Hierarchical Hidden Markov Model of High-Frequency Market Regimes using Trade Price and Limit Order Book Information

Abstract

현대의 시장은 double auction market mechanism에 의해 동작.

이 기제에서는 체결의 확실성과 거래 가격의 매력도 사이의 균형을 무너뜨릴 목적으로 시장 참여자가 매수 매도 주문을 제출.

2.1 Orders and Order Types

day orders의 생명주기는 상당히 짧고, 지배적인 시장 가격에 가깝게 주문이 나가며, 공개적으로 거래되는 거래량의 대부분을 차지.

반면 GTC orders는 현재 시장가격과 차이가 있는 가격에서 거래하고자 하는 투자자에 의해 사용됨.

market orders는 현재 시장가에서 매수 또는 매도하려는 주문.

limit orders는 원하는 가격보다 불리하지 않은 가격에서 거래하려는 주문.

2.2 Trades

모든 거래는 Time&Sales(T&S)에 기록됨.

주문 매치는 exchange mechanism이나 off-the-market을 통해 이루어짐.

exchange mechanism의 경우 매수와 매도에서 나온 주문들은 다른 중개인으로부터 제출된 주문.

off-the-market 거래는 매수와 매도 주문이 같은 중개인으로부터 제출된 경우를 말하며 crosses라고 불림.

2.3 Double Auction Market Mechanism – Limit Order Books

limit orders는 시스템에 의해 바로 성사되지 못 하는 경우가 있는 데, 이 경우 waiting queues에 주문이 쌓이고, 이를 limit order book(LOB) 또는 market depth라 함.

LOB는 bid와 ask stack에 정렬된 매수와 매도 limit orders의 집합체인데 정렬은 limit price와 order arrival time을 기준으로 행해짐.

3.1 Overview

time series 는, 이것의 model이 로 정의되면, linear하다라고 함.

는 의 mean

는 weights defining dynamic structure of with

는 sequence of i.i.d. random variables with mean zero and a well-defined distribution - is a shock at time t

이 모델이 암시하는 것은 data의 linearity에 대한 가정, linear model이 weak stationarity라는 가정임.

가 weak stationary time series라 가정하면, mean of 는 constant이고, 와 사이의 covariance는 constant for integer l 임. 즉, 주어진 lag에 대해 covariance는 constant임.

실제로 대부분의 financial data는 non-linear 함. 이는 linear model의 dependency of residuals로 해석될 수 있고, 비모수적 또는 모수적 검정을 통해 확인될 수 있음. 이러한 검정을 통과하지 못한다는건 선형성 가정의 부적당함과 선형모델의 비유용성 의미함.

많은 비선형 모델은 금융 시계열 모델링에서 유용하지만, 대량의 데이터를 처리하는데 있어 제약이 있음.

관측된 데이터를 바로 모델링하는 것 보다는 상태 전이와 각 상태에서의 관측을 모델링하는 것이 더 나은 결과를 만들 수 있음.

18p 3.2.1 Markov Models

underlying stochastic process가 Markov property를 가지고 있다고 가정된 stochastic models의 일종을 Hidden Markov Model이라 함.

present state와 past states가 주어졌을 때, process의 future states의 conditional probability distribution이 오직 present state에 의존적일 때 stochastic process는 Markov property를 가지고 있다고 함.

Markov process의 가장 간단한 형태인 discrete-time discrete-state Markov chain은 아래와 같음.

는 stochastic process

domain of 는 all possible state, S = {} 으로 구성되어 있음

process는 시간에 따라 진행됨

19p

Markov model은 an initial-state probability distribution 와 state-transition probabilities 가 필요함.

state-transition probabilities는 시간에 따라 변할 수도 있고, 상수일 수도 있음.

상수인 경우 Markov process는 stationary, time invariant로 모델링되고, set of transition probabilities는 a time-independent transition probability matrix, A로 표현됨.

위에서 기술된 Markov model은 상태 S가 직접적으로 관측 가능하고 측정될 수 있을 때 적용 가능함. 여러 제약 조건들로 인해 이 전제 조건이 충족되지 않을 수 있음.

또한 복잡한 물리적 현상의 현실적인 모델은 실제로 사용하기에는 너무 복잡한 state-space domain 정보가 필요함.

반면에 state space를 줄이게 되면 현실을 제대로 반영하지 못하게 됨.

3.2.2 The Structure of HMMs: Topology and Parameters

이런 문제를 해결하기 위해 a layer of hidden variables이 등장함.

20p

– a hidden stochastic process

– an observable stochastic process

– a prior distribution such that

A - a state transition conditional probability function in a matrix form such that

B = - a state dependent conditional probability function of observations such

that

HMM이 1회 순환을 마치면 관측열 O = 를 만들어 냄. 여기서 각 관측 는 V에 속하는 것중 하나임.

T는 the number of observations in the sequence 임.

HMM의 유일한 가정은 의 domain인 state space가 discrete 하다는 것.

21p

N개의 항아리가 있는 방이 있다고 가정.

각 항아리는 많은 수의 색깔이 있는 공을 담고 있음.

M개의 색이 있는 공이 있다고 가정.

첫번째 단계로, 초기 항아리를 무작위로 선택한 후에 공 하나를 복원 추출함.

공이 선택된 항아리에 대해 알고 있는 것이 없지만, 공의 색을 관측함.

관측된 공의 색을 기록됨.

다음 단계로, 항아리 하나를 마찬가지 방법으로 선택하고 공을 뽑아서 색을 기록.

이를 반복하면 a finite observations sequence of colors가 만들어지고, observable output from the HMM으로 모델링될 수 있음.

다양한 HMM 구조가 이 모델에 할당될 수 있지만, 가장 간단한 형태는 각 hidden state가 항아리 하나에 할당되는 것, 각 스텝에서 항아리의 선택은 the state transition matrix of the HMM에 의해 진행되며, color (ball) selection은 각 항아리의 조건부 확률 분포에 기반함.

3.2.3 Learning

HMM을 사용한 모델링은 몇 가지 스텝을 밟게 됨.

첫번째 스텝인 finding state-space는 일반적으로 모델링의 대상이 되는 물리적 현상에 대한 싶은 이해가 요구됨.

일단 hidden states가 결정이 되면, a set of observation states를 정해야 함. same hidden states에서는 작은 편차를 갖고, different hidden states 사이에는 명확하게 다른 properties of observation states를 찾아야 함. 그런 속성을 더 잘 찾을 수록, HMM이 자신의 임무(identify a sequence of hidden states based on a given sequence of observations)를 더 잘 수행할 수 있게 됨.

hidden and observation states가 정해지고 모델의 구조가 결정되면, model parameters를 학습할 차례임.

data set이 완전하다면, estimation of parameters는 maximum likelihood(ML) estimation이나 Bayesian Estimation으로 귀결되지만, 실제로 data set은 불완전함. (sometimes

values are missing from the training data, or variables are simply unobservable (latent) )

22p

hidden variables의 케이스의 경우, ML estimation이 적용될 수 있지만, likelihood function이 missing data를 포함해야 함.

O가 the observed training set이고, H가 the set of hidden variables라면, the log-likelihood function는 아래와 같음.

개체이(가) 표시된 사진

자동 생성된 설명

ML estimate은 argument 이고, 이 인수가 log-likelihood function의 값을 최대화 함.



최대우도추정(Maximum Likelihood Estimation) <https://ratsgo.github.io/statistics/2017/09/23/MLE/>

23p

parameter estimation in presence of hidden variables의 또 다른 표준적인 접근은 Baum-Welch algorithm임.

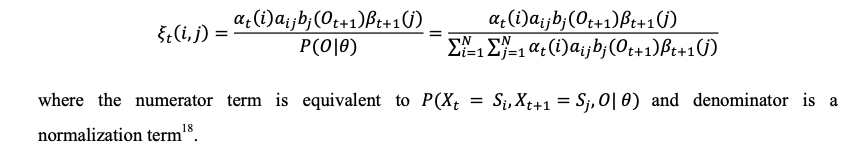
알고리즘의 첫번째 스텝은 expectation step임. current parameter estimates 를 사용해서 values for missing data를 expectation으로서 얻게됨.

두번째 스텝은 maximization step임. the estimated data points from the first step를 통해 initial data set이 보충되고, new best estimates for model parameters, 를 얻는 데 사용됨. 이렇게 새롭게 갖게 되는 model parameters는 다음 반복의 첫번째 스텝에서 사용됨.

a discrete time finite state HMM를 상정해 보자.







24p