

Article

Not peer-reviewed version

Statistics for Hidden Markov Models

[Zigian Huang](#) and [Stanley L. Sclove](#) *

Posted Date: 30 October 2024

doi: 10.20944/preprints202410.2368.v1

Keywords: Hidden Markov Model; Time series segmentation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Statistics for Hidden Markov Models

Ziqian Huang¹ and Stanley L. Sclove^{2,*}

¹ UnitedHealth Group
² Department of Information & Decision Sciences University of Illinois at Chicago
* Correspondence: slsclove@uic.edu

Abstract: This chapter discusses stochastic processes known as “hidden Markov models”. In these models, there are states or phases. In applications to statistical process control, for example, the states could correspond to “in control” or “out of control”. In applications to the market and the economy, there are states of the market conventionally called “Bull” and “Bear”, for up and down markets, resp. The states of the economy are often referred to as *recession*, *recovery*, *expansion*, *contraction*. Questions arise as to how many such states to use, and whether the nature of the states detected by statistical methods correspond to conventional notions about the phases of the market or the economy. The models considered involve a bivariate time series $\{Y_t, S_t, \}$, where Y is observed and S is the state, or label. “Hidden” means that the state is unobserved and must be inferred. For example, Y_t may be the rate of return (ROR) of the S&P 500 index in month t and S_t may be 1 or 0, for Bull or Bear for that month. One definition of Bull and Bear is that $S_t = 1$ if the ROR $Y_t > 0$ and $S_t = 0$, otherwise. In that case, S_t is directly derivable from the data. In other cases, the labels are unobserved and then are estimated from the observed process by more complicated means. The act of estimating S_t for each t , that is, labeling each time point of a time series with the name of its estimated state, is called *time series segmentation*, sequences of 0s or 1s being called *segments*. Thus a *segment (regime, epoch)* is a sequence of time points with the same label. Some definitions of phases are in terms of patterns; others are obtained interms of fitting models in which the states are explicit elements of the model, albeit unobservable. Patterns are considered first, and then hidden state models. Use of model selection criteria for choosing between alternative models is illustrated. For example, it was found that for the stock market use of two states (corresponding to Bull and Bear) is indeed better than three states or a single state.

Keywords: hidden Markov model; time series segmentation

Contents

1	Introduction	2
2	Example: Bull and Bear Markets	2
2.1	Definitions of Bull and Bear Markets	2
2.2	Regressions on Bull3	3
2.2.1	Two Betas, No Alpha	3
2.2.2	Two Betas, One Alpha	4
2.2.3	Two Betas, Two Alphas	5
2.3	Other Models for Bull/Bear	5
2.3.1	Two Means and Two Variances	5
2.3.2	Mixture Model	6
2.3.3	Hidden Markov Model	7
3	Formulation of HMMs	8
3.1	Estimating the Parameters of an HMM.	9
3.2	More on Transition Probabilities	9
3.3	Bull and Bear Portfolios	9

4	Example: GDP – Gross Domestic Product	10
5	Further Reading	10
6	Summary	10
7	References	11

1. Introduction

This chapter relates to a particular type of stochastic process, namely, **hidden Markov models**. Markov models are defined in terms of a set of *states*. For example, in *statistical process control*, "in control" and "out of control" might be characterized by two states, with associated probability distributions. In other settings, "out of control" might be characterized by say, two states, one with a mean that is too high, the other, with a mean that is too low. We shall consider applications to finance and economics. The Markov model includes *transition probabilities* between states. An example of such states in the financial field is a "Bull market" and a "Bear market".

That is to say, in such applications, we speak of *states* or *phases* of the market and the economy. The states of the *economy* are often referred to as *recession*, *recovery*, *expansion*, *contraction*, *level*.

Obviously, then, questions arise as to how many such states to use, and, in the financial example, whether the nature of the states detected by statistical methods correspond to conventional notions about the phases of the market or the economy. We shall illustrate in terms of financial/economic models.

The models considered here involve a bivariate time series $\{Y_t, S_t\}$, where Y is observed and S is the state, or *label*. For example, Y_t may be the rate of return of the S&P 500 index in month t and S_t may be 1 or 0, for Bull or Bear for that month. One definition of Bull and Bear is that $S_t = 1$ if the ROR $Y_t > 0$ and $S_t = 0$, otherwise. In that case, S_t is directly derivable from the data. In other cases, the labels $\{S_t\}$ are unobserved and then are estimated from the observed process $\{Y_t\}$ by more complicated means. The act of estimating S_t for each t , that is, labelling each time point of a time series with the name of its estimated state, is called time series *segmentation*, repeating sequences of 0s or 1s being called *segments*. That is, a segment (*regime*, *epoch*) is a sequence of time points with the same label. Some definitions of phases are in terms of patterns; others are obtained in terms of fitting models in which the states are explicit elements of the model, albeit unobservable ("hidden"). Patterns may be used to indicate (estimate) states, so patterns will be considered first, and then hidden state models.

2. Example: Bull and Bear Markets

2.1. Definitions of Bull and Bear Markets

The popular financial press repeats the conventional wisdom relating to the definition of Bull and Bear markets. One common definition (Vanguard Group; *Investopedia*) is that a Bear market is indicated by a price decline of 20% or more in a key stock market index from a recent peak over at least a two-month period. Others consider a twelve-month period. With this definition, on the average, the market is in a Bear state every four or five years. During an average Bear market, the S&P loses about 25% of its value. It usually takes 11 to 18 months for the market to hit bottom. That would be about two or three to four or five quarters, or one to one and a half years.

The simplest definition of Bull and Bear states would be to label a month as a Bear month if the market index went down in that month and as a Bull month if the market index went up in that month. This is the Up-Down (U/D) method. The U/D method definition of a Bull market indicator variable is

$$\text{Bull}_t = [\text{sgm}(x_t) + 1]/2,$$

where x_t is the market ROR in month t and the sign function is

$$\text{sgm}(x) = -1 \text{ if } x \leq 0 \text{ and } = +1 \text{ if } x > 0.$$

If $\text{sgm}(x_t) = -1$, then $\text{Bull}_t = 0$, and if $\text{sgm}(x_t) = +1$, then $\text{Bull}_t = 1$.

Other definitions of Bull/Bear states can involve a moving window of, say, three months of market RORs since some analysts say that three consecutive down months of a market index constitute a Bear market. This notion can be applied to the data considered in Chapter 4. The idea is that the state should remain Bull until there are three consecutive months with negative market ROR and should remain Bear until there are three consecutive months with positive market ROR. Accordingly, define Bull3, a Bull/Bear indicator built on this idea of agreement of the most recent three months. That is, take Bull3 to be 0 (Bear) if each of the last three months had negative market ROR, equal to 1 (Bull) if each of the last three months had positive market ROR, and equals Bull3_{t-1} , (unchanged), otherwise. That is, the Bull/Bear indicator Bull3_t is defined in terms of Bull_t algebraically as follows.

Initialization:
for $t = 1, 2, 3$, $\text{Bull3}_t = \text{Bull}_t$.

Computation:
for $t = 4, 5, \dots, n$,
 $\text{Bull3}_t = 0$ if $\text{Bull}_t = \text{Bull}_{t-1} = \text{Bull}_{t-2} = 0$,
 $\text{Bull3}_t = 1$ if $\text{Bull}_t = \text{Bull}_{t-1} = \text{Bull}_{t-2} = 1$,
 $\text{Bull3}_t = \text{Bull3}_{t-1}$, otherwise.

In programming this, it is convenient to work in terms of the moving sum $\text{S3}_t = \text{Bull}_t + \text{Bull}_{t-1} + \text{Bull}_{t-2}$. Then $\text{Bull3}_t = 0$ if $\text{S3}_t = 0$, $\text{Bull3}_t = 1$ if $\text{S3}_t = 3$, and $\text{Bull3}_t = \text{Bull3}_{t-1}$ if $\text{S3}_t = 1$ or 2. Table 1 at the end of the chapter shows the RORs and the Bull3 indicator.

2.2. Regressions on Bull3

Next we do some regression analysis before going further into hidden Markov models. Some regressions on Bull3 will be shown, namely:

- two betas, no alpha: regression of fund on market and market*Bull3 (no constant)
- two betas, one alpha: regression of fund on market and market*Bull3 (with constant)
- two betas, two alphas: regression of fund on market, Bull3 and market*Bull3 (with constant)

2.2.1. Two Betas, No Alpha

This is the regression of $Y =$ excess ROR of the fund (in percent) on $X_1 =$ excess ROR of the market and $X_2 = X_1 \times \text{Bull3}$ (without a constant).

Software Output
Five years of monthly data ($n = 60$)
Regression Analysis:
fundROR versus marketROR, marketROR*Bull3 (marketROR times Bull3)
The regression equation is
fitted value of fundROR = 0.975 marketROR + 0.0939 marketROR*Bull3

Predictor	Coef	SE Coef	t	p

(No constant)				

```
marketROR      0.97463  0.02259 43.15  0.000
marketROR*Bull3 0.09388  0.03700  2.54  0.014
```

s = 0.724429

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	2	1674.80	837.40	1595.66	0.000
Residual Error	58	30.44	0.52		
Total	60	1705.24			

The fitted regression, of the form $\hat{Y} = b_1x_1 + b_2x_2$, is

$$\text{fundROR} = 0.975\text{marketROR} + 0.0939 \text{marketROR*Bull3}.$$

This is

$$\begin{aligned} \text{fundROR} &= 0.975 \text{ marketROR, if Bull3} = 0 \\ &= 1.069 \text{ marketROR, if Bull3} = 1. \end{aligned}$$

The standard error of fit is $s_{y.x} = 0.724$ (less than one percent). The variance of Y was $s_y^2 = 1705.24/60 = 28.42$, giving $s_y = 5.331$ (percent).

2.2.2. Two Betas, One Alpha

Next to be discussed is the model with state-dependent slope but common intercept. This is the regression of Y = excess ROR of the fund on X_1 = excess ROR of the market and $X_2 = X_1 \times \text{Bull3}$ (with constant). The fitted regression will be of the form $\hat{Y} = a + b_1x_1 + b_2x_2$.

Software Output

Regression Analysis: fundROR versus marketROR, marketROR*Bull3 \\\

The regression equation is

fitted value of fundROR = 0.171 + \, 0.983 marketROR + \, 0.0729 marketROR*Bull3.

Predictor	Coef	SE Coef	t	p
Constant	0.17081	0.09671	1.77	0.083
marketROR	0.98350	0.02275	43.24	0.000
marketROR*Bull3	0.07286	0.03824	1.91	0.062

s = 0.711544 R-Sq = 98.3% R-Sq(adj) = 98.2%

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	2	1676.16	838.08	1655.32	0.000
Residual Error	57	28.86	0.51		
Total	59	1705.02			

The fitted regression $\hat{Y} = a + b_1x_1 + b_2x_2$ is

fitted value of fundRORfund = 0.171 + 0.983 market + 0.0729 market*Bull3.

This is

fitted value of fundROR = $0.173 + 0.983 \text{ marketROR}$, if Bull3 = 0 = $0.173 + 1.056 \text{ marketROR}$, if Bull3 = 1.

The standard error of fit is 0.7115.

2.2.3. Two Betas, Two Alphas

This is the regression of Y = excess ROR of the fund on $X_1 = \text{Bull3}$, $X_2 = \text{excess ROR of the market}$ and $X_3 = X_2 \times \text{Bull3}$ (with constant), that is, the regression of ROR of the fund on the three variables ROR of the market, Bull3 and the product ROR of the market $\times$ Bull3 (with constant).

Software Output

Regression Analysis: fundROR versus marketROR, Bull3, marketROR*Bull3

The regression equation is

fitted value of fundROR

$$= 0.002 + 0.975 \text{ marketROR} + 0.224 \text{ Bull3} + 0.0778 \text{ marketROR} \times \text{Bull3}.$$

Predictor	Coef	SE Coef	t	p
Constant	0.0018	0.1957	0.01	0.993
marketROR	0.97472	0.02440	39.94	0.000
Bull3	0.2236	0.2251	0.99	0.325
marketROR*Bull3	0.07775	0.03856	2.02	0.049

s = 0.711626 R-Sq = 98.3% R-Sq(adj) = 98.2%

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	3	1676.66	558.89	1103.62	0.000
Residual Error	56	28.36	0.51		
Total	59	1705.02			

The fitted model is

fitted value of fundROR = $0.002 + 0.9747 \text{ marketROR} + 0.224 \text{ Bull3} + 0.0778 \text{ marketROR} \times \text{Bull3}$. This is
 fitted value of fund ROR = $0.002 + 0.9747 \text{ marketROR}$, if Bull3 = 0 = $(.002 + .224) + (0.9747 + 0.0778) \text{ marketROR}$

$$= .226 + 1.0525 \text{ market}, \text{ if Bull3} = 1.$$

The standard error of fit is 0.7116, about the same as the 0.7115 obtained with the preceding model.

2.3. Other Models for Bull/Bear

2.3.1. Two Means and Two Variances

A model with separate Bull and Bear distributions can be considered. The parameters of the two distributions can be estimated from the months with Bull3 = 0 and those with Bull3 = 1. The results are shown below. The mean and standard deviation for Bear months are mean = -3.34% and standard deviation = 7.53%; those for Bull months are mean = +0.988% and standard deviation = 3.635%, resp. Thus as expected there is a negative mean and relatively large standard deviation for Bear months and

a positive mean and relatively small standard deviation for Bull months. The size of the coefficient of variation (relative standard deviation) is $7.53/3.34 = 2.25$ for Bear months and $3.635/0.988 = 3.68$ for Bull months.

Descriptive Statistics: marketROR

Bull3	n	Mean	SE Mean	StDev	Min	Q1	Median	Q3	Max
0	16	-3.34	1.88	7.53	-18.62	-9.10	-2.42	1.03	8.96
1	44	0.988	0.548	3.635	-8.570	-0.938	1.065	3.265	8.38

The preceding is predicated upon a given indicator for Bull/Bear states. Indicators such as Bull and Bull3 seem somewhat arbitrary. What about a Bull4, based on a moving window of width 4, or a Bull2, based on a moving window of width 2? Why not use a more flexible method that lets the data do more to speak for themselves? To do this, we consider other descriptions, which model Bull and Bear as hidden (latent) states, to be detected from the data.

2.3.2. Mixture Model

Another model that incorporates hidden states and fits state-specific distributions is the *finite mixture model*. (See, e.g., McLachlan and Peel 2000). This model is for any number K of states and observations x_t that may be vectors. The p.d.f. is of the form

$$f(x_t) = \pi_1 f_1(x_t) + \pi_2 f_2(x_t) + \cdots + \pi_k f_k(x_t) + \cdots + \pi_K f_K(x_t).$$

The p.d.f.s f_k are called *component* p.d.f.s. or *state-conditional* p.d.f.s. Sometimes the states are called *classes* and the corresponding distributions are called *class-conditional* distributions. The mixture probabilities π_k of course must add to 1. When there are just two states, it is often convenient to index them by $k = 0, 1$. Then the p.d.f. is

$$f(x_t) = \pi_0 f_0(x_t) + \pi_1 f_1(x_t),$$

where the mixture probabilities π_0 and π_1 add to 1. The two class-conditional distributions could be multivariate Normal. Of course, a special case is the univariate case in which the vector x is just a scalar x . The class-conditional densities could, for example, be Normal, with

$$f_k(x) = \phi(x; \mu_k, \sigma_k^2), \quad k = 0, 1,$$

where $\phi(x; \mu, \sigma^2)$ is the p.d.f. of the Normal distribution with mean μ and variance σ^2 . The probabilities $\pi_k, k = 0, 1$ are prior probabilities of the distributions given by f_0 and f_1 . The posterior probability of state k , given x , is

$$\pi(k|x) = \pi_k f_k(x) / f(x) = \pi_k f_k(x) / (\pi_0 f_0(x) + \pi_1 f_1(x)), \quad k = 0, 1.$$

In the Gaussian case this is

$$\pi(k|x) = \pi_k \phi(x; \mu_k, \sigma_k^2) / [\pi_0 \phi(x; \mu_0, \sigma_0^2) + \pi_1 \phi(x; \mu_1, \sigma_1^2)], \quad k = 0, 1.$$

Denote estimates of $\pi_k, \mu_k, \sigma_k^2, \pi(k|x)$ by $p_k, m_k, v_k, p(k|x)$, respectively. Estimation starts with initial guesses $p_k^{(0)}, m_k^{(0)}, v_k^{(0)}, k = 0, 1$. Then it alternates between estimating the posterior probabilities and the distributional parameters. (This alternation is an example of the EM algorithm; see Dempster,

Laird and Rubin 1978; McLachlan and Krishnan 1997). At stage $N(N = 0, 1, 2, \dots)$ set the updated, stage $N + 1$ estimates p'' in terms of the current estimates p for states $k = 0, 1$ to

$$p'(k|x) = \frac{p_k \phi(x; m_k, v_k)}{p_0 \phi(x; m_0, v_0) + p_1 \phi(x; m_1, v_1)},$$

the updated estimates of the prior probabilities, means and variances being

$$p'_k = \sum_{t=1}^n p'(k|x_t) / n,$$

$$m'_k = \sum_{t=1}^n w'(k|x_t) x_t, \quad v'_k = \sum_{t=1}^n w'(k|x_t) (x_t - m'_k)^2,$$

where the new weights $w'(k|x_t)$ are proportional to the estimated posterior probabilities,

$$w'(k|x_t) = p'(k|x_t) / \sum_{t=1}^n p'(k|x_t).$$

Iterations are continued until adequate precision is reached.

2.3.3. Hidden Markov Model

Another model with hidden (latent) states is the *hidden Markov model* (HMM). An HMM also consists of state-conditional probability functions, e.g., Normal distributions with different means and different variances, but in addition there is a matrix of transition probabilities. The entry π_{jk} of the matrix gives the probability of transition to state k , given that the process was in state j in the preceding time period,

$$\pi_{jk} = \Pr\{S_t = k | S_{t-1} = j\}, \quad j, k = 1, 2, \dots, K \text{ states},$$

where $\{S_t\}$ is the unobservable *state process*. Each $S_t = 1, 2, \dots, \text{or } K$. The p.d.f. for an HMM is analogous to that of a mixture model, but with transition probabilities taking the place of mixture probabilities:

$$f(x_t | S_{t-1} = j) = \pi_{j1} f_1(x_t) + \pi_{j2} f_2(x_t) + \dots + \pi_{jk} f_k(x_t) + \dots + \pi_{jK} f_K(x_t).$$

For modeling Bull and Bear states, we take $k = 0, 1$ and use Normal p.d.f.s, obtaining the p.d.f.

$$f(x_t | S_{t-1} = k) = \pi_{k0} \phi(x_t; \mu_0, \sigma_0^2) + \pi_{k1} \phi(x_t; \mu_1, \sigma_1^2), \quad k = 0, 1,$$

where S_t denotes the hidden (unobservable) state, 0 for Bear, 1 for Bull, at time t .

Remarks. (i) Here, an HMM is being applied to RORs. It is often noted that the distribution of RORs is not Gaussian. Note, however, that though the marginal distribution of x_t may not be Gaussian, the component densities $f_k(x_t)$ in a finite mixture model or HMM may be.

(ii) In the above formulation, the transition probabilities are taken as stationary (not time-dependent, i.e., the corresponding stochastic process is homogeneous), but they may instead be modeled in terms of time-varying covariates, such as the risk-free rate. HMMs for Bull and Bear Markets. Maheu and McCurdy (2000) used a Markov-switching model that incorporates duration dependence to capture nonlinear structure in both the conditional mean and the conditional variance of stock returns. Their data are monthly returns including dividends (1802 to 1925 from Schwert 1990; 1926 to 1995:12 from CRSP, the Center for Research in Security Prices). The model sorts returns into a high-return stable state and a low-return volatile state. These can be labeled as Bull and Bear states, resp. The authors' method identifies all major stock-market downturns in over 160 years of monthly data. Bull markets have a

declining hazard function although the best market gains come at the start of a Bull market. Volatility increases with duration in Bear markets. Allowing volatility to vary with duration captures volatility clustering. According to their method the market spent 90% of the time as a Bull market and only 10% in a Bear market. Lunde and Timmermann (2004) studied time series dependence in the direction of stock prices by modeling the (instantaneous) probability that a Bull or Bear market terminates as a function of its age and a set of underlying state variables, such as interest rates. A random walk model is rejected both for bull and bear markets. Although it fits the data better, a generalized autoregressive conditional heteroscedasticity model (GARCH) is also found to be inconsistent with the very long bull markets observed in the data. The strongest effect of increasing interest rates is found to be a lower bear market hazard rate (chance of ending, as time goes on) and hence a higher likelihood of continued declines in stock prices. Huang (2011) and Huang and Sclove (2011) segmented the time series of S&P500 monthly RORs using an HMM with Gaussian p.d.f.s with differing means and variances. Three values of the number K of states were tried, $K = 1, 2, 3$. With K states, the number of free parameters is $2K$ means and variances and $K(K - 1)$ free parameters in the transition probability matrix, a total of $2K + K(K - 1) = K(K - 1 + 2) = K(K + 1)$. This is 6 for $K = 2$, 12 for $K = 3$, and just 2 for $K = 1$. A conclusion that $K = 1$ fits best corresponds to no state-switching and might justify fitting a single ARIMA model over all time points. A conclusion that $K = 2$ corresponds to the conventional use of two states, Bull and Bear, particularly if one state has a positive mean ROR and the other has a negative mean. A conclusion that $K = 3$ would suggest another state, perhaps intermediate between Bull and Bear.

BIC (Schwarz 1978) was used to compare models. BIC_k is an expansion of $-2 \ln(\text{posterior probability of Model } k)$, and so can be put onto a scale of posterior probability. That is, the posterior probability of Model k is approximately proportional to $\exp(-BIC_k/2)$. (This assumes equal prior probabilities of $1/K$ but the expression can be adjusted for unequal prior probabilities.) That is, the posterior probability of Model k is approximately

$$\exp[-BIC_k/2] / \sum_{j=1}^K \exp(-BIC_j/2).$$

It was found that the conventional number of states, $K = 2$, has the best (lowest) BIC and hence the highest posterior probability, with a larger variance corresponding to the Bear state.

3. Formulation of HMMs

Next the likelihood function for HMMs will be discussed.

Joint Distribution. The joint distribution of the observations in a time series may be built up from

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2|x_1)f(x_3|x_2, x_1) \cdots f(x_n|x_n, x_{n-1}, \dots, x_2, x_1).$$

In the case of a first order Markov process, this becomes

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2|x_1)f(x_3|x_2) \cdots f(x_n|x_n).$$

The joint distribution for an HMM takes account of both the observations and the hidden states. The observation vector x_t

is (y_t, s_t) . Thus,

$$\begin{aligned} f(x_t | x_{t-1}) &= f(y_t, s_t | y_{t-1}, s_{t-1}) = f(y_t | s_t, y_{t-1}, s_{t-1})f(s_t | y_{t-1}, s_{t-1}) \\ &= f(y_t | s_t)f(s_t | s_{t-1}) = f_{s_t}(y_t)\pi_{s_{t-1}, s_t}. \end{aligned}$$

Here we have used the facts of the HMM model formulation that

$$f(y_t | s_t, y_{t-1}, s_{t-1}) = f(y_t | s_t) = f_{s_t}(y_t)$$

and

$$f(s_t | y_{t-1}, s_{t-1}) = f(s_t | s_{t-1}) = \pi_{s_{t-1}, s_t}.$$

For a given state sequence $s_0, s_1 s_2 \dots, s_n$, the probability is

$$p(s_0) \pi_{s_0, s_1} f(y_1 | s_1) \pi_{s_1, s_2} f(y_2 | s_2) \dots \pi_{s_{n-1}, s_n} f(y_n | s_n).$$

This involves also an initial distribution $p(s_0)$, $s_0 = 1, 2, \dots, K$. The marginal distribution of the observable sequence $y_1, y_2, \dots, y_n$ is a sum of this over all such possible state sequences, since the event that $y_1, y_2, \dots, y_n$ occurs is the union of their occurrence over possible state sequences.

The Likelihood. The likelihood is the joint distribution of the observations, with the observations taken as given and the parameters taken as variables. The likelihood is to be maximized over the parameters, both the distributional parameters and the transition probabilities.

3.1. Estimating the Parameters of an HMM.

The parameters may be estimated by direct maximization of the likelihood or by iterative numerical methods such as the EM (Expectation/Maximization) algorithm; see, *e.g.*, McLachlan and Krishnan (1997).

In the HMM model, it is natural to alternate between estimating the distributional parameters and the transition parameters. First, make an initial guess of the states. For example, in the case of a scalar observation and two states, estimate low values as State 1 and high values as State 2. Then estimate the distributional parameters by method of moments, or maximum likelihood (giving an EM algorithm). Then, for time t , given that $\hat{s}_{t-1} = r_{t-1}$, relabel according to $\hat{s}_t = \arg \max_s \hat{\pi}_{r_{t-1}, s} \hat{f}(x|s)$. Then re-estimate the distributional parameters. *Etc.*

A special case of the EM algorithm is the Baum-Welch algorithm (Baum and Welch; Baum, Petrie, Soules and Weiss 1970). In the Baum, Petrie, Soules and Weiss article (1970), there is a reference to a paper by Baum and Welch submitted to the Proceedings of the National Academic of Sciences (PNAS), but the archives of PNAS reveal no publication by Baum and Welch. The Baum-Welch algorithm has become an important tool in many fields, first especially in speech recognition. The algorithm estimates the probability distribution $\pi_t(k) = \Pr\{S_t = k\}$ over the states of the hidden Markov chain at each time point. Though formulated before any general formulation of the EM algorithm, the Baum-Welch algorithm is an EM algorithm. We do not elaborate further on the Baum-Welch algorithm here. For further discussion of it, see, *e.g.*, McLachlan and Krishnan (1997).

3.2. More on Transition Probabilities

As above, let $\pi_{jk} = \Pr\{S_t = k | S(t-1) = j\}$. These are the *transition probabilities*. The process is *homogeneous* if these are constant across time. The *transition probability matrix* is $\mathbf{T} = [\pi_{jk}]$. The distribution across states at time t is the vector $\mathbf{v}_t = (p_1 p_2 \dots p_S)'$, where S is the number of states. Then to move forward one time period, $\mathbf{v}'_t = \mathbf{T} \mathbf{v}'_{t-1}$.

Having estimated the transition probabilities, one may solve for the steady-state, long run distribution across the states. If $\mathbf{T}$ is the transition probability matrix and $\mathbf{v}$ is the steady-state distribution, then to find $\mathbf{v}$, one solves the set of simultaneous linear equations $\mathbf{v}'\mathbf{T} = \mathbf{v}'$, *i.e.*, $\mathbf{T}'\mathbf{v} = \mathbf{v}$, which is the homogeneous set of equations $(\mathbf{T}' - \mathbf{I})\mathbf{v} = \mathbf{0}$,

3.3. Bull and Bear Portfolios

Having segmented the relevant time series of market indicators into states such as Bull and Bear, one may then estimate the mean vector and covariance matrix of a portfolio of selected stocks and

form an optimal Bull portfolio and an optimal Bear portfolio. The estimate of an optimal weight vector w is a function of the estimate m of the mean vector and S of the covariance matrix, say $w^* = H(m, S)$. There are estimates m_0, S_0 and m_1, S_1 for Bull and Bear states, resp.;

$$w_0^* = H(m_0, S_0), w_1^* = H(m_1, S_1).$$

One would try to predict whether or not there will be a change in the state of the market, and rebalance or not, accordingly. A decision risk analysis should be performed to see if the expected increase in portfolio ROR would compensate for transaction costs incurred in rebalancing. The transaction costs might be proportional to the amount of buying and selling across the m stocks, that is, proportional to $\sum_{a=1}^m |w_{1a}^* - w_{0a}^*|$, where w_{ka}^* is the weight of asset a in the portfolio for state k , for $k = 0, 1$.

4. Example: GDP – Gross Domestic Product

Next, a macro-economic example, GDP, will be considered further. Gross domestic product is the total monetary or market value of all the finished goods and services produced within the country's borders in a specific time period. GDP is similar to GNP (Gross National Product), except the GNP includes imports and exports, whereas GDP excludes them. This modeling could be applied to GNP as well.

In terms of time series analysis, a seasonal ARIMA model for, say, log quarterly GDP could be discussed. It is interesting to compare this with a model with different states, although even if it happened that a single ARIMA model fit better, one might still want to fit a model with states, which labels the time points. Chen and Sclove (2011) fit a state model to GDP. The time series of quarterly growth rates of US GDP, from 1947 through the third quarter of 2010, was segmented by hidden Markov models (HMMs). HMMs with several states were fit and compared with a single distribution for the growth rate. State-conditional Normal distributions with different means and variances were fit for different numbers of states. The extent to which states correspond to recession, recovery, expansion, and contraction was assessed. The HMMs were scored by BIC (Schwarz 1978). Some comparison was made to ARIMA models involving regular and quarterly differencing and regular and quarterly autoregression of log GDP. Components of GDP were also fit with HMMs, with a view toward determining which components are leading or lagging indicators of the state of overall GDP.

5. Further Reading

The model selection criterion BIC was formulated by Schwarz (1978); the paper by Kashyap (1982) is also very helpful. To begin a study of HMMs, it is helpful to begin with the Viterbi algorithm, which finds the most likely state sequence. See Viterbi (1968) and Forney (1978). Then for further study on the HMM, one may refer, for example, to McLachlan and Krishnan (1997) or Zucchini and MacDonald (2009) or Zucchini, MacDonald and Langrock (2016).

6. Summary

Stochastic processes, such as time series, may have *states*. These are phases through which the process passes over time. Examples are Bull and Bear states of the stock market, and Recession, Recovery, Expansion, and Contraction in the economy. A segment (regime, epoch) is a sequence of time points having the same state. Segmentation involves labeling each time point with its most likely state, or giving a probability distribution over the states for each time point.

Hidden Markov models involve state-conditional p.d.f.s and transitions between states, with the corresponding transition probabilities. The state process is a Markov chain with a transition probability matrix. Alternative, competing models with different numbers of states and different state-conditional distributions can be compared with criteria such as BIC (Bayesian Information Criterion). The values of BIC for the alternative models can be put on a scale of posterior probability.

Table 1. Excess monthly RORs (%) of market and fund, with Bull/Bear state defined as Bull3. S3 is the moving sum of 3.

Month market			fund	Bull	S3	Bull3	Month market			fund	Bull	S3	Bull3
5 yrs ago	J	2.16%	2.45%	1	*	1	3	J	-1.13%	-1.80%	0	1	0
	F	-0.32%	-0.50%	1	*	1	yrs	A	1.07%	1.52%	1	1	0
	M	0.73%	0.75%	1	3	1	ago,	S	-9.61%	-9.98%	0	1	0
	A	0.82%	1.35%	1	3	1	cont'd	O	-18.62%	-18.24%	0	1	0
	M	-3.53%	-3.75%	0	2	1		N	-7.80%	-8.80%	0	0	0
	J	-0.39%	-0.78%	1	2	1		D	0.78%	1.35%	1	1	0
	J	0.09%	0.22%	1	2	1	2	J	-8.97%	-9.69%	0	1	0
	A	1.69%	1.96%	1	3	1	yrs	F	-11.67%	-9.11%	0	1	0
	S	2.03%	2.16%	1	3	1	ago	M	8.18%	7.58%	1	1	0
	O	2.69%	3.13%	1	3	1		A	8.96%	8.61%	1	2	0
4 yrs ago	N	1.22%	1.31%	1	3	1		M	5.16%	5.51%	1	3	1
	D	0.85%	1.40%	1	3	1		J	0.00%	0.31%	1	3	1
	J	0.98%	1.14%	1	3	1		J	7.14%	8.69%	1	3	1
	F	-2.63%	-2.99%	0	2	1		A	3.29%	3.64%	1	3	1
	M	0.58%	0.71%	1	2	1		S	3.50%	4.34%	1	3	1
	A	3.83%	4.51%	1	2	1		O	-2.00%	-2.37%	0	2	1
	M	2.81%	3.27%	1	3	1		N	5.57%	4.67%	1	2	1
	J	-2.18%	-1.73%	0	2	1		D	1.76%	2.91%	1	2	1
	J	-3.65%	-4.22%	0	1	1	1	J	-3.77%	-4.13%	0	2	1
	A	0.93%	0.94%	1	1	1	yr	F	2.80%	4.51%	1	2	1
3 yrs ago	S	3.19%	4.16%	1	2	1	ago	M	5.70%	6.28%	1	2	1
	O	1.15%	1.29%	1	3	1		A	1.45%	1.32%	1	3	1
	N	-4.78%	-3.23%	0	2	1		M	-8.57%	-8.55%	0	2	1
	D	-1.12%	-0.56%	0	1	1		J	-5.55%	-5.81%	0	1	1
	J	-6.54%	-7.32%	0	0	0		J	6.64%	7.09%	1	1	1
	F	-3.71%	-1.97%	0	0	0		A	-4.87%	-5.05%	0	1	1
	M	-0.70%	-1.36%	0	0	0		S	8.38%	8.96%	1	2	1
	A	4.54%	4.38%	1	1	0		O	3.61%	2.45%	1	2	1
	M	0.92%	1.92%	1	2	0		N	-0.24%	0.59%	0	2	1
	J	-9.14%	-9.15%	0	2	0		D	6.31%	7.31%	1	2	1

References

1. Baum, Leonard E., and Welch, Lloyd (ca. 1970). A statistical estimation procedure for probabilistic functions of finite Markov processes. Submitted for publication to *Proceeding of the National Academy of Sciences, U.S.A.*

2. Baum, Leonard E., Petrie, Ted, Soules, George, and Weiss, Norman (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, **41**, 164-171.

3. Chen, Yu, and Sclove, Stanley L. (2011). Segmenting the time series of quarterly GDP using a hidden Markov model. Presentation at the Joint Statistical Meetings (American Statistical Association and related societies), Miami Beach, FL.

4. Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, **39**, 1-38.

5. Forney, G. D. (1973). The Viterbi algorithm. *Proceedings of the IEEE*, **61**, 268-278.

6. Huang, Ziqian (2011). Segmentation of financial and marketing data: mixture logit model and hidden Markov model. PhD dissertation, Department of Information & Decision Sciences, University of Illinois Chicago.

7. Huang, Ziqian, and Sclove, Stanley L. (2011). Segmenting the time series of a market index using a hidden Markov model. Working Paper, Dept. of Information & Decision Sciences, University of Illinois at Chicago.

8. Investopedia (2011). Bear market.
www.investopedia.com/terms/b/bearmarket.asp#axzz1gXPbjeE2
9. Kashyap, Rangasami L. (1982). Optimal choice of AR and MA parts in autoregressive moving average models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **4**, 99-104.
10. Lunde, Asger, and Timmermann, Allan (2004). Duration dependence in stock prices: an analysis of Bull and Bear markets. *Journal of Business and Economic Statistics*, **22**, 253-273.
11. Maheu, John M., and McCurdy, Thomas H. (2000). Identifying Bull and Bear markets in stock returns. *Journal of Business and Economic Statistics*, **18**, 100-112.
12. McLachlan, Geoffrey, and Krishnan, Thiriyambakam (1997). *The EM Algorithm and Extensions*. Wiley, New York.
13. McLachlan, Geoffrey, and Peel, David (2000). *Finite Mixture Models*. Wiley, New York.
14. Schwarz, Gideon (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.
15. Schwert, G. W. (1990). Indexes of U.S. stock prices from 1802 to 1987. *Journal of Business*, **63**, 399-426.
16. Vanguard Group (2011). Staying calm during a Bear market.
retirementplans.vanguard.com/VGApp/pe/PubVgiNews?ArticleName=Stayingcalmbearmkt
17. Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, **13**, 260-269.
18. Zucchini, Walter, and MacDonald, Iain L. (2009). *Hidden Markov Models for Time Series: an Introduction using R*. Chapman & Hall/CRC (Taylor & Francis Group), Boca Raton, FL.
19. Zucchini, W., MacDonald, I.L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R, 2nd ed.* Chapman & Hall/CRC (Taylor & Francis Group), Boca Raton, FL.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.