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

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

(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

- Market: 10 stocks from SET50 (Thailand) and KOSPI50 (Korea)

- Input Data: mid-point closing price, bid size, ask size

- Scope of data: 1\textsuperscript{st} Oct 2016 to 31\textsuperscript{st} 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
  – Back testing: 1st Nov 2016 to 31st Jan 2017
## Selected Stocks: SET50

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Company Name</th>
<th>Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVANC.BK</td>
<td>Advance Info Service PCL</td>
<td>Information &amp; Communication</td>
</tr>
<tr>
<td>BANPU.BK</td>
<td>Banpu PCL</td>
<td>Energy &amp; Utilities</td>
</tr>
<tr>
<td>BCP.BK</td>
<td>Bangchak Petroleum PCL</td>
<td>Energy &amp; Utilities</td>
</tr>
<tr>
<td>CPF.BK</td>
<td>Charoen Pokphand Foods PCL</td>
<td>Food and Beverage</td>
</tr>
<tr>
<td>DTAC.BK</td>
<td>Total Access Communication PCL</td>
<td>Information &amp; Communication</td>
</tr>
<tr>
<td>IRPC.BK</td>
<td>IRPC PCL</td>
<td>Energy &amp; Utilities</td>
</tr>
<tr>
<td>IVL.BK</td>
<td>Indorama Ventures PCL</td>
<td>Petrochemicals &amp; Chemicals</td>
</tr>
<tr>
<td>PTTEP.BK</td>
<td>PTT Exploration and Production PCL</td>
<td>Energy &amp; Utilities</td>
</tr>
<tr>
<td>TCAP.BK</td>
<td>Thanachart Capital PCL</td>
<td>Banking</td>
</tr>
<tr>
<td>TRUE.BK</td>
<td>True Corporation PCL</td>
<td>Information &amp; Communication</td>
</tr>
</tbody>
</table>
## Selected Stocks: KOSPI50

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

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

\[
OIR = \frac{V_t^B}{V_t^A + V_t^B}
\]

- \(V_t^B\) = 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

Before the model training, data needs to be discretized:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Price movement</th>
<th>Order Imbalance Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0% &lt;</td>
<td>&lt; 25% percentile</td>
</tr>
<tr>
<td>2</td>
<td>0% &lt;</td>
<td>25% percentile ≤ OIR ≤ 0.75% percentile</td>
</tr>
<tr>
<td>3</td>
<td>0% &lt;</td>
<td>OIR &gt; 75% percentile</td>
</tr>
<tr>
<td>4</td>
<td>≥ 0%</td>
<td>&lt; 25% percentile</td>
</tr>
<tr>
<td>5</td>
<td>≥ 0%</td>
<td>25% quantile ≤ OIR ≤ 0.75% quantile</td>
</tr>
<tr>
<td>6</td>
<td>≥ 0%</td>
<td>OIR &gt; 75% quantile</td>
</tr>
</tbody>
</table>

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

\[ S_1 \rightarrow S_2 \rightarrow S_3 \]

- From \( S_1 \) to \( S_1 \): 70%
- From \( S_1 \) to \( S_2 \): 20%
- From \( S_2 \) to \( S_3 \): 10%
Case 1: Discrete Trading signal

Symbol          | Return | Probability |
----------------|--------|-------------|
1               | 0% <   | 5%          |
2               | 0% <   | 5%          |
3               | 0% <   | 5%          |
4               | ≥ 0%   | 30%         |
5               | ≥ 0%   | 30%         |
6               | ≥ 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

\[
\text{Hit Ratio} = \frac{\text{# correct positive movement signals}}{\text{#signals results in non–zero movement}}
\]

• t-test: \( H_0: \text{Hit ratio} = 0.5, H_a: \text{Hit ratio} > 0.5 \)
Performance evaluation

• Benchmark: SET and KOSPI total return Index
• Sharpe’s ratio
• Jenson’s Alpha
Result
Predictability
## Hit ratio: SET50, Discrete case

<table>
<thead>
<tr>
<th></th>
<th>3 states</th>
<th>4 states</th>
<th>5 states</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5 min</strong></td>
<td>78.61%</td>
<td>80.60%</td>
<td>83.38%</td>
</tr>
<tr>
<td><strong>10 min</strong></td>
<td>72.21%</td>
<td>72.41%</td>
<td>67.15%</td>
</tr>
<tr>
<td><strong>30 min</strong></td>
<td>71.81%</td>
<td>62.89%</td>
<td>62.20%</td>
</tr>
</tbody>
</table>
Observations

• 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

<table>
<thead>
<tr>
<th></th>
<th>3 states</th>
<th>4 states</th>
<th>5 states</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 min</td>
<td>72.71%</td>
<td>58.93%</td>
<td>71.57%</td>
</tr>
<tr>
<td>10 min</td>
<td>66.16%</td>
<td>69.61%</td>
<td>59.41%</td>
</tr>
<tr>
<td>30 min</td>
<td>60.00%</td>
<td>43.04%</td>
<td>45.04%</td>
</tr>
</tbody>
</table>
Observations

• Similar pattern is observed, 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
Profitability
## Jenson’s Alpha

### SET50, Discrete case

<table>
<thead>
<tr>
<th></th>
<th>3 states</th>
<th>4 states</th>
<th>5 states</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 min</td>
<td>2.87%***</td>
<td>3.57%***</td>
<td>4.73%***</td>
</tr>
<tr>
<td></td>
<td>(0.00538)</td>
<td>(0.00493)</td>
<td>(0.00399)</td>
</tr>
<tr>
<td>10 min</td>
<td>1.06%***</td>
<td>0.14%</td>
<td>0.41%</td>
</tr>
<tr>
<td></td>
<td>(0.00233)</td>
<td>(0.00175)</td>
<td>(0.00343)</td>
</tr>
<tr>
<td>30 min</td>
<td>0.77%***</td>
<td>0.05%</td>
<td>0.06%</td>
</tr>
<tr>
<td></td>
<td>(0.00157)</td>
<td>(0.00137)</td>
<td>(0.00134)</td>
</tr>
</tbody>
</table>

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

• 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

<table>
<thead>
<tr>
<th></th>
<th>3 states</th>
<th>4 states</th>
<th>5 states</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 min</td>
<td>-0.09%</td>
<td>0.05%</td>
<td>0.03%</td>
</tr>
<tr>
<td></td>
<td>(0.00062)</td>
<td>(0.00082)</td>
<td>(0.00154)</td>
</tr>
<tr>
<td>10 min</td>
<td>-0.10%</td>
<td>-0.28%***</td>
<td>-0.17%**</td>
</tr>
<tr>
<td></td>
<td>(0.00101)</td>
<td>(0.00089)</td>
<td>(0.00077)</td>
</tr>
<tr>
<td>30 min</td>
<td>-0.04%</td>
<td>-0.08%</td>
<td>-0.17%*</td>
</tr>
<tr>
<td></td>
<td>(0.00035)</td>
<td>(0.00083)</td>
<td>(0.00095)</td>
</tr>
</tbody>
</table>

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

• 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

• 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

• 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