Well I’ll be *damned* - Insights into the predictive value of pedigree information in horse racing

Timothy Baker
Ming-Chien Sung
Johnnie E V Johnson
Tiejun Ma
Centre for Risk Research
University of Southampton, UK
Overview

- Pedigree and theories
- Information sources and cognitive tools
- UK Bookmaker market and suitable race types
- Methodology & Results, including the ability of bettors to incorporate pedigree information
- Implications
Pedigree

Source: https://relatedhowagain.wordpress.com/category/relationships/page/16/
Breeding and the analysis of pedigree

Theories include:

- Dosage
- Inbreeding/line breeding
- Sex balancing
- Nicking
- Rasmussen factor
- X-factor

https://derbywinner.net/2014/09/21/il-tesio-great-expectations/
Dosage theory

- Prevalent in form guides
- Influential sires - *chef de races*
- DI indicative of speed/stamina balance
- Issues – Retrospective changes
  - Bias towards sire-line
  - No control group
  - Limited to four generations
Dosage Theory - Nyquist

- Sire never won a race at 10 Furlongs (2000m)
- Same great damsire as
  * American Pharoah
  * A Shin Hikari

Kentucky Derby winner
Nyquist
Dosage Index – 7.00
Source: http://tinyurl.com/hysx9zq
## Analysing information from a range of sources

<table>
<thead>
<tr>
<th>NYQUIST ♂ (USA) b. 2008</th>
<th>IN EXCESS (IRE) br. 1987</th>
<th>SIBERIAN EXPRESS (USA) gr. 1981</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNCLE MO (USA) b. 2008</td>
<td>IN EXCESS (IRE) br. 1987</td>
<td>SIBERIAN EXPRESS (USA) gr. 1981</td>
</tr>
<tr>
<td>INDIAN CHARLIE (USA) b. 1995</td>
<td>SOVIET SOJourn (USA) b. 1989</td>
<td>KANTADO (IRE) b. 1976</td>
</tr>
<tr>
<td>PLAYA MAYA (USA) dkb/br. 2000</td>
<td>ARCH (USA) dkb/br. 1995</td>
<td>LEO CASTELLI (USA) b. 1984</td>
</tr>
<tr>
<td></td>
<td>DIXIE SLIPPERS (USA) dkb/br. 1995</td>
<td>POLITICAL PARFAIT (USA) dkb/br. 1984</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEEKING GABRIELLE (USA) ch. 2007</th>
<th>STORM CAT (USA) br. 1983</th>
<th>STORM BIRD (CAN) b. 1978</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORESTRY (USA) b. 1996</td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
<tr>
<td></td>
<td>STORM CAT (USA) br. 1983</td>
<td>STORM BIRD (CAN) b. 1978</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEEKING REGINA (USA) ch. 1992</th>
<th>SEEKING THE GOLD (USA) br. 1985</th>
<th>MR. PROSPECTOR (USA) b. 1970 [BC]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fulbright Scholar (USA) ch. 1985</td>
<td>Fulbright Scholar (USA) ch. 1985</td>
<td>Mr. Prospector (USA) b. 1970 [BC]</td>
</tr>
<tr>
<td></td>
<td>Fulbright Scholar (USA) ch. 1985</td>
<td>Mr. Prospector (USA) b. 1970 [BC]</td>
</tr>
<tr>
<td></td>
<td>Fulbright Scholar (USA) ch. 1985</td>
<td>Mr. Prospector (USA) b. 1970 [BC]</td>
</tr>
<tr>
<td></td>
<td>Fulbright Scholar (USA) ch. 1985</td>
<td>Mr. Prospector (USA) b. 1970 [BC]</td>
</tr>
<tr>
<td></td>
<td>Fulbright Scholar (USA) ch. 1985</td>
<td>Mr. Prospector (USA) b. 1970 [BC]</td>
</tr>
</tbody>
</table>

Source: [www.pedigreequery.com/nyquist](http://www.pedigreequery.com/nyquist)
14:00 Ludlow 8 May 2016
Mike Rhodes Shrewsbury Honda Centre Novices' Hurdle (Class 4) (4YO plus)
Winner £3,899 - 4 run - Draw Advantage: None

Weighed In

<table>
<thead>
<tr>
<th>No.</th>
<th>Form</th>
<th>Horse</th>
<th>Age</th>
<th>Jockey / trainer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>311-1</td>
<td>RED TORNADO (FR) (19)</td>
<td>4-11-6</td>
<td>Harry Skelton</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ch g Dr Fong - Encircle (Spinning World)</td>
<td>127</td>
<td>D Skelton</td>
</tr>
<tr>
<td>2</td>
<td>878-1</td>
<td>APACHE PEARL (IRE) (95)</td>
<td>5-11-0</td>
<td>Jake Greenall</td>
</tr>
<tr>
<td></td>
<td></td>
<td>br g Indian Danehill - Pearl Buttons (Alflora)</td>
<td>W Greatrex</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3-1</td>
<td>RACING SPIRIT (301) (30F)</td>
<td>4-10-10</td>
<td>Lee Edwards</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ch g Sir Percy - Suertuda (Domedriver)</td>
<td>Kevin Frost</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5-1</td>
<td>SANAJA (191)</td>
<td>4-10-3</td>
<td>Non Runner</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ch m Pivotal - Sanjida (Polish Precedent)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8-1</td>
<td>RIGHT MADAM (IRE) (19)</td>
<td>4-10-3h</td>
<td>Jack Quinlan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b f Jeremy - Mawaared (Machiavellian)</td>
<td>Sarah Hollinshead</td>
<td></td>
</tr>
</tbody>
</table>

2015: (6 ran) Are They Your Own, Age 7, Wgt 11-4, SP 14/1 - Nick Slatter / F O'Brien
Forecast: 4/5 Red Tornado, 9/4 Sanajja, 10/1 Racing Spirit, 12/1 Right Madam, 12/1 Apache Pearl
(Note: non-runner may affect prices)
Decision making under uncertainty

Source: www.wordcloud.com
Setting – UK Bookmaker market

- Bookies dominate >80% market share
- Market maker
- Fixed odds
- Over-round varies

Photo by Will Palmer
http://tinyurl.com/gr5ua7
Can pedigree information help predict the winner?

- Pedigree information potentially more valuable for:
  - Maidens
  - Untried distance increases (Gramm and Ziemba, 2008, 2012)

- Sire & dams performance (if raced)

- Performance of sire & dams offspring
Methodology

- Maiden races – restricted to Over-round >8%
- Preference variables
- Conditional Logit Modelling
- Kelly wagering strategy

Secretariat
Heart size: 22 pounds (>2.5 times larger than average)

Photo by tccana
https://www.flickr.com/photos/10386256@N03/877252285/
Results – Pedigree only

- Pedigree information can be valuable
- Performance of sire, dams, damsires offspring better predictor
- Dam-line provides more insight

Pseudo $R^2 = 0.1022$ (variation explained by model)

18 significant variables (P<0.05)
Results – Pedigree + odds

- Pseudo $R^2$ including odds = 0.27444
- Pseudo $R^2$ odds only = 0.27414
- Difference of 0.0003
Results & Conclusion

- Only one factor not discounted in the starting price odds
- Performance of sires’ offspring of a similar age/sex
- Bettors good at discounting pedigree info
  - High level of experience and expertise
  - Regularly assessing events in probabilities
  - Regular & immediate feedback plus incentive
Wagering – 2013 to 2014 (3154 races)
Summary

- Pedigree has some predictive value
- Bettors are good at using this information
- Bookie market semi-strong form efficient
- A positive ROI is achievable, thanks to FLB
- Naturalistic settings make for well calibrated decisions
Thank you
Interested in pedigree?
Check out Boojums bonanza (blog by David Dink)

https://ddink55.wordpress.com/

Centre for Risk Research
University of Southampton

tob1g13@soton.ac.uk
m.sung@soton.ac.uk