Hierarchical Hidden Markov Models in High-Frequency Stock Markets

Luis Damiano with Michael Waylandt and Brian Peterson
R/Finance 2018 | 2018-06-02
Agenda

- Motivation (30"")
- Hierarchical Hidden Markov Models (2’)
- Features (3’)
- Application (7’)
- Takeaway (1’)

R/Finance 2018 | Chicago, IL | 3/49
Motivation
Motivation = problem

Identify and predict price trends systematically in a profitable way
What we know = stylized facts

- Market behavior is complex and partially unknown
- Non-linear interactions between price and volume
- Multi-resolution: short-term trends within long-term trends
- High-frequency: noisy and large datasets need fast online computations
One approach (among many)

Ensemble of statistical and machine learning techniques

1. Create **intermediate** indicator variables
2. Combine into **discrete features** using technical analysis rules
3. **Build a hierarchy** to link all the features in a logical way
4. Apply clustering with **Markovian memory** (a parsimonious way to model non-linear correlations)
Hierarchical Hidden Markov Models
Why Hierarchical?

HMM cannot capture **multi-scale dynamics**.

- Recursive hierarchical generalization of the HMM.
- Systematic unsupervised approach for complex multi-scale structure.
- Motivated by multiplicity of **length scales** and the different **stochastic levels**.
- Inference on **correlation over long periods** via higher levels of hierarchy.
Figure 1: Hierarchical Hidden Markov Model for price and volume. Top states $z_1^1$ and $z_2^1$ represent bulls and bears.

1See a complete description in the write-up (see last slides).
Features
Sequence of triples \( \{y_k\} \)

\[ y_k = (t_k, p_k, v_k), \]

where \( t_k \leq t_{k+1} \) is the time stamp in seconds, \( p_k \) is the trade price and \( v_k \) is the trade volume.

In other words: tick-by-tick trade price and size, or L1 data.
How to make useful features?

[...] some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used. (Domingos 2012)

What would make features strong?

- **Underlying theory**: representative of our beliefs about how markets work (interactions between price and volume)
- **Empirical support**: when applied on real data, results are consistent with empirical evidence
- **Statistical properties**: captures non-linearities in a simple, parsimonious, and tractable way
- **Noise reduction**: by discretization
- **Computational complexity**: reduce dataset size
(1) Identify local extrema, where \( e_n \) is the price at the extreme.

(2) Create intermediate variables and features\(^2\):

- \( f_n^0 \) direction: up/down.
- \( f_n^1 \) price trend: up/down/no trend.
- \( f_n^2 \) volume trend: volume strengthens/weakens/is indeterminant.

\(^2\)See the appendix for a formal definition of the variables.
(3) Combine into 18 meaningful features linked hierarchically by the model.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Zig-zag</th>
<th>Price trend</th>
<th>Volume trend</th>
<th>Market State</th>
<th>Feature</th>
<th>Zig-zag</th>
<th>Price trend</th>
<th>Volume trend</th>
<th>Market State</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_1$</td>
<td>Up</td>
<td>+1</td>
<td>Up</td>
<td>Strong  +1</td>
<td>Bull</td>
<td>Dn</td>
<td>-1</td>
<td>Up</td>
<td>Weak  -1</td>
</tr>
<tr>
<td>$U_2$</td>
<td>Up</td>
<td>+1</td>
<td>Dn</td>
<td>Strong  +1</td>
<td>Bull</td>
<td>Dn</td>
<td>-1</td>
<td>Dn</td>
<td>Weak  -1</td>
</tr>
<tr>
<td>$U_3$</td>
<td>Up</td>
<td>+1</td>
<td>Up</td>
<td>Indet  0</td>
<td>Bull</td>
<td>Dn</td>
<td>-1</td>
<td>Up</td>
<td>Indet 0</td>
</tr>
<tr>
<td>$U_4$</td>
<td>Up</td>
<td>+1</td>
<td>No</td>
<td>Strong  +1</td>
<td>Bull</td>
<td>Dn</td>
<td>-1</td>
<td>No</td>
<td>Weak  -1</td>
</tr>
<tr>
<td>$U_5$</td>
<td>Up</td>
<td>+1</td>
<td>No</td>
<td>Indet  0</td>
<td>Local</td>
<td>Dn</td>
<td>-1</td>
<td>No</td>
<td>Indet 0</td>
</tr>
<tr>
<td>$U_6$</td>
<td>Up</td>
<td>+1</td>
<td>No</td>
<td>Weak  -1</td>
<td>Bear</td>
<td>Dn</td>
<td>-1</td>
<td>No</td>
<td>Strong  +1</td>
</tr>
<tr>
<td>$U_7$</td>
<td>Up</td>
<td>+1</td>
<td>Dn</td>
<td>Indet  0</td>
<td>Bear</td>
<td>Dn</td>
<td>-1</td>
<td>Dn</td>
<td>Indet 0</td>
</tr>
<tr>
<td>$U_8$</td>
<td>Up</td>
<td>+1</td>
<td>Up</td>
<td>Weak  -1</td>
<td>Bear</td>
<td>Dn</td>
<td>-1</td>
<td>Up</td>
<td>Strong  +1</td>
</tr>
<tr>
<td>$U_9$</td>
<td>Up</td>
<td>+1</td>
<td>Dn</td>
<td>Weak  -1</td>
<td>Bear</td>
<td>Dn</td>
<td>-1</td>
<td>Dn</td>
<td>Strong  +1</td>
</tr>
</tbody>
</table>
Figure 2: Tick by tick trades from SPY 2018-01-04 16:39:00/2018-01-04 16:41:00.
Figure 3: Extrema extracted from SPY 2018-01-04 16:39:00/2018-01-04 16:41:00.
Figure 4: Features extracted from SPY 2018-01-04 16:39:00/2018-01-04 16:41:00.
Application
Back tested on 12 stocks\(^3\), 17 days, 7 configurations:
\[12 \times 17 \times 7 = 1,428\] out of sample daily returns.

- For most stocks, HHMM outperforms buy & hold (B&H).
- Returns virtually uncorrelated with B&H.
- Sometimes HHMM offers less variance than B&H (further research needed).

\(^3\)Namely BBDb, BCE, CTCa, ECA, G, K, MGa, NXY, SJRb, SU, TCKb, TLM (all from Toronto Stock Exchange).
Figure 5: Equity curves for twelve stocks.
We now test the model against more relevant data: current, larger datasets from different assets in more competitive and liquid markets.\(^4\) A total of 55 million observations.

- Does the model generalize well?
  - Will the model structure be representative of the behaviour of other assets and markets?
  - Will the model perform similarly in different contexts?
  - Will significantly larger datasets pose new computational challenges?

\(^4\)Namely EFA, GLD, SPY, XLB, XLE, XLF, XLI, XLK, XLP, XLU, XLV, XLY. L1 data for 15 trading days each.
If not, ...

- What part of the model does not generalize?
- What can we learn from the deviances?
- What should we address next?
Latent state distinction - Hypothesis

Has the model learnt two distinct latent states?

- In financial terms: Do returns vary in each state?
- In statistical terms:
  - Are the conditional (given the latent state) and unconditional distributions of returns different?
  - Alternatively, do latent states contain information about the returns?

⚠ Note: Informativeness (i.e. the ability to extract latent information from observations) does not guarantee profitability.
Figure 6: Distribution of features from GLD 22017-12-29 14:30:00/2018-01-05 21:30:00 (in sample).
- Tayal (2009) finds that the relative frequency of the conditional returns is significantly different from the relative frequency of the unconditional returns.
- In our new application, there is enough evidence to argue that return characteristics vary per state as well.

---

5 Statistical tests are reported in the appendix.
Does the bullish regime have a greater mean return than the bearish regime?

- In **financial terms**: Are observed mean returns logically consistent with estimated states?
- In **statistical terms**: Is the mean return in the bullish state greater than the mean return in the bearish state?
Regime return characteristics - Results

- In-sample
  - Tayal (2009) finds strong in-sample evidence in favor of the hypothesis for the most liquid half of Canadian stocks.
  - In our new application, we also find in sample that the mean return in the bull state is greater than the mean return in the bear state.

- Out-of-sample:
  - Tayal (2009) finds strong evidence to answer the question positively for most Canadian stocks.
  - In our new application, no stock has statistically larger out-of-sample returns in bull states.
    - States are interchanged out-of-sample!
    - Some rather strong limitations to t-test assumptions apply (further research on a better comparison methodology needed).

---

6 Statistical tests are reported in the appendix.
Does the bullish regime have a positive mean return? Does the bearish regime have a negative mean return?

- **In financial terms**: Does the model capture runs and reversals correctly?
- **In statistical terms**: Is the mean return in the bullish state greater than zero? Is the mean return in the bearish state less than zero?
Regime return characteristics - Results

- In-sample:
  - Tayal (2009) finds strong evidence to answer the question positively for all Canadian stocks.
  - In our new application, all stocks have statistically positive (negative) in-sample returns in bull (bear) states.

- Out-of-sample
  - Tayal (2009) finds strong evidence in favor of the hypothesis for the most liquid half of Canadian stocks.
  - In our new application, none has statistically positive (negative) returns in bull (bear) states.
    - There seems to be a misclassification problem in top states.
    - Some rather strong limitations to t-test assumptions apply (further research on a better comparison methodology needed).

7 Statistical tests are reported in the appendix.
An informative model is not be profitable per se.

Our workflow:

1. Construct features from observed trade series.
2. Use features to make on-line inference about the latent states.
3. Use filtered states as a trading signal.
   - Go long when top level state switches to bullish (a run).
   - Go short when top level state switches to bearish (a reversal).
   - We trade with a one-tick lag because zig-zags are observed after completion.
   - We assume that we trade the next price (no fees).
Figure 7: Out-of-sample equity line (SPY 2018-01-02 14:30:00/2018-01-02 21:30:00).
Figure 8: Out-of-sample equity line (GLD 2018-01-05 14:30:00/2018-01-05 21:30:00).
Figure 9: Out-of-sample equity line (GLD 2018-01-08 14:30:00/2018-01-08 21:30:00).
Trading strategy - Example

Figure 10: Out-of-sample equity line (GLD 2018-01-02 14:30:00/2018-01-02 21:30:00).
Conclusions

- In sample, **the model shows a good fit** in both the original and the new applications.
  - Estimated bull and bear markets show the expected properties.

- Out of sample, **the model does not generalize well**.
  - Although the model learns distinct states, bull and bear out-of-sample returns do not exhibit reasonable characteristics.
  - Trading performance deteriorates along with the number of trades, a hint of bias.
Further research (1)

- Possible improvements:
  - The model should account for **bid-ask bounce**. In the proposed implementation, a bounce may trigger a trade.
  - More realistic feature engineering rules: volume bars (Easley, Lopez de Prado, and O’Hara 2012) and trade imbalance (Cont, Kukanov, and Stoikov 2014).
  - **More stable regimes**. With the current specification, top state has a median duration of 3 ticks. Market regimes are short lived.
  - The $\alpha$ threshold (change in volume) should be estimated to allow for a **smoother transition among features**. The suggestion that $\alpha = 0.25$ may not produce reasonable zig-zags outside the original application.
On the computational side, more relevant datasets are larger than the original application. Fully Bayesian inference is unreasonable as of today.

Further research is needed on either:

1. More efficient learning algorithm.
2. More efficient implementations of current algorithms.
Our fully-reproducible implementation is available in GitHub.

- L1 (tick by tick) data for 12 stocks (CC-BY-NC).\textsuperscript{8}
- R code for feature engineering and analysis (GNU-GPL 3).
- Stan code for Bayesian inference (GNU-GPL 3).
- Write-up with details about our replication (CC-BY).

---

\textsuperscript{8}Thomson Reuters has generously agreed to allow us to make the data available under the CC-BY-NC license. Please see the LICENSE file.

R package to run full Bayesian inference on Hidden Markov Models (HMM) using the probabilistic programming language Stan. By providing an intuitive, expressive yet flexible input interface, we enable non-technical users to carry out research using the Bayesian workflow.
Appendix
Feature engineering rules (1)

\[ f^0_n = \begin{cases} 
+1 & \text{if } e_n \text{ is a local maximum (positive zig-zag)} \\
-1 & \text{if } e_n \text{ is a local minimum (negative zig-zag)}, 
\end{cases} \]

\[ f^1_n = \begin{cases} 
+1 & \text{if } e_{n-4} < e_{n-2} < e_n \land e_{n-3} < e_{n-1} \text{ (up-trend)} \\
-1 & \text{if } e_{n-4} > e_{n-2} > e_n \land e_{n-3} > e_{n-1} \text{ (down-trend)} \\
0 & \text{otherwise (no trend)}. 
\end{cases} \]
Feature engineering rules (2)

\[ \nu_n^1 = \frac{\phi_n}{\phi_{n-1}}, \quad \nu_n^2 = \frac{\phi_n}{\phi_{n-2}}, \quad \nu_n^3 = \frac{\phi_{n-1}}{\phi_{n-2}}, \quad \tilde{\nu}_n^j = \begin{cases} +1 & \text{if } \nu_n^j - 1 > \alpha \\ -1 & \text{if } 1 - \nu_n^j > \alpha \\ 0 & \text{if } |\nu_n^j - 1| \leq \alpha \end{cases} \]

\[ f_n^2 = \begin{cases} +1 & \text{if } \tilde{\nu}_n^1 = 1, \tilde{\nu}_n^2 > -1, \tilde{\nu}_n^3 < 1 \text{ (volume strengthens)} \\ -1 & \text{if } \tilde{\nu}_n^1 = -1, \tilde{\nu}_n^2 < -1, \tilde{\nu}_n^3 > -1 \text{ (volume weakens)} \\ 0 & \text{otherwise (volume is indeterminant).} \end{cases} \]
- Tayal (2009) finds that the relative frequency of the conditional returns is significantly different from the relative frequency of the unconditional returns.
- In our new application, there is enough evidence to argue that return characteristics vary per state as well.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>EFA</th>
<th>GLD</th>
<th>SPY</th>
<th>XLB</th>
<th>XLE</th>
<th>XLF</th>
<th>XLI</th>
<th>XLK</th>
<th>XLP</th>
<th>XLU</th>
<th>XLV</th>
<th>XLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>0.2980</td>
<td>0.2496</td>
<td>0.3160</td>
<td>0.2851</td>
<td>0.3144</td>
<td>0.2851</td>
<td>0.3160</td>
<td>0.2851</td>
<td>0.3144</td>
<td>0.2851</td>
<td>0.3144</td>
<td>0.2851</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 1: Two-sample Kolmogorov-Smirnov test. Null: the empirical cumulative conditional and unconditional distributions of out-of-sample returns are drawn from the same distribution. Alternative: two-sided.
Tayal (2009) finds strong in-sample evidence in favor of the hypothesis for the most liquid half of Canadian stocks.

In our new application, we also find **in sample** that the mean return in the bull state is greater than the mean return in the bear state.

<table>
<thead>
<tr>
<th></th>
<th>EFA</th>
<th>GLD</th>
<th>SPY</th>
<th>XLB</th>
<th>XLE</th>
<th>XLF</th>
<th>XLI</th>
<th>XLK</th>
<th>XLP</th>
<th>XLU</th>
<th>XLV</th>
<th>XLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>337.34</td>
<td>164.01</td>
<td>619.79</td>
<td>110.67</td>
<td>474.35</td>
<td>585.09</td>
<td>143.77</td>
<td>158.13</td>
<td>105.79</td>
<td>484.78</td>
<td>108.22</td>
<td>72.92</td>
</tr>
<tr>
<td>p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\hat{\mu}<em>{\text{bull}} - \hat{\mu}</em>{\text{bear}}$</td>
<td>0.37</td>
<td>0.22</td>
<td>0.31</td>
<td>0.19</td>
<td>0.48</td>
<td>0.49</td>
<td>0.24</td>
<td>0.2</td>
<td>0.16</td>
<td>0.5</td>
<td>0.15</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table 2: Two-sample unpaired t-test. Null: the mean of the distribution of out-of-sample bull returns is less or equal the mean of bear returns. Alternative: mean return conditional on bull state is greater than conditional on bear state. Some limitations to the test assumptions apply.
Regime return characteristics - Out-of-sample

- Tayal (2009) finds strong evidence to answer the question positively for most Canadian stocks.
- In our new application, no stock has statistically larger out-of-sample returns in bull states versus bear states.
  - States are interchanged out-of-sample!.
  -⚠️ Some rather strong limitations to t-test assumptions apply (further research on a better comparison methodology needed).

<table>
<thead>
<tr>
<th></th>
<th>EFA</th>
<th>GLD</th>
<th>SPY</th>
<th>XLB</th>
<th>XLE</th>
<th>XLF</th>
<th>XLI</th>
<th>XLK</th>
<th>XLP</th>
<th>XLU</th>
<th>XLV</th>
<th>XLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>-27.86</td>
<td>-0.09</td>
<td>-29.71</td>
<td>-4.06</td>
<td>-18.94</td>
<td>-46.81</td>
<td>-7.23</td>
<td>-9.03</td>
<td>-6.06</td>
<td>-26.47</td>
<td>-1.07</td>
<td>-1.15</td>
</tr>
<tr>
<td>p-value</td>
<td>1.00</td>
<td>0.53</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.86</td>
<td>0.87</td>
</tr>
<tr>
<td>$\hat{\mu}<em>{\text{bull}} - \hat{\mu}</em>{\text{bear}}$</td>
<td>-0.2</td>
<td>0.00</td>
<td>-0.09</td>
<td>-0.06</td>
<td>-0.11</td>
<td>-0.25</td>
<td>-0.11</td>
<td>-0.11</td>
<td>-0.11</td>
<td>-0.18</td>
<td>-0.01</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

Table 3: Two-sample unpaired t-test. Null: the mean of the distribution of out-of-sample bull returns is less or equal the mean of bear returns. Alternative: mean return conditional on bull state is greater than conditional on bear state. Some limitations to the test assumptions apply.
Regime return characteristics - In-sample results

- Tayal (2009) finds strong evidence to answer the question positively for all Canadian stocks.
- In our new application, all stocks have statistically positive (negative) in-sample returns in bull (bear) states.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>EFA</th>
<th>GLD</th>
<th>SPY</th>
<th>XLB</th>
<th>XLE</th>
<th>XLF</th>
<th>XLI</th>
<th>XLK</th>
<th>XLP</th>
<th>XLU</th>
<th>XLV</th>
<th>XLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_{\text{bear}} )</td>
<td>-238.68</td>
<td>-128.83</td>
<td>-455.47</td>
<td>-77.95</td>
<td>-345.89</td>
<td>-396.8</td>
<td>-131.58</td>
<td>-108.3</td>
<td>-88.14</td>
<td>-354.95</td>
<td>-70.24</td>
<td>-50.28</td>
</tr>
<tr>
<td>p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \hat{\mu}_{\text{bear}} )</td>
<td>-0.18</td>
<td>-0.13</td>
<td>-0.16</td>
<td>-0.09</td>
<td>-0.25</td>
<td>-0.22</td>
<td>-0.2</td>
<td>-0.09</td>
<td>-0.11</td>
<td>-0.27</td>
<td>-0.06</td>
<td>-0.06</td>
</tr>
<tr>
<td>( t_{\text{bull}} )</td>
<td>239.17</td>
<td>101.76</td>
<td>421.09</td>
<td>78.62</td>
<td>324.62</td>
<td>430.27</td>
<td>58.11</td>
<td>115.36</td>
<td>58.54</td>
<td>330.23</td>
<td>82.33</td>
<td>53.24</td>
</tr>
<tr>
<td>p-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( \hat{\mu}_{\text{bull}} )</td>
<td>0.2</td>
<td>0.09</td>
<td>0.15</td>
<td>0.1</td>
<td>0.22</td>
<td>0.27</td>
<td>0.04</td>
<td>0.11</td>
<td>0.05</td>
<td>0.23</td>
<td>0.09</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 4: One-sample t-test. Null: the distribution mean of out-of-sample bearish (bullish) returns is greater (less) or equal than zero. Alternative: the mean is less (greater) than zero. Some limitations to the test assumptions apply.
Regime return characteristics - Out-of-sample

- Tayal (2009) finds strong evidence in favor of the hypothesis for the most liquid half of Canadian stocks.
- In our new application, none has statistically positive (negative) returns in bull (bear) states.
  - There seems to be a misclassification problem in top states.
  - ! Some rather strong limitations to t-test assumptions apply (further research on a better comparison methodology needed).

<table>
<thead>
<tr>
<th>Symbol</th>
<th>EFA</th>
<th>GLD</th>
<th>SPY</th>
<th>XLB</th>
<th>XLE</th>
<th>XLF</th>
<th>XLI</th>
<th>XLK</th>
<th>XLP</th>
<th>XLU</th>
<th>XLV</th>
<th>XLY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{bear}$</td>
<td>19.95</td>
<td>1.01</td>
<td>20.78</td>
<td>2.97</td>
<td>14.75</td>
<td>32.4</td>
<td>7.31</td>
<td>6.61</td>
<td>4.23</td>
<td>17.76</td>
<td>0.2</td>
<td>1.2</td>
</tr>
<tr>
<td>P-value</td>
<td>1</td>
<td>0.84</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.58</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}_{bear}$</td>
<td>0.1</td>
<td>0.01</td>
<td>0.05</td>
<td>0.03</td>
<td>0.06</td>
<td>0.12</td>
<td>0.08</td>
<td>0.06</td>
<td>0.06</td>
<td>0.08</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>$t_{bull}$</td>
<td>-19.44</td>
<td>0.88</td>
<td>-21.23</td>
<td>-2.77</td>
<td>-12.04</td>
<td>-33.8</td>
<td>-2.91</td>
<td>-6.16</td>
<td>-4.33</td>
<td>-19.66</td>
<td>-1.31</td>
<td>-0.42</td>
</tr>
<tr>
<td>P-value</td>
<td>1</td>
<td>0.19</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.91</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}_{bull}$</td>
<td>-0.1</td>
<td>0.01</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.05</td>
<td>-0.13</td>
<td>-0.03</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.09</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Table 5: One-sample t-test. Null: the distribution mean of out-of-sample bearish (bullish) returns is greater (less) or equal than zero. Alternative: the mean is less (greater) than zero. Some limitations to the test assumptions apply.

