

# Composite Ordinal Forecasting in Horse Racing - An Optimization Approach

*Alan Tse*

*Associate Professor*

*Department of Marketing*

*The Chinese University of Hong Kong*

*Chan-ieong Chan*

## Abstract

Using horse racing data in Hong Kong as an example, this paper looks into the properties of an optimization model for making composite ordinal forecasts based on minimization of the absolute error of the joint distribution of the errors of twelve forecasters of race outcomes. It was found that the optimization model is not only sound theoretically, but it is also robust, and can handle situations when data are sparse. **KEYWORDS:** *Horse racing, composite forecasting, integer programming*

## Introduction

Horse racing has been in Hong Kong for more than 30 years, and is a major industry in the territory. Average turnover per race was HK\$111 million (about US\$14 million), and on several occasions, turnover exceeded \$1 billion (about US\$0.128 billion) for a single meeting. On average, payments of duty and taxation to the Hong Kong government constitute about 12.3 percent of the total direct and indirect taxes received by the government.

The horse racing industry is a rapidly growing industry. The annual growth rate of betting turnover in the years from 1988 to now is about 20%. Nowadays, there are at least 12 newspapers devoted solely to horse racing. In fact, it is one of the most popular hobbies among people in Hong Kong.

Given that horseracing is such a big and growing industry in Hong Kong, it is of practical and academic importance to look into the problem of determining the best bet in a race. In general, punters would bet based on the tips provided by newspapers, although some might perform analyses looking at variables like the previous performance of the horse, the current status of the horse, and so on. However, the latter approach would be very difficult for an average laymen punter because of the huge number of variables and the technicalities involved. Hence, the newspaper is a major source of information for a punter when placing a bet. The decision problem therefore becomes:

1. How should we use the forecasts made by the newspapers? What are the mathematical issues involved since all these forecasts are ordinal forecasts?
2. Should we somehow combine the forecasts of all the newspapers, or should we ignore some of them?
3. If we are to combine the forecasts, how should we weigh the individual forecasts to come up with a composite forecast the error of which would be minimized?

To address the above important issues, the authors constructed a model based on an optimization algorithm that attempts to provide an optimal composite forecast based on a weighted average of individual forecasts. The distinguishing feature of the model is that it accepts ordinal rankings as input and produces an ordinal forecast. In addition, the model is capable of determining the optimal number of forecasters to be included in the composite forecast.

## Models of Composite Forecasting

In the horse racing decision-making situation, information can be obtained from various sources. This information can be analyzed using systematic techniques based on mathematical models. Intuitively, a better forecast could be obtained by combining the forecasts made by a number of individual forecasters (Batchelor & Dua, 1995). Lots of studies have been done on combining forecasts to improve prediction accuracy. Barnard (1963) found that a simple average of individual forecasts outperformed the individual components. Bates and Granger (1969) studied how the weights should be optimally allocated to the individual components. Newbold and Granger (1974) and Bates and Granger (1969) have shown that combining forecasts usually provides a result that is superior to individual forecasts. Mahmoud (1984) provided an extensive review of composite forecasting.

**Given that horseracing is such a big and growing industry in Hong Kong, it is of practical and academic importance to look into the problem of determining the best bet in a race.**

As a practical example, Newbold and Zumwalt (1987) found that combined forecasting gives better prediction for earnings per share of utility stock.

Several criteria can affect the choice of a composite model (Moriarty, 1990; Chandrasekharan, Moriarty & Wright, 1994; Harvey, Leybourne & Newbold, 1998). First, the model must be robust so that it is not affected by extreme observations in the data. Second, implementation of the model must be easy. Third, it must allow maximum likelihood estimation of the parameters. Fourth, the model must accommodate situations of small sample size. In the sequel, the major composite models are reviewed using the above criteria and justifications are provided for the development of the author's model.

### **Equally Weighted Approach**

The simplest approach is to take the arithmetic average of the individual forecasts. It has been found that the approach produces estimates that are robust and accurate in many decisional problems (Makridakis & Winkler, 1983; Figlewski & Ulrich, 1984; Clement & Winkler, 1986).

The problem with the approach is it does not have any theoretical backing. Furthermore, it does not allow a decision-maker to make use of available information to improve the forecast. In fact, Ashton and Ashton (1985), Makridakis and Winkler (1983) and Clement and Winkler (1986) have shown that incorporation of a less accurate forecast would degrade the performance of this approach. In fact, the method represents only an ad hoc approach towards composite forecasting.

### **Minimization of Error Approach**

This approach represents a group of methods that minimize forecast errors which follow an assumed distribution, and which takes into consideration the dependencies among, as well as the accuracy of, the models (Newbold & Granger, 1974; Winkler, 1981). Typical approaches include:

### **The Bayesian Approach**

A composite forecast was developed by Morris (1973) using a Bayesian approach by assuming that the errors follow a multivariate normal distribution, and imposing the constraint that the sum of the weights of the individual forecaster add up to one. The advantage of the approach is that it rewards accurate and independent models (Freeling, 1981). The problem with the model is that negative weights may appear which are counter-intuitive. Another problem is that the performance of the model is generally disappointing (Gupta & Wilton, 1987).

### **Regression Approach**

Basically, the regression approach (Aksu & Gunter, 1992; Harvey, Leybourne & Newbold, 1998) looked at the composite forecasting problem as estimating the coefficients or weights attached to each individual forecaster using the individual forecasts as the independent variables and the actual results of the race as the de-

pendent variable. This approach is somewhat ad hoc because all the data involved are ordinal data. The assignment of numerical values, such as 1 for the best bet, 2 for the second best bet and so on are arbitrary. In general, the normality assumption required for the error terms will be violated.

An alternative to bypass the ordinal forecasting problem is to use payoff data instead of the ordinal forecasts and race results. Nevertheless, this approach is problematic because the regression problem will then be turned into a problem of estimating the payoffs of the individual forecasters.

In general, the regression approach is not robust because the minimization of the square of the error term will inevitably lead to estimators that are strongly affected by extreme observations. Other reasons for non-robustness are data scarcity, instability or nonstationarity (Bunn, 1975; Dickinson, 1975; Newbold & Granger, 1974). Although the implementation of the model is relatively easy, maximum likelihood estimation of the model is not possible. Meanwhile, a large sample is required for this approach to work, and its performance has often been outperformed by the equally weighted approach (Gupta & Wilton, 1987).

**Although tracks in Hong Kong pay based on the first three places, for simplicity only forecasts that predict accurately the first two places are included in estimating the weights of importance of forecasters.**

## **Outperformance Approach**

The outperformance approach (Bunn, 1975) weights a model by the proportion of times the model has outperformed all others to date. Although the model is simple, the only information it uses to update the priors is the performance of the single model compared to all the others, and it does not record by how much one model is better than the others (Gupta & Wilton, 1987). More importantly, the weights are found to be unstable because of the pairwise comparison rule used (Gupta & Wilton, 1987).

## **The Odds-Matrix Method**

The Odds-Matrix method (Gupta & Wilton, 1987) uses a matrix of pairwise odds on performance to derive the weights. The method is simple, robust, permits updating of the weights, and does not require a lot of data. However, the approach is rather ad hoc without sound theoretical backing, especially when experts had to derive the odds themselves given little past information available.

## **Majority Rule Approach**

A simple approach to combining forecasts is to select the horse tipped by most forecasters. This approach, however, cannot guarantee that the composite forecast has the minimum error. In fact, very often, consensus does not exist. The model is

ad hoc and does not allow maximum likelihood estimation of the parameters. Finally, the robustness of the approach is doubtful.

Taking into account the limitations of the various models described above, the following optimization model is devised that attempts to produce a composite forecast that is robust, accepts ordinal input data, requires little input information and allows maximum likelihood estimation of parameters.

## The Optimization Model

The following notations were used in the formulation of the model:

1. Subscript  $i$  indexes for the horse,  $i = 1, \dots, h$ ;
2. Subscript  $j$  indexes for the forecaster,  $j = 1, \dots, f$ ;
3. Subscript  $g$  indexes for the race,  $g = 1, \dots, m$ ;
4.  $R_{ij}^g$  is the forecasted rank assigned to horse  $i$  by forecaster  $j$  for race  $g$ . It can take one of the following three values:

- $R_{ij}^g = 1$  if  $i$  is the first ranked horse in  $j$ 's forecast;
- $R_{ij}^g = 2$  if  $i$  is the second ranked horse in  $j$ 's forecast; and
- $R_{ij}^g = 3$  if otherwise.

$R_{i^*}^g$  is the actual ranking of horse  $i$  in race  $g$ . Since the newspapers only report the first three ranked horses,  $R_{i^*}^g$  can take three values only (that is, 1, 2, and 3 as  $R_{ij}^g$ ).

Although tracks in Hong Kong pay based on the first three places, for simplicity only forecasts that predict accurately the first two places are included in estimating the weights of importance of forecasters. This makes the computation a lot less tedious and the program much smaller in size.

The constructs used by the authors in the model are as follows:

1.  $w_j$  is the weight of importance of forecaster  $j$ ;
2.  $P_{ij}^g$  is the score assigned to horse  $i$  by forecaster  $j$  for race  $g$  such that  $P_{ij}^g > P_{i'j}^g$  if  $R_{ij}^g < R_{i'j}^g$  ..... (1)
3.  $S_i^g$  is the combined score assigned to horse  $i$  for race  $g$  defined as:

$$S_i^g = w_j P_{ij}^g \quad \dots \dots \dots (2)$$

Given  $R_{i^*}^g$  for  $g = 1, \dots, m$  and  $i = 1, \dots, h$ , the winning horses are known in each race. For a particular  $g$ , we have:

$$\begin{aligned} S_i^g - S_{i^*}^g + \epsilon_1 &\geq 0 \\ S_{i^*}^g - S_i^g + \epsilon_2 &\geq 0 \\ \dots &\dots \dots \dots (3) \\ \dots &\dots \dots \dots \\ S_{i^*}^g - S_{i^{h-1}}^g + \epsilon_{h-1} &\geq 0 \end{aligned}$$

where  $\epsilon_d$  is the noise term not captured in the model.

Let the  $\epsilon$ s be independent and follow a Laplace distribution with mean zero and variance equal to a constant  $\sigma$ . Then,

$$f(\epsilon_d) = (1/2a) \exp(-|\epsilon_d|/2a) \text{ where } \sigma^2 = 2a^2 \quad \dots \dots (4)$$

Then the joint likelihood function is given by L where:

$$L = \pi(1/2a)^{m(f-1)} \exp(-|g\epsilon d|/2a) \dots\dots\dots (5)$$

Therefore, maximizing the joint likelihood function is equivalent to minimizing the sum of absolute error. That is, formulating in the form of an optimization problem, we have the following objective function and constraints:

Objective function:  $\min |g\epsilon d|$

such that:

$$\begin{aligned} & {}_gS_i - {}_gS_j + {}_g\epsilon_1 \geq 0 \\ & {}_gS_j - {}_gS_1 + {}_g\epsilon_2 \geq 0 \\ & \dots\dots\dots \\ & {}_gS_j - {}_gS_{i'} + {}_g\epsilon_{h-1} \geq 0 \\ & w_1 + w_2 + \dots\dots + w_f = 1 \end{aligned} \dots\dots\dots (6)$$

However, since if  ${}_gS_i - {}_gS_j$  is positive,  ${}_g\epsilon_d$  will be set to be zero by the minimization algorithm, and when  ${}_gS_i - {}_gS_j$  is negative,  ${}_g\epsilon_d$  will always be positive, the model can be simplified as below:

Objective function:  $\min g\epsilon d$

such that:

$$\begin{aligned} & {}_gS_i - {}_gS_j + {}_g\epsilon_1 \geq 0 \\ & {}_gS_j - {}_gS_1 + {}_g\epsilon_2 \geq 0 \\ & \dots\dots\dots \\ & {}_gS_j - {}_gS_{i'} + {}_g\epsilon_{h-1} \geq 0 \\ & w_1 + w_2 + \dots\dots + w_f = 1 \end{aligned} \dots\dots\dots (7)$$

To solve the problem of selecting forecasters, the following constraints can be added:

$$\begin{aligned} w_1 &\leq I_1 \\ w_2 &\leq I_2 \\ &\dots\dots\dots \\ w_f &\leq I_f \end{aligned} \dots\dots\dots (8)$$

where  $I_1, \dots, I_f$  are integer variables (that is, 0 or 1) representing the f forecasters; and

$$I_1 + I_2 + \dots\dots + I_f = C \dots\dots\dots (9)$$

Hence, by varying C from 1 to f, we can select the optimal number of forecasters that minimize the sum of absolute errors.

As far as the implementation of the model is concerned, we do not have information to identify  $w_i$  and  ${}_gP_{ij}$  given only  ${}_gR_{ij}$  and  ${}_gR_{i^*}$ . Therefore, to avoid overparameterization, the following simple scale was used. That is:

- ${}_gP_{ij} = 1$  if forecaster j ranks horse i to be number 1;
- ${}_gP_{ij} = 2$  if forecaster j ranks horse i to be number 2 and
- ${}_gP_{ij} = 3$  if otherwise.

Of course, there are other possible scales for  ${}_gP_{ij}$ . Nevertheless, by assuming that these different scales are linear combinations of one another, the scale used will not affect the validity of the model.

The optimization model detailed above should perform better than the other models for the following reasons. First, the model is a very robust model because it assumes that the error terms follow a Laplace distribution. It will not be affected

greatly by extreme observations, instability of the dependent variable and redundancy of information sets (Moriarty, 1990). Implementation is also relatively simple using common Linear Programming Algorithms and Mixed Integer Algorithms. Software for implementing these algorithms is available and familiar to most researchers. Furthermore, the model also allows maximum likelihood estimation of the parameters. Finally, the model uses the minimum amount of information, as the only required input are the rankings produced by the forecasters and the racing results in the sample period.

## Results

Using the horse racing data in the 1993 race season, the model was implemented using Lindo (Schrage, 1991), an optimization modeling software system, with the results shown in Table 1.

Table 1. Results of Using the Optimization Model

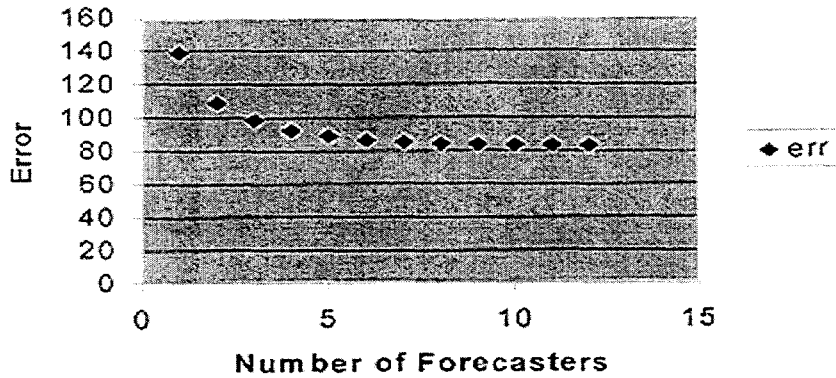
| N  | w <sub>1</sub> | w <sub>2</sub> | w <sub>3</sub> | w <sub>4</sub> | w <sub>5</sub> | w <sub>6</sub> | w <sub>7</sub> | w <sub>8</sub> | w <sub>9</sub> | w <sub>10</sub> | w <sub>11</sub> | w <sub>12</sub> | err     |
|----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|---------|
| 1  | 0              | 1              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 139.000 |
| 2  | 0              | 0              | 0.5            | 0              | 0              | 0              | 0.5            | 0              | 0              | 0               | 0               | 0               | 109.000 |
| 3  | 0              | 0              | 0              | 0              | 0              | 0.333          | 0.333          | 0              | 0              | 0.333           | 0               | 0               | 98.333  |
| 4  | 0              | 0              | 0.25           | 0              | 0              | 0.25           | 0.25           | 0              | 0              | 0.25            | 0               | 0               | 92.250  |
| 5  | 0              | 0              | 0.2            | 0.2            | 0              | 0.2            | 0.2            | 0              | 0              | 0.2             | 0               | 0               | 88.800  |
| 6  | 0              | 0.167          | 0.167          | 0.167          | 0              | 0.167          | 0.167          | 0              | 0              | 0.167           | 0               | 0               | 86.333  |
| 7  | 0              | 0.146          | 0.146          | 0.146          | 0.098          | 0.146          | 0.146          | 0              | 0              | 0.171           | 0               | 0               | 85.077  |
| 8  | 0              | 0.134          | 0.137          | 0.125          | 0.094          | 0.138          | 0.118          | 0.096          | 0              | 0.159           | 0               | 0               | 84.162  |
| 9  | 0              | 0.110          | 0.135          | 0.116          | 0.125          | 0.116          | 0.115          | 0.077          | 0.059          | 0.146           | 0               | 0               | 83.687  |
| 10 | 0              | 0.086          | 0.126          | 0.085          | 0.087          | 0.116          | 0.094          | 0.086          | 0.052          | 0.181           | 0               | 0.086           | 83.395  |
| 11 | 0.047          | 0.115          | 0.128          | 0.075          | 0.100          | 0.116          | 0.086          | 0.070          | 0.073          | 0.128           | 0               | 0.061           | 83.261  |
| 12 | 0.055          | 0.105          | 0.114          | 0.078          | 0.098          | 0.114          | 0.068          | 0.082          | 0.060          | 0.141           | 0.042           | 0.042           | 83.171  |

N is the number of forecasters to be retained in the composite forecasting model.  
W<sub>i</sub> is the weight assigned to forecaster i in the composite forecasting model.  
err refers to the forecasting error.

The result shows that the larger the number of forecasters included in the composite forecast, the lower the amount of absolute error. Plotting the absolute error against the number of forecasters as shown in the diagram below, it is clear

that the optimal number of forecasters is about 4 or 5, as indicated by the elbow in the diagram.

Figure 1. Number of Forecasters vs. Error



The forecasting error of each individual forecaster is given in Table 2 below. Clearly, the composite forecast is better than each forecaster acting individually, since even the combined forecast of only two forecasters - forecasters three and seven (see Table 1) - lowers the absolute error substantially from 139 to 109.

Table 2. Accuracy of Individual Forecasters

| Forecaster Number | Forecasting Error |
|-------------------|-------------------|
| 1                 | 155               |
| 2                 | 139               |
| 3                 | 142               |
| 4                 | 150               |
| 5                 | 152               |
| 6                 | 159               |
| 7                 | 142               |
| 8                 | 147               |
| 9                 | 156               |
| 10                | 147               |
| 11                | 152               |
| 12                | 154               |



## Discussion

The results of this research show that the optimization model described in this paper is a good model for predicting the outcome of a horse race. The model is easy to implement, robust and uses very little information.

The area in which the model would be found useful is not restricted to horse racing. Other potential areas of application include for example, forecasting the market price of stocks or other financial instruments and in various gaming situations as well. In general, the model would be very useful for laypersons who do not have the skill or resources to perform the necessary complex analysis needed to come up with a forecast. The model can be applied so long as forecasting data made by other experts are available.

The model will be particularly useful if the forecasters give ordinal forecasts. Most forecasting models available today require interval-scaled data which are very hard to come by in situations like horse racing.

## References

- Aksu, C. & S. I. Gunter. (1992). An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. *International Journal of Forecasting*, 8 (1), 27-43.
- Ashton, A. H. & R. H. Ashton. (1985). Aggregating subjective forecasts: Some empirical results. *Management Science*, 31 (12), 449-508.
- Barnard, G. A. (1963). New methods of quality control. *Journal of the Royal Statistical Society A*, 126, 255-9.
- Batchelor, R. & P. Dua. (1995). Forecaster diversity and the benefits of combining forecasts. *Management Science*, 41 (1), 68-75.
- Bates, J. M. & C. W. J. Granger. (1969). The combination of forecasts. *Operations Research Quarterly*, 20, 451-68.
- Bunn, D. W. (1975). A Bayesean approach to the linear combination of forecasts. *Operations Research Quarterly*, 26, 325-29.
- Chandrasekharan, R., M. M. Moriarty & G. P. Wright. (1994). Testing for unreliable estimators and insignificant forecasts in combined forecasts. *Journal of Forecasting*, 13 (7), 611-24.
- Clement, R. T. & R. L. Winkler. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics*, 4 (1), 39-46.
- Dickinson, J. P. (1975). Some comments on the combination of forecasts. *Operations Research Quarterly*, 26, 205-10.
- Figlewski, S. & T. Urich. (1983). Optimal aggregation of money supply forecasts: Accuracy, profitability and market efficiency. *Journal of Finance*, 38, 695-710.
- Freeling, A. N. S. (1981). Reconciliation of multiple probability assessments. *Organizational Behaviour and Human Performance*, 28, 395-414.

- Granger, C. W. J. & P. Newbold. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics*, 5, 35-47.
- Gupta, S. & P. C. Wilton. (1987). Combination of forecasts: An extension. *Management Science*, 33 (3), 356-72.
- Harvey, D. I., S. J. Leybourne & P. Newbold. (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics*, 16 (2), 254-59.
- Mahmound, E. (1984). Accuracy in forecasting: A survey. *Journal of Forecasting*, 3, 139-59.
- Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen & R. Winkler. (1982). The accuracy of extrapolation (time series) methods: Results of forecasting competition. *Journal of Forecasting*, 1, 111-53.
- Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen & R. Winkler. (1983). *The accuracy of major extrapolation (time series) methods*. London: Wiley.
- Makridakis, S. & R. L. Winkler. (1983). Averages of forecasts: Some empirical results. *Management Science*, 29, 987-96.
- Moriarty, Mark M. (1990). Boundary value models for the combination of forecasts. *Journal of Marketing Research*, 27 (4), 402-417.
- Morris, P. A. (1973). Combining expert judgments: A Bayesean approach. *Management Science*, 23, 679-93.
- Newbold, P. & C. W. J. Granger. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society*, 137, 131-46.
- Newbold, P. & J. K. Zumwalt. (1987). Combining forecasts to improve earnings per share prediction: An examination of electric utilities. *International Journal of Forecasting*, 3, 229-38.
- Schrage, L. E. (1991). *LINDO: An optimization modeling system*. South San Francisco, CA: Scientific Press.
- Winkler, R. L. (1981). Combining probability distributions from dependent information sources. *Management Science*, 2, 479-88.