The Impact of Publicly Available Information on Betting Markets: Implications for Bettors, Betting Operators and Regulators

Ming-Chien Sung and Johnnie Johnson

The 6th European conference on Gambling Studies and Policy Issues

University of Southampton
UK

7 July, 2005
Table of content

- Background
- Literature
- Why UK horseracing market
- Research designs
- Findings
- Conclusions
• ‘A market can be considered informational efficient if all available information is fully reflected in market prices’ (Fama, 1970).

• ‘Efficient markets are those, in which current market prices reflect all available information. Hence it is impossible to make unusual or excess profits by using the available information’ (Ross et. al., 1988).

• ‘Purchase or sale of any security at the prevailing market price is never a positive-NPV transaction’ (Brealey and Myers, 1996).
Background - Decision making & hypothesis

Model developing

Decision making

All existing information

Published Information

Historical information

Weak form

Semi-strong form

Strong form

Availability

Heuristic
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Market type</td>
<td>Parimutuel system</td>
<td>Parimutuel system</td>
<td>Parimutuel system</td>
<td>Parimutuel system</td>
<td>Bookmaker markets</td>
<td>Bookmaker markets</td>
</tr>
<tr>
<td>Number of races</td>
<td>200 races</td>
<td>2,000 races</td>
<td>3,198 races</td>
<td>600 races</td>
<td>200 races</td>
<td>16,836 races</td>
</tr>
<tr>
<td>Data source</td>
<td>Five U.S. tracks</td>
<td>Hong Kong</td>
<td>Hong Kong</td>
<td>Hong Kong</td>
<td>Australia</td>
<td>35 U.K. racetracks</td>
</tr>
<tr>
<td>Number of variables</td>
<td>10</td>
<td>20 plus public’s implied probability</td>
<td>More than 20 plus public’s implied probabilities</td>
<td>14 plus public’s implied probability</td>
<td>12 plus starting prices</td>
<td>66 fundamental plus 4 odds-related variables</td>
</tr>
<tr>
<td>Variable types</td>
<td>Basic horse, jockey, context variables</td>
<td>Sophisticated horse, workout, jockey, context, and interaction variables</td>
<td>Sophisticated context preference, current condition, past performance and recent workout of horses and past performance of jockeys</td>
<td>Sophisticated horse and jockey variables</td>
<td>Basic horse and context independent variables for single previous and current starts</td>
<td>Sophisticated horse, jockey, trainer, race context and interaction variables</td>
</tr>
</tbody>
</table>
Features of UK horseracing market

- Multiple betting media

- Various topology of racetracks

- Open population of horses
Research designs

- **Data:** Computer Raceform
- **Time span:** May 96’ ~ Aug 00’
- **Geographic spread:** 35 British racetracks
- **Data structure:**

<table>
<thead>
<tr>
<th>Period</th>
<th>Races</th>
<th>Horses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training sample</td>
<td>16 May 96’ – 15 May 00’</td>
<td>16,836</td>
</tr>
<tr>
<td>Holdout sample</td>
<td>16 May 00’ – 13 Aug 00’</td>
<td>1,929</td>
</tr>
<tr>
<td>Total</td>
<td>18,765</td>
<td>201,354</td>
</tr>
</tbody>
</table>
Research designs-continued

• Attributes: 66 fundamental attributes + 4 odds-related attributes

• Missing value indicators:

\[
y_{ij} = \beta_A x_{ij}^A + \beta_B M_{ij}^B x_{ij}^B + \beta_C \left(1 - M_{ij}^B\right) + \varepsilon_{ij}
\]

• Conditional logit model:

\[
U_{ij} = \beta' x_{ij} + \varepsilon_{ij}
\]

\[\varepsilon_{ij} \text{ Follows logistic distribution}\]

\[
P_{i*} = \Pr ob \left( U_{i*} \geq U_{ij} , \text{ for all other } j \neq * \right)
\]

\[
\Pr ob(y_{i*} = 1) = \frac{\exp(\beta' x_{i*})}{\sum_{j_i} \exp(\beta' x_{ik})}
\]

\[
L(\beta') = \prod_{i=1}^{n} \Pr ob_i \left( y_{i*} = 1 \right) = \prod_{i=1}^{n} \frac{\exp(\beta' x_{i*})}{\sum_{k=1}^{j_i} \exp(\beta' x_{ik})}
\]
## Findings - Summary statistics

<table>
<thead>
<tr>
<th>Summary statistics</th>
<th>Testing Global Null Hypothesis: all BETA = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L(\theta = 0)$</td>
<td>-61,242</td>
</tr>
<tr>
<td>$L(\theta = \hat{\theta})$</td>
<td>-31,737</td>
</tr>
<tr>
<td>Adj $R^2$</td>
<td>0.48157</td>
</tr>
</tbody>
</table>

| $LL$ ratio statistic | 59,011.6 |
| Degree of freedom    | 70       |
| Critical value of $\chi^2_{70}$ (0.05) | 90.5      |
## Findings-Overall model fit tests

<table>
<thead>
<tr>
<th>Models</th>
<th>Summery statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Odds-related model</td>
</tr>
<tr>
<td>Number of predictors</td>
<td>4</td>
</tr>
<tr>
<td>$L(\theta = 0)$</td>
<td>-61,242.40</td>
</tr>
<tr>
<td>$L(\theta = \hat{\theta})$</td>
<td>-32,081.65</td>
</tr>
<tr>
<td>Adj $\hat{R}^2$</td>
<td>0.47614</td>
</tr>
</tbody>
</table>

\[
\Delta R^2_{C-P} = R^2_C - R^2_P = 0.48157 - 0.47614 = 0.005
\]

\[LL \text{ ratio} = 690.11, \chi^2_{66}(0.05) = 85.965\]
Findings-significance of attributes

- The log of normalised SP probability is the most significant attribute in the model.

- 34 of the 70 attributes are significant. Most of these significant attributes are derived from sophisticated transformations (such as, nonlinear transformation of age, and the average speed rating of the past runs of each horse in current going).

- The missing indicator for horse-related attributes is highly significant.

- Identify factors associated with jockeys and trainers which are significant.

- Interaction terms between odds and Shin’s z value and number of runners are significant.
## Findings - Wagering Simulations

<table>
<thead>
<tr>
<th>Wagering strategy</th>
<th>No. of races bet</th>
<th>No. of races won</th>
<th>Amount bet</th>
<th>Profits</th>
<th>Rate of return</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) £1 bet on each horse</td>
<td>1,929</td>
<td>316</td>
<td>20,190</td>
<td>-6,307</td>
<td>-31.2%</td>
</tr>
<tr>
<td>(2) Return £1 on each horse if the horse wins</td>
<td>1,929</td>
<td>19</td>
<td>2,344</td>
<td>-415</td>
<td>-17.7%</td>
</tr>
<tr>
<td>(3) £1 bet on the favourite of each race</td>
<td>1,929</td>
<td>681</td>
<td>2,102</td>
<td>-93.83</td>
<td>-4.46%</td>
</tr>
<tr>
<td>(4) £1 bet on the model winner of each race</td>
<td>1,929</td>
<td>644</td>
<td>1,929</td>
<td>50.28</td>
<td>2.61%</td>
</tr>
<tr>
<td>(5) Kelly strategy based on one-stage model</td>
<td>1,559</td>
<td>531</td>
<td>122</td>
<td>20.05</td>
<td>16.38%</td>
</tr>
</tbody>
</table>
Findings - Kelly wagering strategy

Kelly Wagering Strategy based on the holdout sample

Log of Cumulative Wealth

1,929 races (20,190 runners)
Thank you!

Presenter: Ming-Chien Sung
Centre for Risk Research
School of Management
University of Southampton
Highfield, Southampton
SO17 1BJ
United Kingdom
Tel: +44 23 8059 7995
Fax: +44 23 8059 3844
Email: ms9@soton.ac.uk

University of Southampton
UK

7 July, 2005