Combining forecasts to predict the outcome of horseraces

The 15th International Conference on Gambling & Risk Taking, Las Vegas, USA

27 – 31 May 2013

Section 2-2-F

Stefan Lessmann, Ming-Chien Sung,

Johnnie Johnson, Tiejun Ma, Chung-Ching Tai
Outline

- Introduction
  - Competitive Event Prediction
  - Research Objective
- Methodology
  - Modeling Competitive Events
  - Forecast Combination
- Evaluation
  - Experimental Design
  - Empirical Results
- Conclusions
Competitive Event (CE) Prediction

- **Definition**
  
  *Scenario where multiple participants compete against each other for some reward*

- **Prediction task**
  
  - Estimate participants’ chances of winning
  - Evaluate the relative importance of factors that govern CE outcome

- **Examples (e.g., elections, sports events)**
Competitive Event (CE) Prediction

- Several statistical methods have been employed to predict probabilities in CEs (e.g., ANN, SVM, decision trees etc.).
  
  \[ \text{But fail to account for the intensity of competition.} \]

- Pooling of statistical forecasts is effective in many other domains.
  
  \[ \text{But the combination of statistical forecasting models in CEs has been neglected.} \]
Research objective

Develop a methodology for combining model-bases predictions in CEs.
Contributions

- Demonstrate how a library of diverse and accurate base forecasts can be constructed in CEs.
- Establish that average-based forecast pooling (employed in many other domains) is ineffective in CEs.
- Develop a mechanism for forecast combination (stacking) which meets the requirements and exploits the peculiarities of CEs.
Modeling Competitive Events

Choice modeling approach: *Conditional logit regression*

- Interpretation: View competitors as alternatives within a choice set and the winner as the participants whose credentials have resulted in it being the preferred alternative.

- Formula:

\[ p_i^j = \frac{\exp(\beta \cdot x_i^j)}{\sum_{i=1}^{m_j} \exp(\beta \cdot x_i^j)} \]

- \( p_i^j \) Winning probability of participant \( i \) in event \( j \)
- \( x_i^j \) Participant characteristics (i.e., independent variables)
- \( \beta \) Regression coefficients to be estimated
- \( m_j \) Number of participants in event \( j \)

[McFadden, 1974]
Modeling Competitive Events

- Conditional logit regression:
  Account for competition element within CEs.

\[
p_i^j = \frac{\exp(\beta \cdot x_i^j)}{\sum_{i=1}^{m_j} \exp(\beta \cdot x_i^j)}
\]

... normalized by the strengths of its opponents in event \( j \)

[McFadden, 1974]
Forecast Combination

- Essence of forecasting ensemble
  - Build a (large) library of **strong** and yet **diverse base models**
  - **Combine** predictions in some manner
Forecast Combination

- **Base models generation**: Three-level approach

1. **Define** surrogate measures of **event outcome** to translate prediction tasks into ‘ordinary’ modeling objectives (continuous: ‘finishing position in horseracing’ or discrete: ‘win or loss’)

2. Forecast resulting dependent variables with alternative **prediction methods** (regression & classification)

3. Vary **meta-parameter settings** of these methods
Two Forecast Combination Schemes

- **Average-based combination**
  - Problem: Complicated by surrogate objectives
  - Overcome: Develop forecast calibration algorithm (Platt, 2000)

- **Stacking-based combination**
  - ‘Learn’ combination rule empirically
  - CL can be employed to combine base forecasts: **LLR-based selection**

Data

\[ \hat{y}(x) = \sum_{t=1}^{T} w_t \hat{y}_t(x) / T \]

Note: \( T = 571 \)

**Weighted function**

\[ f \left( \sum_{t=1}^{T} \hat{y}_t(x) \cdot \beta_t \right) \]

**CL function**
Research question

- To assess the accuracy of composite forecasts resulting from average- and stacking-based pooling mechanisms.
Horseracing data

- Many similarities with wider financial markets
  - Ease of market entry
  - Numerous diverse participants
  - Widespread availability of information
  - Multiple factors affect assets’ values
  - Similar behavioral biases among traders
- Difficult benchmark
  - Renowned as efficient markets
  - Number of participants varies between events.

Betting markets are routinely used to shed light on decision maker’s behavior in wider financial markets
Experimental Design

- Data & variables (Bolton & Chapman, 1986)
  - 4,276 horseraces run in Hong Kong (55,690 runners)
  - Past performance (runners/jockeys) & race conditions

- Model evaluation
  - Split-sample setup (65% : 35%)
  - 5-fold cross validation on in-sample data

- Measures of forecasting performance
  - Coefficient of determination, $R^2$
  - Rate of return when betting on model predictions using Kelly’s (1956) investing strategy.

- Base models

By varying dependent variable measures, predictions methods, and meta-parameter values, a library of 571 individual base models is produced.
## Empirical Results

- Average-based forecast combination

<table>
<thead>
<tr>
<th>Ensemble member</th>
<th>$R^2$</th>
<th>Rate of return</th>
<th>$p$-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark model (conditional logit)</td>
<td>CL base model</td>
<td>0.1532</td>
<td>10.84</td>
</tr>
<tr>
<td></td>
<td>Track probabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple average</td>
<td>All base models</td>
<td>0.1003</td>
<td>-8.44</td>
</tr>
<tr>
<td></td>
<td>Track probabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal trimmed simple average</td>
<td>CL base model</td>
<td>0.1531</td>
<td>10.80</td>
</tr>
<tr>
<td></td>
<td>Track probabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td>CL base model</td>
<td><strong>0.1538</strong></td>
<td><strong>11.25</strong></td>
</tr>
<tr>
<td></td>
<td>Track probabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Support vector regression</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Statistical test of $H_0$: return > 0

Forward-selection of base models (Caruana et al., 2006). Weights are decided by no. of times the models enter the ensemble.
Empirical Results

- Stacking-based forecast combination (conditional logit stacking model)

<table>
<thead>
<tr>
<th>Model</th>
<th>Ensemble size*</th>
<th>$R^2$</th>
<th>$p$-value**</th>
<th>Rate of return</th>
<th>$p$-value***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark model</td>
<td>2</td>
<td>0.1532</td>
<td>/</td>
<td>10.84</td>
<td>0.1386</td>
</tr>
<tr>
<td>CL stacking model: LLR-based variable selection</td>
<td>5</td>
<td>0.1543</td>
<td>0.0432</td>
<td>20.31</td>
<td>0.0149</td>
</tr>
<tr>
<td>CL stacking model: best models per modeling objective</td>
<td>5</td>
<td>0.1538</td>
<td>0.1933</td>
<td>16.16</td>
<td>0.0505</td>
</tr>
<tr>
<td>CL stacking model: best models per method</td>
<td>10</td>
<td>0.1540</td>
<td>0.6561</td>
<td>17.67</td>
<td>0.0399</td>
</tr>
</tbody>
</table>

* Number of base models selected for the ensemble
** LLR-test of benchmark model vs. ensemble model
*** Statistical test of $H_0$: return > 0
Conclusions

• Forecast combination improves accuracy

• Standard combination scheme (averages) less suitable due to competition

• Stacking through 2\textsuperscript{nd} stage cond. logit model superior

• Novel analytical tool to study competitive events
Questions & comments?

Prof. Ming-Chien Sung
Centre for Risk Research
Southampton Management School
University of Southampton
Email: ms9@soton.ac.uk