&gt;10/29, 11/1/2001 two leading credit rating agencies downgraded credit rating&lt;br&gt;11/8/2001 announcement of intention to restate financial statements for 1997-2001 to reduce previously reported net income&lt;br&gt;11/21/2001 credit rating = “junk” status&lt;br&gt;12/2/2001 filed for bankruptcy</td>
</tr>
<tr>
<td></td>
<td>(various days)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/31-11/13/2001</td>
<td></td>
</tr>
<tr>
<td>Western Digital</td>
<td>8/7/1998</td>
<td>8/10/1998 announcement of an agreement with IBM concerning component supply, technology, and design licensing&lt;br&gt;11/16/1999 release of quarterly report detailing poor performance and 37% decrease in revenue&lt;br&gt;1/10-20/2000 announcement of new C.E.O. announcement of shift in strategic focus and resources announcement of net loss for Q4 of 1999</td>
</tr>
<tr>
<td></td>
<td>11/11-15/1999</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/10-20/2000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5/2-3/2001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11/4-12/1998</td>
<td></td>
</tr>
<tr>
<td>Newmont Mining</td>
<td>9/21-30/1998</td>
<td>data only available since 2002</td>
</tr>
<tr>
<td>Sun Microsystems</td>
<td>1/19-23/2001</td>
<td>no attributable cause found</td>
</tr>
<tr>
<td>Adaptec</td>
<td>1/12-13/1998</td>
<td>no attributable cause found</td>
</tr>
</tbody>
</table>

Table 3.5: Outliers detected from Biplots
Figure 3.14: Biplot on 06-02-2000
Figure 3.15: Left panel: Daily (shifted) stock returns for 5/31/2000 - 6/12/2000.

Figure 3.16: Biplot on 01-11-2000
Chapter 4

Conclusions and Further Work

4.1 Conclusions

We have developed a feasible and accurate estimation procedure for a multivariate time series whose components are both continuous and categorical in the specific case of a logistic smooth transition regression (LSTR) model whose transition variable is related to a categorical variable. The estimation procedure was investigated both with simulation and a real-world example. The simulation study showed us the accuracy of the procedure when the model is known. The real-world example demonstrated the feasibility of our procedure and obtained similar results to previously studied non-linear models for the same data. The LSTAR model had a superior fit for the in-sample observations, but the forecast evaluation criteria demonstrated equivalent forecasting performance for the LSTAR, LSTR, and regime switching models. Even though we
see no advantage in our method for the unemployment example, our method proves its importance when a categorical (instead of continuous) variable governs the nonlinear behavior. Since this methodology has been developed in this contribution, there are no methods for comparison other than ones using continuous transition variables.

We have also developed a regime-switching Multivariate Generalized Autoregressive Conditional Heteroscedasticity (MGARCH) model that captures the dynamics of financial returns. This model has provided us with an understanding of changing market conditions which is useful for portfolio diversification assessment. Additionally, we have been able to detect unusual events in the data, often forecasting unusual changes in the system, by coupling our model with other multivariate methods for outlier detection. Our method could be used as a key tool for risk assessment in financial investment. While we forecast an unusual change, we do not necessarily know the direction in which the change will occur. Even though a direction of change cannot be predicted, this is still very useful information to the investor as one could purchase options on both sides of the security (strike prices below and above the current price). Options are inexpensive compared to the price of the stock and can have large payoffs if a large movement in the stock takes place before the option expires. Additionally, the model developed can be easily applied to the entire market with the main computational challenge seeming to be computational time.
4.2 Further Work

The model estimation procedure developed for the LSTR model whose transition is related to a categorical variable is a very specific case of nonlinear behavior and multiple data types. This work could be extended to additional models by developing estimation procedures for more general systems with nonlinear behavior and multiple data types. Some additional work also needs to be done with regard to model specification when applying the model to real data. In our case, we conducted a simulation study and compared our model with one previously developed. In reality, one would need to conduct some additional steps such as specifying the autoregressive order of the model and testing for linearity against STR or STAR nonlinearity.

A further extension of the evolving structure work would be to create a portfolio allocation method which capitalizes on these results. After examining the estimate of the conditional covariance for changing structure using multivariate techniques, the investor can rebalance the portfolio if necessary in order to obtain the proper level of diversification. Note that this is a higher-level approach to portfolio management than ones in which a rebalancing effort between multiple stocks takes place (which undoubtedly involves more transaction costs). We refer to Engle and Sheppard (2001) for the minimum variance portfolio weighting. The time-varying weights of the minimum variance portfolio are calculated as:

\[ w_t = \frac{H_t^{-1}1}{C_t} \]

where
\[ C_t \equiv 1' H_t^{-1} 1 \]

and where \( 1 \) is a \( K \) by 1 vector of ones. This method of portfolio rebalancing involves trading on multiple assets instead of the broader approach mentioned above. The broader approach would include securities such as exchange traded funds based on specific sectors.
Appendix A

Multivariate Generalized

Autoregressive Conditional

Heteroscedasticity (MGARCH)

model development

Suppose we have a vector stochastic process \( \{ y_t \} \) with dimension \( K \times 1 \). We model this process as:

\[
y_t = \mu_t(\theta) + \epsilon_t,
\]

where \( \mu_t(\theta) \) is the mean vector conditional on the information up to time \( t \) and

\[
\epsilon_t = H_t^{1/2}(\theta) z_t,
\]
where $H_t^{1/2}(\theta)$ is a $K \times K$ positive definite matrix. Also, $z_t$ has the following two properties:

$$E[z_t] = 0$$

$$\text{Var}[z_t] = I_K,$$

where $I_K$ is the identity matrix of order $K$. The matrix $H_t^{1/2}$ is any $K \times K$ positive definite matrix such that $H_t$ is the variance matrix of $y_t$ conditional on the information up to time $t$.

Some MGARCH models that are direct generalizations of the univariate GARCH model include the VEC model of Bollerslev et al. (1988), and the BEKK model of Engle and Kroner (1995) which is a special case of the VEC model. Due to the high number of unknown parameters, these models are rarely used when the number of series is larger than 3 or 4 (Bauwens et al. (2004)).

There is another group of MGARCH models with fewer parameters due to the common dynamic structure they impose on all the elements of $H_t$. These are called factor models because they assume that the comovements of the series are driven by a small number of common underlying variables which are called factors. These were introduced by Engle et al. (1990) but there exist several variants of the model. One is the full-factor multivariate GARCH (FF-MGARCH) model of Vrontos et al. (2003) where:

$$H_t = W\Sigma_t W',$$
and $W$ is a $K \times K$ triangular matrix with ones on the diagonal and the matrix

$\Sigma_t = \text{diag}(\sigma^2_{1,t}, \ldots, \sigma^2_{K,t})$ where $\sigma^2_{i,t}$ is the conditional variance of the $i$th factor. So, the $i$th element of $W^{-1} \epsilon_t$ can be separately defined as any univariate GARCH model.

Linear combinations of several univariate models (such as the orthogonal GARCH (O-GARCH) and latent factor models) are another group of MGARCH models. In the O-GARCH model (Kariya (1988), Alexander (1997)) the $K \times K$ time-varying variance matrix $H_t$ is generated by $m \leq K$ univariate GARCH models. The O-GARCH(1,1,m) is defined as:

$$V^{-1/2} \epsilon_t = u_t = \Lambda_m f_t,$$

where $V = \text{diag}(v_1, v_2, \ldots, v_K)$, with $v_i$ the population variance of $\epsilon_{it}$ and $\Lambda_m$ is a matrix of dimension $K \times m$ given by:

$$\Lambda_m = P_m \text{diag} \left( l_1^{1/2}, \ldots, l_m^{1/2} \right),$$

$l_1 \geq \cdots \geq l_m > 0$ being the $m$ largest eigenvalues of the population correlation matrix of $u_t$, and $P_m$ the $K \times m$ matrix of associated (mutually orthogonal) eigenvectors. The vector $f_t = (f_{1t}, \ldots, f_{mt})'$ is a random process such that:

$$E_{t-1} [f_t] = 0$$

(A.1)

$$Var_{t-1} [f_t] = \Sigma_t = \text{diag} (\sigma^2_{f_{1t}}, \ldots, \sigma^2_{f_{mt}})$$

(A.2)

$$\sigma^2_{f_{it}} = (1 - \alpha_i - \beta_i) + \alpha_i f^2_{i,t-1} + \beta_i \sigma^2_{f_{i,t-1}} \quad i = 1, \ldots, m.$$ 

The $t - 1$ subscripts in Equations A.1 and A.2 denote the fact that the expectation and variance are conditional on the information at time $t - 1$. Consequently,
\[ H_t = \text{Var}_{t-1}(\epsilon_t) = V^{1/2}V_tV^{1/2} \]

where

\[ V_t = \text{Var}_{t-1}(u_t) = \Lambda_m \Sigma_t \Lambda_m'. \]

Note that the order of the O-GARCH model above \((1,1,m)\) could be generalized to an order of \((p,q,m)\) by allowing each \(\alpha_i\) to be a vector of length \(p\) and each \(\beta_i\) a vector of length \(q\).

Nonlinear combinations of univariate GARCH models are a group of MGARCH models which allow modelers to specify individual conditional variances and the conditional correlation matrix of the series. This group includes the conditional correlation model of Bollerslev (1990) which is defined as follows:

\[ H_t = D_tRD_t = (\rho_{ij} \sqrt{h_{ii} h_{jj}}), \]

where

\[ D_t = \text{diag} \left( h_{1i}^{1/2}, \ldots, h_{Ki}^{1/2} \right), \]

(A.3)

\( h_{ii} \) can be defined as any univariate GARCH model, and

\[ R = (\rho_{ij}) \]

is a symmetric positive definite matrix with \(\rho_{ii} = 1, \forall i\).

The assumption that the conditional correlations are constant is relaxed in Christodoulakis and Satchett (2002), Engle (2002), and Tse (2002) and is called a dynamic conditional correlation (DCC) model. Christodoulakis and Satchett (2002) use the Fisher transformation of the correlation coefficient which guarantees the positive definiteness of
the conditional correlation matrix, but its limitation is that it is only a bivariate model. One DCC model which is useful for high dimensional data sets is the one of Tse (2002) and is defined as:

\[ H_t = D_t R_t D_t, \]

where \( D_t \) is defined in Equation A.3 and \( h_{iit} \) can be defined as any univariate GARCH model, and

\[ R_t = (1 - \theta_1 - \theta_2) R + \theta_1 \Psi_{t-1} + \theta_2 R_{t-1}. \]  \hspace{1cm} (A.4)

In Equation A.4, \( \theta_1 \) and \( \theta_2 \) are non-negative parameters satisfying \( \theta_1 + \theta_2 < 1 \), \( R \) is a symmetric \( K \times K \) positive definite parameter matrix with \( \rho_{ii} = 1 \) and \( \Psi_{t-1} \) is the \( K \times K \) correlation matrix of \( \epsilon_t \) for \( \tau = t - M, t - M + 1, \ldots, t - 1 \). The \( j \)th element is:

\[ \psi_{j,j,t-1} = \frac{\sum_{m=1}^{M} u_{i,t-m} u_{j,t-m}}{\sqrt{\left( \sum_{m=1}^{M} u_{i,t-m}^2 \right) \left( \sum_{j=1}^{M} u_{j,t-m}^2 \right)}}, \]

where \( u_{it} = \frac{\epsilon_{it}}{\sqrt{h_{iit}}} \). The matrix \( \Psi_{t-1} \) can be expressed as:

\[ \Psi_{t-1} = B_{t-1}^{-1} L_{t-1} L_{t-1}' B_{t-1}^{-1}, \]

where \( B_{t-1} \) is a \( K \times K \) diagonal matrix with \( i \)th diagonal element given by \( \left( \sum_{h=1}^{M} u_{i,t-h}^2 \right)^{1/2} \)

and \( L_{t-1} = (u_{t-1}, \ldots, u_{t-M}) \) is a \( K \times M \) matrix, with \( u_t = (u_{1t}, \ldots, u_{Kt})' \). Also, \( M \geq K \) in order to ensure the positivity of \( R_t \).

The DCC model of Engle (2002) has the same form for \( H_t \) (see Equation 3.1) but with
\[ R_t = \text{diag}\left(q_{11,t}^{-1/2}, \ldots, q_{KK,t}^{-1/2}\right)Q_t \text{ diag}\left(q_{11,t}^{-1/2}, \ldots, q_{KK,t}^{-1/2}\right), \]

where the \( K \times K \) symmetric positive definite matrix \( Q_t = (q_{ij,t}) \) is given by:

\[ Q_t = (1 - \alpha - \beta) \bar{Q} + \alpha u_{t-1} u_{t-1}' + \beta Q_{t-1}, \]

with \( u_{it} = \frac{\epsilon_{it}}{\sqrt{h_{it}}} \). Also, \( \bar{Q} \) is the \( K \times K \) unconditional variance matrix of \( u_t \), and \( \alpha \) and \( \beta \) are nonnegative scalar parameters satisfying \( \alpha + \beta < 1 \).

The MGARCH model of Klaassen (1999) also allows the correlations between the series to be time-varying using a two-step approach. In step one, he removes all unconditional correlations by taking principal components of the data. The conditional means and variances of each principal component are specified by a univariate GARCH model. In the second step, the inverse of the principal components construction is used to transform the conditional moments of the principal components into the conditional mean and variance of the data series themselves. This MGARCH model is easy to estimate since we only estimate several univariate GARCH models.
Appendix B

Multiple step-ahead Forecasting in LSTR model

The multiple step-ahead forecasting results are included in Table B.1 and shown in Figures B.1-B.4. We computed the forecasts using the naive method. According to van Dijk et al. (2002), a better method would be a Monte Carlo or bootstrap method. Hence, our forecasting results are less desirable than theirs (as demonstrated in Table B.1).
<table>
<thead>
<tr>
<th>Steps ahead</th>
<th>LSTAR MPE</th>
<th>LSTAR MSPE</th>
<th>LSTR MPE</th>
<th>LSTR MSPE</th>
<th>Regime MPE</th>
<th>Regime MSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.02</td>
<td>0.05</td>
<td>0.02</td>
<td>0.05</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>-0.04</td>
<td>0.07</td>
<td>0.04</td>
<td>0.09</td>
<td>-0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>3</td>
<td>-0.07</td>
<td>0.1</td>
<td>0.06</td>
<td>0.15</td>
<td>-0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Recession/contraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.01</td>
<td>0.05</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>-0.03</td>
<td>0.11</td>
<td>0.05</td>
<td>0.1</td>
<td>-0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>3</td>
<td>-0.04</td>
<td>0.14</td>
<td>0.18</td>
<td>0.2</td>
<td>-0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>Expansion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.03</td>
<td>0.04</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>-0.05</td>
<td>0.06</td>
<td>0.04</td>
<td>0.09</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>3</td>
<td>-0.09</td>
<td>0.08</td>
<td>0.05</td>
<td>0.14</td>
<td>0</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Table B.1: Forecast Evaluation for Monthly US Unemployment Rate
Figure B.1: Two steps ahead forecasts for LSTR model for December 1989 - October 1999 compared to actual values
Figure B.2: Two steps ahead forecasts for LSTR model for December 1989 - October 1999 with 95% Confidence Intervals compared to actual values.
Figure B.3: Three steps ahead forecasts for LSTR model for December 1989 - September 1999 compared to actual values
Figure B.4: Two steps ahead forecasts for LSTR model for December 1989 - September 1999 with 95% Confidence Intervals compared to actual values
Bibliography


Jenkins, G. J. (1979). *Practical Experiences with Modelling and Forecasting Time Series*. GJP Ltd., Channel Islands, UK.


