A regime switching model: Estimation, robustness, and empirical evidence

Xin Guo

Abstract. This paper is a summary of two of our recent work on regime switching models.

The first (joint work with D. Chan) concerns the problem of applying regime switching models to analyze financial time series data. Within a Bayesian framework, an estimation methodology to fit the model and model selection strategies to determine the optimum number of states are proposed. A case study on AT&T stock price data is reported. Based on the empirical study, a notion of “regime shift detection” in financial time series data and a detection method based on our estimation algorithm are developed.

The second (joint work with G. Yin) is the probabilistic view of the estimation procedure. It deals with the rate of convergence and error bounds for the Wonham filter with random parameters. The basic idea is to construct approximate filters when only noisy/simulated values are available. These suboptimal filters are proved to converge to the desired Wonham filter under simple ergodicity conditions.

Finally, The pros and cons of the regime switching model and further research problems are discussed.

1. Introduction

Let \( X(t) \) be the price of a single stock at time \( t \). Consider the following dynamics:

\[
  dX(t) = X(t)\mu_{J(t)}dt + X(t)\sigma_{J(t)}dW(t),
\]

where \( W(t) \) is a Weiner process, and \( J(t) \) is a Markov process taking discrete values in \( \{1, \ldots, S\} \) and independent of \( W(t) \), with \( S \) being the total number of states. For each state \( j \), there is a drift parameter \( \mu_j \) and a volatility parameter \( \sigma_j \).

The above model is referred to by several names such as the “Regime/Markov switching model”, the “Markov modulated geometric Brownian motion model”, etc. This model has been well-studied in several contexts. To cite a few instances, see: [Ham89] and [Nef84] for earlier applications in the study of GNP; [DKR94] for hedging issues under a mean variance criterion; [DKR94] and [Guo01a] for European option pricing problems; [Guo99], [Guo01b], and [GZ02] for extensive study on related infinite time horizon optimal stopping problems; [Tim00] on the

1991 Mathematics Subject Classification. Primary; Secondary.

Key words and phrases. Hidden Markov model, Bayesian, Model selection, Recursive algorithm, Regime shift detection, Wonham filter, Error Bounds, Rate of convergence.
moments of regime switching models; [YZZ02] for numerical algorithm for European option pricings and volatility smiles with finite-state Markov chains; [Hel03] for applying linear programming techniques in optimal stopping time; and [BE02] for numerics of American options.

Despite these work on pricing and hedging issues, there has been no thorough study on the validation/invalidation of the model from an empirical/statistical viewpoint. This is an important omission for several reasons: (1) A good way to accept or reject a model is to test it on real data; (2) Empirical studies not only provide a better understanding of the model but also help to uncover latent deficiencies, thereby providing a guiding principle to derive better alternative models; and (3) Analytical tractability can be fully exploited only when sufficient estimation procedures are available.

This paper is a summary of two of our recent work on statistical issues for regime switching models. The first [CG02] is devoted to the issue of parameter estimation and empirical study. The second [GY03] is a probabilistic view of the estimation procedure and deals with the rate of convergence and error bounds for the related Wonham filter with random parameters.

2. Parameter estimation and empirical evidence

Given the model in the form of equation (1.1) and the discrete feature of financial time series data, a natural statistical problem is the estimation of the states of the Markov chain $J(t)$ together with other parameters. In other words, assuming that the stock price is observed at discrete time intervals $t = 1, 2, \ldots, n$, and defining $y_t = \log X(t) - \log X(t - 1)$, the model of focus then becomes

$$y_t = \mu_{J_t} + \sigma_{J_t} e_t, \quad e_t \sim N(0, 1),$$

where $J_t$ indicates the state of the hidden Markov process at discrete time $t$, such that $J_t$ takes values in $\{1, 2, \ldots, S\}$. (Here we slightly abuse the usage of $\mu$.)

In statistical literature, a model of the form (2.1) falls under a bigger umbrella of a more generic class of models called hidden Markov models (HMMs). HMMs are mixture models, where the mixture variable, in this case $J(t)$, is unobserved and follows a finite state Markov distribution. An HMM consists of an unobservable finite number of states and an observable output. Each state produces an output with a certain probability and the (hidden) states form a first-order Markov chain, with movement between the states governed by transition probabilities. Such models are popular far beyond the field of mathematical finance due to the wide range of phenomena it can model with a good degree of success. For example, it has found applications in speech recognition [Gal02], protein sequencing [BTB00], image analysis [AEH99], etc. In particular, [Rab89] gives an excellent account of the HMM and its applications in speech recognition.

Methods for parameter estimation of general HMMs can be divided broadly into two groups—recursive methods and non-recursive methods. The most popular non-recursive method is based on the EM (Expectation Maximization) algorithm (see [DLR77]). In particular, when the number of states in $J(t)$ is known, either through experimentation or real-world interpretation, EM and its online/recursive variants are popular, especially in the area of signal processing and network performance analysis (see for example [KY02] and [KR98] and the reference therein). [AT02] compares several non-recursive methods for parameter
estimation in HMMs, including maximum likelihood, which can be viewed as the baseline standard and a number of EM-based methods such as maximum pseudo-likelihood. In the application of HMMs to the analysis of financial time series data, finding efficient algorithms for robust parameter estimation has remained largely open.

Recursive methods used for parameter estimation in HMMs include the Kalman filter (cf. [KA89]) and the auxiliary particle filter [PS99]. With the significant advances in Markov chain Monte Carlo (MCMC) techniques over recent years, recursive methods are increasingly being implemented within a simulation-based framework. MCMC techniques now allow the practitioner to perform Bayesian analysis of many types of HMMs. For example, [CK94] developed a forward-filtering, backward-sampling Gibbs algorithm for state-space models, and [Sco02] developed a recursive Gibbs algorithm for Markov modulated Poisson processes.

The recursive MCMC approach has a number of advantages over the EM algorithm. The first is that the output of the MCMC simulation can be applied to determine the most likely transition path of the data points, whereas an additional optimization routine such as the Viterbi algorithm (cf. [For73]) is required with the EM algorithm. The MCMC output can also be used to determine the optimum number of states in the HMM. This problem of determination of state space size will be discussed further later in the paper.

In [CG02], we work within a Bayesian framework and use the recursive approach of [CK94] for parameter estimation. We provide an estimation methodology to fit the above proposed models, together with model selection strategies to determine the optimum number of states for the HMMs.

2.1. Prior specification and label switching. Let \( q = (q_{ij}) \) be the matrix of one-step transition probabilities such that

\[
q_{ij} = p(J_t = j \mid J_{t-1} = i),
\]

and

\[
\mu = (\mu_1, \ldots, \mu_S)', \\
\sigma^2 = (\sigma_1^2, \ldots, \sigma_S^2)', \\
\theta = (\mu, \sigma^2, q), \\
y^t = \{y_1, \ldots, y_t\}.
\]

We use

\[
J^t = \{J_1, \ldots, J_t\}
\]

to represent values up to time \( t \) and therefore \( y^n \) represents the complete set of data.

This model shares some of the computational and inferential difficulties associated with estimating mixture models within a Bayesian framework ([GCR00] provides a good exposition of some of these difficulties and details a few strategies for handling them). One of the main issues is an identifiability problem known as label switching. Namely, a permutation of the state labels may result in the same value for the likelihood, leading to the identifiability problem. A suggested approach around this is to introduce constraints in the prior specification of the parameters. The constraints can be introduced by either assigning zero mass on regions where the constraints are violated or by reparametrizing the model (see [GCR00] and [Sco02] for a full discussion of constraints for parameters). In our
model, we introduce an ordering constraint such that \( \mu_1 < \cdots < \mu_S \), which ensures that identifiability of the state labels is not a problem.

A further issue is the possibility that no observation is assigned to a state, leading to a collapsed state which leads to sampling difficulties when trying to draw values for the parameters of that state. We choose to impose the additional constraint that in each iterate of the sampling scheme, each state must have at least five observations assigned to it. These two additional constraints in the prior specification have implications for the sampling scheme, which we will address in the next section.

The rest of the prior specification is:

\[
p(\theta) = \prod_{i=1}^{S} p(\mu_i)p(\sigma_i^2)p(q_i),
\]

with

\[
\sigma_i^2 \sim \Gamma^{-1}(\cdot | \alpha_0, \beta_0),
\]

\[
\mu_i \sim \phi(\cdot | 0, c),
\]

and

\[
q_i \sim \text{Dirichlet}(\cdot | \gamma_1, \ldots, \gamma_S),
\]

where \( q_i \) is the \( i \)-th row of \( q \), \( \phi(\cdot | m, s^2) \) is a Normal density with mean \( m \) and variance \( s^2 \) and \( \Gamma^{-1}(\cdot | \alpha, \beta) \) is an inverse Gamma density with parameters \( \alpha \) and \( \beta \). Such priors for the parameters of a Gaussian mixture are quite common in the literature. With the stock price data we apply our model to, setting \( \alpha_0 = 2 + 10^{-6} \), \( \beta_0 = 10^{-3} + 10^{-9} \), \( c = 100 \) and \( \gamma_i = 1, i = 1, \ldots, S \) gives fairly uninformative priors for all the parameters.

2.2. The recursive algorithm for parameter estimation. In the description of the simulation scheme below, we assume that \( S \), the number of states in the finite state Markov chain is known and fixed. The dimension of the parameter vector \( \theta \) is dependent on \( S \), but \( S \) is left out in the notation below for convenience. We have

\[
p(y_t | y_t^{-1}, J^n, \theta) = p(y_t | J_t, \theta) = \phi(y_t | \mu_t, \sigma_t^2).
\]

The simulation scheme can be summarized as follows:

0. Initialize \( \theta \) and \( J^n \) to appropriate starting values
1. Generate \( q \) from \( p(q | \mu, \sigma^2, J^n, y^n) \)
2. Generate \( \sigma^2 \) from \( p(\sigma^2 | \mu, q, J^n, y^n) \)
3. Generate \( \mu \) from \( p(\mu | \sigma^2, q, J^n, y^n) \)
4. Generate \( J^n \) from \( p(J^n | \mu, \sigma^2, q, y^n) \)

Here, we cycle across the steps (1) to (4), updating \( \theta \) and \( J^n \) after each complete cycle. This is an example of a Gibbs sampling scheme (see [GS90]).

In particular, the approach we adopted to generate \( J^n \) is the so-called forward-backward (FB) Gibbs sampling method [Sco02]. An alternative method is to generate \( J_t \) one at a time from the conditional density

\[
p(J_t | y^n, J_{t\neq t}, \theta).
\]

We referred to this approach as the direct Gibbs (DG) sampling method. FB is an \( O(S^2n) \) algorithm while DG is an \( O(Sn) \) algorithm which begs the question of why
we want to use FB instead of DG. The main reason is that FB has much better convergence properties due to the fact it samples all of $J^n$ jointly. Another reason is that DG is more prone to getting trapped in local modes.

Given the state labels $J^n$, it is straightforward to generate the other parameters, $\mu$, $\sigma^2$, and $q$ in steps (1) to (3). Let $y_i = \{y_t : J_t = i\}$ be the vector of observations that has been assigned to the $i$-th state, $n_i$ be the number of observations in $y_i$, and $i_k$ be a unit vector of length $n_i$. The conditional densities are derived as:

$$p(\mu \mid \sigma^2, q, J^n, y^n) = \prod_{i=1}^{S} \phi(\mu_i \mid \tilde{\mu}_i, B_i) ,$$

$$p(\sigma^2 \mid \mu, q, J^n, y^n) = \prod_{i=1}^{S} \Gamma^{-1}(\sigma^2_i \mid \alpha_i, \beta_i) ,$$

and

$$p(q \mid \mu, \sigma^2, J^n, y^n) = \prod_{i=1}^{S} \text{Dirichlet}(q_i \mid n_{i1} + \gamma_1, \ldots, n_{iS} + \gamma_S) ,$$

where $n_{ij}$ is the number of transitions from state $i$ to $j$, $\alpha_i = \alpha_0 + n_i/2$, $\beta_i = \beta_0 + (y_i - 1_{iii})^2/(y_i - 1_{iii})^2/2$, $\tilde{\mu}_i = (1/n_i/\sigma^2_i)^{-1}(1_{ii}y_i/\sigma^2_i)$ and $B_i = (1/n_i/\sigma^2_i)^{-1}$.

In order to accommodate the two additional constraints in the prior specification, the Gibbs sampling steps (3) and (4) can be embedded within a larger Metropolis–Hastings step, where we reject both $\mu$ and $J^n$ if the constraints are violated.

From an initial state $\theta^{[0]}$, $J^{n[0]}$, the simulation scheme is first run for a number of iterates to allow convergence to the invariant distribution. This is known as the burn-in phase and after a sufficient number of burn-in iterations, the iterates are assumed to be drawn from the posterior density $p(\theta, J^n \mid y^n)$. From hereon, samples $\theta^{[1]}, \ldots, \theta^{[M]}$ and $J^{n[1]}, \ldots, J^{n[M]}$ are collected which are then used for inference.

In the Bayesian approach, inference regarding the functional $f(\theta)$ is based on its posterior distribution, i.e., $p(f(\theta) \mid y^n)$. In particular, the posterior mean $E(f(\theta) \mid y^n)$ is often taken as the estimate of $f(\theta)$. Frequently, the posterior mean of $f(\theta)$ cannot be evaluated explicitly, but can be estimated from the simulation output,

$$E(f(\theta) \mid y^n) \approx \frac{1}{M} \sum_{m=1}^{M} f(\theta^{[m]}) .$$

Our estimates of the parameters based on the simulation output are

1. $\tilde{\mu}_i = \frac{1}{M} \sum_{m=1}^{M} \mu^{[m]}_i$
2. $\tilde{\sigma}^2_i = \frac{1}{M} \sum_{m=1}^{M} \sigma^2_i^{[m]}$
3. $\tilde{q} = \frac{1}{M} \sum_{m=1}^{M} q^{[m]}$
4. $\tilde{p}(J_i \mid y^n) = \frac{1}{M} \sum_{m=1}^{M} J_i^{[m]}$

2.3. Determining the state size $S$. An important question of interest is what is an appropriate value for $S$, the number of states in the hidden Markov process. This is a model selection problem of which much has been written about in the literature (see [BA98] for an overview).

One approach is to compare the marginal likelihood (ML) for each model with a different number of states. The reason that one should avoid comparing the
likelihood values directly (with $p(y^n \mid J^n, \theta_s, S = s)$ being the likelihood for the model with state size $s$) is that the values do not take into account the fact that as $S$ increases, the extra parameters in $\theta_s$ provide a better fit to the data. The consequence is a biased vote towards choosing more complex models. The marginal likelihood given by

$$p(y^n \mid S = s) = \int p(y^n \mid J^n, \theta_s, S = s) d(J^n, \theta_s),$$

can be compared directly for models with different number of states, as the model complexity has been taken into account through the integrating out of the parameters. This integration is in general difficult to perform analytically, but [Chi95] provides a numerical method for estimating the marginal likelihood using the output of the simulation scheme. This method is computationally intensive, requiring additional simulation runs to estimate the marginal likelihood. Scott [Sco02] also describes a method for model selection based on the simulation output. By bounding the number of states $S$, he provides a method for calculating the posterior probability $p(S \mid y^n)$. For a flat prior on $S$, he compares the averaged likelihood values for the different models and does not take into account the model complexity. Therefore, in our empirical study, we perform model selection based on the Bayesian Information Criteria (BIC),

$$BIC = -2 \log(\hat{l}) + k_s \log(n),$$

where $\hat{l}$ is the likelihood evaluated at the mode of the parameters and $k_s$ is the number of parameters in the model. BIC compares the (negative) likelihood, but penalizes for increased model complexity. BIC is related to the Schwartz criterion by a factor of $-\frac{1}{2}$. For small sample sizes, BIC tends to choose less complex models, but the data sets in the empirical study have many observations, so this is not an issue.

2.4. Case study with the AT&T stock. We apply the switching model to AT&T stock price data, using the estimation method outlined in the previous section. The data consists of the closing stock price for over ten years from the period 4th December 1990 to 25th October 2002. The data has 3001 observations which reduced to 3000 after taking the log differences to get daily returns. This data has been adjusted to account for dividends and share splits.

We fit a series of five models to the data, with the number of states in the HMM increasing from 1 to 5. In each model, the sampling scheme was ran for 400 iterations before a sample of 2000 iterations were collected for inference.

Table 1 summarizes the log likelihood values evaluated at the parameter modes and BIC values obtained for the five different models. As expected, the log likelihood increases with the number of states in the HMM, with a large jump when we go from one state to two states. According to the BIC values, a switching model with three states is the strongest candidate. We examine the results from this model further.

Table 2 contains the parameter estimates for the three state HMM. We see that the three states can be characterized as negative returns, neutral returns and positive returns, with the model spending roughly the same amount of time in each state. State two actually has a very slight positive return, but this is so small, we effectively refer to it as a neutral return. Estimates of the volatility reflects what we
A REGIME SWITCHING MODEL: ESTIMATION, ROBUSTNESS, AND EMPIRICAL EVIDENCE

Table 1. The log likelihood values evaluated at the parameter modes and BIC values obtained for the five different models.

<table>
<thead>
<tr>
<th>S</th>
<th>K_1</th>
<th>log(ℓ)</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>7175.86</td>
<td>-14335.71</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>7738.59</td>
<td>-15429.15</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>7797.82</td>
<td>-15499.56</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>7799.22</td>
<td>-15438.31</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>7817.10</td>
<td>-15394.02</td>
</tr>
</tbody>
</table>

Table 2. Parameter estimates for the three state hidden Markov model.

<table>
<thead>
<tr>
<th>State One</th>
<th>State Two</th>
<th>State Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{\mu} )</td>
<td>(-0.0024)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>(0.0011)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>( \pi )</td>
<td>(0.3233)</td>
<td>(0.3494)</td>
</tr>
</tbody>
</table>

Table 3. Parameter estimates for the two state hidden Markov model.

The results from the two state HMM is illuminating. Apart from being the simplest HMM we can fit, the two state HMM is also very easy to interpret. Parameter estimates for the two state model are summarized in Table 3.

In interpreting the HMMs and what each hidden state means, we find that the two state model actually provides a better interpretation of the process driving the stock price than the three state model, even though the model selection process chose the three state model. This is evidenced by the fact that during the first half of the period, AT&T was experiencing positive returns and not neutral returns as implied by the three state model. This again illustrates the general consensus that it is difficult to devise an optimum model selection procedure. Care needs to be taken that the resulting HMM makes real-world sense.

2.5. Regime change recognition. Looking at the structure of the state indicator \( J^n \) can also provide much useful information about the model process and which state is driving the stock price at a particular time. As an output of the simulation, we have an estimate of \( \hat{p}(J_t | y) \), which allows the user to trace the optimal state path taken by the data. For each observation, we can assign it to the state where \( \hat{p}(J_t | y) \) is highest.

The top panel in Figure 2 displays the assignment of the daily returns to the two states. We find that for the first half of the period, the stock price is driven by
state two, while in the second half of the period, the stock price is driven largely by state one.

Figure 1 is of interest as it shows that up until around the 1500th data point (around early November 1996), the stock price has been driven largely by state two, one of neutral returns (or slight positive returns), with low volatility. From this point onwards, the stock price experiences high volatility with swings between high positive and negative returns.

A question of interest is how consistent the change is in sentiment occurring around the middle of the period and whether the model still picks this up when the number of forward observations is reduced. To answer this, we fit a two state
HMM to the first 2000 and 1700 observations respectively. The lower two panels in Figure 2 shows the assignment of the observations to the two states in each case. The plots show that there is evidence of a sentiment change around 1490-th data point (end of October and early November of 1996), though in the case of having only 1700 observations, this evidence is slight.

Interestingly, the official announcement of the AT&T–Lucent split came on November 15th, 1996.
The detailed case study on AT&T stock prices reveals an interesting issue that we think is somehow unique for financial time series data—the “regime shift detection” problem. The traditional concept of “change point” refers to a particular point up until which a sequence of data of interest follows one type of dynamics and after which the data switches and follows another fashion. In contrast, our previous pictures shows that in the financial market, a given pattern change is more gradual and takes time before its pattern is more sustainable. This is intuitive—there are always two to three different driving forces in presence, notably bullish, bearish, and neutral groups. The overall market change is caused by the relative strength change among these groups and not the complete disappearance or appearance of a single force. As the data above shows, the regime switching model capture this feature well, and the recursive algorithm can be a promising tool in identifying this type of regime change.

3. Wonham Filter with random parameters

A different way to look at the estimation problem is via filtering. A well-known result is the Wonham filter which deals with the estimation of a Markov process with noisy observations. A crucial assumption in the Wonham filter is the apriori knowledge of the prior on the state variables. One possible solution is to replace the states in the filter by their simulated/approximated values. Such an approach raises questions such as the rate of convergence and error estimates. This issue was addressed in [GY03].

3.1. Wonham filter. Consider a probability space \((\Omega, \mathcal{F}, P)\) and \(t \in [0, T]\) for some \(T > 0\). Suppose that \(\alpha(t)\) is a finite-state continuous time Markov process with state space \(\mathcal{M} = \{z^1, \ldots, z^m\}\) and generator \(Q = (q^{ij}) \in \mathbb{R}^{m \times m}\), so that the transition probabilities are

\[
P^{ij}(h) = P(\alpha(t + h) = z^j \mid \alpha(t) = z^i),
\]

and

\[
P^{ij}(h) = \begin{cases} 
1 - q^i h + o(h), & i = j, \quad h \to 0 \\
q^j h + o(h), & i \neq j, \quad h \to 0,
\end{cases}
\]

where

\[q^i = \sum_{i \neq j} q^{ij}.
\]

Let us assume that the Markov process \(\alpha(t)\) is observed with the observation process \(y(t)\) such that

\[
\begin{aligned}
\frac{dy(t)}{dt} &= \alpha(t) dt + \sigma(t) dw(t), \\
y(0) &= 0 \text{ w.p. 1,}
\end{aligned}
\]

where \(w(\cdot)\) is a standard 1-dimensional Brownian motion that is independent of \(\alpha(t)\), and \(\sigma(\cdot) : \mathbb{R} \to \mathbb{R}\), with \(\sigma(t) \geq c\) for all \(t \in [0, T]\) and some \(c > 0\), is a continuously differentiable function.

In this framework, one of the classical results known as the Wonham filter concerns estimating \(\alpha(t)\) based on the observation \(y(\cdot)\). When the values of the states \(z^1, \ldots, z^m\) and the generator \(Q\) are known a priori and fixed, the Wonham filter [Won65] provides the optimal filter in the sense of mean square error.
Given (3.3), define
\[ p(t) = (p^1(t), \ldots, p^m(t)) \in \mathbb{R}^{1 \times m}, \]
with
\[ p^i(0) = p^i_0. \]

It was proved in [Won65] that this conditional density (posterior probability) provides the minimal mean square error, and satisfies the following system of stochastic differential equations
\[ dp^i(t) = \sum_{j=1}^{m} p^j(t) q^{ij} dt - \sigma^{-2}(t) p^i(t) \mathbf{1}(z^i - \overline{\mathbf{1}}(t)) p^i(t) dt + \sigma^{-2}(t) [z^i - \overline{\mathbf{1}}(t)] p^i(t) dy(t), \quad i = 1, \ldots, m, \]
where
\[ \overline{\mathbf{1}}(t) = \langle p(t), z \rangle \overset{\text{def}}{=} \sum_{i=1}^{m} z^i p^i(t), \quad z = (z^1, \ldots, z^m)', \]
and \( v' \) denotes the transpose of \( v \).

Adopting a vector notation, define
\[ A(t) \overset{\text{def}}{=} \text{diag}(z^1 - \overline{\mathbf{1}}(t), \ldots, z^m - \overline{\mathbf{1}}(t)) = \begin{pmatrix} z^1 - \overline{\mathbf{1}}(t) \\ \vdots \\ z^m - \overline{\mathbf{1}}(t) \end{pmatrix}. \]

Then the Wonham filter can be rewritten as
\[ \begin{align*}
dp(t) &= p(t) Q dt - \sigma^{-2}(t) p(t) A(t) dt + \sigma^{-2}(t) p(t) A(t) dy(t), \\
p(0) &= p_0.
\end{align*} \tag{3.6} \]

### 3.2. Approximate Wonham filter using \( \{ \tilde{z}_n \} \)

Now, let us assume that \( z^i \)'s are not available, and that only their noise corrupted measurements or observations or distributional information are at our disposal. We assume further that \( \{ q^{ij} \} \) remains unchanged and known a priori.

In particular, we assume that a sequence of observations of the form
\[ \tilde{z}_n = (\tilde{z}_n^1, \ldots, \tilde{z}_n^m)' \in \mathbb{R}^{m \times 1} \quad \text{such that} \quad E \tilde{z}_n = z \]
can be obtained. For example, \( \tilde{z}_n = z + \xi_n \), where \( \{ \xi_n \} \) is a sequence of \( \mathbb{R}^m \)-valued zero mean observation noise satisfying appropriate conditions. We proceed to construct the approximate filter.

First, define
\[ \overline{\mathbf{1}}_n = \frac{1}{n} \sum_{j=1}^{n} \tilde{z}_j. \tag{3.7} \]

Then in lieu of (3.6), we have a sequence of approximations \( p_n(t) \) given by
\[ \begin{align*}
\begin{cases}
dp_n(t) = p_n(t) Q dt - \sigma^{-2}(t) p_n(t) A_n(t) dt + \sigma^{-2}(t) p_n(t) A_n(t) dy(t), \\
p_n(0) = p_0,
\end{cases}
\end{align*} \tag{3.8} \]
where
\[ A_n(t) = \text{diag}(\overline{\mathbf{1}}_n^1 - \overline{\mathbf{1}}_n(t), \ldots, \overline{\mathbf{1}}_n^m - \overline{\mathbf{1}}_n(t)). \]
Since \( \{p_n(t)\} \) is a sequence of approximations of the posterior density \( p(t) \), we may appropriately normalize it to ensure its boundedness [Won65].

Define

\[
e_n(t) = p_n(t) - p(t).
\]

Then, \( e_n(t) \) satisfies

\[
de_n(t) = e_n(t)Qdt - \sigma^{-2}(t)\overline{\pi}_n(t)\vec{\pi}(t)p_n(t)A_n(t)dt
- \sigma^{-2}(t)\overline{\pi}(t)e_n(t)A_n(t)dt
- \sigma^{-2}(t)\pi(t)p(t)[A_n(t) - A(t)]dt
+ \sigma^{-2}(t)e_n(t)A_n(t)\mu(t)dt
+ \sigma^{-2}(t)p(t)[A_n(t) - A(t)]\mu(t),
\]

with \( e_n(0) = 0 \).

To obtain the desired limit result, we impose the following conditions.

\( (A1) \) \{\( \vec{z}_n \)\} is a stationary ergodic sequence that satisfies \( E\vec{z}_n = z \) and is uniformly bounded. The sequence \( \{\vec{z}_n\} \) is independent of the Markov chain \( \vec{\alpha}(\cdot) \) and the Brownian motion \( \mu'(:,\cdot) \).

We have

**THEOREM 1.** **Under** \( (A1) \),

\[
\sup_{0 \leq t \leq T} E|e_n(t)|^2 \to 0 \quad \text{as } n \to \infty.
\]

It is well-known that convergence in \( L_2 \) implies convergence in probability. Thus, the following is immediate.

**COROLLARY 2.** **Under** \( (A1) \), for any \( \eta > 0 \),

\[
\lim_{n \to \infty} P(|e_n(t)| \geq \eta) = 0.
\]

Next, define \( e^n_\kappa(t) = n^\kappa e_n(t) \) for any \( 0 < \kappa \leq 1/2 \). Then the following estimates hold.

**THEOREM 3.** **Under** \( (A1) \),

\[
\sup_{0 \leq t \leq T} E|e^n_\kappa(t)|^{2\kappa} = \begin{cases} 
  o(1), & 0 < \kappa < 1/2, \\
  O(1), & \kappa = 1/2,
\end{cases} \quad \text{as } n \to \infty.
\]

Finally, we have error bounds for higher moments.

**THEOREM 4.** **Assume** \( (A1) \).

(i) **For any positive integer** \( \ell > 1 \),

\[
\sup_{0 \leq t \leq T} E|e^n_\kappa|^{2\ell} = \begin{cases} 
  o(1), & 0 < \kappa < 1/2, \\
  O(1), & \kappa = 1/2,
\end{cases} \quad \text{as } n \to \infty.
\]

(ii) **For** \( \kappa = 1/2 \), **denote** \( \vec{e}_n(t) = e_n^{1/2}(t) \). **Then**

\[
\sup_{0 \leq t \leq T} E e^{||\vec{e}_n(t)||} = O(1), \quad \text{as } n \to \infty.
\]

Similar results may be obtained for the error bound estimates with the generator \( Q \) not known apriori. For proofs and more details, including the extension to higher dimensional case, see [GY03].
4. Discussion

We have applied a Markov switching model to the analysis of stock price data, in an attempt to understand the underlying market forces driving the stock price. The states of the hidden Markov process could possibly represent different forces at work—for example, periods of negative and positive sentiment affecting the share price.

The method we have used for the estimation of the parameters of the switching model deals with the issue of label switching and model determination, with a BIC approach to determine the number of states in the model. However, as the case study involving AT&T stock data shows, reliance on a single criterion may not always be the best strategy. Interpretability of the resulting models is also an important consideration when trying to determine the appropriate (HMM) model to use. This reveals one of the limitations of the regime switching model in that the model itself may not be rich enough to capture the complete picture of the complex dynamics of the market and subjective input is needed in its implementation.

Using an MCMC approach for estimation of the parameters allows us to estimate $p(J_t \mid y^n)$ as well. These probabilities reveal a great deal about the structure of the underlying process driving the stock price. By assigning observations to states according to the probabilities, we can see what sentiment is driving the stock at that particular time. In some cases, this can be used as a basis for change point analysis. A sharp change in the assignment of the observations to a state, as in the AT&T example highlights that negative sentiment has largely driven the stock price since the time of the AT&T–Lucent split.

The results of the empirical study shows switching as a promising model for analyzing stock price data. As such, it is a useful tool for the practitioner who seeks to gain insight into the underlying process driving stock price data, especially in the context of “regime shift detection” as discussed earlier. Moreover, the methodology developed here can be extended to studies of mean-reverting models [FPS00] and more complex type of HMMs.

However, the systematic study on the “regime shift detection” is still in its infancy and shares the similar difficulty as in the change point detection problem. Proper choices of the time scale and the number of forward data are the two biggest challenges.

References


[Guo01b] Information and option pricing, Quantitative Finance 1 (2001), 38–44.
[Hel03] K. Helmes, Optimal stopping and linear programming, These proceedings, 2003.

School of ORIE, Cornell University, Ithaca, NY 14853.
E-mail address: xinguo@orie.cornell.edu