CAPTURING ORDER IMBALANCE WITH HIDDEN MARKOV MODEL: A CASE OF SET50 AND KOSPI50

Po-Lin Wu, Thammasat University
Wasin Siwasarit, Thammasat University

Motivation

"How long does it take to remove serial dependence over a daily horizon? The pattern of intra-day serial dependence reveals that it takes more than 5 minutes but less than 60 minutes"

Chordia et al (2005)

Motivation

"Prices were closer to random walk benchmarks in the more liquid regime than in others one … these findings indicate that liquidity stimulates arbitrage activity, which, in turn, enhances market efficiency"

Chordia et al(2008)

Objectives of this Study

THAMMASAT BUSINESS SCHOOL

(1) Use Hidden Markov Model to capture the states of order imbalance of selected stocks from SET50 and KOSPI50 in a consistent and confident manner

(2) Build and back-test strategies by using the signals generated from proposed models

(3) Compare the result between markets, across different frequencies and with traditional buy-and-hold strategies

Outline

- Scope of data
- Methodology
- Result
- Conclusion and Recommendation

Scope of Data

THAMMASAT BUSINESS SCHOOL

- Market: 10 stocks from SET50 (Thailand) and KOSPI50 (Korea)
- Input Data: mid-point closing price, bid size, ask size
- Scope of data: 1st Oct 2016 to 31st Jan 2017
- Frequency of data: Intra-day data
  - Interval: 5 minutes, 10 minutes and 30 minutes

Scope of Data

- Stocks Selection Criteria
  - **Stocks that are consistently listed on SET50 and KOSPI50 Index during the period from 1st January 2012 to 31st July 2016**
  - **Select 10 stocks with highest 250 days average daily volume turnover**
- Data for initial-training and back-testing
  - **Initial Training: 1st Oct 2016 to 31st Oct 2016**
  - **Back testing: 1st Nov 2016 to 31st Jan 2017**

Selected Stocks: SET50

THAMMASAT BUSINESS SCHOOL

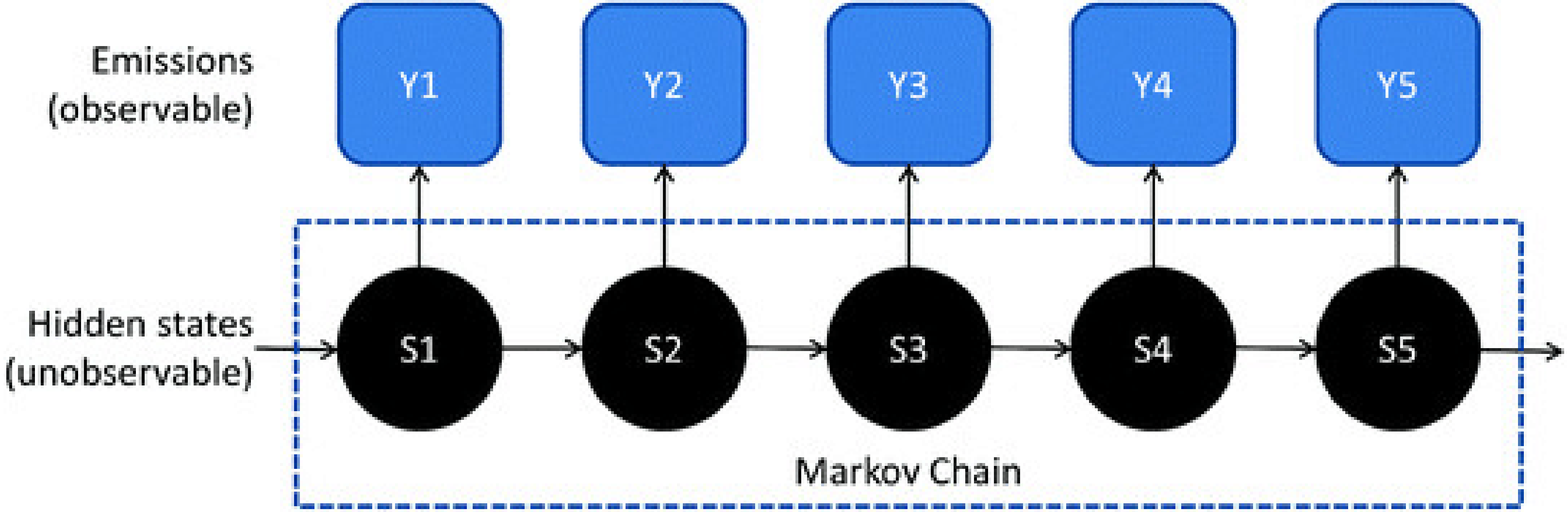
Ticker	Company Name	Sector
ADVANC.BK	Advance Info Service PCL	Information & Communication
BANPU.BK	Banpu PCL	Energy & Utilities
BCP.BK	Bangchak Petroleum PCL	Energy & Utilities
CPF.BK	Charoen Pokphand Foods PCL	Food and Beverage
DTAC.BK	Total Access Communication PCL	Information & Communication
IRPC.BK	IRPC PCL	Energy & Utilities
IVL.BK	Indorama Ventures PCL	Petrochemicals & Chemicals
PTTEP.BK	PTT Exploration and Production PCL	Energy & Utilities
TCAP.BK	Thanachart Capital PCL	Banking
TRUE.BK	True Corporation PCL	Information & Communication

Selected Stocks: KOSPI50

TICKER	Company Name	Sector
034220.KS	LG Display Co, Ltd	Electrical & Electronic Equipment
066570.KS	LG Electronics Inc	Electrical & Electronic Equipment
051910.KS	LG Chem Co, Ltd	Chemicals
005490.KS	POSCO	Iron & Metal Products
006400.KS	Samsung SDI Co, Ltd	Electrical & Electronic Equipment
009150.KS	Samsung Electro Mechanics Co Ltd	Electrical & Electronic Equipment
010140.KS	Samsung Heavy Industry Co, Ltd	Transport Equipment
000880.KS	Hanwha Corp	Finance
000720.KS	Hyundai Engineering & Construction Co Ltd	Construction
009540.KS	Hyundai Heavy Industry Co, Ltd	Transport Equipment

Hidden Markov Model

Basic Idea: Hidden Markov Model



Komorowski (2016)

Possible states of risky assets

- Three fundamental states:
  - Asset is not adjusted to positive information
  - Asset is not adjusted to negative information
  - Asset price is in equilibrium
  - Other possible states

- This study considers models from 3 states to 5 states

Order Imbalance

- **3 indicators from previous literature (Chordia et al, 2005)**
  1. **Number of buy order less number of sell order**
  2. **Number of buy-initiated shares purchased less number of seller-initiated shares sold**
  3. **Dollars paid by buy initiators less dollars received by sell initiators**

Order Imbalance

THAMMASAT BUSINESS SCHOOL

- In this study, we define the order imbalance indicator as:

$$\text{OIR} = \frac{V_t^B}{V_t^A + V_t^B}$$

$$V_t^B = size\ of\ bid\ order\ at\ best\ price\ at\ time\ t$$
$$V_t^A = size\ of\ ask\ order\ at\ best\ price\ at\ time\ t$$

Generating trading signal

- We plan to generate trading signal by 2 approaches:
  - Discrete Case
  - Continuous Case

Data Discretization

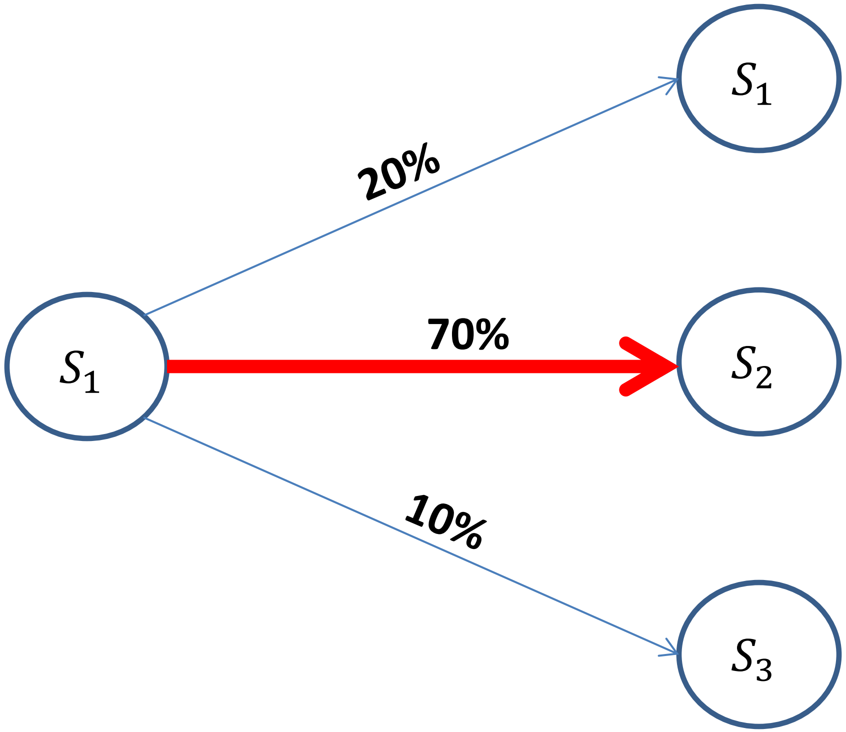
THAMMASAT BUSINESS SCHOOL

- **Before the model training, data needs to be discretized:**

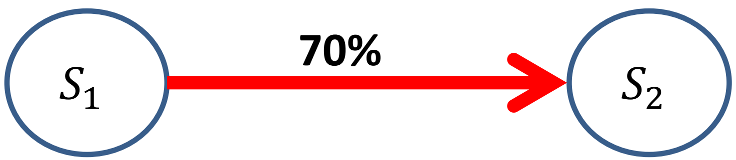
Symbol	Price movement	Order Imbalance Ratio
1	$0\% <$	$< 25\%\ \boldsymbol{percentile}$
2	$0\% <$	$25\%\ \boldsymbol{percetile} \leq \mathbf{OIR} \leq \mathbf{0.75}\%\ \boldsymbol{percentile}$
3	$0\% <$	$\mathbf{OIR} > 75\%\ \mathbf{percentile}$
4	$\geq 0\%$	$< 25\%\ \mathbf{percentile}$
5	$\geq 0\%$	$25\%\ \mathbf{quantile} \leq \mathbf{OIR} \leq \mathbf{0.75}\%\ \mathbf{quantile}$
6	$\geq 0\%$	$\mathbf{OIR} > 75\%\ \mathbf{quantile}$

Be noted: the percentile is computed by averaging the percentile of each selected stocks during the initial training period

Generating trading signal



Case1 :Discrete Trading signal



Probability of emission:

Symbol	Return	Probability
1	$0\% <$	5%
2	$0\% <$	5%
3	$0\% <$	5%
4	$\geq 0\%$	30%
5	$\geq 0\%$	30%
6	$\geq 0\%$	25%

$$P(q_{t+1} = S_2 | q_t = S_1) P(r \geq 0 | S_2)$$

If $P \geq 80\%\ threshold$, then signal is generated

Intuition behind threshold

- This study aims to capture the price movement in a confident manner:
  - **Confidence in transition: 90%**
  - **Confidence in observing the desire movement: 90%**

- The joint-probability gives us an approximate number of 80% for our threshold value

Trading strategy

1. **Generate a list of signals that we should enter long position**
2. **Liquidate any stocks that are not in the list**
3. **If there is any remaining wealth, allocate wealth equally to all stocks in the list**
4. **If current interval is the end of the day, then re-train the models**
5. **Proceed to next period**

The study will consider bi-directional transaction cost at 0.05%

Performance evaluation

- Hit ratio: how well the signal is able to predict positive return for each individual stock

$$\boldsymbol{Hit\ Ratio = \frac{\#\ correct\ positive\ movement\ signals}{\#signals\ results\ in\ non-zero\ movement}}$$

- t-test: $\boldsymbol{H_0\!: Hit\ ratio = 0.5, H_a\!: Hit\ rato > 0.5}$

Performance evaluation

- Benchmark: SET and KOSPI total return Index
- Sharpe's ratio
- Jenson's Alpha

Result

Predictability

Hit ratio: SET50, Discrete case

THAMMASAT BUSINESS SCHOOL

	3 states	4 states	5 states
5 min	78.61%	80.60%	83.38%
10 min	72.21%	72.41%	67.15%
30 min	71.81%	62.89%	62.20%

Observations

THAMMASAT
BUSINESS SCHOOL

- As frequency decreases, the predictability of the signals decreases.

- At highest frequency, model with higher number of states achieve higher hit ratio.

Hit ratio: KOSPI50, Discrete case

	3 states	4 states	5 states
5 min	72.71%	58.93%	71.57%
10 min	66.16%	69.61%	59.41%
30 min	60.00%	43.04%	45.04%

Observations

- Similar pattern is obersved, as frequency decreases, the hit ratio decreases

- Compare to Thai market:
  - hit ratio is lower across all frequency and models
  - Though having longer trading hours, number of signals is significantly lower

THAMMASAT BUSINESS SCHOOL

Profitability

Jenson’s Alpha SET50, Discrete case

	3 states	4 states	5 states
5 min	2.87%*** (0.00538)	3.57%*** (0.00493)	4.73%*** (0.00399)
10 min	1.06%*** (0.00233)	0.14% (0.00175)	0.41% (0.00343)
30 min	0.77%*** (0.00157)	0.05% (0.00137)	0.06% (0.00134)

Be noted: 0.05% bi-directional transaction cost assumed. Number of stars represent the level of significance, with *** represents 0.01 significance, ** represents 0.05 significance and * represents 0.1 significance. The HAC standard errors are presented below the estimators.

Observation

THAMMASAT
BUSINESS SCHOOL

- Outperform the market in the highest frequency due to high predictability

- Lower predictability in lower frequency cases result in insignificant alpha

Jenson's Alpha
KOSPI50, Discrete case

	3 states	4 states	5 states
5 min	-0.09% (0.00062)	0.05% (0.00082)	0.03% (0.00154)
10 min	-0.10% (0.00101)	-0.28%*** (0.00089)	-0.17%** (0.00077)
30 min	-0.04% (0.00035)	-0.08% (0.00083)	-0.17%* (0.00095)

Be noted: 0.05% bi-directional transaction cost assumed. Number of stars represent the level of significance, with *** represents 0.01 significance, ** represents 0.05 significance and * represents 0.1 significance. The HAC standard errors are presented below the estimators.

Observation

THAMMASAT
BUSINESS SCHOOL

- Strategy could not outperform the market or generate significant alpha due to lower predictability

- Strategy make a loss in most cases due to transaction cost and incorrect predictions

Conclusion

Performance of models

THAMMASAT
BUSINESS SCHOOL

- Appropriate number of states should be further investigated.
  - At 5 minute frequency, the 5 states model achieves decent hit-ratio
  - As frequency decreases, the 3 states model seems to outperform

Effect of frequency on performance of models

- Observed from both market, the higher the frequency, the higher the predictability. The same goes for profitability.

- Consistent with the literature by Chordia et al (2005): Price adjustment to information occurs on the intra-day level and predictability tends to dispear when frequency decreases.

Effect of market liquidity on performance of models

THAMMASAT BUSINESS SCHOOL

- In market with higher liquidity, the model is less consistent and confident:
  - Generate less signals
  - Achieve lower hit ratio

- Consistent with previous literature (Chordia et al, 2008): In a high liquidity environment, the cost of trading is lower (ex. bid-ask spread). Hence, investors have more incentives to exploit the deviation of asset price from equilibrium. This enhances the speed of price adjustment to new information.

Limitation

- Short period of study
- Only consider stocks that satisfied the defined criteria
- For trading strategy:
  - No short selling
  - Assume we can trade at mid-price.
  - Assume we can trade at interval closing price
  - Assume no limitation on number of shares we can buy.

Recommendation

- Focus on data of higher frequency
- Re-consider the data discretization method
- Re-consider the method of constructing order imbalance indicator