ABSTRACT

Title of thesis: COMPARING REGIME-SWITCHING
MODELS IN TIME SERIES: LOGISTIC
MIXTURES: vs. MARKOV SWITCHING

Dimitrios V. Paliouras
Master of Science, 2007

Thesis directed by: Professor Benjamin Kedem
Department of Mathematics

The purpose of this thesis is to review several related regime-switching time series models. Specifically, we use simulated data to compare models where the unobserved state vector follows a Markov process against an independent logistic mixture process. We apply these techniques to crude oil and heating oil futures prices using several explanatory variables to estimate the unobserved regimes. We find that crude oil is characterized by regime switching, where prices alternate between a high volatility state with low returns and significant mean reversion and a low volatility state with positive returns and some trending. The spread between one-month and three-month futures prices is an important determinant in the dynamics of crude oil prices.

COMPARING REGIME-SWITCHING MODELS IN TIME
SERIES: LOGISTIC MIXTURES vs. MARKOV SWITCHING

by

Dimitrios V. Paliouras

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2007

Advisory Committee:
Professor Benjamin Kedem, Chair/Advisor
Professor Paul Smith
Professor Paul Zantek

Acknowledgements

I would like to thank Dr. Benjamin Kedem, my thesis advisor, for his help and guidance during this research. I also benefitted greatly from the many wide ranging and thought provoking discussions in his Time Series RIT. I would also like to thank the committee members Dr. Paul Smith and Dr. Paul Zantek, and also Dr. Gloria Gonzalez-Rivera. My family provided me with invaluable support while I juggled the competing demands of work and graduate school. Lastly, I am very grateful to have attended the Applied Mathematics and Scientific Computation program. The courses that I took here challenged me and proved enormously helpful towards my professional career.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

EM	Expectation-Maximization algorithm
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GLM	Generalized Linear Model
LMARX	Logistc Mixture Autoregressive with Exogenous Variables
LMX	Logistic mixtures model
MFP	Markov model with fixed transition probabilities
MLE	Maximum likelihood estimation
MS-GARCH	Markov-Switching GARCH
MTV	Markov model with time-varying transition probabilities
SETAR	Self-Exiting Threshold Autoregressive
O-U	Ornstein-Uhlenbeck process

Chapter 1

Introduction

Economists have found evidence of occasional sudden breaks in many economic time series. For example, currency exchange rates often move abruptly as governments devalue due to speculative pressure and deteriorating economic conditions. Commodity prices, such as oil, change in response to shocks from exogenous geopolitical events or supply disruptions due to weather related catastrophes like hurricane Katrina, and financial markets can shift abruptly in response to financial crises. Perhaps the best example of this is economic growth, where the rate of growth of the economy alternates between periods of high growth (economic expansion) and periods of declining or negative growth (recession). Not surprisingly, a vast amount of research has come from business cycle researchers but significant methodological contributions have also come from engineering applications.

In contrast to linear models that assume stationary distributions (such as ARIMA models), regime-switching models are based on a mixture of parametric distributions whose mixture probabilities depend on unobserved state variable(s). A key difference between the various regime-switching models lies in the stochastic structure of the state variables. For instance, the state of the unobserved process can be modeled by a discrete time/discrete space Markov chain, which can have either fixed or time-varying transition probabilities, or by an independent stochastic

state variable.

There is an extensive literature both in methodological issues and applications in such diverse areas like business cycle economics, financial economics, meteorology, epidemics, tracking and speech recognition applications. This presentation surveys some of the literature related to regime-switching models and tests some of these models using simulated and real-world data. In particular we compare the performance of Markov-switching time series models against independent logistic mixtures, and we apply these modeling techniques in investigating the dynamics of crude oil prices.

1.1   Survey of Literature

Some of the earliest applications of hidden Markov models to time series appear to have been by Lindgren [1], Cosslett and Lee [2], and Quandt and Goldfeld [3]. More recently, Hamilton's Markov-switching model [4] has become popular among econometricians after Hamilton applied it to model the probability of a recession in the U.S. economy. In this model, the economy alternates between two unobserved states of high growth and slow growth according to a Markov chain process. The model assumes constant transition probabilities for the unobserved states, which in turn imply constant expected durations in the various regimes. As in most applications, the model assumes a mixture of normal distributions but it can be generalized to other distributions as well. Engle and Hamilton [5] apply the Markov-switching model to U.S. dollar data and find some success in matching the swings in the value

of the dollar. Garcia and Perron [6] use this model in investigating the behavior of interest rates.

The (relative) success of this model in matching the timing of recessions as determined ex-post by the National Bureau of Economic Research spawned further research. Diebold et al. [7] and extended Hamilton's model to time-varying transition probabilities using the EM algorithm of Dempster et al. [8] to maximize the likelihood. This model retains a Markov structure for the unobserved states but allows additional covariates to determine the transition probabilities. The implication of this model is that the expected duration of each regime is not constant. Filardo [9] applies the time-varying Markov switching model using various leading indicators to model monthly industrial production data. Durland and McCurdy [10] allow the inferred number of periods of the unobserved state to influence the dynamics of the latent variable. They find that the probability of remaining in a recession drops the longer that regime has lasted but that the same does not hold true for economic expansions. Raymond and Rich [11] use the fixed- and time-varying Markov models to examining the effect of oil prices on the U.S. economy. They find that oil prices suppress growth in the economy but do not have a strong effect on the transition probabilities of the state variable.

Regime-switching models have also been employed in modeling heteroskedasticity in stock market returns. The ability of these nonlinear models to capture regime-shifts make them a natural alternative to the ARCH and GARCH models proposed by Engle [12] and Bollerslev [13]. Also, a mixture of normal densities can give rise to fat tails, which is a persistent feature of financial data. Cai [14],

Hamilton and Susmel [15], Gray [16], and Klaassen [17] combine the two approaches into a Switching-ARCH model (SWARCH) that embeds ARCH models within different regimes. The advantage of these models is that they capture conditional heteroskedasticity that is long term and that is not captured by GARCH models. However, the key limitation is that without assumptions the number of unobserved states grows very rapidly, so these models can be computationally intensive.

A more parsimonious way of modeling regime-switching in time series was offered by Jeffries and Kedem [18]. They propose a model of mixtures of generalized linear models (GLM) where the regime probabilities are jointly modeled with a logistic regression. The advantage of this approach is that it does away with the Markov structure that was assumed in the aforementioned models. Jeffries and Kedem develop an EM algorithm [8] to estimate the model and show that such estimates are consistent and asymptotically normal. They also develop a likelihood ratio test for the test for determining whether the data come from a single distribution versus a mixture that sidesteps some of the difficulties associated with likelihood ratio tests of mixtures. They apply their model to fitting weather patterns. Wong and Li [19] also use a logistic mixtures model which they call *logistic mixture autoregressive with exogenous variables* (LMARX) and contrast this class of variables against threshold models.

Threshold or *self-exiting threshold autoregressive* (SETAR) models were developed by Tong ([20], [21]). These models don't assume any hidden Markov process. There are many variants of these models but the lack of intuition for this kind of regime switching has probably hindered the use of these models in many applica-

tions.

Yet another strand of models is developed by Terasvirta in [22] and [23] and an interesting application to GARCH modeling is provided by Gonzalez-Rivera in [24].

Most of the aforementioned regime-switching models can easily be generalized to multiple variables. In particular, Krolzig et al. in [25] and [26] develop vector autoregressive models. Also Engel and Hamilton [5] provide an early application of multivariate switching. Despite their intuitive appeal, these models have limitations in practice.

More recent research into regime-switching includes Kim et al. [27] who use the Markov-switching approach to model equity returns and stock market volatility. Ang and Bekaert in [28] and [29] investigate the asymmetric effects of inflation on various interest rates. Bansal and Zhou [30] model the term structure of interest rates and find that the regime-shifts model captures certain yield curve dynamics better than other commonly used models. Ang and Bekaert [28] apply a regime-switching model to asset allocation.

Regime-switching models have been used in other areas and in connection with state-space models. One of the earliest applications of state-space models with regime switching was by Bar-Shalom [31] who applied this model to a tracking problem. Kim [32] and Kim and Nelson [33] present algorithms for Markov-switching state-space models. Shumway and Stoffer [34] show an application to epidemic outbreak where they decompose a time series of influenza incidence rates into a cyclic and outbreak component. Such structural component models are popular with

economists who want to disaggregate time series into trend versus cyclic components. Examples of these include Kim and Murray [35] and Mills and Wang [36]. A related application of Markov-switching state-space models is provided by Chib and Dueker [37].

In addition to the literature described above, useful resources on regime-switching models include Hamilton [38] and [39]; Kedem [40]; Shumway and Stoffer [34] and Kim et al. [33].

Although regime-switching model techniques are relatively new, there is already a vast body of applications using these models – especially in economics. The applications described above highlight the versatility and richness of these models in capturing non-linear relationships. These modeling techniques share many similarities but they must be tailored to particular situations – there is not a one size fits all. Finally, there are still many open methodological questions related to hypothesis testing.

1.2   Overview of Thesis

In Chapter 2 we provide an overview of the various competing models and describe the techniques used to estimate the models. Specifically, we review the logistic mixture model of Jeffries and Kedem [18], the fixed-transition Markov model of Hamilton [4], and the time-varying Markov transition model of Diebold et al. [7]. In Chapter 3 we apply these models to simulated data, in Chapter 4 we apply the models on real-world data from the energy markets, and Chapter 5 concludes.

Chapter 2

Models of Regime-Switching

2.1   Regime-Switching

In generalized linear models (GLM), the conditional distribution of a time series y_t belongs to the exponential family of distributions and takes the canonical form

$$f(y_t; \mu_t, \phi) = \exp\left(\frac{y_t\mu_t - b(\mu_t)}{\alpha_t(\phi)} + c(y_t; \phi)\right) \tag{2.1}$$

for various link functions (see Kedem and Fokianos [40]) and covariates $\theta_t = x_t'\beta$. A particular form of this model is the classical linear regression model given by

$$y_t = x_t'\beta + \epsilon_t, \qquad \epsilon_t \sim \text{i.i.d. } N(0, \sigma^2), \qquad t = 1, \ldots, T, \tag{2.2}$$

where the data are independently and identically normally distributed. The model has been adapted to non-i.i.d. data, such as when the errors are heteroskedastic as in the GARCH models of Engle and Bollerslev ([12],[13]) and many others.

The key assumption in model (2.2), however, is that the parameters are constant through time, which would not be true if some sort of structural break occurred in the series and the model suddenly changed. One alternative is to estimate the model over different sub-samples if the timing of the break is known and test the hypothesis of a structural break using an F-test.[1] For example, one might want to

---

[1]An F-test was developed by Chow to test for structural breaks in regression parameters by

test for the effects of a policy change such as Fed policy under Paul Volcker in the early 1980s.

Another alternative is to make the structural break endogenous to the model since in many cases the timing of the shift may not be known. By making the shift(s) endogenous to the model, we can also make inferences about the process that drives these shifts. Models that shift between various densities allow us to incorporate structural breaks in the estimation procedure. Instead of assuming a single density for the data, regime (or state) switching models assume that the observations come from a mixture of r parametric distributions given by

$$
Y_t \sim \begin{cases} f(y_t|\theta_1, \mathcal{F}_{t-1}) & \text{if } S_t = 1 \\ f(y_t|\theta_2, \mathcal{F}_{t-1}) & \text{if } S_t = 2 \\ \quad\vdots & \quad\vdots \\ f(y_t|\theta_r, \mathcal{F}_{t-1}) & \text{if } S_t = r \end{cases}, \tag{2.3}
$$

where θ_i contains the parameters of the model i and $\theta_i \neq \theta_j$ if $i \neq j$. Here S_t is an *unobserved* discrete state variable that determines the conditional distribution of Y_t, which as the notation suggests, is time dependent. Also, $\mathcal{F}_{t-1} = \sigma(X_t, X_{t-1}, \ldots, X_{t-p}, Y_{t-1}, \ldots, Y_{t-p})$ is the sigma algebra generated by the known vector of exogenous random variables or known functions of random variables, or more simply the information known up to time $t - 1$.

The number of states r is unknown but most applications assume that $r = 2$ or $r = 3$. For the remainder of this project we assume that there are only two states ($r = 2$) which we label 0 and 1. With two states, $S_t \in \{0, 1\}$, the GLM in (2.1)

---
comparing regression sum of squares across two different sub samples [41].

becomes

$$f(y_t|S_t = 1; \mu_{i,t}, \phi_i) = \exp\left(\frac{y_t\mu_{i,t} - b(\mu_{i,t})}{\alpha_t(\phi_i)} + c(y_t; \phi_i)\right). \qquad (2.4)$$

where $\mu_{i,t} = x_t'\beta_i$ is a function of known covariates x_t. Although this is general enough to encompass the normal, binomial, Poisson, and gamma distributions as special cases, for the remainder of this project we will focus on the Gaussian distribution since it is the most widely used in economic applications. The work that follows applies to other densities as well (see Jeffries [18]).

With two states, the linear model in (2.2) becomes

$$y_t = x_t'\beta_{S_t} + \epsilon_t, \qquad \epsilon_t \sim i.i.d.\ N(0, \sigma_{S_t}^2) \qquad (2.5)$$

$$\beta_{S_t} = \beta_1 S_t + \beta_0(1 - S_t) \qquad (2.6)$$

$$\sigma_{S_t}^2 = \sigma_1^2 S_t + \sigma_0^2(1 - S_t), \qquad (2.7)$$

so under regime $S_t = i$ the parameters are given by $\theta_i = (\beta_i, \sigma_i^2)$.

However, the state vector S_t is not known a priori, so we must make distributional assumptions about probability of being in a given state. The various models that follow make different assumptions about the evolution of the random process that determines the regime. Let $p_i = P(S_t = i|\mathcal{F}_{t-1}; \gamma)$ with the restrictions $p_i \geq 0$ and $\sum_{i=1}^r p_i = 1$, where γ contains the parameters associated with the probability law of S_t.

Hence, the mixture density $g(\cdot)$ is really a mixture of the conditional densities, each weighted by the probability of being in that particular state

$$g(y_t; \theta) = \sum_{i=1}^r f(y_t|\theta_i)\, p_i, \qquad (2.8)$$

so that in the case of a mixture of normals the density is

$$g(y_t|\mathcal{F}_{t-1}; \theta, \gamma) = \sum_{i=0,1} (2\pi\sigma_i^2)^{-1/2} \exp\left(\frac{(y - x_t'\beta_i)^2}{2\sigma_i^2}\right) \Pr(S_t = i|\mathcal{F}_{t-1}; \gamma), \qquad (2.9)$$

where $\theta = (\beta_1, \beta_0, \sigma_1^2, \sigma_0^2)'$ is the parameter vector and γ contains additional parameters of the S_t process. Notice that unlike the classical model in (2.2) where the errors are i.i.d., the data in (2.5) are neither independent nor identically distributed.

Consequently, the likelihood function (averaged over the unobserved state vector) is as follows:

$$\log \mathcal{L} = \sum_{t=1}^{T} \left[\log \sum_{i=1}^{r} f(y_t|\theta_i)\, P(S_t = i|\mathcal{F}_{t-1}; \gamma)\right]. \qquad (2.10)$$

So far we assumed that there are two states $(r = 2)$. Theoretically, one could test for the presence of one versus two or a higher number of states. However, standard tests (such as a likelihood ratio test) for testing for the number of regimes do not hold as the Fisher information matrix becomes singular in the presence of unidentifiable parameters (θ_i and γ where i is the second regime). This issue is addressed in Hansen [42], Andrews [43], Andrews and Ploberger [44], and Jeffries [18]. We avoid this issue entirely and assume that there are two states as most applications do.

Another point to note is that depending on how the model is parameterized, autoregressive terms may complicate the likelihood in (2.10). In that case, the current observation will not only depend on the current state but also on past states. We will highlight this issue in more detail in the sections that follow.

In the next few sections we will describe different models for the unobserved state vector S_t and provide algorithms for maximizing (2.10) to derive maximum

likelihood estimates for the model parameters θ and γ and make inferences about the likelihood of each regime. We avoid the issue of hypothesis testing which is a very thorny issue in these types of models.

2.2 Markov with Fixed Transitions

In this model the state variable S_t evolves over time as a discrete time, discrete space Markov process. The transition probability of S_t depends on past values of $S_{t-1}, S_{t-2}, \ldots, S_{t-m}$ and not on past values of y or x. Of course any m-th order Markov process can be converted to a first-order ($m = 1$) process.[2] Thus, from here on we assume that $m = 1$. For a two-state process, we write the transition probabilities as

$$
\begin{aligned}
P(S_t = 1 | S_{t-1} = 1, \mathcal{F}_{t-1}) &= P(S_t = 1 | S_{t-1} = 1) = p(t) = p \\
P(S_t = 0 | S_{t-1} = 0, \mathcal{F}_{t-1}) &= P(S_t = 0 | S_{t-1} = 0) = q(t) = q.
\end{aligned}
\tag{2.11}
$$

The parameters of the model are the $(2N_1 + N_2) \times 1$ vector $\psi = (\theta, \gamma)$ where θ is a $2N_1 \times 1$ vector of the parameters of the two-state linear models (autoregressive parameters and exogenous variables x) and $\gamma_{N_2 \times 1} = (p, q)'$ are the parameters of the Markov chain ($N_2 = 2$). In the case of the Gaussian density the parameter vector is $\psi = (\beta_1', \beta_0', \sigma_1^2, \sigma_0^2, p, q)$. Our goal is to estimate ψ, derive standard error estimates, and derive estimates of each probability $P(S_t = i | \mathcal{F}_{t-1}; \psi)$ and $P(S_t = i | \mathcal{Y}; \psi)$ where $\mathcal{Y} = (y_1, \ldots, y_T)$.

2.2.1 Estimation

Depending on the particular problem, estimation can proceed either by a recursive filter and numerical maximization by a method such as Newton Raphson or by an EM algorithm [8], which is appropriate for maximizing a likelihood with

---

[2]However the cost of this transformation is that the chain now has 2^m elements.

unobserved variables or missing observations. We describe these procedures in the next two sections.

2.2.1.1 Nonlinear Filter

A nonlinear filter is sometimes a more convenient means of estimation especially when data are serially correlated. The exact form of the filter depends on the number of autoregressive terms. In the simplest case where there are no autoregressive lags, the filter proceeds as follows. Starting with $\Pr(S_{t-1} = i|\mathcal{F}_{t-1})$ at the beginning of time t we compute

$$
\begin{aligned}
\Pr(S_t = j|\mathcal{F}_{t-1}) &= \sum_{i=0,1} \Pr(S_t = j, S_{t-1} = i|\mathcal{F}_{t-1}) \\
&= \sum_{i=0,1} \Pr(S_t = j|S_{t-1} = i)\ \Pr(S_{t-1} = i|\mathcal{F}_{t-1})
\end{aligned}
\tag{2.12}
$$

and once y_t is observed[3] the probability terms are updated via Bayes' formula:

$$
\begin{aligned}
P(S_t = j|\mathcal{F}_t) &= Pr(S_t = j|\mathcal{F}_{t-1}, y_t) \\
&= \frac{f(S_t = j, y_t|\mathcal{F}_{t-1})}{f(y_t|\mathcal{F}_{t-1})} \\
&= \frac{f(y_t|S_t = j, \mathcal{F}_{t-1})Pr(S_t = j|\mathcal{F}_{t-1})}{\sum_{i=0,1} f(y_t|S_t = i, \mathcal{F}_{t-1}) \cdot Pr(S_t = i, \mathcal{F}_{t-1})},
\end{aligned}
\tag{2.13}
$$

from which we can compute the likelihood via equation (2.10). A convenient byproduct of this method is that the filtered probabilities are readily available.

Assuming the Markov chain is stationary and ergodic, we can initialize the filter using the unconditional distribution of the first observation

$$
\pi_0 = Pr(S_0) \quad = \quad \frac{1-p}{2-p-q}
\tag{2.14}
$$

---

[3]Note that x_t does not enter in the calculation of the probability terms.

$$\pi_1 = Pr(S_1) \quad = \quad \frac{1-q}{2-p-q}. \tag{2.15}$$

These are the steady state probabilities derived from solving the system

$$\pi = (A'A)^{-1}A'e, \tag{2.16}$$

where

$$A = \begin{pmatrix} I_r & - & \Pi \\ 1 & \dots & 1 \end{pmatrix}_{(r+1) \times r}, \tag{2.17}$$

r denotes the number of states (here $r = 2$), and e is an $(r+1) \times 1$ of ones. The transition matrix Π also contains information about the expected duration of each of the two regimes. Solving the equation $E(D_i) = \sum_{i=1}^{\infty} i \Pr(D = i)$ we find that

$$D_i = \frac{1}{1 - \Pi_{i,i}}. \tag{2.18}$$

Thus, the expected duration of each state remains constant throughout time and the higher the transition probability $\Pi_{i,i}$ the longer the expected duration for that regime.

In the case where there are autoregressive lags, the filter equations become more complicated as the density of the dependent variable y_t depends not just on the current state S_t but also on previous states $S_t, \dots, S_{t-p}$. To see this, consider the simplest case of the AR(1) model, $y_t - \mu_{S_t} = \phi_1(y_{t-1} - \mu_{S_{t-1}}) + \epsilon_{S_t}$, whose conditional density is

$$f(y_t|\mathcal{F}_{t-1}, S_t, S_{t-1}) = (2\pi\sigma_{S_t}^2)^{-1/2} \exp\left(\frac{[(y_t - \mu_{S_t}) - \phi_1(y_{t-1} - \mu_{S_{t-1}})]^2}{-2\sigma_{S_t}^2}\right) \tag{2.19}$$

Thus, in the AR(1) the density of y_t depends on both S_t and S_{t-1}, and similarly for a AR(p) model the density depends on $p + 1$ states. This means that the number

of weighting terms grows exponentially to $2^{(p+1)}$ as more AR terms are added. For instance, the AR(4) model on GNP growth considered in Hamilton [4], there are $2^5 = 32$ possibilities to consider. For the AR(1) model, the filter proceeds as follows. Given $\Pr(S_{t-1}|\mathcal{F}_{t-1})$, we compute

$$\Pr(S_t = j, S_{t-1} = i|\mathcal{F}_{t-1}) = \Pr(S_t = j|S_{t-1} = i)\Pr(S_{t-1}|\mathcal{F}_{t-1}) \qquad (2.20)$$

and augmenting with y_t we compute

$$\Pr(S_t = j, S_{t-1} = i|\mathcal{F}_t) = \frac{\Pr(S_t = j, S_{t-1} = i, y_t|\mathcal{F}_{t-1})}{f(y_t|\mathcal{F}_{t-1})} \qquad (2.21)$$

$$= \frac{f(y_t|S_t = j, S_{t-1} = i, \mathcal{F}_{t-1}) \cdot \Pr(S_t = j, S_{t-1} = i|\mathcal{F}_{t-1})}{\sum_{i=0,1}\sum_{j=0,1} f(y_t|S_t = j, S_{t-1} = i, \mathcal{F}_{t-1}) \cdot \Pr(S_t = j, S_{t-1} = i|\mathcal{F}_{t-1})}, \qquad (2.22)$$

from which we compute the log likelihood as

$$\log L = \sum_{t=1}^{T} \log \left(\sum_{i=0,1}\sum_{j=0,1} f(y_t|S_t = j, S_{t-1} = i, \mathcal{F}_{t-1}) \cdot \Pr(S_t = j, S_{t-1} = i|\mathcal{F}_{t-1}) \right)$$

$$(2.23)$$

and probabilities

$$\Pr(S_t = j|\mathcal{F}_t) = \sum_{i=0}^{1} \Pr(S_t = j, S_t = i|\mathcal{F}_t) \qquad (2.24)$$

and

$$\Pr(S_t = j|\mathcal{F}_{t-1}) = \sum_{i=0}^{1} \Pr(S_t = j, S_t = i|\mathcal{F}_{t-1}). \qquad (2.25)$$

Finally, it is interesting to note the similarity between the filter algorithm for the Markov process and the Kalman filter. Whereas in the Kalman filter the unobserved state variable is a linear function with a normally distributed random error, in the Markov-switching process the state variable is a discrete Markov chain.

2.2.1.2 EM Algorithm

Hamilton [45] develops an EM algorithm for the case of no autoregressive variables and shows that this method is robust to different starting points.[4] In the case of a linear model with no autoregressive parameters, the EM algorithm yields some intuitive results.

The first step is to maximize the expected log likelihood function

$$M(\psi; \mathcal{Y}, \psi^k) = \sum_{S_t=1}^{r} \sum_{t=1}^{T} \log f_{Y,S}(y_t, S_t | \mathcal{F}_{t-1}; \psi) \Pr(S_t | \mathcal{F}_t; \psi^k), \qquad (2.26)$$

where ψ^k are the parameter estimates from the $(k-1)$-th iteration, $f_{Y,S}(\cdot)$ is the joint density of y_t and S_t, and $\mathcal{Y} = (y_1, \ldots, y_T)$. It is important to note that y_t and S_t are independent. The density $f_{Y,S}(\cdot)$ can be decomposed as

$$f_{Y,S}(y_t, S_t | \mathcal{F}_{t-1}; \psi) = f(y_t | S_t, \mathcal{F}_{t-1}) \cdot \Pr(S_t | \mathcal{F}_{t-1}), \qquad (2.27)$$

so (dropping $\mathcal{F}_{t-1}$ and $\mathcal{F}_t$) the likelihood with two regimes becomes

$$M(\psi; \mathcal{Y}, \psi^k) = \sum_{i=0}^{1} \sum_{t=1}^{T} \log \left[f(y_t | S_t = i; \theta_i) \cdot \Pr(S_t = i; \gamma) \right] \Pr(S_t = i | y_t; \psi^k). \quad (2.28)$$

The separation of the parameters makes maximization straightforward:

$$\frac{\partial}{\partial \theta_i} M(\psi; \mathcal{Y}, \psi^k) = \sum_{t=1}^{T} \frac{\partial}{\partial \theta_i} \log f(y_t | S_t = i; \theta_i) \Pr(S_t = i | y_t; \psi^k) = 0, \qquad (2.29)$$

which in the case of normally distributed data leads to:

$$\beta_i^{k+1} = \left(\sum_{t=1}^{T} x_t' x_t \Pr(S_{t=1} = i | \mathcal{Y}; \theta^k) \right)^{-1} \left(\sum_{t=1}^{T} x_t y_t \Pr(S_{t=1} = i | \mathcal{Y}; \theta^k) \right) \quad (2.30)$$

---

[4]Another practical issue with EM estimation is the slow rate of convergence.

$$\sigma_i^{2^{k+1}} = \frac{\sum_{t=1}^{T}(y_t - x_t'\beta_i^k)^2 \Pr(S_t = i|\mathcal{Y};\theta^k)}{\sum_{t=1}^{T} \Pr(S_t = i|\mathcal{Y};\theta^k)} \tag{2.31}$$

$$p_{i,i}^{k+1} = \frac{\sum_{t=1}^{T} \Pr(S_t = j, S_{t-1} = j|\mathcal{Y};\theta^k)}{\sum_{t=1}^{T} \Pr(S_{t=1} = i|\mathcal{Y};\theta^k)} \tag{2.32}$$

The last equation is obtained by Lagrange optimization with the added constraint that $\sum_j p_{i,j} = 1$.

The probabilities $\Pr(S_{t=1} = i|\mathcal{Y};\theta^k)$ are the smoothed probabilities. It is interesting that at each M-step of the algorithm, the solutions for β_i^{k+1} and $\sigma_i^{2^{k+1}}$ are OLS estimates weighted by the probabilities of each regime. Also the formula for the transition probabilities $p_{i,i}^{k+1}$ is simply the empirical transition matrix of the Markov chain using the probabilities as counts.

2.3   Markov Time-Varying Transitions

Diebold et al. [7] extended the fixed transition model to time-varying probabilities:

$$
\begin{aligned}
P(S_t = 1 | S_{t-1} = 1, \mathcal{F}_{t-1}) &= p(t) = \frac{\exp(Z_t' \gamma_0)}{1 + \exp(Z_t' \gamma_0)} \\
P(S_t = 0 | S_{t-1} = 0, \mathcal{F}_{t-1}) &= q(t) = \frac{\exp(Z_t' \gamma_1)}{1 + \exp(Z_t' \gamma_1)},
\end{aligned}
\tag{2.33}
$$

where the vector Z_t [5] and $\gamma_i, i = 0, 1$ are vectors of parameters to be estimated. In the case of no additional covariates, this model trivially reduces to the fixed probability Markov switching model. Another parametrization that has been used in this problem is the cumulative normal density function but it will not be considered here.

2.3.1   Estimation

Diebold et al. use the EM algorithm for estimating the model but this algorithm is considerably more complicated that the one for the fixed-transition Markov process. The parameters in this model are the parameters of the linear model (θ), the parameters of the switching process (γ), and the first observation probabilities $\rho_j = P(S_1 = j)$. Whereas in the fixed-transition model, ρ_j was determined by the steady state probabilities π, in this model ρ_j contains additional parameters that need to be estimated. Rather than outline the steps of this algorithm, we refer the

---

[5]Some authors denote this vector by Z_{t-1} so the time subscript is a matter of style. This vector can contain contemporaneous exogenous variables like x_t but can only contain dependent data through y_{t-1} (does not include y_t).

reader to Diebold et al. [7] for more details.

The non-linear filter approach is much more straightforward. An example of this application is Filardo [9], which uses a non-linear filter to model U.S. industrial production as a function of various leading indicators.

2.4 Logistic Mixtures Model

The logistic mixtures model was developed and analyzed by Jeffries and Kedem in [18] for a mixture of Generalized Linear Models (GLM). Wong and Li [19] also use logistic regression to get a mixture of normally distributed autoregressive models. The general form of the exponential density considered by Jeffries is

$$f(y_t|S_t = i, \mathcal{F}_{t-1}; \beta_i, \phi_i) = \exp([y_t X'_{t,i}\beta_i - b(X'_{t,i}\beta_i)]/\phi_i + c_i(y_t, \phi_i)), \tag{2.34}$$

which for the normal probability density function reduces to

$$f(y_t|S_t = i, \mathcal{F}_{t-1}; \mu_i, \sigma_i) = (2\pi\sigma_i^2)^{-1/2} \exp(\frac{(y_t - \mu_i)^2}{2\sigma_i^2}). \tag{2.35}$$

The probability of the given regime is given by

$$\Pr(S_t = 1|Z_t, \mathcal{F}_{t-1}; \gamma) = \frac{\exp(Z'_t\gamma)}{1 + \exp(Z'_t\gamma)} \tag{2.36}$$

where Z_t is a $k \times 1$ vector of known covariates [6] and γ is an unknown vector of regression parameters. As in the Diebold et al model, it allows the probabilities to be time varying and to depend on additional covariates which may have useful information about the unobserved state but does not restrict the stochastic state variable to be a Markov random variable. [7] This simplifies the estimation considerably even in the presence of autoregressive parameters.

---

[6]This is sometimes denoted as Z_{t-1}. It may contain contemporaneous data x_t but dependent data only through y_{t-1}.

[7]Notice that unlike the Markov model, which had two sets of vectors γ_0 and γ_1, the logistic mixtures model has only one vector γ.

2.4.1 Estimation

The EM algorithm is used to solve for the unobserved parameters by forming the *expected* likelihood

$$
\begin{aligned}
M(\psi, \mathcal{Y}, \psi^k) &= \sum_{t=1}^{T} \log f_{Y,S}(y_t, S_t = 1 | \mathcal{F}_{t-1}; \psi) \cdot \Pr(S_t = 1 | \mathcal{F}_t; \psi^k) + \\
&\quad \sum_{t=1}^{T} \log f_{Y,S}(y_t, S_t = 0 | \mathcal{F}_{t-1}; \psi) \cdot \Pr(S_t = 0 | \mathcal{F}_t; \psi^k) \qquad (2.37)
\end{aligned}
$$

where $\psi = (\beta_1', \beta_0', \phi_1, \phi_0, \gamma)'$ is a vector of the model's parameters, ψ^k is an estimate of these parameters from the previous iteration k, $f_{Y,S}(\cdot)$ is the joint density of y_t and S_t, and $\mathcal{Y} = (y_1, \ldots, y_T)$.

2.4.1.1 E-Step

The E-step consists of computing $M(\psi, \mathcal{Y}, \psi^k)$ given the value of the parameter vector ψ^k from the k-th iteration. Using the fact that y_t and S_t are independent, we expand the joint density into

$$
\begin{aligned}
\log f_{Y,S}(y_t, S_t = i | \mathcal{F}_{t-1}; \psi) &= \log \left(f(y_t | S_t = i, \mathcal{F}_{t-1}; \beta_i, \phi_i) \cdot \Pr(S_t = 1 | \mathcal{F}_{t-1}; \gamma) \right) \\
&= \log f(y_t | S_t = i, \mathcal{F}_{t-1}; \beta_i, \phi_i) + \log \Pr(S_t = 1 | \mathcal{F}_{t-1}; \gamma),
\end{aligned}
$$

and substituting into the expected likelihood we get

$$
\begin{aligned}
M(\psi, \mathcal{Y}, \psi^k) &= \sum_{t=1}^{T} p_t^k \cdot \left(\log f(y_t | S_t = 1, \mathcal{F}_{t-1}; \beta_1, \phi_1) + \log \Pr(S_t = 1 | \mathcal{F}_{t-1}; \gamma) \right) + \\
&\quad \sum_{t=1}^{T} (1 - p_t^k) \cdot \left(\log f(y_t | S_t = 0, \mathcal{F}_{t-1}; \beta_0, \phi_0) + \log \Pr(S_t = 1 | \mathcal{F}_{t-1}; \gamma) \right),
\end{aligned}
$$

where $p_t^k = \Pr(S_t = 1 | y_t, \mathcal{F}_{t-1}; \psi^k)$, which is updated via Bayes' formula

$$
p_t^k = \Pr(S_t = 1 | y_t, \mathcal{F}_{t-1}; \psi^k) = \frac{g(y_t, S_t = 1 | \mathcal{F}_{t-1}; \psi^k)}{g(y_t | \mathcal{F}_{t-1}; \psi^k)}
$$

$$= \frac{f(y_t|S_t = 1, \mathcal{F}_{t-1}; \beta_1^k, \phi_1^k) \cdot \Pr(S_t = 1|\mathcal{F}_{t-1}; \gamma^k)}{\sum_{i=0,1} f(y_t|S_t = i, \mathcal{F}_{t-1}; \beta_1^k, \phi_1^k) \cdot \Pr(S_t = i|\mathcal{F}_{t-1}; \gamma^k)}. \quad (2.38)$$

Notice the distinction between the two probability values

$$\Pr(S_t = 1|y_t, \mathcal{F}_{t-1}; \psi^k) = p_t^k \quad (2.39)$$

$$\Pr(S_t = 1|, \mathcal{F}_{t-1}; \psi^k) = \frac{\exp(Z_t'\gamma^k)}{1 + \exp(Z_t'\gamma^k)}. \quad (2.40)$$

The probability weight p_t^k is conditional on the contemporaneous value of y_t while the latter only contains information up to time $t - 1$.

2.4.1.2 M-Step

The M-step consists of calculating

$$\psi^{k+1} = \arg\max_{\psi} M(\psi, \mathcal{Y}, \psi^k), \quad (2.41)$$

which can be separated into three different maximization problems

$$H_1(\beta_1, \phi_1, p^k) = \sum_{t=1}^{T} p_t^k \cdot \log f(y_t|S_t = 1, \mathcal{F}_{t-1}; \beta_1, \phi_1), \quad (2.42)$$

$$H_0(\beta_0, \phi_0, p^k) = \sum_{t=1}^{T} (1 - p_t^k) \cdot \log f(y_t|S_t = 0, \mathcal{F}_{t-1}; \beta_0, \phi_0), \quad (2.43)$$

$$H_2(\gamma, p^k) = \sum_{t=1}^{T} p_t^k \cdot \log \Pr(S_t = 1|\mathcal{F}_{t-1}; \psi) + (1 - p_t^k) \cdot \log \Pr(S_t = 0|\mathcal{F}_{t-1}; \psi), \quad (2.44)$$

where $p^k = (p_1^k, ..., p_t^k, ..., p_T^k)$. The first two equations (2.43) and (2.42) can be solved for $(\beta_i, \phi_i)'$ using a weighted GLM with prior weights p_t^k and (2.44) can be solved for γ^{k+1} using a logistic regression with dependent variables p_t^k and $1 - p_t^k$ or a Newton-Raphson optimization. For a mixture of two normal densities, the M-step estimates of the linear part are completely analogous to the fixed-transition Markov model as in (2.30).

2.5   Filtered and Smoothed Estimates

In addition to estimating the parameters of the models, one is interested in estimating the probabilities of the unobserved state vector S_t at each time t.

For the Markov models making inferences on S_t requires a recursive procedure. Given the maximum likelihood estimates, computing filtered state probabilities $P(S_t = j|\mathcal{F}_t)$ is a straightforward recursive procedure:

$$P(S_t = j|\mathcal{F}_{t-1}) = \sum_{i=0}^{1} P(S_t = j|S_{t-1} = i) \cdot P(S_{t-1} = i|\mathcal{F}_{t-1}) \tag{2.45}$$

$$P(S_t = j|\mathcal{F}_t) = \frac{f(y_t|S_t = j, \mathcal{F}_{t-1}) \cdot P(S_t = j|\mathcal{F}_{t-1})}{\sum_{j=1}^{1} f(y_t|S_t = j, \mathcal{F}_{t-1}) \cdot P(S_t = j|\mathcal{F}_{t-1})} \tag{2.46}$$

Additionally, we can use the entire sample to estimate the state vector as in Kim [32] and [33]:

$$P(S_t = j, S_{t+1} = k|\mathcal{F}_T) = \frac{P(S_{t+1} = k|\mathcal{F}_T) P(S_t = j|\mathcal{F}_t) P(S_{t+1} = k|S_t = j)}{P(S_{t+1} = k|\mathcal{F}_t)} \tag{2.47}$$

$$P(S_t = j|\mathcal{F}_T) = \sum_{k=0}^{1} \Pr(S_t = j, S_{t+1} = k|\mathcal{F}_T). \tag{2.48}$$

In the logistic mixtures model this is more directly derived from equations (2.39). There are no smoothed probabilities for this model.

Chapter 3

Simulation Results

In this chapter we apply the models to simulated data.

3.1   Simulated Data

We simulate time series data from three different models:

1. Markov model with fixed probability transitions (MFP)

2. Markov model with time-varying transitions (MTV)

3. Independent logistic mixtures model (LMX)

For each model we simulated 200 time series of length $T = 500$[1] from two states $(r = 2)$ and got $y_t^{(1)}$ and $y_t^{(0)}$, the state vector S_t, and the mixed series $y_t = S_t y_t^{(1)} + (1 - S_t) y_t^{(0)}$ for $t = 1, \ldots, T$. Additionally, we simulate an exogenous time series $\xi_t \sim$ i.i.d. $N(0, 1)$ that is common to all models.

The linear part of the models is given by:

$$y_t = \beta_i' x_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_i^2), \qquad i = 0, 1 \tag{3.1}$$

---

[1]We actually a longer time series with $T = 600$ and discarded the first 100 observations.

where $x_t = (1 \quad \xi_t \quad y_{t-1})'$. The linear parameters (β_i and σ_i^2) are constant across all three models and are given by:

$$\beta_0 = \begin{pmatrix} 1.0 \\ 1.0 \\ -0.1 \end{pmatrix}, \quad \beta_1 = \begin{pmatrix} -1.0 \\ 0.8 \\ 0.2 \end{pmatrix}, \quad \begin{pmatrix} \sigma_0^2 \\ \sigma_1^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}. \tag{3.2}$$

Thus, in one regime the model has a positive intercept, low error variance, and a slightly negative autoregressive term. In the second regime, the model has a negative intercept, high residual variance, a lower coefficient of ξ_t, and a positive autoregressive term.

The linear part is again the same for all three models. The only difference between these models is in the γ parameters that characterize the unobserved state vector S_t. These sets of parameters were chosen with two objectives in mind. First, both regimes should occur in the sample so that the sample is not all due to one regime. Second, the regimes should have some persistence across time. A state variable that switches very frequently between two regimes would be very hard to identify, especially for the Markov-switching models.

3.1.1 Markov with Fixed Transitions

The state vector S_t is a discrete time, discrete space Markov chain with fixed transition probabilities

$$
\begin{aligned}
p &= P(S_t = 1 | S_{t-1} = 1) = 0.90 \\
q &= P(S_t = 0 | S_{t-1} = 0) = 0.75.
\end{aligned}
\tag{3.3}
$$

so $\gamma = (2.20, 1.10)'$.[2]

3.1.2 Markov with Time-Varying Transitions

The linear part is again given by (3.1) with the parameters given in (3.2). The difference is in the transition probability matrix, which is time-varying and is parametrized as follows:

$$
\begin{aligned}
p(t) &= P(S_t = 1 | S_{t-1} = 1) = \frac{\exp\left(Z_t' \gamma_1\right)}{1 + \exp\left(Z_t' \gamma_1\right)} \\
q(t) &= P(S_t = 0 | S_{t-1} = 0) = \frac{\exp\left(Z_t' \gamma_0\right)}{1 + \exp\left(Z_t' \gamma_0\right)},
\end{aligned}
\tag{3.4}
$$

where $Z_t = (1 \quad y_{t-1})'$ are the explanatory variables and γ_i are the parameters of the transition matrix

$$
\gamma_1 = \begin{pmatrix} 2.0 \\ 0.7 \end{pmatrix}, \quad \gamma_0 = \begin{pmatrix} 1.0 \\ -0.8 \end{pmatrix}.
\tag{3.5}
$$

Notice again that this model nests the fixed parameter case by choosing the explanatory vector to be $Z_t = (1)'$.

3.1.3 Logistic Mixtures

The probabilities in this model are independent and are given by equation (2.36) with $Z_t = (1 \quad y_{t-1})'$ and

$$
\gamma = \begin{pmatrix} -2 \\ -3 \end{pmatrix}.
\tag{3.6}
$$

---

[2]The two sets of parameters are equivalent under the constraint $P(S_t = i | S_{t-1} = i) = \frac{\exp(\gamma_i)}{1+\exp(\gamma_i)}$.

Notice that this is the same explanatory vector as in the time-varying Markov model.

These coefficients were chosen in order to make the simulations comparable to the simulated sample based on Markov switching. More specifically, we chose the coefficients after some trial and error in order to make the conditional probabilities of state 1 ($\Pr(S_t = 1)$) high when y_t was low.

3.1.4 Summary of Simulated Data

3.1.4.1 Example of One Simulation Run

To get a feel for the data simulated under assumptions 3.2, 3.3, 3.5, and 3.6, we show one sample path from the logistic mixture model. The other two Markov simulations produce results that are qualitatively very similar.

In Figure 3.1 the binary state variable S_t is depicted in the shaded regions (left scale) and the dependent variable y_t is shown on the right scale. State $S_t = 1$ shown in the shaded bars occurs around one-third of the time (34%). Notice that y_t tends to be more negative when $S_t = 1$. This is in line with the assumed parameters where we associated state 1 with a negative mean and a higher variance than state 0. This can more easily be seen in the bottom panel of Figure 3.1, where we plot the cumulative sum $\sum_{s=1}^{t} y_s$. The cumulative sum trends up in the white regions, which correspond to $S_t = 0$, and trends down in the shaded regions that correspond to $S_t = 1$. Finally, the observations of the state vector cluster together across time – when $S_t = 1$ it is more likely to be followed by another observation from that state.

Figure 3.2 summarizes the sample data from the logistic mixture model. The boxplot confirms that y_t behaves differently in the two states. State $S_t = 0$ has a higher median and lower interquartile range than state 1. The scatter plot of the independent series ξ_t against the dependent variable y_t shows almost two linear clusters of data: one cluster (blue x's) corresponds to observations from state 0; the other (red dots) corresponds to state 1. A scatter plot of y_t versus y_{t-1} (not shown here) did not reveal any strong patterns due to the amount of noise in the models. The autocorrelation plot reveals that the data is highly correlated across periods. The one-lag autocorrelation is 0.44 and the correlations are significantly different from zero out to five lags at the 5% confidence level. Finally, the normal plot shows that the sample data deviates from the normal distribution. In addition, a histogram of the data (not shown here) is slightly skewed to the left.

Finally, it is interesting to note that the first-order autocorrelation is much higher that either of the two autoregressive coefficients (-0.10 and 0.20) or their state-weighted sum. Most of the autocorrelation is due to the serial dependence of the state vector.[3] I was not able to derive an analytical solution to determine how much of the autocorrelation in the observed series was due to serial dependence of the state vector and how much was due to the linear part of the model.

---

[3]Even though the actual process is not Markov, we computed the empirical transition rates between the two states to be $\hat{\Pr}[S_t = 1 | S_t = 1] = 0.65$ and $\hat{\Pr}[S_t = 0 | S_t = 0] = 0.82$, which imply expected durations of 2.9 and 5.7 periods, respectively. Thus the process is not very persistent.

3.1.4.2 Summary of All Simulation Runs

Table (3.1) summarizes all 200 simulated series. It shows average values for each model, and p-values are shown in brackets. All three datasets exhibit similar characteristics. They have positive mean, variance of close to 4, negative skewness, and kurtosis that is higher than the normal distribution. The Jarque-Bera statistic [46], which tests for departure from normality, is significant at the 1% confidence level. All three series exhibit positive serial correlation, and the Ljung-Box portmanteau test ($Q(10)$) is significant at the 5% level out to 10 lags. The ARCH test [12], which tests for heteroskedasticity, is not statistically significant for any of the models. On average, the state variable S_t is in state 1 approximately 28%-31% of the time. The state variables are persistent in all three models as evidenced by the empirical transition probabilities, which range from 63% to 90%.

	MFP	MTV	LMX
Mean	0.35	0.18	0.16
Variance	3.76	4.12	4.19
Skewness	-0.55	-0.57	-0.58
Kurtosis	3.66	3.56	3.56
Jarque-Berra	[0.00]	[0.00]	[0.00]
$\rho_1(y_t)$	0.17	0.33	0.41
$\rho_2(y_t)$	0.09	0.23	0.20
$\rho_3(y_t)$	0.08	0.12	0.13
Q(10)	[0.05]	[0.00]	[0.00]
ARCH(10)	[0.33]	[0.17]	[0.18]
$P(S_t = 1)$	0.28	0.33	0.31
$P(S_t = 1|S_{t-1} = 1)$	0.74	0.80	0.63
$P(S_t = 0|S_{t-1} = 0)$	0.90	0.90	0.84

Table 3.1: Descriptive Statistics on Simulated Data.

Average values for 200 simulations under each model using the parameters provided earlier in the section. P-values are shown in brackets. The j-th order residual autocorrelation is denoted by $\rho_j(y_t)$, Q(10) is the Ljung-Box portmanteau test on residuals, and ARCH(10) is the Engle (1982) test for heteroskedasticity using 10 lags of data. MFP and MTV are the Markov models with fixed-transition and time-varying transitions, respectively, and LMX is the logistic-mixtures model.
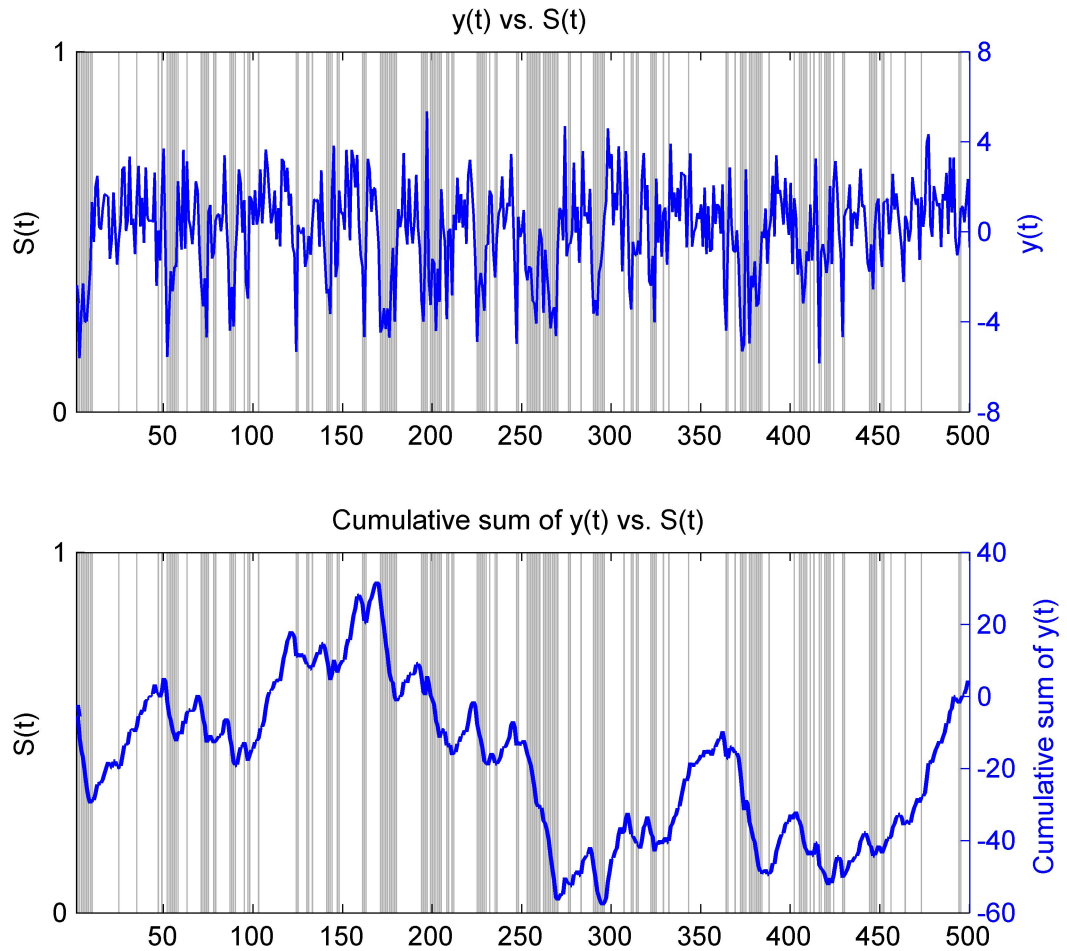
Figure 3.1: Sample Data from Logistic Mixture Model.

Observations where the state variable $S_t = 1$ are denoted by gray bars (left axis) and white gaps correspond to $S_t = 0$. Variables y_t and cumulative sum $\sum_{s=1}^{t} y_s$ are shown on the right axes in blue color.
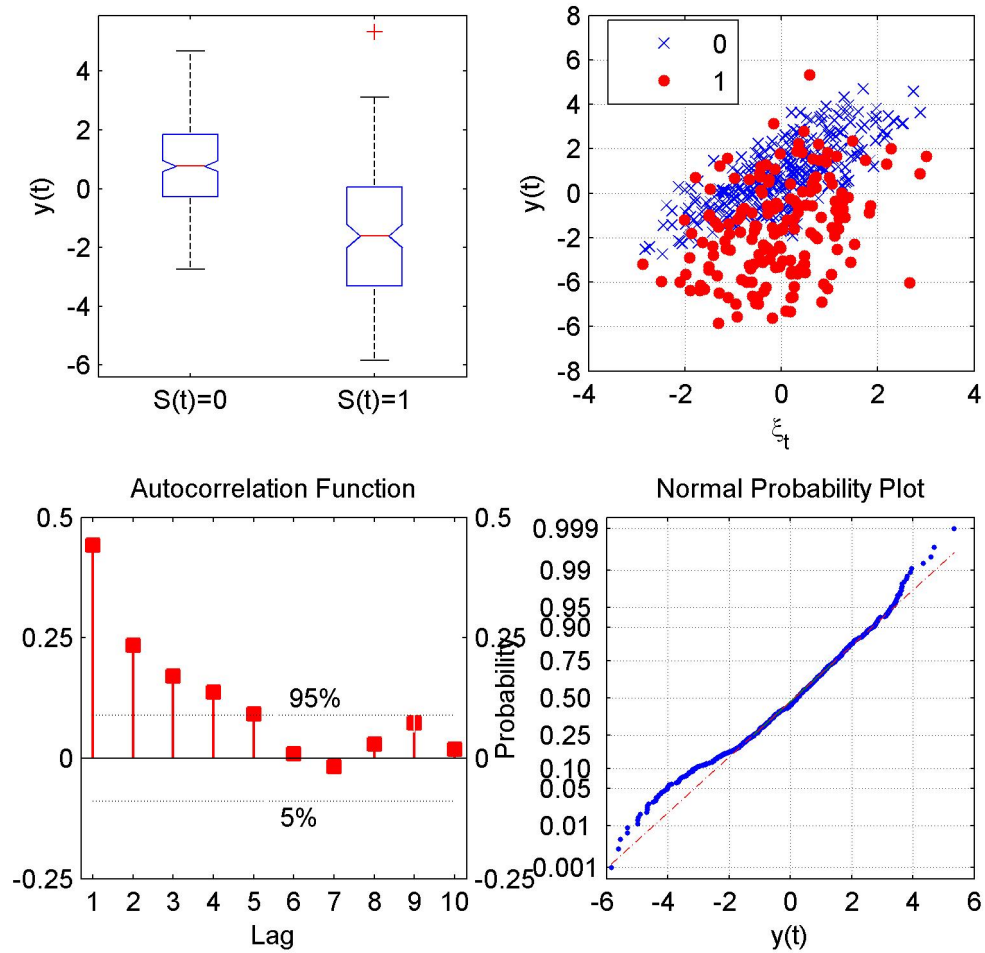
Figure 3.2: Sample Statistics from Logistic Mixture Model.

3.2 Simulation Results

Each of the three models was estimated using the three simulated datasets (for a total of nine data/model combinations). The Markov models were estimated using numerical methods to maximize the likelihood, and the logistic mixture model was estimated using the EM algorithm described previously. Although there is an autoregressive term in the linear part of the model, each obsevation depends on only the contemporaneous state S_t.

Convergence was determined based on both the gradient of the log likelihood $\|\nabla \log \mathcal{L}\|_2^2 \leq \epsilon$ and the change in the likelihood $|\Delta \log \mathcal{L}| \leq \epsilon$ with $\epsilon = 0.01$. [4]

Asymptotic standard errors σ_{ASYMPT} were computed using the Hessian of the likelihood to estimate the Cramér-Rao lower bound. This worked in all but a few cases where the Hessian was ill-conditioned or singular. In the fixed-transition Markov model I also transformed the parameters back to probabilities in the $(0,1)$ interval using the transformation $\theta_j = g(\psi_j) = \frac{\exp(\psi_j)}{1+\exp(\psi_j)}$ for each component j, and I computed the covariance for these parameters using

$$Cov(\hat{\theta}) = \left(\frac{\partial g(\hat{\psi})}{\partial \psi}\right) Cov(\hat{\psi}) \left(\frac{\partial g(\hat{\psi})}{\partial \psi}\right)'. \tag{3.7}$$

A second way of computing standard errors is by using all the simulations to compute to compute a dispersion around the average estimate:

$$\sigma_{SIMUL} = \left(\frac{1}{M-1}\sum_{j=1}^{M}(\psi_j - \bar{\psi})^2\right)^{1/2}. \tag{3.8}$$

---

[4] Initially I also used $\max |\nabla \log \mathcal{L}| \leq \epsilon$ (infinity norm) as an additional criterion for convergence but because it was found to be too restrictive for some models that had otherwise converged to a solution, I dropped it.

This provides a check on the calculations of the standard errors of the estimates based on simulations, but when using real data only asymptotic standard errors would be available.

Before going into the results of all 200 simulations, it is instructive to look at a specific example. In Section 3.1.4 we saw one realization from the logistic mixture sample. Using this series along with the explanatory variables x_t (but ignoring the state variable S_t which would not be observable), we fit the logistic mixture model, by maximizing the likelihood as shown in Section 2.4.1. All the estimated values were close to the true parameters and significantly different from zero. Figure 3.3 displays the fit to this model along with the probabilities of the unobserved state. Here again, we show the cumulative sum of y in order to better visualize the results. The in-sample fit $\hat{y}_t$ traces the movements of the dependent variable but deviates at times. The in-sample root mean squared error of the in-sample prediction (not cumulative) 1.22 – just over half the standard deviation of y.

The two bottom panels compare the estimated probabilities $\Pr[S_t = 1|\mathcal{F}_{t-1}]$ and $\Pr[S_t = 1|\mathcal{F}_t]$ (blue lines) against the true state vector, which is shown as a shaded region when $S_t = 1$. Both estimates do a very good job of matching the unobserved state. That is, when $S_t = 1$ the probabilities rise to near 1 (average estimated probability is 0.89) and are low when $S_t = 0$ (average estimated probability is 0.03). As expected, the updated probability $\Pr[S_t = 1|\mathcal{F}_t]$ has a better fit.

Figure 3.4 graphs the residuals in various forms. The residuals are centered close to zero[5] but they have a wider range in state 1 than state 0 – the variance

---

[5]Actually in $S_t = 1$ the mean of the residuals is slightly positive, while in $S_t = 0$ they are
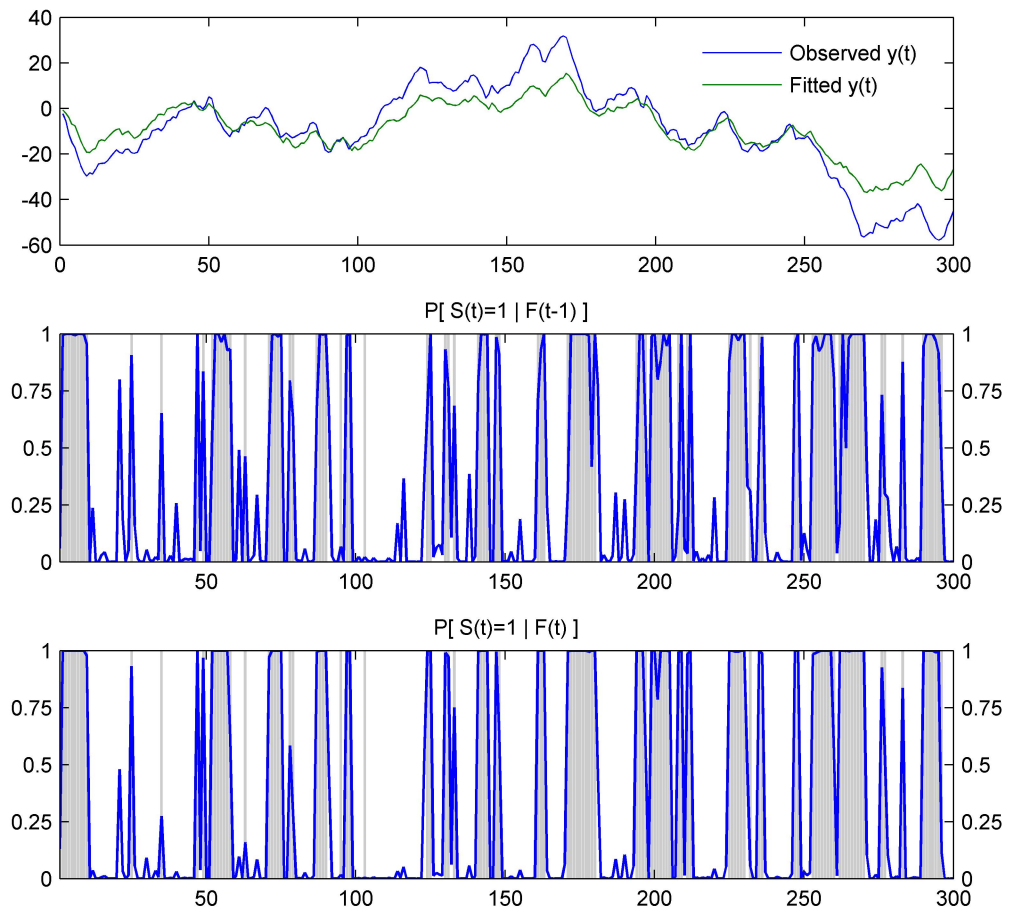
Figure 3.3: Logistic Mixture Model Fit.

Only first 300 of 500 observations are shown for more clarity.

is almost twice as large. This suggests heteroskedasticity across the two states. A scatter plot of the independent variable ξ_t against y_t shows that the residuals are not related to the independent variable or to the state variable. The model residuals have very low autocorrelation so it would seem that the model accounts for most of the autocorrelation of the dependent variable. However, they are not normally distributed as evidenced by the normal quantile plot.

---

slightly negative, but in the pooled data the mean of the residuals is exactly zero.
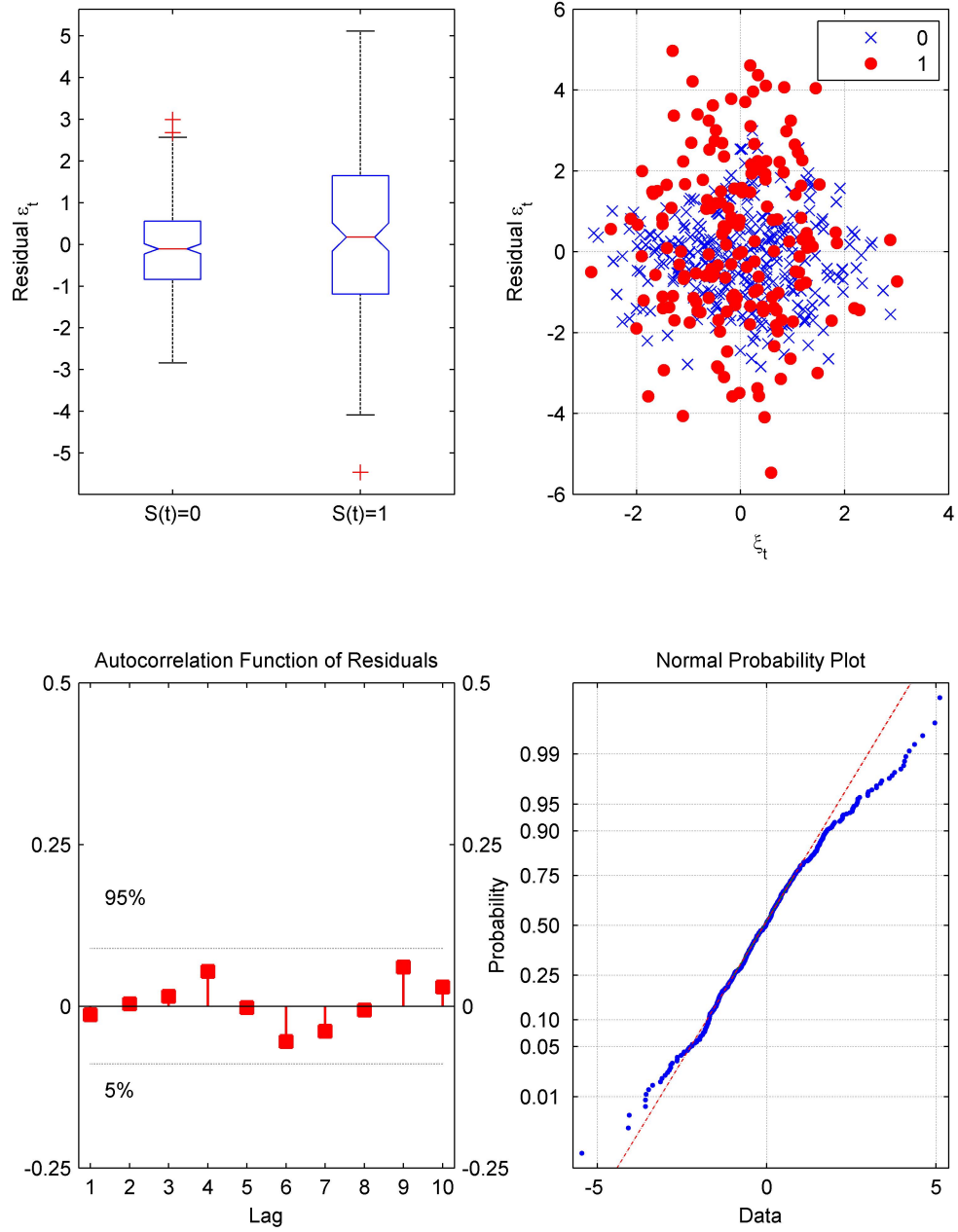
Figure 3.4: Logistic Mixture Model Residual Diagnostics

3.2.1 Parameter Estimates on MFP Data

Table (3.2) shows the estimated parameters of all three models based on the dataset generated by the fixed-transition Markov process (MFP). Estimates are averaged over 200 simulations of each model, of which #Conv iterations converged to a solution. The table shows the average maximum likelihood estimates $(\hat{\psi})$, asymptotic standard errors σ_A, and simulated standard errors σ_S for each of the three models. The true parameter models are shown in the left column. Estimates shown in bold font indicate bias from the true parameter values as measured by the 95% confidence interval $\hat{\psi} \pm 1.96 \frac{\hat{\sigma}_S}{\sqrt{\#CONV}}$. For AIC and BIC the number of times each model achieved the minimum among the three models is shown in brackets. The transition parameters for the LMX model are shown separately and since they have different units they cannot be compared directly to the true parameters.

Not surprisingly, of the three models, the fixed-transition model performs the best since it exactly matches the data generating process. It has a lower average AIC and BIC than either of the other two models, and out of the 200 simulations it had the lowest AIC (BIC) 153 (186) times. Overall, all three models give reasonably accurate predictions on the linear part of the model, except for mis-estimating the intercept of state 1 $(\beta_{0,1})$ and the variance for state 1 (σ_1^2). The MFP model estimates the transition coefficients well (which correspond to transition probabilities of 0.75 and 0.90 for state 0 and 1, respectively).

The MTV model, of which the MFP model is a special case, also comes close to the true parameter values but incorrectly estimates the y_{t-1} term in the transition

matrix to be -0.17 (not statistically significant).

With the exception of the transition parameters of the Markov-switching models, all coefficients were statistically significant. Simulated standard errors tended to be close but slightly larger than asymptotic estimates (except for $\gamma_{0,1}$). Also, there seems to be a lot of bias in the estimates, which is indicated in bold font, and this is true even in the correctly specified model. Bias means that the estimates are statistically different from the true parameters (95% confidence interval using simulated standard errors).

Lastly, there were a few simulations that did not converge under the Markov models. The LM model was considerably faster in converging to a solution (on average it took 6 seconds to converge to a solution) than the Markov models (the MTV model was the most time consuming at 51 seconds). We suspect that the MFP model can be made more efficient by replacing the numerical estimation procedure with the EM algorithm outlined above. However, given that in the MFP model one still has to compute filtered and smoothed probabilities recursively, we think that it would still be slower than the LM algrorithm.

Table (3.3) shows that when the three models were applied to data generated by the fixed-transition Markov process, all three models had similar performance characteristics both in root mean square of the observed series y_t and in mean absolute deviation of the state vector S_t (root mean square error was 1.12). The three state vector inferences differ in the set of information on which they are conditional upon: $\mathcal{F}_{t-1}$, $\mathcal{F}_t$, and $\mathcal{F}_T$. As more information in incorporated into the estimates, the absolute deviation of S_t decreases from 0.32 to 0.20, and 0.17 for the smoothed

	True	MFP Model			MTV Model			LMX Model		
		$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$	$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$	$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$
$\beta_{0,1}$	-1.00	**-1.14**	0.36	0.49	**-1.17**	0.36	0.53	**-1.35**	0.54	0.70
$\beta_{0,2}$	0.80	**0.77**	0.19	0.20	**0.77**	0.19	0.21	**0.76**	0.21	0.22
$\beta_{0,3}$	0.20	**0.18**	0.10	0.11	**0.15**	0.11	0.14	**0.11**	0.14	0.16
$\beta_{1,1}$	1.00	0.99	0.08	0.14	0.99	0.08	0.17	0.98	0.09	0.09
$\beta_{1,2}$	1.00	1.00	0.06	0.06	1.02	0.06	0.20	1.00	0.06	0.06
$\beta_{1,3}$	-0.10	-0.09	0.04	0.06	-0.10	0.04	0.04	-0.10	0.04	0.05
σ_1^2	4.00	**3.59**	0.66	0.73	**3.58**	0.64	0.78	**3.49**	0.86	0.96
σ_0^2	1.00	0.99	0.11	0.15	0.98	0.11	0.16	1.01	0.14	0.17
$\gamma_{0,1}$	1.10	1.03	2.41	1.22	**0.95**	0.60	0.90			
$\gamma_{0,2}$	0.00	0.00	0.00	0.00	**-0.17**	0.31	1.15			
$\gamma_{1,1}$	2.20	2.20	0.32	0.34	2.27	0.45	0.53			
$\gamma_{1,2}$	0.00	0.00	0.00	0.00	0.00	0.29	0.38			
γ_1								-1.16	0.42	0.55
γ_2								-0.47	0.15	0.20
$\log\mathcal{L}$		882.34			881.45			891.55		
AIC		1784.68	[153]		1786.91	[45]		1803.09	[2]	
BIC		1826.83	[186]		1837.48	[9]		1845.24	[5]	
#Conv		195			199			200		
Time (sec)		24			51			6		

Table 3.2: Model estimates on data generated by an MFP process.

estimates. There is only a marginal gain in fit by moving from filtered to smoothed estimates; current data y_t and x_t provide reasonable estimates of the probability of being in a given regime.

We also compared the estimated probabilities within each regime to see if the are more accurate in one state versus the other. When the state was at 1, filtered probabilities ranged from 0.59 to 0.64; on the flip side, when the state was zero the probabilities were 0.14.

I also compared the regime-switching estimates [6] for the linear part of the model against Kalman filter (with the linear part correctly specified). The Kalman filter estimates do not capture the abrupt changes in the parameters well. In particular, the autoregressive parameter is consistently higher than either one of two regimes, however it is believed that a structural type model would fit better.

---

[6]Computed as average of two regimes $\hat{\beta}_t = \sum_{i=1}^{1} \hat{\beta}_i P(S_t = i | \mathcal{F}_t)$.

	MFP	MTV	LMX
RMSE($\hat{y}$)	1.12	1.12	1.13
MAD($S_t = 1\|\mathcal{F}_{t-1}$)	0.32	0.32	0.34
$\bar{P}(S_t = 1\|\mathcal{F}_{t-1})$ when $S_t = 0$	0.22	0.22	0.23
$\bar{P}(S_t = 1\|\mathcal{F}_{t-1})$ when $S_t = 1$	0.42	0.42	0.34
MAD($S_t = 1\|\mathcal{F}_t$)	0.20	0.20	0.21
$\bar{P}(S_t = 1\|\mathcal{F}_t)$ when $S_t = 0$	0.14	0.14	0.13
$\bar{P}(S_t = 1\|\mathcal{F}_t)$ when $S_t = 1$	0.64	0.64	0.59
MAD($S_t = 1\|\mathcal{F}_T$)	0.17	0.17	
$\bar{P}(S_t = 1\|\mathcal{F}_T)$ when $S_t = 0$	0.12	0.12	
$\bar{P}(S_t = 1\|\mathcal{F}_T)$ when $S_t = 1$	0.68	0.68	

Table 3.3: Goodness of fit estimates for MFP Data.

MFP and MTV are the Markov models with fixed-transition and time-varying transitions, respectively, and LMX is the logistic mixtures model. Estimates are averaged over 200 simulations of each model. Smoothed probabilities were only computed for the two Markov-transition models.

3.2.2 Parameter Estimates on MTV Data

Table 3.4 shows the three model estimates using the data generated by the time-varying Markov transition model (MTV). Estimates are averaged over 200 simulations of each model, of which #Conv iterations converged to a solution. The table shows the average maximum likelihood estimates ($\hat{\psi}$), asymptotic standard errors σ_A, and simulated standard errors σ_S for each of the three models. The true parameter models are shown in the left column. Estimates shown in bold font indicate bias from the true parameter values as measured by the 95% confidence interval $\hat{\psi} \pm 1.96 \frac{\hat{\sigma}_S}{\sqrt{\#CONV}}$. For AIC and BIC the number of times each model achieved the minimum among the three models is shown in brackets. The transition parameters for the LMX model are shown separately and since they have different units they cannot be compared directly to the true parameters.

As expected, the MTV model, which matches the data generating process, has the lowest AIC and BIC. The AIC does a better job of identifying the correctly specified model. Simulated standard errors were close but generally higher than asymptotic standard errors. A number of the estimates are biased (shown in bold font), which means that they are statistically different from the true parameter values. In particular, the fixed transition Markov model seems to be more biased.

Table 3.5 shows the filtered and smoothed estimates from these models. As expected, the time-varying transition Markov model fit the data slightly better with the estimated filtered probabilities at 0.10 and 0.76 in states 0 and 1, respectively. The fixed-transition Markov model had only slightly worse performance and the

	True	MFP Model			MTV Model			LMX Model		
		$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$	$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$	$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$
$\beta_{0,1}$	-1.00	**-1.13**	0.31	0.42	**-1.14**	0.38	0.48	**-1.23**	0.47	0.61
$\beta_{0,2}$	0.80	**0.76**	0.16	0.17	0.80	0.17	0.18	0.79	0.18	0.19
$\beta_{0,3}$	0.20	**0.27**	0.09	0.12	**0.18**	0.11	0.14	**0.16**	0.13	0.16
$\beta_{1,1}$	1.00	0.96	0.08	0.40	0.99	0.09	0.08	0.99	0.10	0.10
$\beta_{1,2}$	1.00	**0.96**	0.06	0.19	1.00	0.06	0.06	1.00	0.06	0.06
$\beta_{1,3}$	-0.10	**-0.08**	0.04	0.12	-0.10	0.04	0.05	-0.10	0.05	0.05
σ_1^2	4.00	**3.19**	0.48	0.53	**3.71**	0.59	0.63	**3.60**	0.65	0.68
σ_0^2	1.00	0.98	0.11	0.16	0.99	0.11	0.11	1.00	0.12	0.13
$\gamma_{0,1}$	1.00	1.03	2.41	1.22	1.23	1.04	1.59			
$\gamma_{0,2}$	-0.80				-1.64	2.24	7.49			
$\gamma_{1,1}$	2.00	**2.20**	0.32	0.34	**2.17**	0.42	0.43			
$\gamma_{1,2}$	0.70				0.76	0.36	0.41			
γ_1								-1.07	0.40	0.51
γ_2								-0.97	0.21	0.21
$\log\mathcal{L}$		881.21			872.56			885.86		
AIC		1782.41	[14]		1769.12	[182]		1791.71	[4]	
BIC		1824.56	[86]		1819.70	[101]		1833.86	[13]	
#Conv		192			191			200		
Time (sec)		27			60			5		

Table 3.4: Model estimates on data generated by a MTV process.

	MFP	MTV	LMX
RMSE($\hat{y}$)	1.12	1.12	1.26
MAD($S_t = 1 \vert \mathcal{F}_{t-1}$)	0.27	0.22	0.27
$\bar{P}(S_t = 1 \vert \mathcal{F}_{t-1})$ when $S_t = 0$	0.21	0.16	0.19
$\bar{P}(S_t = 1 \vert \mathcal{F}_{t-1})$ when $S_t = 1$	0.59	0.64	0.56
MAD($S_t = 1 \vert \mathcal{F}_t$)	0.18	0.14	0.17
$\bar{P}(S_t = 1 \vert \mathcal{F}_t)$ when $S_t = 0$	0.14	0.10	0.11
$\bar{P}(S_t = 1 \vert \mathcal{F}_t)$ when $S_t = 1$	0.73	0.76	0.71
MAD($S_t = 1 \vert \mathcal{F}_T$)	0.16	0.13	
$\bar{P}(S_t = 1 \vert \mathcal{F}_T)$ when $S_t = 0$	0.13	0.09	
$\bar{P}(S_t = 1 \vert \mathcal{F}_T)$ when $S_t = 1$	0.75	0.78	

Table 3.5: Goodness of fit estimates for MTV data.

MFP and MTV are the Markov models with fixed-transition and time-varying transitions, respectively, and LMX is the logistic mixtures model. Estimates are averaged over 200 simulations of each model. Smoothed probabilities were only computed for the two Markov-transition models.

logistic mixtures model had comparable fit in terms of S_t but higher mean square error on y_t.

3.2.3 Parameter Estimates on LMX Data

Table 3.6 shows estimates using data from a logistic mixture model (LMX). Estimates are averaged over 200 simulations of each model, of which #Conv iterations converged to a solution. The table shows the average maximum likelihood estimates ($\hat{\psi}$), asymptotic standard errors σ_A, and simulated standard errors σ_S for each of the three models. The true parameter models are shown in the left column. Estimates shown in bold font indicate bias from the true parameter values as measured by the 95% confidence interval $\hat{\psi} \pm 1.96 \frac{\hat{\sigma}_S}{\sqrt{\#CONV}}$. For AIC and BIC the number of times each model achieved the minimum among the three models is shown in brackets. The transition parameters for the LMX model are shown separately and since they have different units they cannot be compared directly to the true parameters.

The LMX model, which matches the data generating process, has the lowest AIC and BIC. The BIC identifies the correctly specified model almost 99% of the time. The simulated standard errors tend to be higher than the asymptotic estimates. Most estimates are statistically significant (from zero), except for the transition parameters of the Markov models. Finally, a number of the estimates are biased (statistically different from the true values), especially in the fixed transition Markov model. The estimates of the transition parameters of the LMX model are statistically significant (from zero) but there is significant bias versus the true parameter values.

Finally, Table 3.7 summarizes the fit of these models. All three models have similar mean squared error with respect to y. The fixed-transition Markov model

	True	MFP Model			MTV Model			LMX Model		
		$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$	$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$	$\hat{\psi}$	$\hat{\sigma}_A$	$\hat{\sigma}_S$
$\beta_{0,1}$	-1.00	**-0.73**	0.26	0.39	**-1.12**	0.43	0.51	**-1.06**	0.42	0.43
$\beta_{0,2}$	0.80	**0.86**	0.15	0.25	0.80	0.16	0.18	0.81	0.16	0.17
$\beta_{0,3}$	0.20	**0.48**	0.08	0.11	0.18	0.13	0.14	**0.17**	0.14	0.13
$\beta_{1,1}$	1.00	**0.81**	0.11	0.83	1.05	0.52	0.63	1.00	0.09	0.09
$\beta_{1,2}$	1.00	**0.93**	0.07	0.30	1.10	0.95	1.19	1.01	0.05	0.06
$\beta_{1,3}$	-0.10	**0.03**	0.06	0.26	-0.18	1.01	1.05	-0.10	0.05	0.05
σ_1^2	4.00	**2.78**	0.43	0.59	**3.74**	0.54	0.63	3.93	0.53	0.55
σ_0^2	1.00	**0.92**	0.13	0.25	0.99	2.35	0.09	0.99	0.09	0.09
$\gamma_{0,1}$		0.88	8.03	2.38	-2.73	40.01	5.28			
$\gamma_{0,2}$					-4.85	4.57	10.10			
$\gamma_{1,1}$		1.32	8.84	2.81	2.14	2.79	3.28			
$\gamma_{1,2}$					3.80	56.79	5.92			
γ_1	-2.00							**-2.15**	0.67	0.68
γ_2	-3.00							**-3.35**	1.16	1.28
$\log \mathcal{L}$		877.65			842.73			842.83		
AIC		1775.30	[1]		1709.46	[19]		1705.65	[180]	
BIC		1817.45	[1]		1760.04	[1]		1747.80	[198]	
#Conv		189			165			198		
Time (sec)		31			63			3		

Table 3.6: Model estimates on data generated by an LMX process.

	MFP	MTV	LMX
RMSE($\hat{y}$)	1.11	1.10	1.10
MAD($S_t = 1 \mid \mathcal{F}_{t-1}$)	0.34	0.12	0.11
$\bar{P}(S_t = 1 \mid \mathcal{F}_{t-1})$ when $S_t = 0$	0.31	0.09	0.08
$\bar{P}(S_t = 1 \mid \mathcal{F}_{t-1})$ when $S_t = 1$	0.58	0.81	0.83
MAD($S_t = 1 \mid \mathcal{F}_t$)	0.26	0.08	0.07
$\bar{P}(S_t = 1 \mid \mathcal{F}_t)$ when $S_t = 0$	0.25	0.06	0.05
$\bar{P}(S_t = 1 \mid \mathcal{F}_t)$ when $S_t = 1$	0.72	0.87	0.89
MAD($S_t = 1 \mid \mathcal{F}_T$)	0.26	0.09	
$\bar{P}(S_t = 1 \mid \mathcal{F}_T)$ when $S_t = 0$	0.25	0.07	
$\bar{P}(S_t = 1 \mid \mathcal{F}_T)$ when $S_t = 1$	0.71	0.85	

Table 3.7: Goodness of fit estimates for LMX data.

MFP and MTV are the Markov models with fixed-transition and time-varying transitions, respectively, and LMX is the logistic mixtures model. Estimates are averaged over 200 simulations of each model. Smoothed probabilities were only computed for the two Markov-transition models.

has the poorest fit to the state vector as it is not able to capture the time-varying properties of the underlying data. The time-varying Markov model does a good job of estimating the state vector probabilities and the logistic mixtures model performs the best. The AIC(BIC) identify the correctly specified model 180(198) times out of 200 simulations.

3.2.4   Model Diagnostics

We ran several tests on the residuals of each model to test the adequacy of each model. Since the results are very similar across all three models, we will focus on analyzing the model residuals on the logistic mixtures dataset.

Table 3.8 summarizes the results, which are averaged over 200 simulations for each model. All three models had very low autocorrelation in the residuals and the Ljung-Box statistic on 10 lags of data was not statistically significant. Looking at individual model results, the model that comes closer to the underlying data generating process will have low (or zero) residual autocorrelations. For instance, if one applies the LMX model to Markov-switching data, residual autocorrelations are noticeable although rarely higher than 0.10.

The Jarque-Bera test overwhelmingly rejected the null hypothesis of normally distributed errors. Even when one model performed better in terms of AIC or fit, residuals were almost always skewed and fat tailed. This suggests that this is not a good diagnostic for regime-switching models.

Finally, an ARCH test was carried out on the residuals. There was no evidence of ARCH effects after the models where estimated.

3.2.5   Summary of Simulation Results

In summary, the models resulted in similar estimates when applied to simulated data. Most estimates are statistically different from zero but there is evidence of bias in the estimates. Bias is particularly a problem for the fixed-transition

	MFP	MTV	LMX
Jarque-Bera	[0.00]	[0.00]	[0.00]
Q(10)	[0.52]	[0.59]	[0.57]
ARCH Test	[0.38]	[0.41]	[0.41]
ρ_1	0.02	0.00	0.00
ρ_2	-0.02	0.00	0.00
ρ_3	-0.02	0.00	0.00

Table 3.8: Model Diagnostics on LMX Data.

P-values are shown in brackets. MFP and MTV are the Markov models with fixed-transition and time-varying transitions, respectively, and LMX is the logistic-mixtures model.

Markov model, which suggests that this is not a good model when there is time variation in the transitions of the states.

Not surprisingly, the model that matched the data generation process worked the best. The Akaike and Bayesian information criteria work well in identifying the correctly specified model.

Under all three models, residuals showed no serial correlation, no evidence of heteroskedasticity (but evidence of heteroskedasticity across the two states), and significant deviations from normality. This suggests that the models correctly pick up the autocorrelation in the data but the usual diagnostic of checking for normally distributed errors might not be relevant to regime-switching models, or the theory is incorrect. A more appropriate set of specification tests are considered in Hamilton [47] and White [48], [49].

Chapter 4

Applications

In this chapter, we apply regime-switching models to crude and heating oil
futures prices.

4.1   Background

Commodity prices are prime examples of assets that are characterized by
regime-switching behavior. There are abrupt price changes due to geopolitical
events, seasonal demands, and supply disruptions, and there are long secular cy-
cles of energy price changes, such as the upswing in energy prices in the last several
years.

There is an extensive literature on modeling commodity prices and some of
this research involves regime-switching. We will briefly describe two papers relevant
to our line of research after giving a brief overview of commodities markets.

4.1.1   Overview of Commodities Markets

A futures contract is an agreement to buy or deliver the underlying asset at
a future point in time. Commodity futures trade on various markets and there are
contracts that range from a one-month horizon to as long as several years. Typically
the spot (i.e., price for immediate delivery) is not observed so it is approximated by

the nearest futures contract (one-month).

The various maturities form what is called the 'term-structure' of commodity prices. This set of prices conveys important information about where the market thinks that prices will be heading in the future. For example, near-term prices, which are affected more by immediate supply and demand imbalances than by longer-term prices, may be higher or lower than long-term prices. There are various terms that are used in the marketplace to describe the term structure of commodity prices. When the near-term futures price is higher than longer-term prices (i.e., downward sloping curve) the market is said to be in 'normal backwardation'. Conversely, when further-out prices are higher than short-term prices the market is said to be in 'contango.' The spread (slope) between the spot price and a future is called the 'basis' and will be used interchangeably with 'convenience yield.' Economists have offered various theories on why commodity prices may be upward sloping or downward sloping but they all agree that the basis conveys information about the relative supply and demand for commodities. With a shortage of shortage in a commodity the near term price will exceed the price of longer-dated futures. Thus, an extremely negative basis is a sign of shortages. On the other hand, an extremely positive basis is associated with oversupply. On average, the basis is negative for crude oil.

To summarize, the basis is an important determinant in the dynamics of commodity futures markets. The goal of this research will be to understand the effect of various term structures on both futures returns and volatility using regime-switching models.

4.1.2   Prior Work

There is an extensive literature on pricing commodities under various assumptions. We will briefly touch on two papers before proceeding to our own model.

An important paper in commodity futures pricing is that by Schwartz [50]. This paper is based on the assumption that log price is characterized by an Ornstein-Uhlenbeck stochastic process

$$dX = \kappa \left(\alpha - X \right) dt + \sigma dz, \tag{4.1}$$

where X is the log spot price ($X = \log P$), α is an average (log) price level to which the model reverts to this price with speed κ. The key feature is that the price follows a mean-reverting Brownian motion process. From this equation, Schwartz derives futures prices. It is important to note that this paper treats both the spot price and convenience yield (basis) as unobservable quantities, which are determined endogenously by the model. The model does not assume any regime-switching, and the parameters are estimated by applying the Kalman filter to the entire set of futures prices. The model produces reasonably accurate estimates of futures prices. The key thing for us is that near prices are modeled by an Ornstein-Uhlenbeck process as in Equation (4.1).

Another paper that is relevant to our analysis is by Fong and See [51] (hereafter FS). This paper applies the Markov-switching GARCH model of Gray [16] to log differences crude oil prices to test for the effect of the basis on volatility. Unlike [50] they don't assume any structure for the stochastic process of prices. Consistent with theoretical explanations, they find that a decrease in basis increases the probability

of being in a high-volatility state (more negative basis $\Leftrightarrow$ near prices higher $\Leftrightarrow$ supply shortages).

4.1.3   Outline of Research

We apply a regime-switching model to one-month crude oil prices and one-month heating oil prices. We estimate two versions for the linear part of the model:

$$d \log P_t = \kappa_{S_t} \left(\alpha_{S_t} - \log P_{t-1} \right) dt + \epsilon_t, \qquad (4.2)$$

and

$$d \log P_t = \alpha_{S_t} + \beta_{S_t} \cdot d \log P_{t-1} + \epsilon_t, \qquad (4.3)$$

where $d \log P_t$ denote log differences of the futures price, and S_t denotes dependence on the regime $S_t \in \{0, 1\}$ and $\epsilon_t \sim i.i.d.\ N\left(0, \sigma_{S_t}^2\right)$.

For the stochastic part of the model, we assume either the time-varying Markov transition model or the independent-switching logistic model with several additional covariates. We assume that two states are appropriate, although one may argue that three states would do better in characterizing the process (for example, very negative, very positive, and normal basis).

We assume that the linear part of the model is only a function of past returns and that the other explanatory variables come through the non-linear part.

There are several major differences between our model versus both Schwartz and Fong. First, Schwartz treats the spot price and convenience yield as unobservable quantities that are determined endogenously by the model. By contrast, both

our model and that of FS take these variables as observable quantities, which are proxied by the nearest contract (one-month) and the slope between month 3 and 1 for the spot price and basis, respectively. Visual inspection of the convenience yield in [50] shows that it is indeed similar to our own calculations.

Second, Schwartz uses the entire term structure to fit his model since the emphasis is on pricing futures at various maturities. Like Fong, we focus on the nearest futures, which for us is the one-month futures contract. [1]

Third, we estimate of model of both the Ornstein-Uhlenbeck form and as a general autoregressive process. Fourth, like FS, we assume regime-switching as do FS but we use a simpler model rather than the more complicated MS-GARCH model that they use. Lastly, we incorporate additional covariates to explain returns and volatility.

4.2   Data and Preliminary Analysis

4.2.1   Data

The data consist of weekly log returns of crude oil and heating oil futures traded on the New York Mercantile Exchange (NYMEX) with expiration terms of one to four months. The data was obtained from the U.S. Department of Energy and NYMEX and is from January 3, 1986 to December 29, 2006 (twenty one years).

Crude oil futures contracts expire on the third business day prior to the 25th calendar day of each month, and heating oil futures expire at the end of every month.

---

[1] Actually Fong and See model the second nearest contract.

So once every month there is a discontinuity in prices when the second nearest month contract rolls into the nearest month. In computing daily log differences we adjusted for the "roll," without which the return is biased downward.

We also computed a *convenience yield* or *basis* by comparing the slope between the first two contracts

$$BASIS_t = \log\left(\frac{F_{t_2}}{F_{t_1}}\right) - (t_2 - t_1)r_t, \tag{4.4}$$

where F_{t_1} is the price of the nearest futures contract and proxies for the spot price, and F_{t_2} is the price of the second nearest contract. So when commodity prices are downward sloping (said to be in normal backwardation), the basis will be negative. The slope is adjusted by financing costs r_t, which is the yield on three-month Treasury bills (pro-rated for the time between the two futures contracts). We ignore storage and insurance costs associated with the spot price of the commodity.

As time passes the maturity of the futures contracts decreases. To adjust for this we created "constant-maturity" prices for the two contracts by weighting by the time to maturity $\tilde{F}_{t_1} = w_1 F_{t_1} + (1 - w_1)F_{t_2}$ and using these constant-maturity prices in equation (4.4). Doing this adjustment had a minor effect on the convenience yield by reducing some outliers in the days prior to expiration.

The explanatory variables consist of average temperatures for the entire United States and Industrial Production, which are proxies for seasonal demand and demand due to economic growth, respectively.

Monthly temperature data was obtained from the National Oceanic and Atmospheric Administration. This series represents the average temperature for each cal-

endar month, and from it we created a series of abnormal temperatures by subtracting the average temperature from 1900 to 1985 (different mean for each month).[2] The idea here is that if temperatures affect energy prices, it is only abnormally high or low temperatures that have an impact on prices (like abnormally cold winter temperatures pushing up the price of heating oil). By using a fixed average temperature for each month we ignore the issue of global warming which has raised temperatures over the past century. More importantly, because the temperature series is monthly, when we combine with weekly price data, there is some look-ahead bias in the weather data, which can range from one to four weeks. This is partly offset by the fact that market participants have access to weather forecasts. It's the best data that was available.

Data on Industrial Production from the Federal Reserve Board is also monthly, and it measures the state of the economy – particularly, industrial demand for energy. We calculate three-month log differences from the seasonally adjusted series. When merged with weekly energy data this variable also suffers from look-ahead bias as the data becomes available with a two-week lag.

Figure 4.1 graphs the data. Crude and heating oil prices are shown in log form to better discern the data. One can see a spike in crude and heating oil prices during 1990 due to the Iraqi invasion of Kuwait, a decrease in prices in the late 1990s, and an upward trend in energy since that time, interrupted only by the 2001-2002 recession, which is reflected in declining industrial production. The basis for crude oil tends to be negative and appears to mean revert over a period of a few months. Extremely

---

[2]An alternative measure of the impact of weather would be "degree-days."

positive values in crude oil basis are associated with oversupply and declining prices but prices seem to reverse in the future. Extreme negative values for basis of crude oil are contemporaneously associated with increasing prices. During the invasion of Kuwait, one-month prices shot up but three-month prices did not rise as quickly as traders anticipated a U.S. victory.

The basis for heating oil follows an entirely different pattern and appears to be more seasonal in the early part of the sample. It has more extreme negative spikes which appear to line up with increasing prices. The graph of abnormal temperatures (adjusted by long-term mean) clearly shows temperatures going up and some of the abnormally low temperatures seem to line up with the negative spikes in heating oil basis. Finally, the graph on industrial production shows three-month percent changes and one can see the last two recessions during 1990-91 and 2001-2002.

4.2.2 Preliminary Analysis

Table 4.1 shows that the average weekly returns for crude and heating oil are 0.20% and 0.12%, respectively, with a large standard deviation of about 5%. Both returns are negatively skewed with high kurtosis and the Ljung-Box statistic for autocorrelations is significant at the 5% confidence level. First-order autocorrelations are -0.04 for both series. The basis for each series tends to be negative and is negatively skewed.

Table 4.2 provides t-tests for returns across different sub samples. For crude oil, when the basis is negative returns average 0.42% versus a loss of -0.10% per week
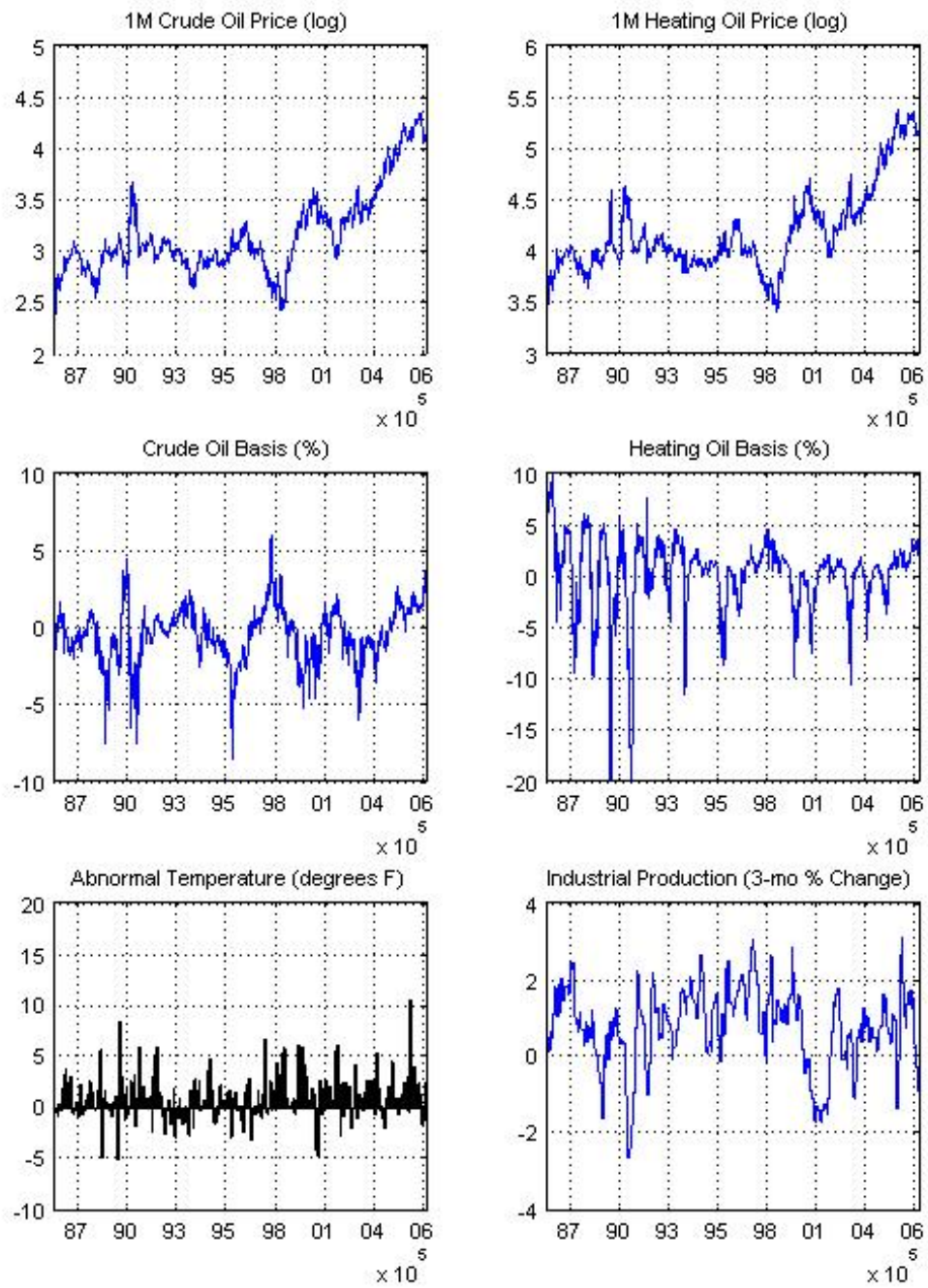
Figure 4.1: Sample Statistics from Logistic Mixture Model

	Mean	Std	Skew	Kurt	Q(10)
Crude Oil	0.20	5.04	-0.44	6.64	0.03
Crude Oil Basis	-0.50	1.81	-0.51	4.30	0.02
Heating Oil	0.12	5.05	-0.31	7.01	0.03
Heating Oil Basis	-0.06	4.12	-2.73	19.69	0.02

Table 4.1: Descriptive Statistics for Energy Data.

Crude oil and heating oil returns are weekly log differences in percentage terms. Basis is in percent and is computed using three-month and one-month constant-maturity futures prices. Q(10) is Ljung-Box p-value for autocorrelations.

when the basis is positive. The null hypothesis of equal means is rejected at the 10% level. Interestingly, volatility is slightly lower when crude oil prices are falling. For heating oil, the pattern is reversed: returns are -0.17% when the basis is negative and +0.27% when the basis is positive but the null hypothesis of equal means cannot be rejected. Using lagged returns or contemporaneous temperature (difference from mean for each month) to divide sample returns, does not result in meaningful differences. Economic growth (IP) appears to differentiate returns. When industrial production is falling/increasing, crude oil returns are -0.52%/+0.38%, and the differences are statistically significant at the 5% confidence level. This makes sense as economic growth creates demand for oil. High/low growth also leads to higher/lower heating oil prices. My original hypothesis was that heating oil would not be affected by economic growth as much but the differences make sense given that heating oil derives from crude.

	Crude Oil		Heating Oil	
	Mean	Std Dev	Mean	Std Dev
$RETURN_{t-1} < 0$	0.01	5.08	0.01	4.71
$RETURN_{t-1} > 0$	0.37	5.00	0.20	5.34
$H_0 : \mu_{<0} = \mu_{>0}$	[0.24]		[0.58]	
$BASIS_{t-1} < 0$	0.42	5.18	-0.17	6.22
$BASIS_{t-1} > 0$	-0.10	4.82	0.27	4.31
$H_0 : \mu_{<0} = \mu_{>0}$	[0.09]		[0.22]	
$TEMPER_{t-1} < 0$	0.08	4.70	0.38	4.57
$TEMPER_{t-1} > 0$	0.25	5.17	0.01	5.23
$H_0 : \mu_{<0} = \mu_{>0}$	[0.60]		[0.23]	
$IP_{t-1} < 0$	-0.52	6.33	-0.88	6.14
$IP_{t-1} > 0$	0.39	4.62	0.38	4.69
$H_0 : \mu_{<0} = \mu_{>0}$	[0.04]		[0.01]	

Table 4.2: T-Tests for Energy Returns.

Crude oil and heating oil returns are weekly log differences in percentage terms. Tests for equality of means assume unequal variances.

4.3 Empirical Results

We experimented with different assumptions for the stochastic part of the model by estimating both Markov (time-varying and fixed) and independent logistic mixtures models. The fixed-transition Markov model proved inadequate because it did not converge in some of the models. The logistic mixtures model proved the most robust in terms of convergence. Both the logistic mixtures and the time-varying Markov model produced similar results with respect to the stochastic component S_t but in the independent switching the results where visibly more noisy. Thus, for the remainder of this section, we will focus on results from the Markov model with time-varying transitions, with transition covariates $Z_t = (BASIS_{t-1}, TEMPER_{t-1}, IP_{t-1})'$.

4.3.1 Ornstein-Uhlenbeck Model

Table 4.3 shows the parameter estimates for the Ornstein-Uhlenbeck model for both crude and heating oil along with p-values for each of the estimated parameters.[3] State 1 has the higher variance. The positive coefficient on κ_1 shows that crude oil reverts to the mean very quickly in the high volatility state (significant at the 5% confidence level). In the low volatility regime there is slight momentum but the estimate is not significant. Heating oil is slightly mean reverting in both periods but he estimates are not significant.

The basis has the expected effect on crude oil prices. In both states it is

---

[3]Standard errors were computed from the standard errors of the switching-regression coefficients using the Delta method.

	Crude Oil		Heating Oil	
	Estimate	P-Value	Estimate	P-Value
κ_1	1.869	0.05	0.257	0.39
κ_0	-0.110	0.38	0.216	0.31
α_1	2.752	0.00	4.770	0.30
α_0	-0.545	0.50	5.056	0.02
σ_1	8.236	0.00	7.298	0.00
σ_0	3.814	0.00	3.503	0.00
γ_1 Basis	-0.925	0.00	-0.029	0.31
γ_1 Temp	0.504	0.01	0.152	0.14
γ_1 IP	1.814	0.00	1.258	0.01
γ_0 Basis	-4.540	0.00	0.139	0.11
γ_0 Temp	1.132	0.01	0.048	0.43
γ_0 IP	10.880	0.01	3.210	0.00
$\log \mathcal{L}$	3225		3154	
AIC	6479		6332	

Table 4.3: Parameter Estimates for Ornstein-Uhlenbeck Model

negative, indicating that a negative basis (prices for nearest contract higher than longer-dated contracts), which is associated with shortages, makes it more likely to transition to a high volatility state. This is true for both 0-to-1 and 1-to-1 transitions but the effect is stronger going from state 0 to 1, which is consistent with the results of [51]. Thus, a negative basis makes it more likely to either enter or remain in the volatile state. For heating oil, the basis has a small effect and actually has the wrong sign in one of the state.

Abnormal temperatures have an effect on oil prices but, surprisingly, they have no effect on heating oil. Higher than normal temperatures lead to higher volatility (could this be due to driving season in summer months?). Industrial Production, which proxies for demand from economic growth, is significant for both crude and heating oil and has a bigger impact on crude oil prices.

Figure 4.2 shows the filtered probability of the high variance state along with the basis, which we believe is a driver for the switches in regime. The figure shows spikes which correspond to the high volatility state interspersed with quiet periods. The model matches some of the decreases in basis, like 1990-91, which corresponds to the first Iraq war, 2000 and 2003. However, the model misses the largest spike in early 1995.

Hypothesis testing of one versus two regimes is complicated by the fact the second state is a nuissance parameter. As an informal test of how the probability $\Pr(S_t|\mathcal{F}_t)$ fits the data, we divided up the sample into two parts depending on whether the probability $\Pr(S_t|\mathcal{F}_t)$ was greater than $\frac{1}{2}$ (high volatility state) or less than one half (low volatility state). Table 4.4 shows a t-test for the hypothesis that

	Mean	Std Dev
$\Pr(S_t|\mathcal{F}_t) > 0.5$	-1.250	9.104
$\Pr(S_t|\mathcal{F}_t) < 0.5$	0.476	3.762
T-test	0.01	

Table 4.4: T-test of Equality of Means Across Two Regimes

the means are the same across these two samples. The null hypothesis that the means are equal is rejected at that 1% level, thus showing that the inferred states really do correspond to different sub samples.

4.3.2 Autoregressive Model

Table 4.5 provides estimates for the autoregressive model of Equation (4.3). The estimated transition parameters are very similar the Ornstein-Uhlenbeck model. In the high volatility state, unconditional returns for crude oil are -0.69% versus a 0.07% increase in the low volatility state. For heating oil the returns are both positive and similar in magnitude. Thus, we see that the high variance state corresponds to negative returns for crude oil.

The AR model results for crude are consistent with Table 4.3. In the high variance state the process is mean reverting but the opposite is true in the low variance state. Heating oil has some momentum in both regimes but this result is not statistically significant.

	Crude Oil		Heating Oil	
	Estimate	P-Value	Estimate	P-Value
$\beta_{1,1}$	-0.695	0.13	0.121	0.40
$\beta_{1,2}$	-0.207	0.01	-0.077	0.14
$\beta_{0,1}$	0.394	0.00	0.208	0.15
$\beta_{0,2}$	0.072	0.01	0.031	0.23
σ_1	8.232	0.00	7.249	0.00
σ_0	3.804	0.00	3.455	0.00
γ_1 Basis	-0.851	0.00	-0.020	0.36
γ_1 Temp	-0.469	0.01	0.145	0.14
γ_1 IP	1.588	0.00	1.056	0.01
γ_0 Basis	-4.397	0.00	0.144	0.07
γ_0 Temp	1.068	0.00	-0.011	0.48
γ_0 IP	10.570	0.00	2.965	0.00
$\log \mathcal{L}$	3221		3153	
AIC	6470		6334	

Table 4.5: Parameter Estimates for AR(1) Model

4.3.3 Summary

In summary, using the basis, temperature deviations, and industrial production within a regime-switching model produces several interesting results for crude oil prices. There is mean reversion in the high volatility state, which is associated with negative returns. The lagged basis, abnormal temperatures, and economic growth all are significant in determining the state of the model. In particular, a negative basis, which is associated with high near term demand relative to supply, makes it more likely to enter and remain in the high volatility state.

The results for heating oil were weak. An informal analysis that does not involve any hypothesis testing of one versus two regimes, shows that the regime-switching models considered here do not characterize heating oil prices.

There are some limitations to this analysis. First, we did not account for seasonal effects in the crude oil market and more importantly in the heating oil market. Heating oil did not turn out to be a significant determinant of the regime that the market is in. However, there might still be seasonal effects in the data. We tried adding a dummy variable for every month but this resulted in too many parameters (number of dummy variables $\times$ 2 regimes). Second, we let all the parameters switch between the two states. It might be the case that only some parameters switch. Third, we modeled future returns as a function of past returns and several explanatory that affect the state of the process. We did not test to see whether these explanatory variables enter through the linear part or the non-linear part of the model. The complications with hypothesis testing for these type of models

would hamper this kind of analysis. Fourth, we were not able to test the direction of causality in these markets as one would, for example, with a vector autoregressive process. Clearly, rising/falling prices for one-month futures lead to lower/higher basis, and we don't know whether it's the basis that affects future prices more than prices themselves. Fifth, we treated each market as separate, but a more realistic analysis might consider several related commodity markets together (for example, since heating oil is related to crude or if two commodities are substitutes of one another) or consider the entire term structure of commodity prices as [50] does.
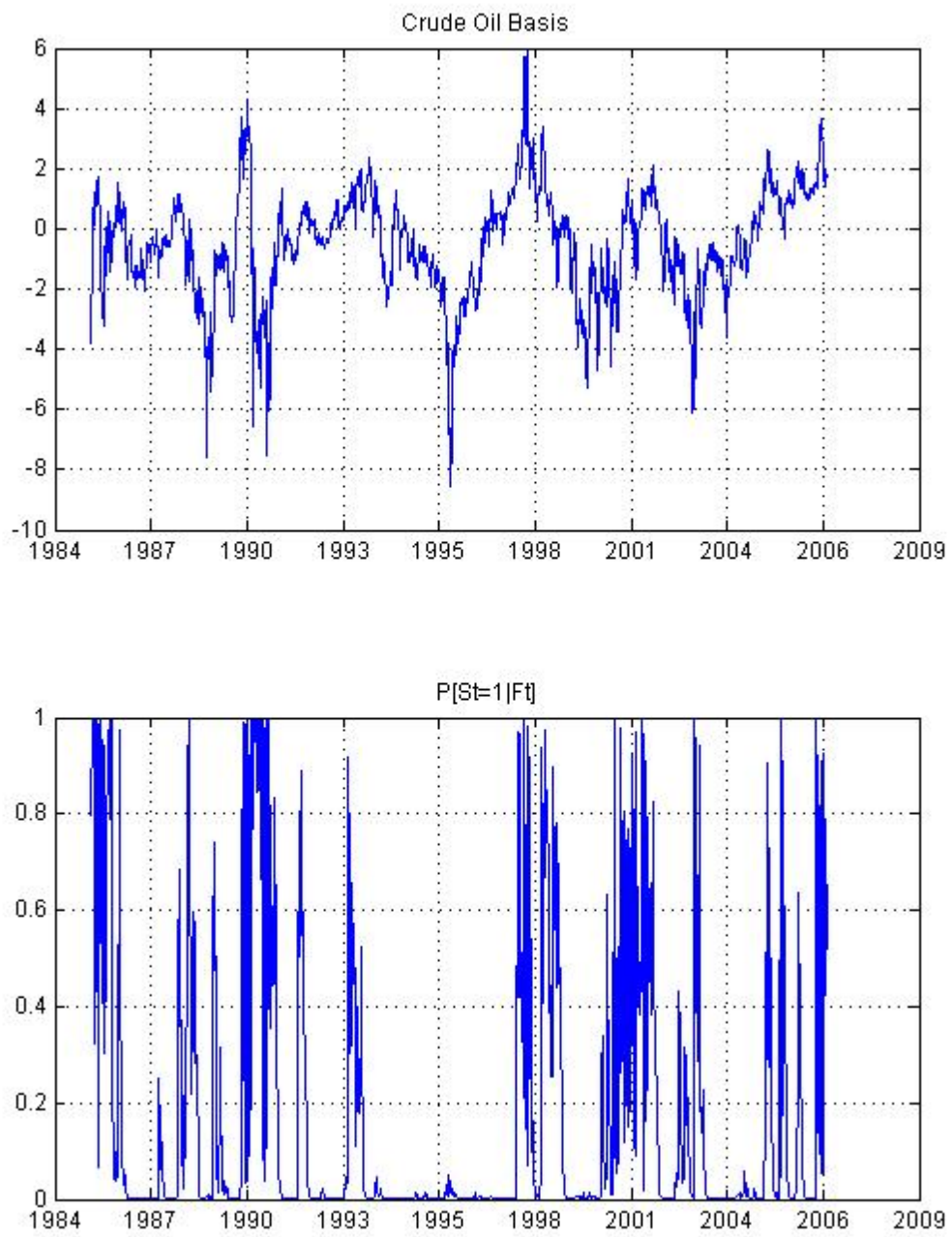
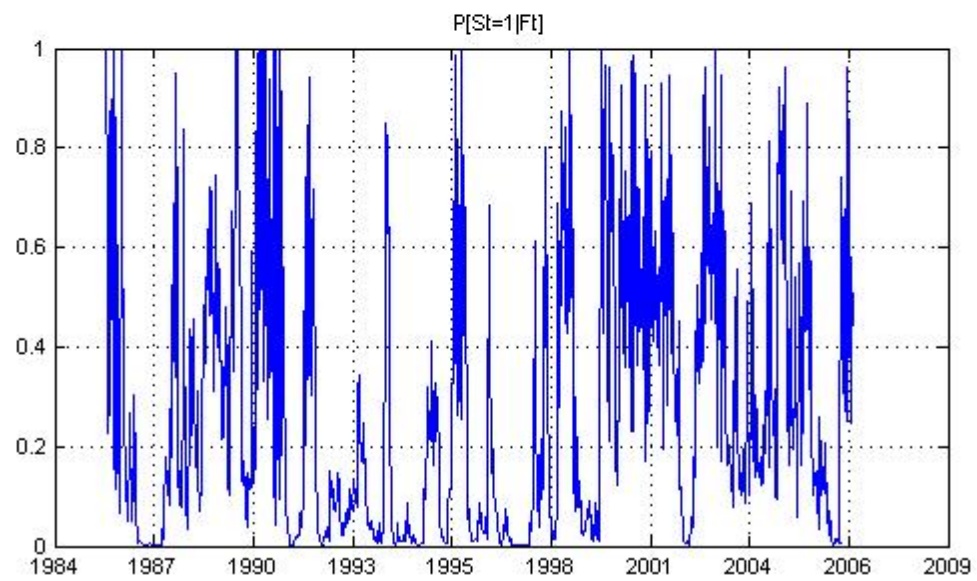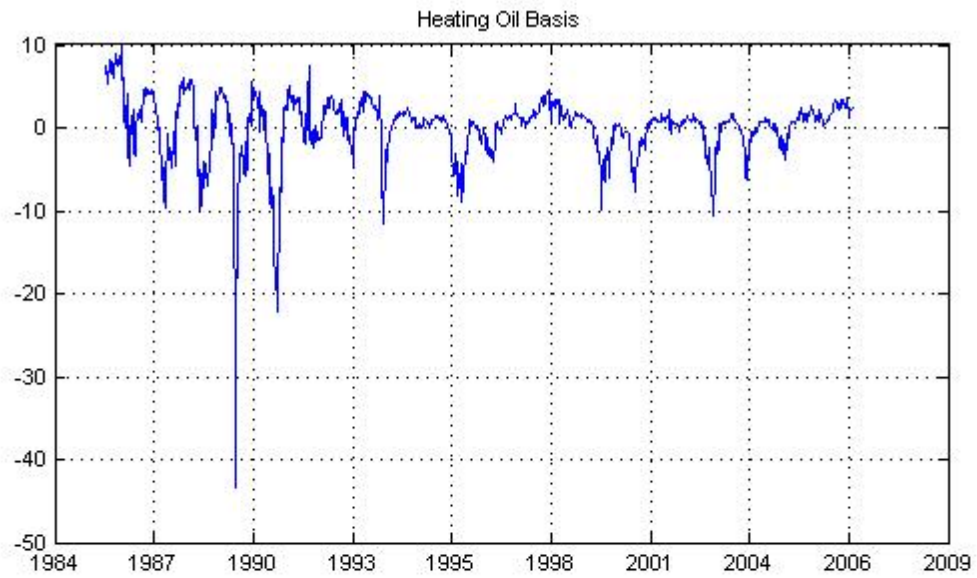Figure 4.2: Crude Oil Regime Inferred from Ornstein-Uhlenbeck Process

Figure 4.3: Heating Oil Regime Inferred from Ornstein-Uhlenbeck Process

Chapter 5

Conclusion

In this thesis, we reviewed some of the vast literature on regime-switching time series models and compared three types of models: models where the the unobserved state vector is a Markov chain (both with fixed and time-varying transition probabilities), and a model with independent probabilities determined via a logistic function. We also discussed the techniques involved in estimating these models, which include the *EM* algorithm and filtering/smoothing and numerical maximization.

In Chapter 3 we applied each one of these models to simulated data. Typically, the correctly specified model had a slight edge in each case but overall the estimates were similar. Generally, simulated standard errors were slightly larger than asymptotic estimates derived from the Fisher information matrix. Bias seems to be a problem (particularly for the fixed-transition Markov model). However, we found that we were able to identify the state that the process was in with reasonably high probability. The Akaike and Bayesian information criteria did a good job of correctly identifying the best model. Convergence was an issue even in simulated data and we found the EM algorithm in the logistic mixture model to be much more efficient in terms of computational time versus filtering and numerical maximization. We used several basic diagnostics on the model residuals, and we think that this is an area that needs more investigation.

In Chapter 4 we applied a regime-switching model to crude oil and heating oil futures. We found some evidence of switching between two regimes in the crude oil markets but not for heating oil. We find that crude oil is characterized by regime switching, where prices alternate between a high volatility state with low returns and significant mean reversion and a low volatility state with positive returns and no mean reversion. The spread between one-month and three-month futures prices is an important determinant in the dynamics of crude oil prices, as are temperatures and economic growth. The filtered probabilities of the high volatility regime appeared to match well against some major events that affected the world energy markets. A more complete analysis that involves the entire term structure of commodities is an obvious next step.

In this analysis we avoided hypothesis testing, which is still an unresolved issue for these kind of models. As we found in our own application with energy prices, we believe that this is one drawback for these models. An interesting line of research that we didn't get to apply here pertains to the specification tests of White in [48] and [49], which are applied to Markov models in [47]. It is necessary to be able to identify a mis-specified model, particularly as these models must be tailored to individual problems. Unlike single-regime linear models where the models usually converge, and where we can easily do hypothesis testing, in regime-switching model convergence is often a problem and hypothesis and specification testing are more complex. Thus these models must be handled with care and must be tailored for specific problems to ensure good results.

Bibliography

[1] G. Lindgren. Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics*, 5:81–91, 1978.

[2] S.R. Cosslett and L.F. Lee. Serial correlation in discrete variable models. *Journal of Econometrics*, 27:79–97, 1985.

[3] R. Quandt. A Markov model for switching regressions. *Journal of Econometrics*, 1:3–16, 1973.

[4] J.D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.

[5] C. Engel and J.D. Hamilton. Long swings in the dollar: Are they in the data and do markets know it? *The American Economic Review*, 1990.

[6] R. Garcia and P. Perron. An analysis of real interest under regime shift. *Review of Economics and Statistics*, 78:111–125, 1996.

[7] F.X. Diebold, J.H. Lee, and G.C. Weinbach. Regime switching with time-varying transition probabilities. In C.P. Hargreaves, editor, *Nonstationary Time Series and Cointegration*. Oxford University Press, 1994.

[8] A.P. Dempster, N.M. Baird, and D.B. Rubin. Maximum likelihood from incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[9] A.J. Filardo. Business-cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics*, 12(3):299–308, 1994.

[10] J.M. Durland and T.H. McCurdy. Duration-dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business and Economic Statistics*, 12(3):179–288, 1994.

[11] J. Raymond and R. Rich. Oil and the macroeconomy: A Markov state-switching approach. *Journal of Money, Credit and Banking*, 29(2):193–213, May 1997.

[12] R. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007, 1982.

[13] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–317, 1986.

[14] J. Cai. A Markov model of switching-regime ARCH. *Journal of Business and Economic Statistics*, 12(3):309–316, 1994.

[15] J. Hamilton and R. Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 50:1269–1288, 1994.

[16] S.F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62, 1996.

[17] F. Klaasen. Improving GARCH volatility forecasts with regime-switching GARCH. In J.D. Hamilton and B. Raj, editors, *Advances in Markov-Switching Models*, pages 223–254. Physica-Verlag, 2002.

[18] N.O. Jeffries. *Logistic Mixtures of Generalized Linear Model Time Series*. PhD thesis, University of Maryland, 1998.

[19] C.S. Wong and W.K. Li. On a logistic mixture autoregressive model. *Biometrika*, 88(3):833–846, 2001.

[20] H. Tong. Threshold models in non-linear time series analysis. In *Lecture Notes in Statistics*. Springer, Berlin, 1983.

[21] H. Tong. *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford, 1990.

[22] T. Terasvirta. Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218, 1994.

[23] T. Terasvirta. Smooth transition regression modeling. In H.Lutkepohl and M. Kratzig, editors, *Applied Time Series Econometrics*. Cambridge University Press, 2004.

[24] G. Gonzalez-Rivera. Smooth-transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 3(2):61–78, 1998.

[25] H.M. Krolzig, M. Marcellino, and G.E. Mizon. A Markov-swithcing vector equilibrium correction model of the UK labour market. In J.D. Hamilton and B. Raj, editors, *Advances in Markov-Switching Models*, pages 92–112. Physica-Verlag, 2002.

[26] M.P. Clements and H.M. Krolzig. Can oil shocks explain asymmetries in the US business cycle? In J.D. Hamilton and B. Raj, editors, *Advances in Markov-Switching Models*, pages 41–60. Physica-Verlag, 2002.

[27] C.J. Kim, J. Piger, and R. Startz. Estimation of Markov regime-switching regression models with endogenous switching. Federal Reserve Bank of St. Louis Working Paper 2003-015C, 2003.

[28] A. Ang and G. Bekaert. The term structure of real rates and expected inflation. *Journal of Business and Economics*, 20(2):163–182, 2002.

[29] Andrew Ang and Geert Bekaert. The term structure of real rates and expected inflation, 2005.

[30] R. Bansal and H. Zhou. Term structure of interest rates with regime shifts. *The Journal of Finance*, LVII(5), 2002.

[31] Y. Bar-Shalom. Tracking methods in a multi-target environment. *IEEE Trans. Automatic Control*, 1978.

[32] C.J. Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 75:1–22, 1994.

[33] C.J. Kim and C.R. Nelson. *State Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT Press, 1999.

[34] R.H. Shumway and D.S. Stoffer. Dynamic linear models with switching. *Journal of the American Statistical Association*, 86(415):763–769, 1991.

[35] C.J. Kim and C.J. Murray. Permanent and transitory components of recessions. In J.D. Hamilton and B. Raj, editors, *Advances in Markov-Switching Models*, pages 19–39. Physica-Verlag, 2002.

[36] T.C. Mills and P. Wang. Plucking models of business cycle fluctuations: Evidence from G7 countries. In J.D. Hamilton and B. Raj, editors, *Advances in Markov-Switching Models*, pages 113–134. Physica-Verlag, 2002.

[37] S. Chib and M. Dueker. Non-Markovian regime switching with endogenous states and time-varying state strengths. Federal Reserve Bank of St. Louis Working Paper 2004-030A, 2004.

[38] J.D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[39] J.D. Hamilton and B. Raj. *Advances in Markov-Switching Models*. Physica-Verlag, 2002.

[40] B. Kedem and K. Fokianos. *Regression Models for Time Series Analysis*. Wiley Series in Probability and Statistics, 2002.

[41] G. Chow. Tests of the equality between two sets of coefficients in two linear regressions. *Econometrica*, 28:561–605, 1960.

[42] B.E. Hansen. Inference when a nuissance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–430, 1996.

[43] D.W.K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993.

[44] D.W.K. Andrews and W. Ploberger. Optimal tests when a nuissance parameter is present only under the alternative. *Econometrica*, 62(6):1383–1414, 1994.

[45] J.D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45:39–70, 1990.

[46] A. Bera and C.M. Jarque. Efficient tests for normality, homoskedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259, 1980.

[47] J.D. Hamilton. Specification testing in Markov-switching time series models. *Journal of Econometrics*, 70:127–157, 1996.

[48] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–26, 1982.

[49] H. White. Specification testing in dynamic models. In T.F. Bewley, editor, *Advances in Econometrics, Fifth World Congress, Vol II*. Cambridge University Press, 1987.

[50] E.S. Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52(3):923–973, 1997.

[51] W.M. Fong and K.H See. Basis variations and regime shifts in the oil futures market. *The European Journal of Finance*, 9:499–513, 2003.