Open access • Journal Article • DOI:10.1080/07474940008836449

Selecting the best of k multinomial parameter estimation procedures using sprt

— Source link

E.S Rosenbloom

Published on: 01 Jan 2000 - Sequential Analysis (Marcel Dekker, Inc.)

Topics: Sequential estimation, Sequential probability ratio test, Multinomial distribution, Estimation theory and Selection (genetic algorithm)

SELECTING THE BEST OF K MULTINOMIAL PARAMETER ESTIMATION PROCEDURES USING SPRT

by

E.S. Rosenbloom

A thesis
submitted to the Faculty of Graduate Studies
in partial fulfillment of the
requirements for the degree of
Master of Science

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

0-612-45119-4

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

Selecting the Best of K Multinomial Parameter Estimation Procedures Using SPRT

BY

E.S. Rosenbloom

A Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

ABSTRACT

An SPRT-like method is developed for the problem of selecting the best of k multinomial parameter estimation procedures when only one observation per the k estimation procedures is possible but the k estimation procedures can be repeated many times.

The multinomial probability mass function considered is

$$f(x_1, x_2, \ldots x_n) = p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n}$$

$$\text{where } \sum x_i = 1 \text{ and } x_i = 0 \text{ or } 1$$

$$\text{and parameters } p_1, p_2, \ldots p_n \text{ satisfy } \sum p_i = 1 \text{ and all } p_i \geq 0.$$

It is assumed that there are two or more procedures for estimating parameters $p_{1,t}$, $p_{2,t}$,..., $p_{n(t),t}$ for each observation t. The number of parameters n(t) can depend upon the observation number t. An example of this would be competing procedures for estimating the probability of a horse winning a race. The parameter $p_{j,t}$ would represent an estimate of the probability of horse j winning race t. Race t can be run only once and hence only one observation can be obtained for the k estimation procedures for race t. However, the k estimation procedures can be repeated for different races. Other examples of competing multinomial parameter estimation procedures would include different methods of estimating the probability of a financial market being up in a given time period or different forecasts of the probability of precipitation.

A new procedure for estimating probabilities at a racetrack is developed and the SPRT-like method is used to compare this new procedure to the existing theory that an entry's probability of winning is equal to the fraction of the win pool bet on that entry.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE

PAGE

Chapter 1

Introduction

The problem considered in this thesis is how to select the best of k multinomial parameter estimation procedures when only one observation per the k estimation procedures is possible but the k estimation procedures can be repeated many times.

The multinomial probability mass function considered is

$$f(x_1, x_2,\ldots,x_n) = p_1^{x_1} p_2^{x_2} \ldots p_n^{x_n}$$

$$\text{where } \sum x_i = 1 \text{ and } x_i = 0 \text{ or } 1$$

and parameters $p_1, p_2,\ldots,p_n$ satisfy $\sum p_i = 1$ and all $p_i \geq 0$.

It is assumed that there are two or more procedures for estimating parameters $p_{1,t}, p_{2,t},\ldots,p_{n(t),t}$ for each observation t. The number of parameters n(t) can depend upon the observation number t. The goal of this thesis is to develop a Bayesian method to select the best estimation procedure. The term "best" is defined as follows:

Definition: An estimation procedure i* will be declared the best of estimation procedures {1,2,...,k } at a significance level α (α is a parameter between 0 and 1, typically .10, .05 or .01) if the posterior probability of estimation procedure i* being correct is above $1-\alpha$ (assuming one of the estimation procedures {1,2,...,k } is correct.)

This is an unusual definition and some may question the assumptions and the philosophy of this definition. This definition assumes that one of the estimation procedures {1,2,...,k} is correct and is generating the correct parameter estimates $p_{1,t}, p_{2,t},\ldots,p_{n(t),t}$ for observation t. In reality, it is possible that none of the parameter estimation procedures are correct. However, this definition is consistent with the philosophy of a simple versus simple hypothesis test of $\theta =$

θ_0 versus $\theta = \theta_1$. It is possible that the parameter θ may be neither θ_0 nor θ_1 but the hypothesis test is used to determine which is the more appropriate hypothesis. Here the definition is used to help determine which of the various parameter estimation procedures is the most appropriate. This definition is also consistent with the Baum and Veeravalli (1994) approach to multihypothesis testing. A second feature of this definition is its Bayesian perspective. The goal is to use the data collected to select the best estimation procedure. Before the data are collected, prior probabilities (which may be uniform) of whether a particular estimation procedure is correct are required. After data are collected, an estimation procedure i^* is selected the best at a significance level α if its posterior probability of being correct is at least $1-\alpha$. Throughout this thesis, significance level is given this Bayesian interpretation.

The original rationale for investigating this problem was competing probability models at a racetrack. For a given race t, bettors can choose between betting entries 1, 2,...,n(t). In this situation $x_{j,t} = 1$ if betting entry j wins race t and $x_{j,t} = 0$ if betting entry j does not win race t. One estimate of the probability of betting entry j winning race t, is the fraction of the win pool bet on entry j. There are a number of studies (Fabricand (1979), Hausch, Ziemba and Rubinstein (1981), Snyder (1978) and Ziemba and Hausch (1984)) that indicate that this is a reasonable estimate of the probability of an entry winning a race. In this thesis a model using both multiple regression and simulation is developed for estimating the probability $p_{j,t}$ of an entry j winning race t. Given that there are now two models or procedures for estimating the probability of entry j winning race t, which of these two procedures is more accurate?

Another example of competing models at the racetrack are the probabilities associated with the host track and the probabilities associated with satellite tracks in simulcasting of racing. It is quite common for races run at one track (called the host track) to be offered to bettors at a remote location (called the satellite track). Sometimes the host track and the satellite track run separate betting pools. In this situation there are two estimates of the probability of entry j winning race t. The fraction of the win pool bet on entry j at the host track is an estimate of $p_{j,t}$. The fraction of the win pool bet on entry j at the satellite track is also an estimate of $p_{j,t}$. In a *Journal of Business* article, Hausch and Ziemba (1990) argue that the "true probabilities" are the probability estimates at the host track. They argue that bettors at the satellite track should use the probabilities established at the host track to exploit anomalies in the odds at the satellite track. However, their rationale for the host track's probability estimates being more accurate is "common sense" rather than statistical. They argue that bettors at the satellite track have more limited access to information relative to those attending the home track and therefore the probability estimates at the satellite track will be less accurate. However, no data is used to support this assumption. An alternative "common sense" argument is that bettors at the satellite track consist mainly of hard core horse players while many bettors at the host track are casual bettors interested in a day of entertainment. If this argument is correct, the probability estimates at the satellite track will be more accurate than the probability estimates at the host track. This thesis will examine an example of host versus satellite pools to determine which probability estimate is more accurate.

A third example of competing multinomial parameter estimation procedures is meteorological services that provide probability of precipitation (PoP) forecasts. The multinomial reduces to the Bernoulli probability mass function. If there is precipitation on day t, then $x_{1,t} = 1$ and $x_{2,t} = 0$, whereas if there is no precipitation on day t $x_{1,t} = 0$ and $x_{2,t} = 1$. Consumers of meteorological services would naturally like to select the most accurate service. The current literature on PoP forecasts has measures of accuracy. Accuracy in PoP forecasts is generally measured by a quadratic scoring rule developed by Brier (1950). The Brier score BS^* is given by

$$BS^* = (1/T)\sum_{t=1}^{T}(\bar{p}_{1,t} - x_{1,t})^2$$

where in a sample of T PoP forecasts, $\bar{p}_{1,t}$ is the PoP forecast for the t-th occasion and $x_{1,t}$ equals one if there was precipitation on the t-th occasion and zero otherwise. The range of BS^* is between 0 and 1 with 0 representing the best score and 1 the worst.

In weather forecasting BS^{**} denotes the average Brier score for a constant forecast of the climatological probability of precipitation (probability of precipitation based on data from an appropriate historical period). Murphy and Brown (1984) define a skill score, SS, of a forecasting procedure by

$$SS = 100(1 - (BS^*/BS^{**}))$$

SS represents the percentage improvement in the average Brier score for the forecasts of interest over the average Brier score for the corresponding climatological forecasts. However, the drawback of the skill score and Brier scores is that they do not provide any measure of statistical significance. The meteorological service with the lowest Brier score or highest SS in a sample of T forecasts is not necessarily best at a significance level α.

The sport of baseball provides a fourth example for competing probability models. What is the probability of a batter getting a hit in his next official at bat? An obvious answer would be the batter's batting average for that season. However, an alternative answer or procedure would be to use the batter's lifetime batting average. A third procedure would be to use the batter's lifetime average against the particular pitcher he is facing at his next at bat. Thus there are at least three reasonable procedures for estimating the probability. Which of these procedures is best?

A final example of competing multinomial parameter estimation procedures would be financial models that forecast the probability of a financial market being up in a given time period. Larsen and Wozniack(1995) developed a logit model to forecast the one month ahead probabilities that the market return on equities will exceed the return on debt. The goal of this model was to develop a strategy for market timing. However, standard financial approach in modeling the price of a financial security at time T is to assume a lognormal distribution

$$\mathbf{P}_T = P_0 \exp((\mu - .5\sigma^2)T + \sigma \mathbf{Z} T^{.5})$$

where

P_0 = price of the security at time 0,

$\mathbf{P}_T$ = price of security at time T,

T = time in years,

μ = continuously compounded growth rate of security per year,

σ = standard deviation of growth rate,

and

$\mathbf{Z}$ = standard normal variate.

Using the lognormal model, the probability that the market return in 1 month $\mathbf{P}_{1/12}/P_0$ exceeds the return on debt d is equal to

$$\Pr(\mathbf{Z} > (12)^{.5}(\log d - (1/12)(\mu - .5\,\sigma^2))/\sigma).$$

A third procedure would be to assume the probability that the market return in 1 month, $\mathbf{P}_{1/12}/P_0$, exceeds the return on debt, d, is constant from month to month and can be estimated by the historical frequency. This may have implications for investors who use market timing strategies. Investors who use market timing want to be in the market when they believe the market will go up and out of the market when they believe it will go down. Hypothetically, three different models might provide estimates of say, .8, .6, and .4 for the market return exceeding the return on debt. However, if it was established that the model that provided the .8 estimate was selected as the best of the three models, an investor might use that in his or her financial decisions.

The instrument recommended in this thesis for selecting the best multinomial parameter estimation procedure is a variation of the sequential probability ratio test or SPRT. A very succinct description of the SPRT is given by Kendall and Buckland (1960):

> *A sequential test for the hypothesis H_0 against an alternative hypothesis H_1, is due to Wald(1944). At the end of each stage in the sampling the probability ratio p_1/p_0 is computed where the suffixes 0 and 1 refer to the null and alternative hypotheses respectively and p is the (known) probability function of all sample members so far drawn. Then if $B < p_1/p_0 < A$ the sampling is continued another stage. But if $p_1/p_0 \leq B$ the null hypothesis (H_0) is accepted, and if $p_1/p_0 \geq A$ the null hypothesis is rejected and the alternative hypothesis (H_1) is accepted. The two constants A and B are determined by reference to prescribed requirements concerning the two types of errors made in testing hypotheses, the rejection of H_0 when it is true and the acceptance of H_1 when it is false.*

In this thesis the SPRT is slightly modified in order to provide a Bayesian interpretation of the stopping rule.

In my view, SPRT or more generally likelihood ratio tests, are ideal in situations where there is a need to select one of two estimation procedures and the consequence of selecting the incorrect procedure is the same for both procedures. Hypotheses can be set up in the form procedure 1 is correct versus procedure 2 correct. In terms of selection of one of the two procedures, this is a more effective technique than a null hypothesis that states the two procedures are equally good versus an alternative hypothesis that states they are not equally good. In addition, it will be shown in Chapter 2 that this can be generalized to the selection of one of k estimation procedures.

Chapter 2
An SPRT Approach

The structure of the multinomial probability mass function considered

$$f(x_1, x_2, ... x_n) = p_1^{x_1} p_2^{x_2} \ldots p_n^{x_n}$$

$$\text{where } \sum x_i = 1 \text{ and } x_i = 0 \text{ or } 1$$

$$\text{and parameters } p_1, p_2, \ldots, p_n \text{ satisfy } \sum p_i = 1 \text{ and all } p_i \geq 0,$$

results in a very simple likelihood function. Defining a discrete random variable **Y** by

$$\mathbf{Y} = y \text{ if } x_y = 1$$

results in the probability mass function

$$f(y) = p_y \text{ for } y = 1,2, \ldots n.$$

If y_t is the observation corresponding to parameter estimates $p_{1,t}, p_{2,t}, \ldots, p_{n(t),t}$ then, assuming independence, the likelihood of $(y_1, y_2, \ldots, y_m)$, $L(p_{1,t}, p_{2,t}, \ldots, p_{n(t),t} ; y_1, y_2, \ldots, y_m)$, would be given by

$$L(p_{1,t}, p_{2,t}, \ldots, p_{n(t),t} ; y_1, y_2, \ldots, y_m) \stackrel{\Delta}{=} L(y_1, y_2, \ldots, y_m) = f_1(y_1) f_2(y_2) \ldots f_m(y_m) = p_{y_1 1} p_{y_2 2} \cdots p_{y_m m} .$$

Given the simple nature of the likelihood function, Wald's (1947) Sequential Probability Ratio Test (SPRT) is a natural technique for choosing between two multinomial parameter estimation procedures. The SPRT is designed to test a simple hypothesis H_0 against a simple alternative hypothesis H_1. The SPRT procedure in an experiment where the data $(y_1, y_2, \ldots, y_m)$ is obtained sequentially is as follows:

1. Two positive constants A and B are chosen ($B < 1 < A$).

2. At any stage m of the experiment, with data ($y_1, y_2, ..., y_m$), the likelihoods of the data under each of the hypotheses, $L_0(y_1, y_2, ..., y_m)$ and $L_1(y_1, y_2, ..., y_m)$, are calculated, as well as the ratio

$$\lambda_m = \frac{L_1(y_1, y_2, ... y_m)}{L_0(y_1, y_2, ... y_m)} .$$

3. If $B < \lambda_m < A$ the experiment is continued by taking an additional observation y_{m+1}. If $\lambda_m \geq A$ the process is terminated with the rejection of H_0 and the acceptance of H_1. If $\lambda_m \leq B$ the process is terminated with the rejection of H_1 and the acceptance of H_0.

Wald proved that this process terminates provided $z_j = \log(\frac{f_1(y_i)}{f_0(y_i)})$ has a positive variance where f_0 and f_1 are the probability mass or density functions under H_0 and H_1, respectively and the z_j's are independent and identically distributed. The situation here is different since the z_j's are not identically distributed. Therefore, there is no guarantee that the process will terminate. In order to guarantee termination in the comparison of parameter estimation procedures, a maximum sample size M is required. If the maximum sample size M is reached the test is inconclusive.

The constants A and B are usually determined by the prescribed requirements of Type I and Type II errors.

Define

α = probability of rejecting H_0 given that H_0 is true (the probability of a Type I error)

and

β= probability of accepting H_0 given that H_1 is true (the probability of a Type II error).

Wald(1947) established the relationship between (α,β) and (A,B). He defined a sample ($y_1, y_2, ..., y_n$), as type 0 if

$$B < \lambda_m < A \text{ for } m=1,2,...,n-1$$

and

$$\lambda_n \leq B.$$

Similarly, a sample $(y_1, y_2, ..., y_n)$, is defined as type 1 if

$$B < \lambda_m < A \text{ for } m=1,2,...,n-1$$

and

$$\lambda_n \geq A.$$

For any type 1 sample $(y_1, y_2, ..., y_n)$, the probability of obtaining such a sample is at least A times as large under H_1 as under H_0.

That is

$$L_1(y_1, y_2, ..., y_n) \geq A\, L_0(y_1, y_2, ..., y_n).$$

Therefore

$$\int_{Type1} L_1(y_1, y_2, ..., y_n) \geq A \int_{Type1} L_0(y_1, y_2, ..., y_n).$$

Or

$$1-\beta \geq A\alpha.$$

Thus an upper bound for A is $\frac{1-\beta}{\alpha}$.

Similarly, for any type 0 sample $(y_1, y_2, ..., y_n)$, the probability of obtaining such a sample is at least B times as large under H_0 as under H_1.

That is

$$B\, L_0(y_1, y_2, ..., y_n) \geq L_1(y_1, y_2, ..., y_n).$$

Therefore

$$B \int_{Type0} L_0(y_1, y_2, ..., y_n) \geq \int_{Type0} L_1(y_1, y_2, ..., y_n).$$

Or

$$B(1-\alpha) \geq \beta.$$

Thus a lower bound for B is $\frac{\beta}{1-\alpha}$.

An actual determination of A and B is a difficult computational problem. Therefore, A and B are almost always approximated by $\frac{1-\beta}{\alpha}$ and $\frac{\beta}{1-\alpha}$ respectively. This will mean that when the SPRT terminates the probability of a Type I error is at most α and the probability of a Type II error is at most β.

If the underlying problem is to select the better of two multinomial parameter estimation models where the consequence of selecting the incorrect procedure is the same for both procedures, it is natural to select $\alpha = \beta$. The resulting hypothesis test concerns

H_1: Parameter estimation procedure 1 is correct

versus

H_2: Parameter estimation procedure 2 is correct.

In this context α is the probability of accepting an hypothesis when the alternative hypothesis is correct. (I have decided to call the hypotheses H_1 and H_2 instead of terms null hypothesis H_0 and an alternative hypothesis H_1 or H_a. The objective is to select the better of two parameter estimation methods and in most situations there is no prior reason to prefer one method over the other).

SPRT-like algorithm for selecting the better of two estimation procedures:

1. A constant α ($0 < \alpha < 1$) and a maximum sample size M are chosen.

2. At any stage m of the experiment, with data ($y_1, y_2, \ldots, y_m$), the ratio

$$\lambda_m = \frac{L_{i*}(y_1, y_2, \ldots, y_m)}{L_j(y_1, y_2, \ldots, y_m)} \text{ for } j \neq i*$$

is calculated where $L_1 = L_1(y_1, y_2, \ldots, y_m)$ and $L_2 = L_2(y_1, y_2, \ldots, y_m)$ are the likelihoods of the data under each of the hypotheses and

$$L_{i*} = L_*(y_1, y_2, \ldots, y_m) = \text{Maximum}\{L_1(y_1, y_2, \ldots, y_m), L_2(y_1, y_2, \ldots, y_m)\}.$$

(Note: There is no need to directly compute L_1 and L_2. These numbers will become extremely small and subject to round-off error. However, it is easy to store L_1/L_2 and update this number with each new data point. Comparing the ratio with 1 will determine which one is L_{i*}.).

3. If $\lambda_m \geq (1-\alpha)/\alpha$. the experiment is terminated with the acceptance of H_{i*} and the rejection of H_j, $j \neq i*$. If $\lambda_m < (1-\alpha)/\alpha$ and $m < M$ the experiment is continued by taking an additional observation y_{m+1}. Finally, if $\lambda_m < (1-\alpha)/\alpha$ and $m = M$, the experiment is terminated with an inconclusive result.

As an illustration of the SPRT-like algorithm, consider the following example: An urn contain 50 white balls and 50 black balls. 10 balls are randomly selected without replacement from the urn. The number of white balls, n_w, and the number of black balls, n_b, in this sample of 10 are recorded. A ball is then randomly selected from this sample of 10.

Let p_w and p_b be the probabilities that this ball is white and black respectively. Consider the following two estimation procedures:

Estimation procedure 1: $p_w = n_w/10$ and $p_b = n_b/10$.

Estimation procedure 2: $p_w = .5$ and $p_b = .5$.

Estimation procedure 1 is the correct estimation procedure. Suppose in the first observation, $n_w = 7$, $n_b = 3$ and a white ball is chosen. Then $L_1(W) = .7$, $L_2(W) = .5$, $i^* = 1$ and $\lambda_1 = (.7/.5)$. Suppose in the second observation $n_w = 4$, $n_b = 6$ and a white ball is again chosen. Then $L_1(W,W) = (.7)(.4)$ or .28, $L_2(W,W) = (.5)(.5)$ or .25 , $i^* = 1$ and $\lambda_2 = (.28)/(.25)$. Suppose in the third observation $n_w = 9$, $n_b = 1$ and the black ball is chosen. Then $L_1(W,W,B) = (.7)(.4)(.1)$ or .028, $L_2(W,W,B) = (.5)(.5)(.5) = .125$, $i^* = 2$ and $\lambda_3 = (.125)/(.028)$. The experiment will continue until $\lambda_m \geq (1-\alpha)/\alpha$.

The SPRT-like algorithm with $\alpha = .05$ and $M = 500$ was simulated for this example with 1000 replications using @Risk 3.5E with a random number seed of 1. The SPRT-like algorithm chose estimation procedure 1 the better of two procedures 962 times and chose estimation procedure 2 the better procedure 38 times. There were no inconclusive results. The mean sample size was 60.41 with a maximum sample size of 371 and a minimum sample size of 5.

A view of the probabilities involved in this SPRT-like approach can be seen through a Bayesian perspective. Since, the underlying problem is to select the better of two parameter selection procedures, it is reasonable to assume that their is no prior preference of one procedure over the other. This would correspond to a uniform prior. That is prior to the collection of data, the probability that procedure 1 is correct is assumed to be equal to the probability of procedure 2 is correct. At the conclusion of the experiment the posterior probability of the procedure with the maximum likelihood being correct would be

$$L_{i^*} / (L_1 + L_2)$$

or

$$\frac{\lambda_m}{1+\lambda_m}.$$

However, since $\lambda_m \geq (1-\alpha)/\alpha$,

$$\frac{\lambda_m}{1+\lambda_m} = \frac{1}{\frac{1}{\lambda_m}+1} \geq \frac{1}{\frac{\alpha}{1-\alpha}+1} = 1-\alpha.$$

Hence the probability that H_{i^*} is true is at least $1-\alpha$ (assuming either H_1 or H_2 is true) while the probability of rejecting the true hypothesis is below α. That is estimation procedure i^* is selected the better of estimation procedures {1, 2} at a significance level α.

This Bayesian perspective provides an attractive interpretation in a non-sequential situation. By a non-sequential situation I mean a situation where n data points (y_1, y_2, ..., y_n) are already available. If $\lambda_n > (1-\alpha) / \alpha$ then procedure i^* would be selected the better procedure at a significance level α. Otherwise, the test is inconclusive and we do not have sufficient evidence to conclude that one of the procedures is better than the other procedure at a significance level α.

This Bayesian view easily generalizes to the selection of the best of k parameter selection procedures. The underlying hypothesis test concerns:

H_1: Parameter estimation procedure 1 is correct

versus

H_2: Parameter estimation procedure 2 is correct

versus

.

.

versus

H_k: Parameter estimation procedure k is correct.

With data $(y_1, y_2, ..., y_m)$, let $L_j = L_j(y_1, y_2, ..., y_m)$ be the likelihood given that H_j is true. Define L_{i^*} = maximum { $L_1, L_2, ..., L_k$ }. Assuming, a uniform prior, the probability that the hypothesis i* with the maximum likelihood is true (assuming one of the k hypotheses is true) is

$$\frac{L_{i^*}}{\sum_{j=1}^{k} L_j}.$$

Alternatively, defining $\lambda_{jm} = L_{i^*}(y_1, y_2, ..., y_m) / L_j(y_1, y_2, ..., y_m)$ for $j \neq i^*$, the probability that the hypothesis i* with the maximum likelihood is true (assuming one of the k hypotheses is true) can be written as

$$\frac{1}{1+\sum_{j\neq i^*}(1/\lambda_{jm})}.$$

The number $1 - \dfrac{1}{1+\sum_{j \neq i^*}(1/\lambda_{jm})}$ can be viewed as the p-value in the test of i^* being the correct estimation procedures versus one of the other k-1 estimation procedures being correct.

If the goal is to select the procedure i* with the maximum likelihood provided the probability that hypothesis i* is true is at least $1 - \alpha$, then this is equivalent to requiring

$$1 + \sum_{j \neq i^*}(1/\lambda_{jm}) \leq \frac{1}{1-\alpha} .$$

The resulting SPRT like algorithm for selecting the best of k estimation procedures would be as follows:

SPRT-like algorithm for selecting the best of k estimation procedures:

1. A constant α ($0 < \alpha < 1$) and a maximum sample size M are chosen.

2. At any stage m of the experiment, with data ($y_1, y_2, ..., y_m$) calculate the ratios

$$\lambda_{jm} = \frac{L_{i^*}(y_1, y_2, \ldots y_m)}{L_j(y_1, y_2, \ldots y_m)} \text{ for } j \neq i^*$$

where $L_j = L_j(y_1, y_2, ..., y_m)$ for j = 1,2,...k are the likelihoods of the data under each of the hypotheses and $L_{i^*} = L_{i^*}(y_1, y_2, ..., y_m) = \text{Maximum } \{L_j(y_1, y_2, ..., y_m) \mid j = 1,2,...k\}$.

(Note: There is no need to directly compute $L_1, L_2, ... L_k$. These numbers will become extremely small and subject to round-off error. However, it is easy to store L_1/L_2, L_2/L_3,.. L_{k-1}/L_k. and update these number with each new data point. Comparing these ratios with 1 will determine which one is L_{i^*}.)

3. If $\phi_m = 1 + \sum_{j \neq i^*} (1/\lambda_{jm}) \leq 1/(1-\alpha)$ the experiment is terminated with the acceptance of H_{i^*}.

If $\phi_m = 1 + \sum_{j \neq i^*} (1/\lambda_{jm}) > 1/(1-\alpha)$ and $m < M$ the experiment is continued by taking an additional observation y_{m+1}. If $\phi_m = 1 + \sum_{j \neq i^*} (1/\lambda_{jm}) > 1/(1-\alpha)$ and $m = M$ the experiment is terminated with an inconclusive result.

The acceptance of H_{i^*} means procedure i^* is selected the best of procedures {1, 2,...k} at a significance level α. Baum and Veeravalli (1994) used the same idea in their paper on sequential multihypothesis testing.

One of the consequences of this SPRT-like approach is that an estimation procedure that gives a parameter estimate of 0 or 1 and is "wrong" will never be chosen as the best procedure. By "wrong I mean giving a parameter estimate of 1 to an outcome that did not occur or giving a parameter estimate of 0 to an outcome that did occur. This may seem "unfair" particularly if its Brier score (which in effect captures mean square error) is quite good. However, I would argue that a model that gives a parameter estimate of 0 or 1 and can be "wrong" should never be viewed as a reasonable model. For example, if an investor believed a particular model was correct and it gives a probability of say 0.97 that the market will be up, the investor may or not invest depending upon his or her risk attitude. However, if the investor believed a particular model was correct and it gives a probability of 1 that the market will be up, the investor should logically and aggressively invest in the market regardless of risk attitude. Therefore, I believe any parameter estimation procedure that gives estimates of 0 or 1 and can be "wrong" should never be chosen as the best procedure. The SPRT-like approach is consistent with this philosophy.

A final point with respect to the issues in this chapter is about the assumption of a uniform prior. There could be a situation where the decision maker had prior views on which estimation procedure is correct. However, the SPRT-like algorithm can easily be modified to handle this situation. For example, in choosing between two estimation procedures, suppose the decision maker believed prior to the collection of any data that the probability of procedure 1 being correct was p and the probability of procedure 2 being correct was 1-p. Presumably $\alpha < p < 1$-

α. After collecting data $(y_1, y_2, \ldots y_m)$, the revised probability of procedure 1 being correct would be $\frac{pL_1(y_1, \ldots y_m)}{pL_1(y_1, \ldots y_m) + (1-p)L_2(y_1, \ldots y_m)}$. Defining $\delta_m = L_1(y_1, y_2, \ldots y_m)/L_2(y_1, y_2, \ldots y_m)$ the revised probability could be written as $p\,\delta_m / (p\delta_m + 1 - p)$. If this number was greater than or equal to $1-\alpha$ procedure 1 would be selected while if this number was less than or equal to α procedure 2 would be selected.

Chapter 3

A Probability Model for the Racetrack

A. The Pari-Mutuel Market

Pari-mutuel betting is the dominant form of wagering in North American thoroughbred racing. Pari-mutuel, originally a French word for "mutual stake" means that winners divide the total amount that is bet (less any deductions) in proportion to the amount each winner wagered. When a bettor makes a wager, the betting facility or track merely acts as the broker for the transaction. For every dollar wagered at the track, the track deducts a commission. The track, the state and the competitors (in the form of purses) share this commission. The commission rate varies from track to track. In North America the commission rate varies from as low as 15% and to as high as 30% but the most common rate is approximately 17%. Thus between 70% to 85% of all money bet is returned to winning bettors.

Bettors have a choice of a number of betting pools for each race. In a pool the bettor is trying to select a winning entry or a winning combination of entries. An entry is usually one horse, but on occasion because of common ownership, two or more horses are "coupled" into a betting entry. A given race may offer:

- Win betting (wagering on an entry to come in first)
- Place betting (wagering on an entry to come in first or second)
- Show betting (wagering on an entry to come in either first, second, or third)
- Quinella betting (wagering on two entries to come in first and second in any order)
- Exacta betting (wagering on two entry to come in first and second in correct order)
- Trifecta or triacta betting (wagering on three entries to come first, second, and third in correct order)

and

- Daily double betting (wagering on the winners of two consecutive races).

Each of these pools form a separate market.

The relative amounts bet on a winning combination and the track's commissions determine the prices paid on winning bets. For example, in the win pool let

W_i = the total amount bet on entry i to win;

$W = \sum_i W_i$ = the win pool at the track;

Q_W = the track's payback proportion on win bets.

The payoff to the bettor of a dollar bet on entry i is 0 if entry i loses. If entry i wins, one first calculates $Q_W\ W/W_i$ and then rounds this number down to the nearest 5 or 10 cents (called breakage). All Canadian tracks round down to the nearest 5 cents while almost all U.S. tracks round down to the nearest 10 cents. For example, if $Q_w\ W/W_i$ equals 4.28, the payoff per dollar bet would be \$4.20 with 10 cent breakage and \$4.25 with 5 cent breakage. The one exception to this rule occurs if the number $Q_w\ W/W_i$ is below 1.05 or 1.10. All jurisdictions in North America have a minimum payoff on a winning bet. Races run in the state of Kentucky have a minimum payoff of \$1.10 per \$1 bet while all other states and provinces have a minimum payoff of \$1.05.

B. The Efficient Market Theory

A theory in research related to the racetrack, is that the win pool at the racetrack is an efficient market. This means an entry's probability of winning a race is equal to the fraction of the win pool bet on that entry (the probability of betting entry j winning race t = $p_{jt} = W_{jt}/W_t$ where W_{jt} is the amount of money bet on j to win while W_t is the total amount of money in the win pool in race t). The rationale for this theory is that data collected over thousands of races indicate that entries that have fraction p of the win pool, as a group win about fraction p of the races. For example Fabricand (1979) provides the following table for the theoretical probability (fraction of the win pool) versus actual probability (fraction of horses that win) for United States thoroughbred tracks for the years 1955 to 1962 (10,000 races).

Table 1

U.S. Tracks 1955-1962

Number of Horses	Average Theoretical Probability	Actual Frequency of Winning
25,044	.025	.014
7,041	.047	.040
12,007	.065	.060
3,617	.081	.082
3,990	.090	.082
4,665	.100	.099
5,154	.114	.110
5,586	.132	.123
3,129	.148	.155
3,296	.162	.161
3,652	.180	.186
3,807	.201	.209
3,623	.228	.230
3,223	.263	.289
1,051	.293	.309
954	.315	.355
874	.341	.379
789	.371	.403
615	.406	.470
470	.449	.513
295	.502	.553
129	.569	.713

Similarly Hausch, Ziemba and Rubinstein (1981) provide the following table for Exhibition Park in 1979:

Table 2

Exhibition Park 1979

Theoretical probability of winning	Number of horses	Average theoretical probability	Actual frequency of winning	Estimated standard error
.000-.025	540	.019	.016	.005
.026-.050	1498	.037	.036	.005
.051-.100	2658	.073	.079	.005
.101-.150	1772	.123	.126	.008
.151-.200	1199	.172	.156	.010
.201-.250	646	.223	.227	.016
.251-.300	341	.272	.263	.024
.301-.350	199	.323	.306	.033
.351-.400	101	.373	.415	.049
.401-.1.000	83	.450	.469	.055

Additional details of this efficient market theory or model can be found in Fabricand [1979], Hausch, Ziemba and Rubinstein [1981], Snyder [1978] and Ziemba and Hausch [1984]. It should be noted that these authors acknowledge a longshot-favorite bias. That is heavy favorites (entries with over 35% of the win pool) tend to be underbet by the public while longshots (horses with less than 5% of the win pool) tend to be overbet by the public. The reason for this may be the risk seeking behavior of bettors at a racetrack. However, in most situations entries that have fraction p of the win pool, as a group win approximately fraction p of their races.

While it might appear reasonable to assign a probability of p to an entry with a fraction p of the win pool, this may not necessarily be the most desirable probability assignment. An analogy would be assigning probabilities for snow on a given day in Winnipeg. If, for example, historical records indicates that Winnipeg receives snow on 10% of all days, assigning a probability of .10 for snow on July 1 is unlikely to be the best probability estimate. Among the group of horses that have say 25% of the win pool it is possible that 50% of them have a .50 probability of winning while the remaining 50% have no chance of winning. This group as a whole will win 25% of their races but assigning a probability of .25 to all of them is again not the best probability estimate.

Thus this efficient market model for the win pool at the racetrack may not be the most appropriate model. However, it is still reasonable to compare any other possible probability model with the efficient market model since a necessary condition for the success of a betting model in pari-mutuel betting is that it produces more accurate probabilities than the efficient market model.

C. Previous Racetrack Research

There has been a considerable amount of previous research on racetrack betting issues. This previous research can be partitioned into five broad areas. One area is the question of market efficiency in the win pool. In finance a market is considered efficient if current security prices reflect all relevant information. In an efficient market an expert should not be able to achieve higher than average returns with regularity. Snyder (1978) investigated whether bets at different odds level yield the same average return. Ali (1979) investigated whether independently derived bets with identical probabilities of winning have the same odds statistically. Asch, Malkiel and Quandt (1984), Hausch, Lo and Ziemba (1994), Hausch, Ziemba and Rubinstein (1981), Fabricant(1965) and Figlewski (1979) have also contributed to this issue.

A second area is various arbitrage or risk-free hedge strategies at the racetrack. This is an attempt to create a wager that will win money regardless of the outcome of the race. Hausch and Ziemba (*INTERFACES* 1990) give an example where a profit in show betting is guaranteed. Hausch and Ziemba (*Journal of Business* 1990) show how under certain conditions profits can be guaranteed in cross-track betting. Cross-track betting involves simulcasting of races with separate pools at different tracks. However, Rosenbloom (1992) found certain flaws with these strategies.

A third area is the question of how a person should bet in the win pool given the probability of each horse winning and the odds on each horse. There has been a number of approaches in the literature to this problem. Isaacs (1953) gives a solution under the assumption that bettor is risk neutral with infinite wealth and is able to be the last bettor in the race. The reason for this last restriction is that in the pari-mutuel betting system the odds are affected by any bet. A more realistic scenario is that the bettor is risk averse with a finite amount of betting wealth and that the bets are not sufficiently large to affect the odds. Rosner (1975) provides a solution to this problem under the assumption that the decision maker's attitude towards risk can be modeled by a logarithm utility function. However, both of these models assume that the bettor knows the exact or true probability of each horse winning the race. Bolton and Chapman (1986) point out that these models may not work very well in practice if only an estimate of the exact probability is available. They suggest modifications to the Rosner model when dealing with approximate probabilities.

A fourth area of research is how to bet in alternative pools such as the place pool, the show pool and exacta pool. Hausch, Ziemba and Rubinstein (1981) presents a model with a positive expectation for betting in place and show pools. Ziemba and Hausch (1987) provides a model with a positive expectation for betting in exacta, quinella and trifecta pools. Both these models require an estimate of the probability of an entry winning the race. The estimate that they suggest for the probability of an entry winning is the fraction of the win pool bet on the entry.

A fifth area of research is how to estimate the probability of an entry winning a race. The only published work in this area is by Bolton and Chapman (1986) who use a multinomial logit model for estimating the probabilities. Their model assumed that the random utility of horse j, U_j, can be written as

$$U_j = V_j + \varepsilon_j$$

where V_j is a deterministic component and ε_j is a random component that reflects the measurement error in the modeling process. If there are N horses in a race, the probability of horse j* winning the race can be written as

$$P_{j^*} = \text{Prob}\,(U_{j^*} \geq U_j,\ j = 1,2,...N).$$

By assuming the error terms ε_j are identically and independently distributed according to a double exponential distribution

$$\text{Prob}\,(\varepsilon_j \leq \varepsilon) = \exp\,(-\exp(-\varepsilon))$$

it can be shown that

$$P_{j^*} = \frac{e^{V_{j^*}}}{\sum_{j=1}^{N} e^{V_j}} \quad \text{for } j^* = 1,2,...N.$$

The V_j's are obtained through a logit model. The explanatory variables for this logit model were chosen because of the availability of data through a newspaper called the *Daily Racing Form*. The *Daily Racing Form* contains data about each horse in every race at a particular track each day as well as data on jockeys. The *Daily Racing Form* contains data on horses such as breeding, post position, weight carried, age, gender, jockey, trainer, owner, workouts, past record, earnings and past performances of the horse. Past performances refers to a detailed description of typically the last ten races run by the horse. This past performance data would include dates, distances, previous jockeys, conditions of the race, fractional times and final times.

Among the explanatory variables chosen in the Bolton-Chapman logit model were the horse's winning percentage over the last two years, the average speed rating over the last four races, the speed rating in the horses last race, the earnings per race in the past year, the post position, the weight carried by the horse, the jockey's winning percentage and number of jockey wins.

Multinomial logit models are often used in marketing research to forecast the probability of a consumer choosing a particular brand j* among competing brands 1,2,...n. Here it is being used to forecast the probability of nature "choosing" horse j* as the winner among competing horses 1,2,...n.

The Bolton-Chapman multinomial logit model is a very clever and innovative approach to estimating the probability of a horse winning. However, in my view there are two flaws with this model. One is the assumption that the ε_j 's are independent. As an example consider the role of post position. Post position was one of the explanatory variables in the Bolton-Chapman model. In the logit model the coefficient of post position is negative. This reflects the fact that horses with inside post positions tend to win a higher percentage of races than horses with outside position. However track conditions can vary dramatically from day to day because of weather and soil conditions. There are days when the track is very biased. This means that one portion of the track (the inside or the outside) is much faster than the other portion. On days when the inside is faster, horses with inside post position will have a higher true probability of winning than what is predicted by the model while conversely horses on the outside will have a lower probability. Similarly on days when the outside is faster, horse with inside post position will have a lower true probability of winning then what is predicted by the model while horses on the outside will have a higher probability. Thus the assumption that the ε_j 's are independent is questionable.

In my view the second flaw in the Bolton-Chapman model involves the issue of time. Since the fastest horse will likely win the race, one might expect previous finishing times would be the major tool in predicting which horse will win. Typically it is not. There are number of reasons why previous final times are a questionable handicapping tool. Races are run at a variety of distances. Shorter races (often called sprints) are run at 4.5 furlongs (a furlong being 1/8 of a mile), 5 furlongs, 5.5 furlongs, 6 furlongs, 6.5 furlongs and 7 furlongs. Longer races (often called routes) are run at 1 mile, 1 mile and 70 yards, 1 and 1/16 miles, 1 and 1/8 miles, 1 and 3/16 miles, 1 and 1/4 miles, 1 and 3/8 miles and 1 and 1/2 miles. Different tracks because of their composition and configuration will result in quite different times for horses of similar ability. For example 1:09 (1 minute and 9 seconds) would be considered an exceptionally fast time for a 6 furlong race at Belmont (a major track in New York) and an average time for a 6 furlong race at Santa Anita (a major track in California). Even at the same track, times can change dramatically from day to day because of moisture, temperature, track maintenance, and wind conditions.

The Bolton-Chapman model used a surrogate for time from the *Daily Racing Form* called speed ratings. In a speed rating a horse's final time for a race is compared to the track's record at that distance. The track record is given a speed rating of 100 while one point is subtracted for every one-fifth of a second slower than the track record. For example if the track record for 6 furlongs is 1:08 and a horse runs a 6 furlong race in 1:11, the horse receives a speed rating of 85. In the Bolton-Chapman model average speed rating accounted for the most variation in the model.

Although speed ratings take into account the fact that some tracks are intrinsically faster than others, there are three major drawbacks to speed ratings. One, the speed rating does not take into account the quality of horses running at a particular track. The track record for a 12 furlong race at Belmont Park was set by a horse called Secretariat (arguably the greatest horse of all time). The track record for a 12 furlong race at Assiniboia Downs (a minor track in Winnipeg) was set by a horse called Baron Hudec (a horse worth about $10,000). A horse running two seconds slower than Secretariat's track record of 2:24 at Belmont would receive a speed rating of 90 while a horse running two seconds slower than Baron Hudec's track record of 2:32 at Assiniboia Downs would also receive a speed rating of 90. However, the performance of the horse at Belmont is almost certainly much superior to the performance of the horse at Assiniboia Downs.

A second drawback is that speed rating is not adjusted for distance. It is much more common to be within 2 seconds of the track record at 6 furlongs than it is to be within 2 seconds of the track record at 12 furlongs. Thus a speed rating of 90 at 12 furlongs is almost always a superior performance than a speed rating of 90 at 6 at 6 furlongs.

Finally, as pointed out earlier, times can change dramatically from day to day because of moisture, temperature, track maintenance, and wind conditions. Thus one day, two seconds off the track record can be a tremendous performance while another day two seconds off the track record may be a mediocre performance. Yet both will receive a speed rating of 90.

D. Beyer Speed Numbers

Starting in 1992, the *Daily Racing Form* began providing the previous 10 Beyer speed numbers of all horses in a race. In explaining the use of Beyer speed numbers the *Daily Racing Form* states:

> *Every performance by every horse in North America is assigned a number which reflects the time of the race and the inherent speed of the track over which it is run, and permits easy comparison of efforts at different distances. A horse who earns a 90 has run faster than one who runs an 80.*

Beyer speed numbers are named after the popular racing columnist for the *Washington Post*, Andrew Beyer. A Beyer speed number is a performance number or a speed rating for a horse which takes into account the quality of the horses at a particular track, the track conditions and the distance. Thus Beyer speed numbers try to correct for the deficiencies in speed ratings.

Beyer speed numbers vary from 0 to about 130. They are reported as integers. The higher the number the better the performance. Beyer speed numbers are calculated by the *Daily Racing Form* using typical winning performances of certain classes of horses. Although neither Andrew Beyer nor the *Daily Racing Form* have ever provided the complete procedure or algorithm for calculation of the Beyer speed numbers, Beyer in his books has described a general approach. As a starting point for calculation of Beyer speed numbers, the average winning performance of races restricted to horses valued between $10,000 and $14,000 (U.S.) is arbitrarily assigned as 85. As an approximate illustration of the logic behind the calculations of Beyer speed ratings, suppose on a particular race day there are three 6 furlong races restricted to horses valued between $10,000 to $14,000 and the winning times are 1:10.80, 1:11 and 1:11.20. The horse that finished at 1:11 would likely be assigned by the *Daily Racing*

Form a Beyer speed number of 85, the horse that finished in 1:10.8 would be assigned a higher Beyer number (likely 87), and the horse that finished at 1:11.20 would be assigned a lower Beyer speed number (likely 83). All other horses in 6 furlong races that day would be assigned Beyer speed numbers in a similar manner based on their final time. Horses that ran a different distance are assigned Beyer speed numbers based on charts for that track which convert times at that distance to the equivalent performance for 6 furlongs.

Thus while the details of the calculation of Beyer speed numbers are likely beyond the grasp of most racetrack bettors, the final Beyer speed number are relatively easy to interpret. In a given race the horse who obtains the highest Beyer speed number will almost always win the race (the rare exception is a horse that finishes first and is disqualified for interfering with other horses). Therefore, forecasting the winner of a race is almost equivalent to forecasting which horse will have the highest Beyer speed number.

A natural statistical question is whether past Beyer speed numbers can be used to forecast future Beyer speed numbers. In particular is it possible to obtain a reasonable estimate of the statistical distribution of a horse's next Beyer speed number? If so, will it be possible to use this to estimate the probability of a horse having the highest Beyer speed number in a race? In the remainder of this chapter I will show that the answers to these questions are positive.

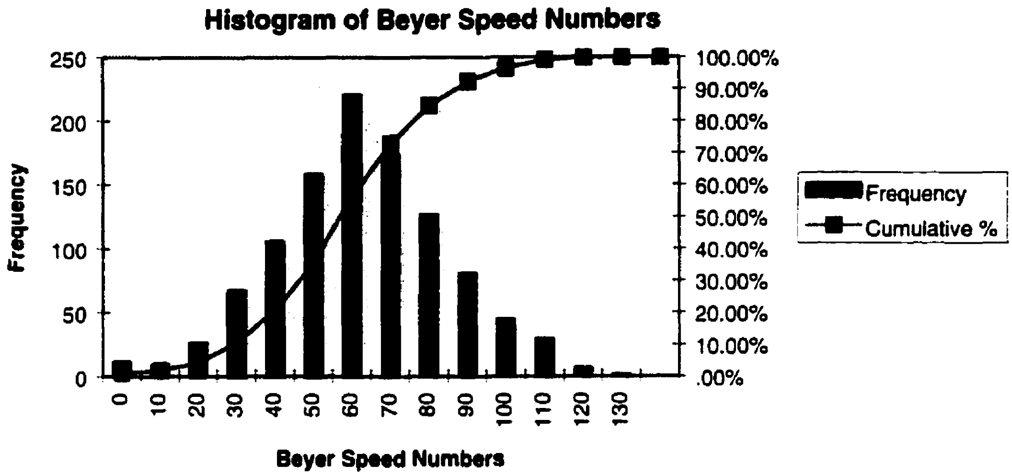
E. A Probability Model for Beyer Speed Numbers

In North America horses can begin racing at the age of 2. However, horses are not fully grown and mature until at least the age of 4. This is usually reflected in their Beyer speed numbers. Beyer (1993) estimates that a horse's Beyer speed numbers usually improve at a rate one to one and a half points a month until the end of their three year old career. After that their Beyer numbers usually do not have any obvious long term trend. North American races are run on two surfaces, dirt and grass. While Beyer speed numbers are calculated for both surfaces, most horses perform better on one of these two surfaces. Therefore, most horses race either almost exclusively on dirt or almost exclusively on grass. The great majority of all North American races are on dirt.

A data base of past performances of thoroughbreds was obtained from the *Daily Racing Form*. The data base consisted of 1056 horses who were racing in either the fall of 1996 and the spring of 1997 at Woodbine, Hollywood Park, Belmont, Assiniboia Downs or Hastings Park. These were all four year olds or older who were racing on dirt and had raced at least 10 times. While not a random sample of all four year or older horses this is a fairly representative sample including horses at the best tracks (Belmont, Hollywood Park), middle level tracks (Woodbine) and lower level tracks (Hastings and Assiniboia Downs).

The data collected on each of these horses were the last 10 Beyer speed numbers as well as an indicator of whether the horse had been laid off before its last race. The *Daily Racing Form* defines a layoff as 45 days without a race. Typically horses during a racing meet will race every two to four weeks. Hence a layoff of 45 days or more is somewhat atypical and may indicate an injury or some physical problem. A histogram of the last Beyer speed numbers of these 1056 horses is as follows:

Figure 1



The Beyer speed numbers vary between 0 to about 130. They are calculated as real numbers but are rounded off to integers in the *Daily Racing Form.* It is reasonable to model Beyer numbers greater than 0 as a continuous random variable. However, a Beyer number recorded as a 0 can represent a number of events. A horse will receive a Beyer number of 0 if he or she ran an unusually slow time or did not finish the race (the jockey fell off or there was an accident or the horse bled or the horse was injured). For older horses (four years old and above) a 0 Beyer speed number usually means the horse did not finish the race. Therefore, the probability of a horse getting a 0 Beyer number should be modeled as non-zero and may be dependent on whether or not the horse has been laid off before its last race. Horses are often laid off because of physical problems.

Of the 1056 horses in the data base, 247 horses had been laid off before their last race while 809 had not. Of the 247 horses who had been laid off, 7 had a Beyer number of 0 in their last race, while of the 809 horses that had not been laid off, only 4 received a Beyer number of 0.

Defining p_0 as the probability of a horse without a layoff getting a 0 Beyer number and p_1 as the probability of a horse with a layoff getting a 0 Beyer number the two-sample hypothesis test of

$$H_0: p_0 = p_1$$

versus

$$H_a: p_0 \neq p_1$$

results in a Z-score of 3.17 and a p-value of .0015. Hence, for horses that have not been laid off , 4/809 (or .005) would be an estimate of the probability of a 0 Beyer number. For horse that have been laid off, 7/247 (or .028) would be an estimate of the probability of a 0 Beyer number.

Removing these eleven horses with 0 Beyer numbers in their last race from our sample means that we have 1045 horses left. For each of these horses we have the following data:

$$(L_{10}, B_{10}, B_9, \ldots B_1)$$

where $L_{10} = 0$ if horse was not laid off before its last race, $L_{10} = 1$ if the horse has been laid off, and B_{10}, B_9,...B_1 the last 10 Beyer speed numbers with B_{10} being the most recent, B_9 being the second most recent etc. Thus the statistical question becomes can L_{10}, B_9, B_8, ...B_1 be used to predict B_{10} ?

A table of Pearson correlations and corresponding p-values are as follows:

Pearson Correlations

	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
B1	1.000	0.744	0.524	0.694	0.660	0.655	0.623	0.604	0.615	0.622
B2		1.000	0.524	0.682	0.679	0.646	0.610	0.606	0.628	0.611
B3			1.000	0.497	0.492	0.512	0.479	0.473	0.474	0.466
B4				1.000	0.698	0.679	0.643	0.642	0.652	0.649
B5					1.000	0.718	0.689	0.645	0.674	0.670
B6						1.000	0.715	0.663	0.676	0.690
B7							1.000	0.693	0.710	0.694
B8								1.000	0.721	0.686
B9									1.000	0.750
B10										1.000

Pearson p-values

	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
B1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B2		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B3			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B4				0.000	0.000	0.000	0.000	0.000	0.000	0.000
B5					0.000	0.000	0.000	0.000	0.000	0.000
B6						0.000	0.000	0.000	0.000	0.000
B7							0.000	0.000	0.000	0.000
B8								0.000	0.000	0.000
B9									0.000	0.000
B10										0.000

An examination of this table reveals that B_{10} is positively correlated with B_9, B_8,... and B_1. In addition, all these correlations are all significant at a .01 level.

It also appears from the correlation matrix that there is a stronger correlation between B_{10} and the more recent data (B_9, B_8, B_7 and B_6) than B_{10} and the less recent data (B_5, B_4, B_3, B_2, and B_1). This suggests using multiple regression to predict B_{10}. B_9, B_8, B_7, and B_6 could be explanatory variables. While it is also possible to use B_5, B_4, B_3, B_2, and B_1 as explanatory variables it was decided to use M, the median of B_5, B_4, B_3, B_2, and B_1 as an explanatory variable. The variable M would capture less recent performance and would be less sensitive to one unusually good or unusually bad past performance. The Pearson correlation between M and B_{10} was .728 and is significantly different from 0 at a .01 level. I also decided to use L_{10} as an explanatory indicator variable. Whether a horse has raced recently my affect his or her performance. In addition, there is likely going to be an interaction effect between L_{10} and the other explanatory variables. Handicapping logic suggests that if the horse has raced recently ($L_{10} = 0$), the most recent race (B_9) will be the most important predictor of future performance. However, if a horse has not raced recently ($L_{10} = 1$), the relative importance of the most recent data is diminished. Thus L_{10}*B_9, L_{10}*B_8, L_{10}*B_7, L_{10}*B_6 and L_{10}*M would be explanatory variables in the multiple regression as well.

The results of the multiple regression are as follows:

Regression Statistics	
Multiple R	0.8221
R Square	0.6758
Adjusted R Square	0.6724
Standard Error	12.1667
Observations	1045

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	11	318811.21	28982.84	195.79	1.10E-243
Residual	1033	152913.89	148.03		
Total	1044	471725.10			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3.11	1.51	2.06	4.01E-02	0.14	6.07
L_{10}	-9.52	3.33	-2.86	4.37E-03	-16.06	-2.98
$L_{10}*M$	0.18	0.09	2.02	4.40E-02	0.00	0.36
$L_{10}*B_6$	-0.03	0.07	-0.47	6.36E-01	-0.17	0.11
$L_{10}*B_7$	0.07	0.08	0.95	3.44E-01	-0.08	0.23
$L_{10}*B_8$	0.08	0.07	1.06	2.90E-01	-0.07	0.22
$L_{10}*B_9$	-0.23	0.07	-3.20	1.39E-03	-0.37	-0.09
M	0.18	0.04	4.23	2.59E-05	0.10	0.27
B_6	0.16	0.04	4.42	1.07E-05	0.09	0.22
B_7	0.13	0.03	3.70	2.29E-04	0.06	0.19
B_8	0.11	0.03	3.22	1.30E-03	0.04	0.17
B_9	0.38	0.03	11.15	2.54E-27	0.31	0.44

Not all the explanatory variables are significant. Dropping the explanatory variables, one by one using backwards elimination results in the following reduced regression model:

Regression Statistics	
Multiple R	0.82161
R Square	0.67504
Adjusted R Square	0.67254
Standard Error	12.16400
Observations	1045

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	8	318435.57	39804.45	269.02	9.63E-247
Residual	1036	153289.53	147.96		
Total	1044	471725.10			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.95	1.50	1.96	4.99E-02	0.00	5.91
B_6	0.15	0.03	4.81	1.76E-06	0.09	0.21
B_7	0.14	0.03	4.57	5.36E-06	0.08	0.20
B_8	0.12	0.03	4.16	3.39E-05	0.06	0.18
B_9	0.37	0.03	11.11	3.58E-27	0.30	0.43
L_{10}	-8.79	3.27	-2.69	7.34E-03	-15.21	-2.37
M	0.17	0.04	4.18	3.14E-05	0.09	0.26
L_{10}*M	0.24	0.07	3.31	9.79E-04	0.10	0.38
L_{10}*B_9	-0.18	0.06	-2.94	3.37E-03	-0.30	-0.06

A comparison of the full model and the reduced model shows an $F_{3,1033}$ of .8459 with a corresponding p-value of .47. Thus the reduced model is appropriate.

Define the random variable **B** to be a horse's next Beyer speed number, and (L, B_{10}, B_9,...B_1) the previous data with respect to that horse. The indicator variable $L = 1$ if the horse has had a layoff before its coming race. $L = 0$, otherwise. The two-sample hypothesis test and the regression results suggest the following probability model:

If $L = 0$ (the horse has not had a layoff) the probability is .005 that $\mathbf{B} = 0$. With probability .995 the distribution of **B** is given by

$$\mathbf{B} = 2.95 + .37B_{10} + .12B_9 + .14B_8 + .15B_7 + .17M + \varepsilon$$

where ε is a random error term and M is the median of $\{B_1, B_2, B_3, B_4, B_5, B_6\}$.

If $L=1$ (the horse has had a layoff of at least 45 days) the probability is .028 that $\mathbf{B} = 0$. With probability .972 the distribution of **B** is given by

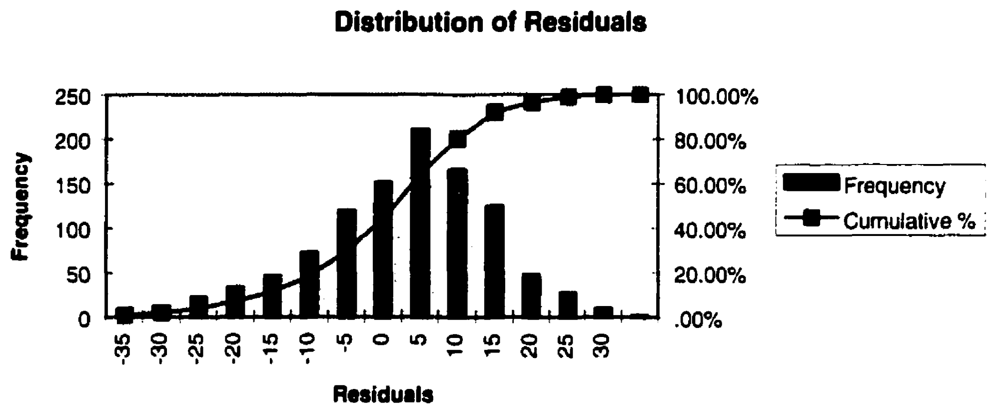
$$B = -5.84 + .19B_{10} + .12B_9 + .14B_8 + .15B_7 + .41M + \varepsilon$$

where ε is a random error term and M is the median of $\{B_1, B_2, B_3, B_4, B_5, B_6\}$.

This model is consistent with fundamental horse race handicapping. If a horse has been racing recently (L = 0), the most recent data point, B_{10}, is a good predictor of future performance. However, if a horse has not raced recently (L=1), the most recent data point, B_{10}, is not as relevant. The horse's historical performance, captured by the variable M, is more relevant. Finally, a layoff is often viewed as negative in terms of a horse's next performance. This is reflected in the intercepts which is 2.95 for L = 0 horses and -5.84 for L = 1 horses.

As for the random error term ε, the histogram of the residuals of the regression is as follows:

Figure 2



In the simulation model in the next section, I decided to use the typical regression assumption of normality for the random error term ε. ε was modeled as normal with mean 0 and standard deviation 12.164 where 12.164 was the standard error of the regression. Other reasonable choices for modeling the random error term were an empirical distribution using the residuals, a truncated normal and the logistic distribution. However, in using simulation to estimate the probabilities of horses winning, the choice of the distribution for the random error term had very little effect upon the estimated probabilities.

F. A Simulation Model for a Race.

The previous section provided a model for a horse's Beyer speed number in his or her next race. Thus if horses 1, 2,...,n compete in a given race, we have a model for their Beyer speed numbers $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_n$. However, will these random variables be independent? Initially, one might suspect that since Beyer speed numbers take into account distance and track conditions, the Beyer speed numbers of horses in the same race would be independent. However, there are races that are run quite tactically. Sometimes the jockey on the horse with the lead is able to have a relatively slow early pace. The final result would be that the race is run in a relatively slow final time for the quality of horses in the race. This would mean relatively lower Beyer speed numbers for all horses in the race. If that is the case, there will be a positive correlation between ε_i and ε_j where i and j are two horses in the same race with ε_i and ε_j being the random error terms for horse i and horse j respectively.

In order to estimate the correlation, the residuals (predicted Beyer speed number from regression model - actual Beyer speed number) were calculated for a pair of horses in the same race in a sample of 100 different races. However, there is no meaningful criterion for assigning one member of the pair to one variable rather than the other. This is analogous to measuring the correlation of a characteristic in twins. A statistic which is used to estimate the correlation in this situation is the sample intraclass correlation r_I. The sample intraclass correlation r_I is defined by

$$r_I = [MS(\text{among classes}) - MS(\text{within classes})] \,/\, [\,MS(\text{among classes}) + (n-1)MS(\text{within classes})]$$

where there are n observation in a class. In our example n = 2. The result was a sample intraclass correlation r_I of .236 with a p-value of .009. A more detailed discussion of intraclass correlation can be found in Steel (1980).

Thus ignoring the 0 Beyer numbers, the Beyer numbers of horses 1,2,...n in a race can be modeled as a with mean vector $\mu = (\mu_1, \mu_2, \ldots \mu_n)^T$ and covariance matrix Σ. The expected Beyer number for horse i, μ_i can be estimated from the regression formula in Section E. The covariance matrix Σ can be estimated by using $(12.164)^2$ for the diagonal elements and .236 $(12.164)^2$ for the off-diagonal elements where 12.164 was the standard error of regression and .236 is the intraclass correlation.

By examining the residuals of the regression I made the assumption that ignoring the 0 Beyer numbers a horse's next Beyer number can be modeled as normal. This does not guarantee that the joint distribution of $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_n$ is multivariate normal since the fact that the marginals are normal does not necessarily establish that the joint distribution is multivariate normal. However, in order to be able to generate random variates in a simulation, I will make the additional assumption that the joint distribution of $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_n$ is multivariate normal. With this assumption the following algorithm for simulating the Beyer numbers of all horses in a horse race is possible.

Step 1: Generate $X_1, X_2, \ldots, X_n$ as multivariate normal with mean μ and covariance matrix Σ. (This can be done by generating $Z_1, Z_2, \ldots, Z_n$ as IID standard Normal and defining $\mathbf{X} = \mu + \mathbf{CZ}$ where $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ and $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)^T$ and Σ is uniquely factored as CC^T where C is lower triangular.)

Step 2: If for horse i, L =0, generate Y_i as Bernoulli (.995). Otherwise (L=1) generate Y_i as Bernoulli (.972).

Step 3: Let $B_i = X_i Y_i$.

The winner of the simulated race is the horse with the maximum B_i. Thus this model provides a tool for estimating the probability of an entry winning a race. If N is the number of replications of the simulated race and N_j is the number of times entry j has the maximum B_i then N_j/N is an estimate of the probability of entry j winning the race.

Chapter 4

A Test of the Beyer Probability Model

The previous chapter developed a simulation model for estimating the probability of an entry winning a race. However, the dominant theory in horse racing is that an entry with fraction p of the win pool has a probability p of winning the race. The SPRT-like algorithm in Chapter 2 provides a method of determining which of these two models is more appropriate.

The two hypotheses tested are

H_1: The probability of entry j winning the race equals the fraction of win pool bet on entry j

versus

H_2: The probability of entry j winning the race equals the fraction of simulated races won by entry. j.

Alternatively these two hypotheses can be stated as

H_1: The probability of betting entry j winning race t is given by $p_{jt} = W_{jt}/W_t$

versus

H_2: The probability of betting entry j winning race t is given by $q_{jt} = N_{jt}/N_t$,

where W_{jt} is the amount bet on entry j in the t-th race, W_t is money bet in the win pool in the t-th race, N_t is the number of replications of race t in the simulation and N_{jt} is the number of times entry j won race t in the simulation.

In conducting the SPRT-like algorithm the following steps were followed in collecting data:

1. α was set equal to .01. Thus experimentation ends when $\lambda_n \geq 99$.

2. Races available from Assiniboia Downs in Winnipeg were selected (either live at Assiniboia Downs or through simulcast from Woodbine, Hastings Park, Del Mar, and Saratoga).

3. Races were selected before the card on weekend days when I could be present at the track. The experiment started on July 26, 1997. In order for a race to be selected:

-The race had to be run on dirt.

- All the horses in the race had to have had at least 10 previous Beyer numbers run as 3 year old horses or older.

4. The betting entry that was declared the winner by the stewards would be treated as the winning betting entry in the experiment. (It is possible, although rare, for the horse that crosses the finish line first (and hence will have the highest Beyer number) to be disqualified for interfering with other horses.)

5. In the rare situation of dead heats, the likelihood under H_0 and H_1 would be calculated as follows:

If betting entries a and b dead heat in race t, $f(x_t \mid H_0) = (p_{at}\, p_{bt})^{.5}$ and $f(x_t \mid H_1) = (q_{at}\, q_{bt})^{.5}$.
(If more than 2 horses dead heat the geometric mean of the probabilities would be used. With modern photographic equipment, dead heats are rare and multiple dead heats are extremely rare).

6. The simulation was performed using @RISK 3.5E, an add-on to Microsoft Excel. The random number seed was set equal to the number 1. There was 1000 iterations for each simulation ($N_t = 1000$).

7. The crowd's probability (H_0) for the winning horse was calculated using the final winning price (which is the return for a \$2 bet.). Because of "breakage" this is an approximation. Breakage means that pari-mutuel payoff is rounded down to the nearest dime. That is a bet that should have paid say \$7.56 is reported as \$7.50. Thus a winning payoff of \$7.50 could have

been any payoff in the interval [\$7.50, \$7.60) if there was no breakage. In order to reduce any bias in estimating the crowd's probability, \$0.05 is added to the winning price. The resulting formula used to estimate the crowd's probability is

$$\text{Crowd's Probability} = 2\,Q\,/\,(\,\text{Win Price} + \$.05)$$

where Q is the fraction of money returned to bettors in the pari-mutuel system. As an example a win price reported as \$7.50 at Assiniboia Downs (with a Q = .82), would correspond by this formula to a crowd probability of .217. If the win payoff without breakage was exactly \$7.50 the crowd probability would be .219 while a win payoff without breakage of say \$7.599 would correspond .216.

After 37 data points λ_n reached 110.16. The relatively rare conditions 4 and 5 of the experimental protocol did not occur during the data collection. Since $\lambda_n \geq 99$, the H_2 assumption should be accepted. It therefore, can be concluded that the probability model is more appropriate than the efficient market hypothesis.

The meteorological statistics such as Brier scores and skill scores are consistent with the above conclusions. The Brier score BS^* for the model was .564 in comparison to the Brier score BS^{**} of .610 for the base or crowd probabilities. The resulting skill score SS was 7.48%.

There was a strong positive correlation between the crowd's probabilities and the probabilities estimated by the model. In most situations both methods agreed on which horse should be the favorite but with somewhat different probabilities. Considering the fact that the crowd has access to the Beyer speed numbers and likely uses them in a reasonable but non-statistical manner, this is not surprising. Although, not the purpose of the study, the results were only partially consistent with the longshot-favorite bias. In most cases when crowd thought a horse had a low probability of winning say .05, the model estimated a much lower probability, say .005. However, there were a number of exceptions such as Race 3 where the crowd estimated a probability of .055 for the winning horse and the model's estimate was .194. However, with heavy favorites (horses with more than 50% of the win pool) the situation was less clear. Often the model predicted slightly lower probabilities than the crowd. In the 3 cases in the experiment when a heavy favorite won (with crowd probabilities of .552, .552 and .521) the model probabilities were .495, .466 and .456 respectively. It may be that the introduction of Beyer speed numbers in the *Daily Racing Form* has changed betting habits. Bettors may be overbetting horses that have the highest Beyer speed numbers and underestimating the natural variation associated with Beyer speed numbers. However, we are dealing with a small sample and a sample restricted to older horses running on dirt.

The data from the experiment was as follows:

Table 3

SPRT for Two Parameter Selection Methods

n	Date	Track	Race	Distance	Entries	Crowd	Model	L1/L2	L2/L1	λ_n
1	7/26/97	WO	3	8.5	8	0.254	0.128	1.98	0.50	1.98
2	7/26/97	WO	4	8.5	6	0.42	0.443	1.88	0.53	1.88
3	7/26/97	WO	11	8.5	8	0.055	0.194	0.53	1.87	1.87
4	7/27/97	AsD	1	8.5	5	0.229	0.266	0.46	2.18	2.18
5	7/27/97	AsD	2	7	9	0.233	0.224	0.48	2.09	2.09
6	7/27/97	AsD	4	6	8	0.201	0.108	0.89	1.12	1.12
7	7/27/97	AsD	6	7	7	0.13	0.195	0.59	1.69	1.69
8	7/27/97	AsD	7	6	7	0.147	0.282	0.31	3.24	3.24
9	7/27/97	AsD	8	8.5	6	0.449	0.257	0.54	1.85	1.85
10	7/27/97	AsD	9	6	10	0.098	0.24	0.22	4.54	4.54
11	7/27/97	Hst	9	8.5	5	0.552	0.495	0.25	4.07	4.07
12	7/27/97	Dmr	8	6	7	0.076	0.194	0.10	10.39	10.39
13	8/2/97	WO	1	7	7	0.363	0.332	0.11	9.50	9.50
14	8/2/97	WO	3	8.5	8	0.125	0.057	0.23	4.33	4.33
15	8/2/97	WO	4	6	6	0.371	0.311	0.28	3.63	3.63
16	8/2/97	Sar	8	9	6	0.239	0.2	0.33	3.04	3.04
17	8/2/97	Hst	3	6.5	6	0.265	0.388	0.22	4.45	4.45
18	8/2/97	Hst	4	6.5	6	0.552	0.466	0.27	3.76	3.76
19	8/2/97	Hst	7	8.5	5	0.362	0.336	0.29	3.49	3.49
20	8/2/97	Dmr	7	6	10	0.134	0.188	0.20	4.89	4.89
21	8/2/97	AsD	7	8	8	0.437	0.352	0.25	3.94	3.94
22	8/4/97	WO	1	7	7	0.263	0.523	0.13	7.83	7.83
23	8/4/97	AsD	4	7	9	0.106	0.188	0.07	13.90	13.90
24	8/4/97	WO	7	7	9	0.13	0.027	0.35	2.89	2.89
25	8/4/97	AsD	5	6	10	0.102	0.041	0.86	1.16	1.16
26	8/4/97	WO	8	8.5	7	0.521	0.456	0.98	1.02	1.02
27	8/4/97	AsD	6	6	5	0.125	0.299	0.41	2.43	2.43
28	8/4/97	WO	9	6	7	0.181	0.238	0.31	3.19	3.19
29	8/4/97	AsD	8	7	8	0.026	0.02	0.41	2.46	2.46
30	8/4/97	AsD	9	6	11	0.051	0.106	0.20	5.11	5.11
31	8/4/97	Hst	1	8.5	6	0.378	0.53	0.14	7.16	7.16
32	8/4/97	Dmr	2	6	6	0.194	0.35	0.08	12.92	12.92
33	8/4/97	Hst	4	6.5	6	0.321	0.176	0.14	7.08	7.08
34	8/8/97	WO	3	8.5	6	0.191	0.308	0.09	11.42	11.42
35	8/8/97	WO	9	7	8	0.113	0.271	0.04	27.39	27.39
36	8/8/97	Sar	8	6	7	0.117	0.167	0.03	39.09	39.09
37	8/9/97	Sar	7	9	5	0.11	0.31	0.01	110.16	110.16

Finally, in order to illustrate how this SPRT-like algorithm can be used to choose between more than two hypotheses, consider the following three hypotheses:

H_1: The probability of entry j winning the race equals the fraction of win pool bet on entry j

versus

H_2: The probability of entry j winning the race equals the fraction of simulated races won by entry j

versus

H_3: The probability of entry j winning the race =1/m where m is the number of entries in the race.

Using the same data set as before φ_n is below 1/(1-.01) after 37 data points. Thus estimation procedure 2 (the simulation model) is selected best at a level α =.01. The calculations needed are demonstrated in the following table:

Table 4

SPRT for Three Parameter Selection Methods

n	Entries	Crowd	Model	Equiprob	L1/L2	L2/L1	L1/L3	L3/L1	L2/L3	L3/L2	φ_n
1	8	0.254	0.128	0.13	1.98	0.50	2.03	0.49	1.02	0.98	2.00
2	6	0.42	0.443	0.17	1.88	0.53	5.12	0.20	2.72	0.37	1.73
3	8	0.055	0.194	0.13	0.53	1.87	2.25	0.44	4.22	0.24	1.98
4	5	0.229	0.266	0.20	0.46	2.18	2.58	0.39	5.62	0.18	1.85
5	9	0.233	0.224	0.11	0.48	2.09	5.41	0.18	11.33	0.09	1.66
6	8	0.201	0.108	0.13	0.89	1.12	8.70	0.11	9.79	0.10	2.00
7	7	0.130	0.195	0.14	0.59	1.69	7.92	0.13	13.36	0.07	1.72
8	7	0.147	0.282	0.14	0.31	3.24	8.15	0.12	26.37	0.04	1.43
9	6	0.449	0.257	0.17	0.54	1.85	21.94	0.05	40.66	0.02	1.59
10	10	0.098	0.24	0.10	0.22	4.54	21.51	0.05	97.58	0.01	1.27
11	5	0.552	0.495	0.20	0.25	4.07	59.35	0.02	241.52	0.00	1.26
12	7	0.076	0.194	0.14	0.10	10.39	31.58	0.03	327.98	0.00	1.13
13	7	0.363	0.332	0.14	0.11	9.50	80.24	0.01	762.24	0.00	1.12
14	8	0.125	0.057	0.13	0.23	4.33	80.24	0.01	347.58	0.00	1.24
15	6	0.371	0.311	0.17	0.28	3.63	178.61	0.01	648.58	0.00	1.28
16	6	0.239	0.2	0.17	0.33	3.04	256.12	0.00	778.30	0.00	1.33
17	6	0.265	0.388	0.17	0.22	4.45	407.23	0.00	1811.88	0.00	1.23
18	6	0.552	0.466	0.17	0.27	3.76	1348.76	0.00	5066.02	0.00	1.27
19	5	0.362	0.336	0.20	0.29	3.49	2441.25	0.00	8510.91	0.00	1.29
20	10	0.134	0.188	0.10	0.20	4.89	3271.28	0.00	16000.52	0.00	1.20
21	8	0.437	0.352	0.13	0.25	3.94	11436.40	0.00	45057.46	0.00	1.25
22	7	0.263	0.523	0.14	0.13	7.83	21054.41	0.00	164955.37	0.00	1.13
23	9	0.106	0.188	0.11	0.07	13.90	20085.91	0.00	279104.48	0.00	1.07
24	9	0.130	0.027	0.11	0.35	2.89	23500.51	0.00	67822.39	0.00	1.35
25	10	0.102	0.041	0.10	0.86	1.16	23970.52	0.00	27807.18	0.00	1.86
26	7	0.521	0.456	0.14	0.98	1.02	87420.49	0.00	88760.52	0.00	1.98
27	5	0.125	0.299	0.20	0.41	2.43	54637.80	0.00	132696.97	0.00	1.41
28	7	0.181	0.238	0.14	0.31	3.19	69226.10	0.00	221073.16	0.00	1.31
29	8	0.026	0.02	0.13	0.41	2.46	14399.03	0.00	35371.71	0.00	1.41
30	11	0.051	0.106	0.09	0.20	5.11	8077.85	0.00	41243.41	0.00	1.20
31	6	0.378	0.53	0.17	0.14	7.16	18320.57	0.00	131154.04	0.00	1.14
32	6	0.194	0.35	0.17	0.08	12.92	21325.15	0.00	275423.48	0.00	1.08
33	6	0.321	0.176	0.17	0.14	7.08	41072.24	0.00	290847.20	0.00	1.14
34	6	0.191	0.308	0.17	0.09	11.42	47068.78	0.00	537485.62	0.00	1.09
35	8	0.113	0.271	0.13	0.04	27.39	42550.18	0.00	1165268.82	0.00	1.04
36	7	0.117	0.167	0.14	0.03	39.09	34848.60	0.00	1362199.25	0.00	1.03
37	5	0.110	0.31	0.20	0.01	110.16	19166.73	0.00	2111408.84	0.00	1.01

Chapter 5

A Test of Cross Track Betting Strategies

The American journalist, politician and scientist Benjamin Franklin said that the only sure things in life are death and taxes. Hausch and Ziemba (1990) claimed that there was another sure thing - risk free wagers in cross-track betting. As explained in their article

Cross-track betting permits bettors to place wagers at their local tracks on a race being run at another track. Since each track operates a separate betting pool, the odds can vary across the tracks. The data suggests that the odds vary, and they often vary dramatically, allowing arbitrage opportunities.

Major races are often simulcast from the host track to many satellite tracks. In the situation where the satellite tracks had separate betting pools Hausch and Ziemba suggested a syndicate have agents at each track. If the agents were in communication they could determine which track provided the best odds on an individual horse. By betting each horse at the track with the best odds it might be possible to create a risk-free hedge or arbitrage opportunity. The example given in their article was the betting on the 1983 Preakness race.

Table 5

Cross Track Betting 1983 Preakness

Horse No.	Highest Win Return (on a $1 bet)	Track	$ amount of wager that will return $1
1	29.40	Louisiana Downs	.0340
2	12.70	Louisiana Downs	.0787
3	34.60	Los Alamitos	.0289
4	169.90	Hollywood	.0059
5	56.90	Louisiana Downs	.0176
6	5.70	Louisiana Downs	.1754
7	10.60	Pimlico	.0943
8	76.60	Louisiana Downs	.0131
9	116.10	Hollywood	.0086
10	2.20	Los Alamitos	.4545
11	40.60	Los Alamitos	.0246
Total			**.9356**

Thus by wagering $0.9356 the syndicate is guaranteed $1.00 regardless of which horse won the race. It must be pointed out that in pari-mutuel betting the bets affect the odds. Therefore, betting $9,356 will return less than $10,000 although likely more than $9,356.

Hausch and Ziemba admit that it is difficult (if not impossible) to successfully implement the above approach. However, they do recommend a strategy for an individual bettor at a satellite track. They suggest that the individual bettor at the satellite track view the odds at the host track and use these "true" win probabilities to search for "overlays" at the satellite track. In gambling the term "overlay" refers to a wager with a positive expectation.

Hausch and Ziemba assume that the betting at the host track provides the true probabilities while the betting at the satellite track provides inaccurate estimates of the probability. They provide a number of reasons for this assumption. First, the efficient market theory on win probabilities is based on host track data. Second, bettors at the host track should be more knowledgeable about the races at their track. They can see the horses in the paddock before the race. They are more likely to be aware of jockey and track biases. Third, bettors at the host track include jockeys, agents, trainers and owners who may have inside information. Finally, the host track usually has much bigger pools than the satellite track and a bigger pool should be more efficient. However, all of these reasons are common sense and had no data to support them. An alternate common sense argument is that bettors at the satellite track tend to be hard core horse racing gamblers while bettors at the host track tend to be casual bettors interested in a day of entertainment. It would follow from that argument that the probabilities determined by bettors at the satellite track are the true probabilities.

I decided to use the SPRT algorithm for the better of two estimation procedures with an $\alpha = .05$. The host tracks were the southern California tracks Hollywood and Santa Anita. The satellite track was Woodbine in Toronto, Canada.

In effect, the underlying assumptions being tested were:

H_1: The probability of entry j winning the race equals the fraction of win pool bet on entry j at the host track.

versus

H_2: The probability of entry j winning the race equals the fraction of win pool bet on entry j at the satellite track.

Data was obtained from the internet after a day of racing. This meant that if there were say 9 races in a given day, the data points would be captured in a block of 9. This leads to the question of what should be done if there is extra data available after the sequential probability ratio test has terminated.

The data from this experiment was as follows:

Table 6

SPRT for Cross Track Betting

n	Date	Race	Host prob	Satellite prob	L1/L2	L2/L1	λ_n
1	12/20/97	1	0.36	0.31	1.15	0.87	1.15
2	12/20/97	2	0.32	0.25	1.44	0.69	1.44
3	12/20/97	3	0.39	0.41	1.39	0.72	1.39
4	12/20/97	4	0.32	0.31	1.41	0.71	1.41
5	12/20/97	5	0.25	0.23	1.58	0.63	1.58
6	12/20/97	6	0.46	0.50	1.45	0.69	1.45
7	12/20/97	7	0.48	0.51	1.37	0.73	1.37
8	12/20/97	8	0.46	0.53	1.18	0.84	1.18
9	12/20/97	9	0.26	0.24	1.31	0.76	1.31
10	12/20/97	10	0.15	0.14	1.38	0.73	1.38
11	12/21/97	1	0.16	0.19	1.17	0.85	1.17
12	12/21/97	2	0.11	0.11	1.13	0.89	1.13
13	12/21/97	3	0.55	0.51	1.20	0.83	1.20
14	12/21/97	4	0.20	0.25	0.97	1.03	1.03
15	12/21/97	5	0.10	0.09	1.12	0.89	1.12
16	12/21/97	6	0.10	0.14	0.81	1.23	1.23
17	12/21/97	7	0.26	0.24	0.87	1.15	1.15
18	12/21/97	8	0.19	0.17	0.96	1.04	1.04
19	12/21/97	9	0.29	0.22	1.23	0.81	1.23
20	12/21/97	10	0.51	0.44	1.43	0.70	1.43
21	12/22/97	1	0.04	0.03	2.12	0.47	2.12
22	12/22/97	2	0.13	0.15	1.84	0.54	1.84
23	12/22/97	3	0.15	0.15	1.75	0.57	1.75
24	12/22/97	4	0.43	0.38	2.00	0.50	2.00
25	12/22/97	5	0.33	0.34	1.95	0.51	1.95
26	12/22/97	6	0.21	0.23	1.80	0.55	1.80
27	12/22/97	7	0.32	0.32	1.80	0.56	1.80

28	12/22/97	8	0.16	0.22	1.31	0.76	1.31
29	12/26/97	1	0.24	0.22	1.40	0.71	1.40
30	12/26/97	2	0.32	0.29	1.57	0.64	1.57
31	12/26/97	3	0.46	0.44	1.62	0.62	1.62
32	12/26/97	4	0.25	0.20	2.03	0.49	2.03
33	12/26/97	5	0.16	0.16	2.12	0.47	2.12
34	12/26/97	6	0.17	0.19	1.92	0.52	1.92
35	12/26/97	7	0.32	0.32	1.92	0.52	1.92
36	12/26/97	8	0.08	0.09	1.78	0.56	1.78
37	12/26/97	9	0.26	0.27	1.74	0.58	1.74
38	12/27/97	1	0.17	0.18	1.65	0.61	1.65
39	12/27/97	2	0.06	0.05	1.69	0.59	1.69
40	12/27/97	3	0.27	0.30	1.53	0.65	1.53
41	12/27/97	4	0.19	0.14	2.05	0.49	2.05
42	12/27/97	5	0.51	0.51	2.05	0.49	2.05
43	12/27/97	6	0.25	0.26	1.92	0.52	1.92
44	12/27/97	7	0.19	0.14	2.64	0.38	2.64
45	12/27/97	8	0.29	0.27	2.84	0.35	2.84
46	12/27/97	9	0.39	0.38	2.94	0.34	2.94
47	12/28/97	1	0.08	0.08	3.07	0.33	3.07
48	12/28/97	2	0.06	0.04	4.62	0.22	4.62
49	12/28/97	3	0.03	0.03	5.15	0.19	5.15
50	12/28/97	4	0.13	0.10	6.33	0.16	6.33
51	12/28/97	5	0.23	0.19	7.56	0.13	7.56
52	12/28/97	6	0.31	0.25	9.29	0.11	9.29
53	12/28/97	7	0.51	0.55	8.71	0.11	8.71
54	12/28/97	8	0.32	0.36	7.66	0.13	7.66
55	12/28/97	9	0.09	0.08	8.53	0.12	8.53
56	12/29/97	1	0.20	0.30	5.73	0.17	5.73
57	12/29/97	2	0.41	0.39	6.08	0.16	6.08
58	12/29/97	3	0.27	0.30	5.41	0.18	5.41
59	12/29/97	4	0.11	0.12	4.96	0.20	4.96
60	12/29/97	5	0.55	0.59	4.61	0.22	4.61
61	12/29/97	6	0.19	0.21	4.26	0.23	4.26
62	12/29/97	7	0.29	0.36	3.45	0.29	3.45
63	12/29/97	8	0.03	0.02	4.30	0.23	4.30
64	12/31/97	1	0.19	0.20	4.18	0.24	4.18
65	12/31/97	2	0.16	0.15	4.32	0.23	4.32
66	12/31/97	3	0.19	0.18	4.55	0.22	4.55
67	12/31/97	4	0.24	0.27	4.00	0.25	4.00
68	12/31/97	5	0.13	0.12	4.56	0.22	4.56
69	12/31/97	6	0.20	0.22	4.29	0.23	4.29
70	12/31/97	7	0.03	0.03	3.76	0.27	3.76
71	12/31/97	8	0.16	0.16	3.60	0.28	3.60
72	1/1/98	1	0.10	0.08	4.60	0.22	4.60
73	1/1/98	2	0.07	0.06	5.25	0.19	5.25
74	1/1/98	3	0.08	0.09	4.61	0.22	4.61
75	1/1/98	4	0.21	0.19	4.98	0.20	4.98
76	1/1/98	5	0.08	0.09	4.47	0.22	4.47
77	1/1/98	6	0.23	0.25	4.09	0.24	4.09
78	1/1/98	7	0.33	0.35	3.91	0.26	3.91

79	1/1/98	8	0.05	0.05	3.94	0.25	3.94
80	1/1/98	9	0.19	0.16	4.61	0.22	4.61
81	1/2/98	1	0.32	0.30	4.97	0.20	4.97
82	1/2/98	2	0.04	0.03	5.85	0.17	5.85
83	1/2/98	3	0.16	0.15	5.99	0.17	5.99
84	1/2/98	4	0.12	0.11	6.47	0.15	6.47
85	1/2/98	5	0.19	0.19	6.44	0.16	6.44
86	1/2/98	6	0.16	0.17	5.97	0.17	5.97
87	1/2/98	7	0.36	0.38	5.66	0.18	5.66
88	1/2/98	8	0.19	0.18	5.77	0.17	5.77
89	1/3/98	1	0.06	0.07	5.20	0.19	5.20
90	1/3/98	2	0.12	0.15	4.17	0.24	4.17
91	1/3/98	3	0.11	0.12	3.93	0.25	3.93
92	1/3/98	4	0.11	0.13	3.36	0.30	3.36
93	1/3/98	5	0.30	0.31	3.24	0.31	3.24
94	1/3/98	6	0.21	0.24	2.90	0.34	2.90
95	1/3/98	7	0.10	0.09	3.19	0.31	3.19
96	1/3/98	8	0.19	0.18	3.33	0.30	3.33
97	1/3/98	9	0.11	0.09	4.22	0.24	4.22
98	1/4/98	1	0.18	0.20	3.78	0.26	3.78
99	1/4/98	2	0.06	0.07	3.21	0.31	3.21
100	1/4/98	3	0.38	0.43	2.80	0.36	2.80
101	1/4/98	4	0.09	0.08	2.94	0.34	2.94
102	1/4/98	5	0.07	0.05	4.36	0.23	4.36
103	1/4/98	6	0.08	0.06	6.09	0.16	6.09
104	1/4/98	7	0.30	0.29	6.33	0.16	6.33
105	1/4/98	8	0.09	0.08	7.32	0.14	7.32
106	1/4/98	9	0.12	0.11	7.49	0.13	7.49
107	1/7/98	1	0.14	0.13	7.81	0.13	7.81
108	1/7/98	2	0.23	0.19	9.44	0.11	9.44
109	1/7/98	3	0.43	0.40	10.27	0.10	10.27
110	1/7/98	4	0.21	0.24	8.97	0.11	8.97
111	1/7/98	5	0.07	0.08	8.52	0.12	8.52
112	1/7/98	6	0.25	0.27	7.83	0.13	7.83
113	1/7/98	7	0.22	0.27	6.34	0.16	6.34
114	1/7/98	8	0.31	0.26	7.43	0.13	7.43
115	1/8/98	1	0.04	0.05	5.97	0.17	5.97
116	1/8/98	2	0.28	0.30	5.59	0.18	5.59
117	1/8/98	3	0.13	0.10	6.94	0.14	6.94
118	1/8/98	4	0.23	0.19	8.20	0.12	8.20
119	1/8/98	5	0.19	0.21	7.65	0.13	7.65
120	1/8/98	6	0.51	0.46	8.62	0.12	8.62
121	1/8/98	7	0.19	0.21	7.77	0.13	7.77
122	1/8/98	8	0.24	0.23	7.97	0.13	7.97
123	1/9/98	1	0.41	0.39	8.45	0.12	8.45
124	1/9/98	2	0.43	0.37	9.88	0.10	9.88
125	1/9/98	3	0.63	0.57	10.93	0.09	10.93
126	1/9/98	4	0.08	0.07	12.49	0.08	12.49
127	1/9/98	5	0.03	0.05	9.06	0.11	9.06
128	1/9/98	6	0.12	0.11	10.63	0.09	10.63
129	1/9/98	7	0.02	0.03	7.94	0.13	7.94

130	1/9/98	8	0.23	0.22	8.27	0.12	8.27
131	1/10/98	1	0.29	0.28	8.59	0.12	8.59
132	1/10/98	2	0.22	0.21	8.94	0.11	8.94
133	1/10/98	3	0.39	0.37	9.48	0.11	9.48
134	1/10/98	4	0.05	0.05	10.10	0.10	10.10
135	1/10/98	5	0.05	0.05	9.92	0.10	9.92
136	1/10/98	6	0.11	0.13	8.85	0.11	8.85
137	1/10/98	7	0.10	0.10	8.40	0.12	8.40
138	1/10/98	8	0.11	0.10	8.83	0.11	8.83
139	1/10/98	9	0.07	0.08	7.67	0.13	7.67
140	1/11/98	1	0.09	0.09	7.11	0.14	7.11
141	1/11/98	2	0.28	0.30	6.66	0.15	6.66
142	1/11/98	3	0.33	0.33	6.64	0.15	6.64
143	1/11/98	4	0.15	0.13	7.72	0.13	7.72
144	1/11/98	5	0.25	0.22	8.74	0.11	8.74
145	1/11/98	6	0.16	0.18	7.96	0.13	7.96
146	1/11/98	7	0.11	0.11	7.97	0.13	7.97
147	1/11/98	8	0.68	0.66	8.19	0.12	8.19
148	1/11/98	9	0.31	0.29	8.81	0.11	8.81
149	1/14/98	1	0.43	0.43	8.88	0.11	8.88
150	1/14/98	2	0.17	0.19	8.30	0.12	8.30
151	1/14/98	3	0.17	0.19	7.51	0.13	7.51
152	1/14/98	4	0.12	0.14	6.49	0.15	6.49
153	1/14/98	5	0.05	0.05	7.03	0.14	7.03
154	1/14/98	6	0.11	0.12	6.70	0.15	6.70
155	1/14/98	7	0.16	0.14	7.73	0.13	7.73
156	1/14/98	8	0.30	0.27	8.46	0.12	8.46
157	1/15/98	1	0.15	0.14	8.90	0.11	8.90
158	1/15/98	2	0.51	0.43	10.60	0.09	10.60
159	1/15/98	3	0.27	0.26	10.85	0.09	10.85
160	1/15/98	4	0.13	0.13	11.06	0.09	11.06
161	1/15/98	5	0.04	0.03	14.46	0.07	14.46
162	1/15/98	6	0.03	0.04	12.98	0.08	12.98
163	1/15/98	7	0.12	0.11	14.33	0.07	14.33
164	1/15/98	8	0.20	0.20	14.39	0.07	14.39
165	1/16/98	1	0.11	0.12	12.44	0.08	12.44
166	1/16/98	2	0.22	0.20	13.79	0.07	13.79
167	1/16/98	3	0.15	0.17	12.18	0.08	12.18
168	1/16/98	4	0.25	0.23	12.85	0.08	12.85
169	1/16/98	5	0.18	0.17	13.11	0.08	13.11
170	1/16/98	6	0.16	0.18	11.71	0.09	11.71
171	1/16/98	7	0.07	0.08	9.72	0.10	9.72
172	1/16/98	8	0.19	0.19	9.57	0.10	9.57
173	1/17/98	1	0.32	0.28	10.88	0.09	10.88
174	1/17/98	2	0.13	0.13	10.84	0.09	10.84
175	1/17/98	3	0.36	0.41	9.54	0.10	9.54
176	1/17/98	4	0.33	0.29	10.88	0.09	10.88
177	1/17/98	5	0.10	0.08	13.44	0.07	13.44
178	1/17/98	6	0.29	0.30	12.77	0.08	12.77
179	1/17/98	7	0.05	0.04	15.43	0.06	15.43
180	1/17/98	8	0.68	0.69	15.20	0.07	15.20

181	1/17/98	9	0.48	0.47	15.70	0.06	15.70
182	1/18/98	1	0.16	0.13	19.50	0.05	19.50
183	1/18/98	2	0.26	0.36	13.99	0.07	13.99
184	1/18/98	3	0.28	0.29	13.58	0.07	13.58
185	1/18/98	4	0.04	0.04	11.76	0.09	11.76
186	1/18/98	5	0.48	0.44	12.86	0.08	12.86
187	1/18/98	6	0.46	0.47	12.56	0.08	12.56
188	1/18/98	7	0.02	0.02	11.65	0.09	11.65
189	1/18/98	8	0.21	0.22	11.20	0.09	11.20
190	1/18/98	9	0.20	0.16	13.93	0.07	13.93
191	1/23/98	1	0.17	0.22	11.21	0.09	11.21
192	1/23/98	2	0.39	0.42	10.52	0.10	10.52
193	1/23/98	3	0.09	0.09	10.01	0.10	10.01
194	1/23/98	4	0.07	0.06	12.60	0.08	12.60
195	1/23/98	5	0.26	0.25	13.10	0.08	13.10
196	1/23/98	6	0.18	0.22	11.00	0.09	11.00
197	1/23/98	7	0.10	0.10	11.28	0.09	11.28
198	1/23/98	8	0.21	0.21	11.45	0.09	11.45
199	1/24/98	1	0.19	0.25	8.69	0.12	8.69
200	1/24/98	2	0.17	0.18	8.50	0.12	8.50
201	1/24/98	3	0.17	0.18	8.07	0.12	8.07
202	1/24/98	4	0.25	0.24	8.39	0.12	8.39
203	1/24/98	5	0.26	0.25	8.86	0.11	8.86
204	1/24/98	6	0.48	0.48	8.89	0.11	8.89
205	1/24/98	7	0.39	0.39	8.99	0.11	8.99
206	1/24/98	8	0.46	0.48	8.53	0.12	8.53
207	1/24/98	9	0.09	0.07	10.15	0.10	10.15
208	1/25/98	1	0.23	0.19	11.84	0.08	11.84
209	1/25/98	2	0.20	0.16	14.38	0.07	14.38
210	1/25/98	3	0.22	0.19	16.31	0.06	16.31
211	1/25/98	4	0.05	0.04	18.40	0.05	18.40
212	1/25/98	5	0.05	0.05	19.07	0.05	19.07
213	1/25/98	6	0.31	0.31	18.71	0.05	18.71
214	1/25/98	7	0.04	0.07	12.12	0.08	12.12
215	1/25/98	8	0.43	0.41	12.87	0.08	12.87
216	1/25/98	9	0.13	0.12	14.16	0.07	14.16

With $\alpha = .05$, the sequential probability ratio test should end when λ_n reaches 19. λ_n reached 19 after 212 data points which means the assumption of the host track probabilities being correct should be accepted at a significance level of .05. However, because of the way the data was collected, 4 additional data points were available at the end of the experiment. It seems wrong to ignore relevant data. However, if data points 213, 214, 215 and 216 are included, λ_n drops to 14.16 and the results are no longer significant at a 5% level. If a selection must be made (with no prior views on which hypothesis is correct), the hypothesis with the higher likelihood would be selected. In this case, the assumption that the host track probabilities are correct is selected. With a λ_n = 14.16, the posterior probability that the host track's probabilities are correct (assuming either the host track's or the satellite track's probabilities are correct) is λ_n / $(1+\lambda_n)$ or .93. Finally, with respect to the definition in Chapter 1, the host track procedure for estimating parameters is selected the better of the host and the satellite procedures at a significance level of .07. Thus although this is not a proof of the Hausch-Ziemba conjecture, the results are at least consistent with the views of Hausch and Ziemba.

In order to avoid a future ambiguous outcome, I would required a stopping rule of $\lambda_m \geq (1-\alpha)/\alpha$ after a block of data has been captured.

Chapter 6
Conclusions

This thesis developed a method for selecting the best of k multinomial parameter estimation procedures when only one observation per the k estimation procedures is possible but the k estimation procedures can be repeated many times. The inspiration for this method was the Sequential Probability Ratio Test (SPRT). In my view an SPRT-like approach is ideal in situations where one of k procedures must be selected and the consequence of selecting an incorrect procedure is the same for all procedures.

Chapters 4 and 5 demonstrated the use of this SPRT-like procedure. Both chapters examined the SPRT-like approach with respect to racetrack models. However, in my view the most important future application of this procedure would be in testing financial models that forecast probabilities in financial markets. Market timing is a controversial area in financial research. This SPRT-like approach provides an objective test to determine which probability model developed for market timing is best.

Chapter 3 developed a probability model for the racetrack. The SPRT-like approach in Chapter 4 established that assuming one of the two models was correct, the Beyer probability model should be selected over the efficient market model. However, horse racing is so complex, it is doubtful that any model will ever be able to provide the "true probabilities".

Nevertheless, the fact that a statistical model provides better probabilities than the crowd's probabilities is important. In pari-mutuel betting, the crowd's probabilities determine the odds. A necessary condition for a winning betting strategy is to have better probabilities than the crowd.

The results of this chapter are relevant to other research problems associated with the race track. Given an estimate of the probability of each horse winning and the odds on each horse, how should a decision maker bet the race? As pointed out earlier, Isaacs (1953), Rosner (1975), and Bolton and Chapman (1986) have provided models for this problem. The results of this chapter provide a tool for implementing these models.

There are a number of models (Hausch, Ziemba and Rubinstein (1981), Ziemba and Hausch (1984) and Ziemba and Hausch (1985), Asch and Quandt (1987)) which require estimates of the win probabilities in order to exploit anomalies in other pools at the race track. These models all used the fraction of the win pool as the estimate of the true probability of an entry winning a race. These models should perform better with a more accurate estimate of the win probabilities.

Chapter 5 examined the Hausch and Ziemba (1990) assumption on cross track betting strategies. The results were consistent with the Hausch and Ziemba assumption.

The SPRT-like approach in Chapter 2 was developed for discrete random variables. A natural extension of this work would be in continuous random variables. For example suppose there were k procedures for generating a probability density function for next month's return in a particular stock. Which of these procedures is correct? Again this is a situation where one observation per the k procedures is possible but the k procedures can be repeated each month.

A second area of future research is the probability model for the racetrack in Chapter 3. Simulation was used to estimate the parameters. Simulation can only approximate the expected value of a parameter. If an analytic solution can be developed an exact solution can be obtained. However, it should be noted that even with an exact solution to the parameters, the model will not necessarily be generating the "true probabilities". While I believe the best type of model for the race track is one which uses the future Beyer number as the response variable, different analysts might choose different explanatory variables and develop a better model. However, the SPRT-like procedure does provide a technique for comparing that model with the model developed in this thesis.

References

Ali, M.M., "Probability and Utility Estimates for Racetrack Bettors, " *Journal Political Economy* 85 (August 1977), 803-815.

Asch, P., B. G. Malkiel and R. E. Quandt, "Market Efficiency in Racetrack Betting," *Journal of Business* 57 (April 1984), 166 - 175.

Asch, P. and R. E. Quandt, "Efficiency and Profitability in Exotic Bets," *Econimica* 59 (August 1987), 289-298.

Azzalini, A., *Statistical Inference Based on the Likelihood* Chapman and Hall, London (1996).

Baum,C.W. and V.V. Veeravalli, "A Sequential Procedure for Multihypothesis Testing", *IEEE Transactions on Information Theory* Vol. 40, No. 6 (1994), 1994-2007.

Beyer, A., *Beyer on Speed* Houghton Mills, Boston (1993).

Brier, G.W., "Verification of Forecasts Expressed in Terms of Probability", *Monthly Weather Review*, 78 (1950), 1-3.

Bolton, R.N. and R.G. Chapman, "Searching for Positive Returns at the Track: A Multinomial Logit Model for Handicapping Horse Races", *Management Science* Vol. 32, No. 8 (1986), 1040-1060.

Fabricant, B.F. *The Science of Winning*, Van Norstrand Reinhold, New York (1979).

Figlewski, S., "Subjective Information and Market Efficiency in a Betting Market," *Journal of Political Economy* Vol. 87 No. 1 (1979) 75-88.

Hausch, D. B., and W. T. Ziemba, "Arbitrage Strategies for Cross - Track Betting on Major Horse Races," *Journal of Business* 63 (January 1990), 61- 78.

Hausch, D. B., and W. T. Ziemba, "Locks at the Racetrack," *Interfaces* 20 (May - June 1990), 41- 48.

Hausch, D.B., W.T. Ziemba and M. Rubinstein, "Efficiency in the Market for Race Track Betting", *Management Science* Vol. 27, No. 12 (1981), 1435-1452.

Isaacs, R., "Optimal Horse Race Bets", *American Mathematical Monthly* Vol. 60 May (1953), 310- 315.

Johnson, N.L., "Sequential Analysis: a Survey," *Royal Statistical Society Journal* Vol. 124 Part 3 (1961), 372-411.

Kendall, M.G. and W.R.Buckland, *A Dictionary of Statistical Terms,* Hafner, New York (1960).

Larsen, G.A. Jr. and G.D. Wozniak, "Market Timing Can Work in the Real World," *Journal of Portfolio Management* 21 (Spring 1995), 74-81.

Mood, A.M. and Graybill, F.A., *Introduction to the Theory of Statistics* second edition, McGraw-Hill, New York (1963).

Murphy, A.H. and B.G. Brown, " A Comparative Evaluation of Objective and Subjective Weather Forecasts in the United States," *Journal of Forecasting* Vol. 3 (1984) 369-393.

Rosenbloom, E.S., "Picking the Lock: a Note on "Locks at the Racetrack"," *Interfaces* Vol. 22 No. 2 (1992) 15-17.

Rosner, B. "Optimal Allocation of Resources In a Pari-Mutuel Setting", *Management Science* Vol. 21 No. 9 (1975) 997-1006.

Snyder, W. "Horse Racing: Testing the Efficient Market Model", *Journal of Finance* 33 (1978) 1109- 1118.

Steel, R.G.D. *Principles and Procedures of Statistics: a Biometrical Approach* McGraw-Hill, New York (1980).

Wald, A. *Sequential Analysis* John Wiley & Sons, London (1947).

Ziemba, W.T. and D.B. Hausch *Beat the Racetrack* Harcourt Brace Jovanovitch, San Diego (1984).

Ziemba, W.T. and D.B. Hausch *Betting at the Racetrack* Dr. Z Investments, Inc, Vancouver (1985).