International Journal of Forecasting 8 (1992) 595–611
North-Holland

Combining vector forecasts to predict thoroughbred horse race outcomes

Edna M. White and Ronald Dattero
Department of Decision and Information Systems, College of Business, Florida Atlantic University, Boca Raton, FL 33431-0991, USA

Benito Flores
Department of Business Analysis and Research, College of Business Administration, Texas A&M University, College Station, TX 77843-4217, USA

Abstract: A large body of empirical studies has shown that a forecast developed by combining individual base forecasts performs surprisingly well. Previous work on the combination of forecasts has been confined to the area of time series forecasting. This work extends the combination of forecasts technique into the domain of forecasting one-time competitive events, specifically the scaled, relative finishing position of horses in thoroughbred sprint races. The present research develops a framework for the selection of the base forecasts and selects 12 base forecasts for analysis. The performance of the combination of the base forecasts is assessed on a sample of sprint races. Results of the analysis strongly suggest that the combination approach is both appropriate and effective. Some differences in results between this work and previous work in the time series domain suggest promising avenues for future research.

Keywords: Forecast combination, Event forecasting, Thoroughbred racing.

1. Introduction

In forecasting, it would, of course, be most desirable to use the method that gives the most accurate forecast. Unfortunately, the forecasting method that *always* gives the most accurate forecast has yet to be found. An intuitively attractive alternative is to draw on a variety of sources rather than relying solely on one forecasting method. This alternative was supported by some results of the Makridakis et al. (1982) M-competition. In this well-known competition, the accuracy of many forecasting methods was compared for a large number of data series. Two of the forecasting methods tested were combinations of forecasts. These combinations performed quite well compared with the individual forecasting methodologies involved in the combinations.

Results from other previous research show that the combination of base forecasts is a good, robust approach to forecasting that is likely to provide a good, though not optimal, forecast [Makridakis and Winkler (1983), Winkler and Makridakis (1983)]. A wide variety of base forecasts should be included in the combination – that is, forecasts that are based on different techniques or information [Makridakis and Winkler (1983) and Winkler and Makridakis (1983)].

Correspondence to: E.M. White, Department of Decision and Information Systems, College of Business, Florida Atlantic University, P.O. Box 3091, Boca Raton, FL 33431-0991, USA. Tel: (407) 367-3924; Fax: (407) 367–3978.

Researchers suggest combining judgmental (subjective) and statistical forecasting methods. Lawrence, Edmundson and O'Connor (1986) report that the combination of judgmental methods performed better than the objective ones. (Admittedly, if the majority of base forecasts are very inaccurate, the combination will also be poor; however, in such cases, selection of any of the individual forecasts is also likely to result in a poor forecast.)

The number of base forecasts included need not be large; typically, researchers have reported the highest rate of gains from combining only two to four base forecasts [Makridakis and Winkler (1983) and Winkler and Makridakis (1983)]. With larger numbers of base forecasts, there are still incremental gains; however, the improvement is at a decreasing rate.

Procedures for combining the individual forecasts range from the simple average [Makridakis and Winkler (1983) and Armstrong (1984)] to sophisticated weighting methods [for example, Reeves and Lawrence (1982)], and to the application of Bayesian analysis [Bunn (1981)]. A simple average is somewhat unattractive since it implies either that all of the base methods provide equal expected accuracy or that the accuracy of a forecasting method cannot be reliably rated prior to the event. However, the majority of empirical efforts have found little, if any, improvement over the simple average [Flores and White (1988)].

Previous work on the combination of forecasts has been confined to the area of time series forecasting. The present research extends the combination of forecast approach into the domain of predicting one-time, competitive events. This type of forecasting is useful in a wide variety of situations; for example, forecasting market shares for an industry, predicting sales for a set of competing products and estimating the relative importance of a set of factors in facility location.

2. The forecasting situation

The forecasting situation examined in this paper is that of thoroughbred sprints (6 to 7 furlongs). The study was confined to a single type of race in order to obtain a homogeneous sample. Sprints were selected since they are somewhat more popular than the longer route races. It should be emphasized that the purpose of the current research is to explore the use of the combination approach in a non-time series context, not to produce a handicapping methodology. Thoroughbred racing was selected as the domain area for the following reasons:

- The event is complex, but clearly defined.
- A large database of repetitive, unique events exists and is readily accessible.
- A large number of competing philosophies for handicapping races exist and can provide the basis for a large number of independent forecasts.
- Measures for rating 'practical' success of forecasts are supplied by the competitive nature of a horse race.

A first step in this study was the selection of the measure to be forecasted. The answer that comes most readily to mind is probably 'which horse will win'. However, in addition to predicting which horse will come in first, it is (at a minimum) also desirable to predict which horses will come in second and third.

More importantly, it is widely agreed that in thoroughbred racing there is a strong element of random chance, noise, or 'racing luck' present in every race, making it impossible to predict every winner with perfect accuracy. Therefore it is extremely important to have a forecast that predicts the outcome for each horse relative to the rest of the field. Clearly a horse that is predicted to win 'by a nose' is a more risky bet than one that is predicted to win by several lengths (a horse length is usually estimated to be about 10 ft and a nose at 0.5 ft).

For the present study, a 0–100, scaled vector showing the finishing order for the race was selected as the forecast measure. One factor in this decision was the fact that race results always report the number of lengths separating the horses at the finish. For example, in Exhibit 1 the column labeled 'lengths back to next' gives the reported results of the third race at Calder Race Course on 19 August 1990.

Exhibit 1 shows how the scaled finish vector that will be forecasted is developed. By accumulating the reported lengths back, the number of lengths ahead of the horse in last place can be found. Finally, this vector is scaled from 0

Exhibit 1

Order of finish	Horse	Lengths back to next	Lengths ahead of last	Scaled vector
1	Legend One	1.00	10.50	100
2	Burst With Pride	1.50	9.50	90
3	Winning Partners	1.75	8.00	76
4	Doonascus Princess	2.50	6.25	60
5	Mrs. Bacalaito	3.75	3.75	36
6	Belle's Tiny Tank	–	0.00	0

to 100, giving the scaled finish vector. With this methodology, although the actual values are lost, the scaled finish vector preserves the order of finish and gives a measure of relative performance.

The next section of the paper describes and justifies the base forecasts that were included in the present study. Next, the performance measures and methodology used in evaluating the effectiveness of the combination approach in forecasting the scaled finish vector are described. The final section reports the results and draws some tentative conclusions on the results to date.

3. Specification of the base forecasts

The first criterion for selection of the forecasts to be combined was that they should reflect a wide range of handicapping viewpoints. Ord (1988, p. 397) reports that selecting the forecasts to be included in a combination is an unsettled and difficult issue. Based on previous research, Ord concludes that one effective, though difficult to obtain, characteristic is that the forecasts should be uncorrelated. The present study selected base forecasts based on diverse handicapping methodologies in the hope that the resulting base forecasts would be fairly independent. A second criterion was that the base forecasts must allow for replication from publicly available sources. The final criterion was that the base forecasting methods must be capable of producing a numeric vector that could be scaled to forecast the scaled vector of relative finish positions.

All the selected methods rely solely on information that is available from the *Daily Racing Form*, a newspaper published for each racing day and containing information on the past performances of each horse as well as information and predictions for each race. As a result, the base forecasting methods selected may not be the very best available; in fact, many of these forecasts are somewhat simplistic – though better than random guessing. Although more sophisticated (and more accurate) handicapping analyses are available, the selected base forecasts are well-suited to the current work where interest is focused on evaluating the performance of the combination approach rather than on successful handicapping. Twelve base forecasting methods arising from four major 'philosophies' of handicapping were selected for inclusion. These philosophies and methods are described briefly below. Formulas for calculating each base forecast are given in Exhibit 2.

3.1. Forecasts based on class

The class of a horse relative to the class of the other horses in the race is generally agreed to be an important factor; however, there is little agreement on how class should be defined. Mitchell (1989) states that '. . . class is a qualitative concept. Class refers to brilliance, determination, courage, competitiveness, and the willingness to prevail.' Quinn (1987) argues that class should be defined even more broadly:

> . . . operational definitions of class have concentrated on a particular aspect of the concept, specifically on those attributes that are concrete and objective and therefore more readily assessed – consistency, earnings, weight, speed, and rates of speed (pace). If operational definitions ignore or discount the less tangible aspects of class – endurance, willingness, determination, and courage – and they do, the practices that emerge from them will likely prove at least partially inadequate and disappointing.

Exhibit 2
The base forecasting methods.

Class measures

All class measures are determined using data for the most recent year provided there are at least three races. For horses with less than three races in the most recent year, data are totaled for the two most recent years.

1. Earnings per Start (EPS): EPS = *Earnings*/*Starts*
2. Win Percent (W%): W% = *Wins*/*Starts*
3. In Money Percent (IM%): IM% = (*Wins* + *Seconds* + *Thirds*)/*Starts*
4. Mitchell Class (MC):

$$MC = 0.4\ EPS_{60-100} + 0.3\ W\%_{60-100} + 0.3\ IM\%_{60-100}$$

where

EPS_{60-100} = EPS for entries scaled from 60 to 100;
$W\%_{60-100}$ = W% for entries scaled from 60 to 100; and
$IM\%_{60-100}$ = IM% for entries scaled from 60 to 100.

5. Average Purse Value (APV):

$$APV = Earnings/(0.6 Wins + 0.2 Seconds + 0.1 Thirds)$$

Speed Measures

The 'most recent race' (r-1) used in some of the speed and pace measures is defined here as the most recent completed sprint of the last four races (unless exceptional circumstances were noted in the performance data). If there were no sprints, the most recent route was used.

6. Speed Rating + Track Variant (SRTV): (for most recent race$_{(r-1)}$)

$$SRTV = Speed\ Rating_{r-1} + Track\ Variant_{r-1}$$

7. Average Speed (AVS): $AVS = (SRTV_{r-1} + SRTV_{r-2} + SRTV_{r-3})/3$

where

$SRTV_{r-2}$ and $SRTV_{r-3}$ are the SR + TV for the two next most recent races

8. Ability Rating (AR) $AR = MAX(SRTV_{r-1} + SRTV_{r-2} + SRTV_{r-3})$

Pace Measures

9. Ronnie's (RON): RON = SR + PACE + ADJ

where (all measures are on the most recent race)

$SR = SpeedRating$

for routes ($D > 7$): PACE = 5(46.4 – *1st call*)
for sprints ($D \leq 7$): PACE = 5(46 – *2nd call*)

ADJ depends on length of most recent race

Distance in furlongs (D)	ADJ
4.5 or 5	0
5.5	2
6 or 6.5	10
7	13
>7	15

Exhibit 2 (continued)

10. Unadjusted Factor W (FW) $FW = 0.75EP + 0.25F_3$

where (all measures on most recent race)

T_1, T_2, T_F are the 1st, 2nd and final calls, respectively;
LB_1, LB_2, LB_F are lengths back at the 1st, 2nd and final calls, respectively;
D = length of most recent race (in furlongs)

for sprints:
$D_3 = 660\ (D-4)$
$EP = (2640 - (10LB_2))/T_2$
$F_3 = (D_3 - 10(LB_F - LB_2))/(T_F - T_2)$
for routes:
$T_2' = T_1 - 0.4$
$T_F' = T_2 - 0.4$
$EP = (2640 - (10LB_1))/T_2'$
$F_3 = (1320 - (10(LB_2 - LB_1))/(T_F' - T_2')$

Expert Opinion

11. Consensus Totals (CON):

CON = Sum of *Score* for each horse

Score (as defined by *Daily Racing Form*)	*Experts included*
7 points for 'best bet'	*Trackman*
5 points for first	*Handicap*
2 points for second	*Analyst*
1 point for third	*Hermis®*
0 points for no mention	*Sweep™*
	Reigh Count®

12. Post Time Odds (PTO): PTO: *Post Time Odds*

PTO = 1/(1 + ODDS)

where ODDS are the betting odds of horse to win given in a ratio to one

Note: Italicized measures are taken directly from the *Daily Racing Form*.

One of the constraints on the base forecasting methods is that they require no opinion and permit no experimenter subjectivity. Therefore the class measures included focus on what Quinn calls the operational aspects of class. The five forecasts based on class handicapping are:

1. Earnings per Start (EPS)
   Total earnings/Number of races entered.
2. 'Win' Percent (W%)
   Number of races won/Number of races entered.
3. 'In Money' Percent (IM%)
   Number of races won, place, or show/Number of races entered.
4. Mitchell Class (MC)
   A measure developed by Mitchell (1989) to reflect the relative class of each horse.
5. Average Purse Value (APV)
   Gives an estimate of the value of the purses for the races in which the horse has been entered by adjusting earnings using the approximate percentages of the total purse paid to the first, second, and third place finishers. Only 0.9 is accounted for in this formula as most races pay a small amount to other finishers.

3.2. Forecasts based on speed

Speed handicapping is relatively simple concept; the horse with the minimum estimated finish time is predicted to be the winner. Beyer (1975) states that speed handicappers 'believe that a horse can be measured by how fast he runs. Speed handicappers perform various arcane calculations to translate a horse's ability into a number. If an animal earns a figure of 99, he is superior to a rival who earns a 92. His age, sex, class, breeding, and even his name is irrelevant.'

The *Daily Racing Form* reports the performance for the most recent 10 races for each horse including a speed rating for the horse and a track variant reflecting the condition of the racetrack on the day the race was run. The track variant is typically added to the speed rating producing an adjusted speed rating [Quinn (1988)]. The adjusted speed rating forms the basis for the three speed forecasts described below.

6. Speed Rating Plus Track Variant (SRTV)
   The sum of the speed rating and track variant of the most recent sprint.
7. Average Speed Rating (AVS)
   The average of the most recent three SRTV's.
8. Ability Rating (AR)
   The maximum Speed Rating Plus Track Variant of the last three races.

3.3 Forecasts based on pace

Pace handicapping, although based on the speed of a horse, is more complex than speed handicapping. Determining the pace of a race quickly becomes complicated as pace is '. . . concerned with the multitudinous relations among final times, fractional times, and running styles. . .' [Brohamer (1991)]. Two simple measures of pace are included. Although more complicated to compute than the other base forecasts, they are still relatively simplistic pace calculations and can be computed using the past performance data available in the *Daily Racing Form*.

9. Ronnie's (RON)
   A simple pace measure developed by the second author that summarizes in a simple manner many methodologies that adjust the speed rating by pace.
10. Factor W (FW) (Unadjusted)
   A measure of average pace developed by Howard Sartin [Brohamer (1991)]. The measure used here does not include any adjustments.

3.4. Forecasts based on subjective opinion

Research on the combination of forecasts for time series data suggests the desirability of including subjective as well as objective forecasts into the combination. To include the subjective element here, two final base forecasts were included. The first of these, the consensus of experts, is provided for each race by the *Daily Racing Form*. The scoring method used (see Exhibit 2) is the one used by the *Daily Racing Form*. The last base forecast views post time odds (for win betting) as a consensus ranking of the horses by all bettors. The handicappers included in the consensus picks are clearly recognized as expert by the *Daily Racing Form*. Post time odds is frequently used as a measure of the probabilities of winning for each horse. Ziemba and Hausch (1987), in fact, provide substantial evidence that the post time odds are very good estimates of the true probabilities, although they have found that even better estimates can be obtained by correcting for the 'favorite-long shot' bias [Ziemba and Hausch (1986)].

11. Consensus Total (CON)
12. Post Time Odds (PTO)

3.5. Example of base forecasts

Exhibit 3 shows the calculated base forecasts and normalized forecasts for the example race. In Exhibit 3 the post time odds (PTO) have been converted to the implied probabilities: for each horse, the implied probability is the reciprocal of the odds plus one, where the odds are given in a ratio to one. The implied probabilities for all horses entered in a race will sum to a value greater than one (usually around 1.25) owing to the 'track take' from the betting pool.

Finally, each base forecast is scaled from 0 to 100 reflecting the relative finish vector to be forecast. Accuracy measures can then be determined for each forecast. The accuracy measures will be discussed in the following subsection.

3.6. Discussion of base forecasts

The suggested approach of using normalized vectors to predict a scaled vector of relative competitive results has, potentially, both advan-

Exhibit 3
Example race – the base forecasts.

Order of finish	Horse	Lengths ahead of last	Scaled vector
1	Legend One	10.50	100
2	Burst With Pride	9.50	90
3	Winning Partners	8.00	76
4	Doonascus Princess	6.25	60
5	Mrs. Bacalaito	3.75	36
6	Belle's Tiny Tank	0.00	0

	EPS	W%	IM%	MC	APV	SRTV	AVS	AR	RON	FW	CON	PTO
Base forecasts												
Finish												
1	1762	0.11	0.44	92	15860	101	101	101	93	54.68	13	0.286
2	1630	0.18	0.18	89	14938	98	97	100	94	54.51	4	0.185
3	1961	0.00	0.50	88	19613	101	98	103	91	55.24	21	0.345
4	173	0.00	0.00	60	3	84	94	104	76	53.38	5	0.278
5	1123	0.13	0.25	83	11231	90	87	91	83	54.14	1	0.056
6	342	0.00	0.11	64	30800	88	90	92	90	53.95	0	0.095
Scaled base forecasts												
Finish												
1	89	61	89	100	51	100	100	77	94	70	62	80
2	81	100	36	91	48	82	71	69	100	61	19	45
3	100	0	100	87	64	100	83	92	83	100	100	100
4	0	0	0	0	0	0	51	100	0	0	24	77
5	53	69	50	71	36	35	0	0	39	41	5	0
6	9	0	22	13	100	24	20	8	78	31	0	14
Accuracy measures												
Forecasting accuracy												
Error												
Mean	5	22	11	0	10	3	6	3	−5	10	25	8
MAD	22	36	31	20	44	19	15	24	27	30	33	26
MSE	785	2007	1310	841	2969	788	360	703	1629	1144	1559	806
Handicapping accuracy												
Win	0.00	0.00	0.00	1.00	0.00	0.50	1.00	0.00	0.00	0.00	0.00	0.00
Dutch	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00
Top 3	3.00	2.00	2.00	3.00	2.00	3.00	3.00	2.00	3.00	3.00	2.00	2.00

tages and disadvantages. There are three major potential disadvantages to this approach. First, information is lost when scaling the results; the scaled values do not indicate whether a few yards or half the track separate the first horse from the last place horse. Second, only the top 4 or 5 finish positions may accurately reflect the ability of the horses since it seems likely that horses approaching the finish line in, say, sixth and seventh places will not duel as fiercely as the horses in first and second places. Finally, in many cases the vectors being used as base forecasts are being reinterpreted. For example, post time odds reflect the bettors' consensus estimate as to the likelihood of each horse winning the race. The assumption that a horse that is second most likely to win is also the horse most likely to come in second is subject to debate. Mitchell (1989) provides some evidence contradicting this assumption and Brohamer (1991) argues (from the philosophy of pace handicapping) that the second place horse is often a horse with different running style than the winner rather than the second-best horse in the race. However, Harville (1973) and Ziemba and Hausch (1987) use this assumption in their mathematical analyses with little difficulty. In fact, Ziemba and Hausch (1986, 1987) have provided substantial data in support of this assumption.

There are also three potential major advantages to this approach. First, scaling all values results automatically in equal ranges for the forecasts and the values being forecast; this is particularly important when incorporating subjective forecasts which often show much lower variability than the actual series. Second, the vector of results takes into account the one-time, competitive nature of the event being forecast. Third, the method allows the drawing on a wide variety of philosophies. By developing vectors that give relative 'goodness' ratings for each horse and then scaling the values, one can use forecasts in any units. The next step, and the focus of this paper, is the averaging of these scaled forecasts to produce a combined forecast vector.

4. Data collection and methodology

In order to evaluate the potential usefulness of this approach in the context of forecasting a vector of relative, scaled performance by competitors in one-time events, a detailed empirical evaluation of the suggested base forecasts and their combinations was undertaken for 25 races. This section describes the data collection and methodology used.

In order to have a fairly homogeneous set of races to forecast, the races included in the study were arbitrarily confined to non-maiden (a maiden horse is one that has never won a race), mid-level (bottom-level claiming races were omitted as were graded stakes), sprints (6, 6.5, and 7 furlong races) in which information on at least one prior race was available for all horses. All races analyzed were run at Calder racetrack in Florida during June, July and August of 1990. For the present study, 25 races matching that criteria were selected at random for inclusion in the analysis. In the present study the only combination method examined is simple averaging.

4.1. Accuracy measures

The present study has somewhat of a dual nature. On the one hand, it is designed to study the general question of the appropriateness and effectiveness of the combination approach in forecasting one-time competitive events. On the other hand, the study is being performed in the context of thoroughbred racing where the most pertinent measures of effectiveness are concerned with identifying winning horses and the payoffs associated with various betting strategies. To reflect this dual nature, two sets of performance measures were selected: one reflecting standard forecasting accuracy measures and the other handicapping accuracy. The measures used are described and justified below. An exact definition of each measure is given in Exhibit 4.

4.2. Forecasting accuracy measures

A large body of research has explored various aspects of the combination of forecasts. Typically, evaluation of forecast performance has been done by measuring the reduction in average variability of the forecasting error [usually measured using MAD (mean absolute deviation), MSE (mean square error) or MAPE (mean absolute percentage error)] as the number of base forecasts included in the combination is increased.

The following three measures were used to assess forecasting accuracy: Mean Error (ME), Mean Absolute Deviation (MAD), and Mean Square Error (MSE). These measures were selected as those most commonly used in the combination of time series forecasts in order to obtain comparable results. One difference is that although the present work uses only a simple unweighted average to combine the base forecasts, ME is included as an accuracy measure. ME is not found in previous work using a simple average to combine since the average error of all possible combinations is simply a mean of means and thus mean error remains constant regardless of how many base forecasts are included in the combination. Here, however, the combinations were rescaled after averaging so that the mean error is also changed.

4.3. Handicapping accuracy measures

The handicapping accuracy measures are designed to give a measure of how accurately the forecasting method picks winning horses. These

Exhibit 4
Accuracy measures.

Forecasting accuracy measures

(1) *Mean Error (ME)*

$$\mathrm{ME} = \Sigma(\mathrm{AS} - \mathrm{FS})/n,$$

where

AS = finish position (scaled 0 to 100);
FS = forecasted finish position (scaled 0 to 100);
n = number of horses in race.

(2) *Mean Absolute Deviation (MAD)*

$$MAD = \Sigma|AS - FS|/n.$$

(3) *Mean Square Deviation (MSE)*

$$\mathrm{MSE} = \Sigma(\mathrm{AS} - \mathrm{FS})^2/n.$$

Handicapping accuracy measures

(1) *Picking Winner (WIN)*

Win = [1 if correctly identified winner, 0 if not] * *Adjw*,

where

Adjw = 1/number of horses with forecasted scaled value of 100.

(2) *Dutching (DUTCH)*

Dutch = [1 if winner was either of two highest scored horses, 0 if not] * *Adjd*,

where

Adjd = 2/number of ties in two highest scores.

(3) *Top 3 (TOP3)*

Top3 = [number of top 3 forecasted horses that did finish in top 3] * *Adj3*,

where

Adj3 = 3/number of ties in three highest scores.

simple measures reflect the accuracy of the ranking adjusted for tied rankings.

(1) *Picking Winner (WIN)*. Whether or not the winner was correctly identified. The average of this value indicates the average percent of winners selected.

(2) *Dutching (DUTCH)*. Dutching is a common method of selecting horses on which to bet. The two most likely horses to win are identified and both bet to win (often with the relative amounts of the wagers in proportion to the potential payoffs). The average of this value indicates the average percent of races where the winner was one of top two selections.

(3) Top *3* (TOP*3*). The top three horses ('in the money') are of interest in a handicapping context since horses can be bet to win, place, or show as well as perfecta (first two horses in order) and trifecta (all three horses in order) betting.

4.4. Methodology

In order to evaluate the combination of forecasts methodology, the same general procedure that was employed in previous research on time series data is used here. The set of available base forecasts (n) gives an upper bound on the number of possible forecasts to be included. For all sizes less than or equal to n, all possible combinations of base forecasts are selected, combined, and evaluated on accuracy. The average of these accuracy measures is found for each size and used to evaluate the effects of the combination methodology. In the present research, there is a additional step since the results must also be averaged over a number of races ($N = 25$). The general methodology of data collection is:

(1) The results of each of the $N = 25$ races were entered and scaled (see Exhibit 3).

(2) The data required for the base forecasts were entered (see Exhibit 2).

(3) The 12 base forecasts were scaled and accuracy measures found (see Exhibit 3).

(4) All combinations ('n choose x') were found for $n = 12$ and $x = 1$ to 12 using a simple unweighted average to combine the base forecasts. For each possible combination:

- (i) scaled base forecast vectors were averaged;
- (ii) averaged vectors were rescaled;
- (iii) accuracy measures were calculated and stored.

(5) Average accuracy measures were calculated for each number of combinations (x).

(6) Average (for each x) accuracy measures were calculated for the N races in the analysis.

Exhibit 5 shows the resulting summary accuracy results for all sizes of combinations of base forecasts for the example race. It should be noted that each entry in the first row of combinations of base forecasts for the example race. It should be noted that each entry in the first row is the average of 12 combinations (12 choose 1) and each entry in the sixth row is the average of 924 combinations (12 choose 6) while the twelfth row is the one result of averaging all 12 base forecasts.

5. Analysis and results

A summary of the results for this study of $N = 25$ races is shown in Exhibit 6. The six different measures of accuracy are shown for the individual base forecasts and for all possible combination sizes. In Exhibit 6, each entry in the first row is the average of 12 combinations (12 choose 1) for 25 races and each entry in the sixth row is the average of 924 combinations for 25 races while the twelfth row is the average of the 25 forecasts determined by averaging all 12 base forecasts for each race. The effects on the accuracy measures of increasing the number of base forecasts in the combination are also shown graphically in Exhibit 7. These results are discussed in detail below.

5.1. Forecasting accuracy measures

The results of the present study are very

Exhibit 5
Example race – accuracy measures of combinations of size x ($N = 1$).

	Forecasting accuracy			Handicapping accuracy		
x	ME	MAD	MSE	WIN	DUTCH	TOP3
1	8.2	27.3	1241	0.21	0.75	2.50
2	6.8	24.6	979	0.23	0.95	2.71
3	7.6	23.2	879	0.23	1.00	2.91
4	7.9	22.2	824	0.28	1.00	2.99
5	8.5	21.9	809	0.34	1.00	3.00
6	9.0	21.7	800	0.36	1.00	3.00
7	9.4	21.6	798	0.36	1.00	3.00
8	9.8	21.5	798	0.33	1.00	3.00
9	10.1	21.5	803	0.28	1.00	3.00
10	10.5	21.6	809	0.18	1.00	3.00
11	10.7	21.6	806	0.08	1.00	3.00
12	10.9	21.8	797	0.00	1.00	3.00

similar in pattern to the results of previous studies with time series data in regard to measures of error variability. Average MAD and MSE both show a monotonic decrease as the number of base forecasts (x) being combined is increased. As with previous work, the greatest rate of decrease is with a small number of forecasts; after combinations of size 3 or 4, the improvement, though consistent, is at a decreasing rate (see Exhibit 7). The combination approach gives less variable forecasting errors than any of the base forecasts; the 12 forecasts combined has an average MAD of 27.87, while the minimum average MAD for the individual base forecasts is 30.5 (see Exhibit 6). MSE gives similar results, the 12 forecasts combined show a MSE of 1242 compared with a *minimum* MSE of 1352 for the base forecasts. These results strongly support the effectiveness of the combination of forecasting approach in the domain of one-time, competitive events.

Accuracy as measured by mean error, however, shows a different pattern. As previously discussed, combinations of time series forecasts made with a simple unweighted average can never demonstrate improvement (or deterioration) in accuracy with combinations. Here, however, the combined forecasts are rescaled so that the mean of the combinations is not identical to the mean of the individual forecasts. The mean error shows an intriguing pattern (which can also be seen in some of the handicapping accuracy measures): the change is no longer a monotonic function of x. For the mean error, there is a large drop in the average error when two forecasts are combined, a third forecast decreases mean error slightly, then as more forecasts are included, the mean error increases quite consistently, though the average error for all 12 forecasts combined is less than the average for the individual forecasts.

In examining the base forecasts, there is a wide divergence in mean error. However, the mean errors seem to vary by the handicapping 'philosophies' previously discussed. The forecasts based on speed and on pace analysis (SRTV, AVS, AR, RON, and FW) show by far the lowest mean error (in fact, their average mean errors are all less than the lowest average mean error found with combinations). All other base forecasts show mean errors higher than with the highest mean combination error. The subjective methods (CON and PTO) appear to be particularly biased. This pattern suggests that in the domain of forecasting competitive events, it may be possible to specify systematic patterns of performance behavior of the individual base forecasts and thus by either weighting or by selecting the base forecasts attain consistent improvement in the combined forecasts. A detailed analysis of this pattern is beyond the scope of the present study and must await future research.

The results on forecasting accuracy strongly support the usefulness of the combination of forecasts of scaled vectors to predict a scaled vector in the domain of competitive one-time events. Increasing the number of base forecasts in the combination usually enhanced average accuracy and only decreased it slightly in mean error when larger numbers of base forecasts were included. The practical question of how good this methodology is in the specific area of handicapping thoroughbred races still remains and is discussed in the next two sections.

5.2. *Handicapping accuracy measures*

Measures of handicapping accuracy will be discussed in order of least difficult (correctly identifying the top 3 horses in any order) to most difficult (correctly picking the winner). With the Top3 measure, the combination approach showed consistent positive results for identifying horses 'in the money'. The best individual forecast was the post time odds (PTO) which correctly identified an average of 1.96 of the top three horses. On average, the 12 base forecasts identified 1.61 of the 'in the money' horses. As more base forecasts were included in the combination, the average number correctly identified increased monotonically to an average of 2.00 of the 3 top horses identified correctly. These results can be compared with chance. The 25 races had an average of 7.72 horses per race. Rounding to 8 and using the hypergeometric distribution shows that selecting 3 horses at random would result in an average of 1.125 'in the money' horses. All but Average Purse Value (APV) give much better than chance results.

With Dutching, the forecasting task becomes somewhat more difficult. Here, success is mea-

Exhibit 6
Summary of results, $N = 25$.
A. Average accuracy measures for all combinations of size x from 12 forecasts

x	ME	MAD	MSE	'Winner'	'Dutching'	Number 'in money'
1	7.5	32.5	1,754	0.30	0.48	1.61
2	6.1	30.3	1,522	0.31	0.53	1.69
3	6.1	29.4	1,433	0.39	0.55	1.75
4	6.1	28.8	1,376	0.42	0.57	1.78
5	6.1	28.5	1,339	0.45	0.57	1.81
6	6.3	28.3	1,314	0.42	0.69	1.83
7	6.4	28.1	1,295	0.37	0.58	1.86
8	6.5	28.1	1,281	0.36	0.59	1.88
9	6.6	28.0	1,268	0.35	0.64	1.90
10	6.7	27.9	1,258	0.34	0.59	1.92
11	6.8	27.9	1,251	0.34	0.60	1.94
12	7.0	27.9	1,242	0.32	0.64	2.00

B. Average accuracy measures for 12 base forecasts

	EPS	W%	IM%	MC	APV	SRTV	AVS	AR	RON	FW	CON	PTO
Forecasting accuracy												
Error												
ME	10.8	15.1	4.9	8.6	14.0	−2.1	−2.3	0.0	−1.2	2.7	24.7	15.3
MAD	33.6	35.9	31.0	31.1	41.6	30.5	30.5	31.1	28.8	31.9	35.2	28.3
MSE	1825	2082	1626	1692	2553	1608	1567	1613	1403	1717	2007	1352
Handicapping accuracy												
Win	0.24	0.42	0.26	0.40	0.12	0.28	0.32	0.13	0.30	0.40	0.36	0.40
Dutch	0.48	0.48	0.47	0.52	0.20	0.41	0.59	0.29	0.51	0.52	0.69	0.61
Top 3	1.72	1.37	1.58	1.40	1.20	1.62	1.78	1.67	1.68	1.56	1.81	1.96

sured by whether or not the winner is one of the two top picks by the forecast. Increasing the number of base forecasts included reflects (though irregularly) the same pattern as the mean error: rapid improvement as x is increased, then a drop off in performance. With Dutching, however, there is a strong upsurge in improvement at $x = 6$, a smaller surge at $x = 9$ (see Exhibit 7), and overall improvement due to combining. The performance of the forecasts at Dutching can also be compared with chance; with an average of 7.72 horses per race, randomly selecting two horses would result in selecting the winner 25.91% of the time.

A more practical touchstone for the value of the combinations is provided by Howard Sartin, developer of the Sartin Methodology [Brohamer (1991)]. Sartin has identified winning 63% of races using the Dutching strategy as a benchmark for success. The average percent Dutching success for the base forecasts was 48% ranging from a low of 20% (less than chance) for Ability Rating (AR) to a high of 69% for the consensus rating (CON). Unadjusted Factor W (FW), based on the Sartin Methodology, averaged 52%. (In the Sartin Methodology, Factor W includes subjective adjustments incorporating the expertise of the handicapper.) Combining of the forecasts reached a maximum of 69% with 6 forecasts, then declined. The combination of all 12 values, however, shows that the winner was one of the top two horses selected in an average of 64% of the races. The combination method shows promise in supporting Dutching of bets.

Predicting the winning horse is the most difficult of the three handicapping measures. A random selection of the winner in the 25 races would be expected to result in a 12.95% success rate. Previous research also suggests expected success rates for some of the base forecasts. Post time odds are often used to provide a probability measure as to the likelihood of a horse winning a race. Much previous research has shown that the horse with the lowest post time odds ('favorite') wins approximately 36% of the time [Ali (1977) and Asch, Malkiel and Quandt (1982)]. From the *Daily Racing Form*, favorites won approximately 35% of the races run at Calder in 1989

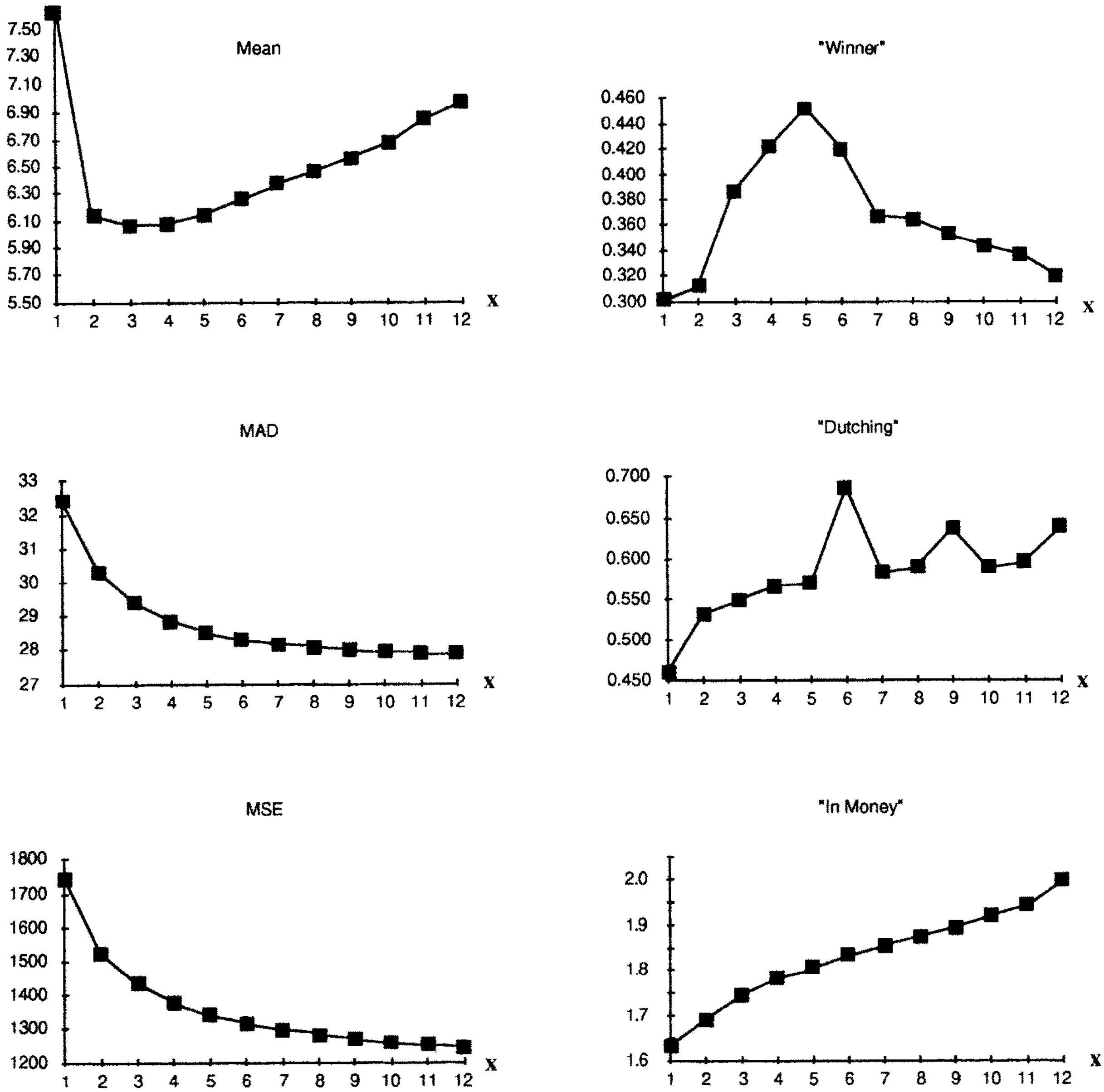

Exhibit 7. Effects of combination on accuracy.

but were only winning at about a 31% rate in the summer of 1990. As Exhibit 6 demonstrates, this sample has favorites (see PTO) winning at 40%, showing that the favorite won slightly more often than usual in the 25 races of the study. Davidowitz (1983) reports that the horse with the highest speed rating wins about 35%. Since SRTV and AVS are not based on sophisticated methods of rating speed, their lower win rates of 28% and 32%, respectively, is reasonable. Two of the base forecasts, APV and AR, performed at about the level of chance. The other base forecasts performed noticeably better than chance with four of the methods predicting at least 40% of the winners in the 25 races.

In predicting winners with combined forecasts, the general pattern of quick improvement for a small number of base forecasts followed by a gradual deterioration in performance as the number of base forecasts is increased can be seen in Exhibit 7. This same pattern can be seen in the mean forecast error and, to a lesser extent, in Dutching. Further research must be conducted to explore the intriguing non-monotonic pattern. Overall, for predicting winners, there is a small average increase from 30% for the individual forecasts to 32% for all 12 combined (see Exhibit 6). The average percent of correctly identified winners peaks at $x = 5$ with an impressive average of 45% winners correctly identified.

5.3. Betting applications

The previous analysis was concerned with assessing the accuracy of forecasts in two contexts – forecasting accuracy and handicapping accuracy – and the effects of combination on accuracy. This subsection reports a preliminary analysis of the effects of combination on the betting applications of the forecasts and combinations of forecasts. Because of the parimutuel structure of thoroughbred racing, accuracy, even handicapping accuracy, will not necessarily result in a positive return on money bet. For example, it is well known and documented that, although the favorite wins about 33% of the time, betting on the favorite to win is a losing proposition as the small payoffs for winning favorites cannot counterbalance the money lost on the 67% of losing favorites plus the track take (typically 15–20%).

A large number of bets (and combinations of bets) and betting strategies exist and most have their opponents and proponents. For this preliminary study three simple betting strategies were selected for analysis: (1) Flat bet to win; (2) Dutching with flat bet; and (3) Kelly betting. These strategies are described in more detail below.

(1) *Flat bet to win*. In each race, a set amount ($100 in the study) is bet 'to win' on the horse predicted to win. If two or more horses are tied in the forecast, the set amount is divided equally among them. An infinite bankroll is assumed.

(2) *Dutching with flat bet*. In each race, a set amount ($100 in the study) is divided evenly and bet 'to win' on the two top horses. If three or more horses are tied in forecast, the set amount is divided equally among them. An infinite bankroll is assumed.

(3) *Kelly betting*. This system was developed by Kelly (1956) and has gained acceptance as a method to determine how much to wager on each horse in each race. In Kelly betting, the estimated probability of each horse winning is compared with the probability implied by the post time odds. The better has an 'edge' if the estimated probability is greater than the implied probability and should bet on that horse, proportional to the edge. The formula are: Let

p_i = the estimated probability that horse$_i$ will win

(probabilities estimated from forecasted vector are computed by dividing each value of the vector by the total of vector) and

ODDS_i = the post time odds of horse$_i$.

Then the edge is computed:

$$\mathrm{EDGE}_i = (\mathrm{ODDS}_i * p_i) - (1 - p_i).$$

To determine the amount to bet on each horse$_i$:

if $\mathrm{EDGE}_i > 0$, bet$(\mathrm{EDGE}_i/\mathrm{ODDS}_i) *$ bankroll;
if $\mathrm{EDGE}_i \leq 0$, do not bet horse$_i$.

The study uses a $1000 bankroll for each race; the amount bet and the number of bets will vary for each race.

Both of the first two betting strategies are concerned only with the rank of the horses in the forecasted vector. In order to assess the effectiveness of the complete vector of the suggested forecast, Kelly betting was adopted as the third strategy. Since Kelly betting uses the Post Time Odds to determine the betting strategy, PTO cannot sensibly be included as one of the base forecasts in the combinations. The following analysis is based on the remaining 11 base forecasts.

Exhibit 8 summarizes the results of the analysis on betting strategies. The performance of the individual base forecasts for the 25 races is shown in tabular form at the bottom of the page. Results for PTO for the win and Dutching bets (strategies 1 and 2) are displayed here, but PTO is not included in the rest of the analysis. As with the previous analysis, all possible combinations of size 1 to 11 were formed for each race and the results of each of the three betting strategies

Exhibit 8
Results of different betting strategies with 11 base forecasts, $N = 25$.
A. Average return for all combinations of size x from 11 base forecasts (PTO not included)

	Return from			Kelly betting	
x	Win	Dutching	Kelly	No. of Bets	% Bet
1	1.06	0.97	0.93	3.0	0.27
2	0.98	1.01	1.00	3.3	0.20
3	0.97	0.99	1.03	3.0	0.20
4	1.07	1.02	1.05	3.3	0.17
5	0.96	0.99	1.07	3.4	0.16
6	1.06	1.02	1.08	3.4	0.16
7	1.05	1.00	1.10	3.4	0.15
8	1.00	1.00	1.11	3.4	0.15
9	0.99	0.96	1.13	3.4	0.15
10	0.90	0.94	1.17	3.4	0.15
11	0.85	0.92	1.26	3.3	0.15

B. Average wagering return for all 12 individual base forecasts

	EPS	W%	IM%	MC	APV	SRTV	AVS	AR	RON	FW	CON	PTO
Return												
Bet												
Win	0.75	1.25	0.92	1.19	1.36	0.93	1.29	0.40	1.06	1.56	0.94	1.04
Dutch	0.74	0.80	1.06	0.87	0.77	0.74	1.70	0.52	0.96	1.24	1.23	0.97
Kelly	0.53	0.53	0.91	0.43	0.95	1.08	1.51	0.71	1.29	1.21	1.08	none
Data on Kelly betting												
No. of bets	3.20	3.00	3.40	3.40	3.10	3.00	3.30	3.00	3.00	3.20	2.00	
% bet	0.25	0.32	0.26	0.25	0.40	0.23	0.22	0.22	0.23	0.23	0.21	

found. The return for each trial was determined by dividing dollars won by dollars bet for each trial (a return of less than 1 shows a loss, of more than one a profit). Exhibit 8 shows the average return as more base forecasts are included in the combination. With the first two strategies a flat amount was bet on each race. With Kelly betting the number of bets and amount bet varies; in fact, some races may not be bet at all. The second set of columns shows average information on Kelly betting: the proportion of races in which a bet was made (playable races), the number of bets per race, and the average percent of a set amount bet per race.

These results are somewhat different from those of handicapping accuracy (see Exhibit 6) since they reflect the effect of the three betting strategies. APV (Average Purse Value), for example, correctly identified only 12% of the winners (3 of 25 races), yet, because one of the correct forecasted winners was a long-shot, APV returned 136% (36% profit). On the other hand, because the 25 races selected had, by chance, more than the average number of favorites winning, PTO correctly identified 40% of the winners (10 of 25 races). However, since PTO, by definition, is betting on the favorite, PTO returned only 104%.

Since only 25 races are in the sample, it would be premature to draw firm conclusions about the performance of any of the individual base forecasts or combinations with the betting strategies. However, by looking at all possible combinations, it is possible to assess the effectiveness of the use of the combination approach with the three betting strategies of this analysis. Exhibit 8 shows the average return for the three strategies for all combinations of size x from the 11 remaining base forecasts. Combination, at least with these 11 base forecasts, does not appear to enhance the return with the betting strategies of Flat win or Dutching betting. In fact, overall performance deteriorates: with win betting the average return of the individual base forecasts was 106%, with all 11 combined, average return fell to 85%; similarly, with Dutching return fell from 97% to 92%. Both strategies show the same non-monotonic pattern as was seen with handicapping accuracy (see Exhibit 7), suggesting that perhaps the combination approach would be more effective with a larger sample size or with a different selection of base forecasts.

With Kelly betting, however, the combination approach shows a monotonic and fairly large increase in performance as the number of included forecasts is increased (see Exhibit 8). The average return of the individual base forecasts was 93% (7% loss), with all 11 combined, average return was 126%. Increasing the number of base forecasts has little effect on the number of bets per race; however, the average percent bet per race decreased monotonically from \$270 (27% of the bankroll) for the individual forecasts to \$150 for all 11 combined. This suggests that the combination approach is successful in producing a vector of relative values which, as the number of base forecasts is increased, are converging to a reasonable estimate of the relative likeliness of each horse winning the race.

6. Conclusions and future research

There are three major original contributions of this research. First, this research explores the use of the combination approach to non time series data. The results strongly support the combination of forecasting one-time competitive events. The study found that as the number of base forecasts increased, variability of the error (measured by MSE and MAD) decreased monotonically. These results parallel the findings of research studying the combination of time series forecasts. As with previous work in the time series domain, the greatest decrease in variability was found with a small number of base forecasts; for a larger number of forecasts, improvement, though consistent was at a decreasing rate. Second, the research examines situations where forecasts and outcomes are vectors rather than scalars. This provides a natural extension to previous work on the combination of forecasts. The results strongly support the success of the combination approach to a forecasted vector since all applications of the forecasted vector that incorporated the entire vector showed enhanced monotonic improvement as the number of base forecasts was increased. Third, it uses indexed values for both forecasts and outcomes which provides a way of combining different metrics. Previous work tended to restrict the combination to forecasts that are based on the same basic numeric scale. This research applies the forecasting methodology to a situation where there is no 'natural' number to forecast, yet the forecasts used provide reasonable support for decision-making.

This study suggests that the combination approach is, at least potentially, of value in betting strategies, such as Kelly betting, that require estimation of the relative likelihood of each horse winning the race. In such situations the analysis demonstrates that increasing the number of base forecasts enhances the accuracy of such estimation. Work needs to be done to produce a simplified, robust method of using these estimated data in actual wagering situations.

In betting strategies focusing on identifying the winning horse (or top 2 horses), the study did not show the combination approach as effective. An exploratory analysis including only the most successful base forecasts, however, showed improvement with combinations of a small number of forecasts. Further research is needed to see if it is possible to increase the usefulness of the combination approach by careful selection of the base forecasts or by weighting the base forecasts to reflect their success at winner selection.

Within the domain of thoroughbred racing additional study is needed to refine the method. For example, are there more effective methods for computing the base forecasts than scaling from 0 to 100? Is a simple average the best combination method? Should more, or less, or different base forecasts be included in the combinations?

References

Ali, M.M., 1977, "Probability and utility estimates for racetrack bettors", *Journal of Political Economy*, 85, 803–815.

Armstrong, J.S., 1984, "Forecasting by extrapolation: Conclusions from 25 years of research", *Interfaces*, 14, No. 6, 52–66.

Asch, P., B.G. Malkiel and R.E. Quandt, 1982, "Racetrack betting and informed behavior", *Journal of Financial Economics*, 10, 187–194.

Beyer, A., 1975, *Picking Winners: A Horseplayer's Guide* (Houghton Mifflin, Boston).

Brohamer, T., 1991, *Modern Pace Handicapping* (William Morrow, New York).

Bunn, D., 1981, "Two methodologies for the linear combination of forecasts", *Journal of the Operational Research Society*, 32, 213–222.

Davidowitz, S., 1983, *Betting Thoroughbreds* (Dutton, New York).

Flores, B.E. and E.M. White, 1988 "A framework for the combination of forecasts", *Journal of the Academy of Marketing Science*, 16, Nos. 3 & 4, 95–103.

Harville, D.A., 1973, "Assigning probabilities to the outcomes of multi-entry competitions", *Journal of the American Statistical Association*, 68, 312–316.

Kelly, J. L, Jr., 1956, "A new interpretation of the information rate", *Bell System Technical Journal*, July, 917–926.

Lawrence, M., R. Edmundson and M. O'Connor, 1986, "The accuracy of combining judgmental and statistical forecasts", *Management Science*, 32, No. 2, 1521–1532.

Makridakis, S. and R. Winkler, 1983, "Averages of forecasts: Some empirical results", *Management Science*, 29, No. 9, 987–996.

Makridakis, S. et al., 1982, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition", *Journal of Forecasting*, 1, No. 2, 111–154.

Mitchell, D., 1989, *Winning Thoroughbred Strategies* (William Morrow, New York).

Ord, J.K., 1988, "Future developments in forecasting – the time series connection", *International Journal of Forecasting*, 4, 389–401.

Quinn, J., 1987, *Class of the Field* (William Morrow, New York).

Quinn, J., 1988, *The ABC's of Thoroughbred Handicapping* (William Morrow, New York).

Reeves, G. and K. Lawrence, 1982, "Combining multiple forecasts given multiple objectives", *Journal of Forecasting*, 1, No. 3, 271–279.

Winkler, R. and S. Makridakis, 1983, "The combination of forecasts: Some empirical results", *Journal of the Royal Statistical Society, Ser. A*, 146, 150–157.

Ziemba, W.T. and D.B. Hausch, 1986, *Betting at the Racetrack* (Norris M. Strauss, New York).

Ziemba, W.T. and D.B. Hausch, 1987, *Dr. Z's Beat the Racetrack* (William Morrow, New York).

Biographies: Edna M. WHITE is an Associate Professor in the Department of Decision and Information Systems Department at Florida Atlantic University. She received her Ph.D. degree in Operations Management from the University of Minnesota. She has published several articles in the area of combination of forecasts. She is currently active in the areas of time series forecasting, quality control, the management of quality, and improving her handicapping skills.
Ronald DATTERO is an Associate Professor in the Department of Decision and Information Systems at Florida Atlantic University. He received his Ph.D. from Purdue University's School of Industrial Engineering. His research focuses on supporting decision-making through the use of operations research, forecasting, database, and artificial intelligence techniques. He regularly applies these techniques to 'parimutuel investing' at thoroughbred race tracks.
Benito FLORES is Associate Professor of the Department of Business Analysis and Research at Texas A&M University. His research has been published in journals such as *Operations Research*, *Management Science*, *Omega*, the *Journal of Operations Management* and the *Journal of Forecasting*. His areas of research interest include forecasting and time series analysis, strategic planning, inventory systems, and international business.