Learning and Efficiency in a Gambling Market

James D. Dana, Jr. • Michael M. Knetter
Kellogg Graduate School of Management, Northwestern University, Evanston, Illinois 60208
Department of Economics, Dartmouth College, Hanover, New Hampshire 03755, and NBER

We present a statistical model which uses data on National Football League games and betting lines to study how agents learn from past outcomes and to test market efficiency. Using Kalman Filter estimation, we show that teams' abilities exhibit substantial week-to-week variation during the season. This provides an ideal environment in which to study how agents learn from past information. While we do not find strong evidence of market inefficiency, we are able to make several observations on market learning. In particular, agents have more difficulty learning from "noisy" observations and appear to weight recent observations less than our statistical model suggests is optimal. (Market Efficiency; Optimal Learning; Kalman Filter)

1. Introduction
The efficiency of asset markets is a widely studied empirical issue in the economics and finance literature. Unfortunately, tests of the efficient markets hypothesis, most of which focus on the stock market or the foreign exchange market, are hard to interpret since they are joint tests of market efficiency and a particular model of market equilibrium. Difficulties arise in part because many fundamentals may be unobservable to the econometrician, making the specification and identification of structural models problematic. Furthermore, there is no point in time at which market participants receive an objective signal about the fundamental value of the asset since most financial assets are infinitely-lived. Thus, it is difficult to make inferences about the learning behavior of investors from tests of market efficiency in asset markets. This paper proposes a method for studying pointspread gambles on NFL games which avoids many of the ambiguities of efficient market tests and enables us to draw inferences about market learning.

The market for pointspread gambles has a number of features that make it well-suited to studying how agents learn and testing efficiency. Most important, the market offers us the opportunity to learn about teams' abilities (the fundamentals) on the basis of objective, but noisy, weekly observations. Second, in contrast to other asset and wagering markets, pointspread wagers have similar horizons and identical risk and return characteristics, enabling us to test efficiency without estimating or making explicit assumptions about the shape of the utility function. Third, the weekly observations of relative team ability are quite noisy. This makes it difficult for market participants to process the information, increasing the importance of learning and the likelihood of market inefficiencies, as well as suggesting where inefficiencies may be found. Finally, other than the house cut, there are relatively low entry and exit costs for bettors in the market, so barriers to entry in the form of capital requirements could not explain unexploited profit opportunities in equilibrium.

Our starting point is a model in which each game outcome is a realization of a random variable whose


2 Ali (1977) has shown that wagerers overbet the longshot in horse racing, but this may reflect violations of the risk neutrality assumption as opposed to violations of market efficiency. See Thaler and Ziemba (1988) for an excellent survey of pari-mutuel betting markets.

3 Exit and entry barriers are much larger where betting on sporting events is illegal.
mean is determined by the true abilities of the teams and the home field advantage. In the most basic model, teams’ true abilities are unobservable but stationary, and can be estimated by ordinary least squares. However, we specify a more general model in which teams’ abilities follow a random walk over time, and we estimate them using the Kalman filter. The variance of the random walk component is estimated by maximum likelihood. Estimates of teams’ abilities are then used to predict future game outcomes and conduct betting simulations.

Two extensions of the model are also considered. First we account for unsystematic sources of noise in game outcomes, i.e., fumbles, interceptions, and penalties, that are measurable ex post. While there may be some systematic (team-specific) component to these variables, it is small relative to the noise component. Furthermore, these three sources of randomness explain over 40% of the variance in game outcomes. By controlling for these sources of noise in estimation, we obtain new estimates of team abilities and can analyze how bettors’ learn in the presence of unsystematic noise.

Second, we modify the model to account for lopsided outcomes. We define a lopsided outcome as one in which one team wins by a large margin. In these games, we hypothesize that the losing team may exert less than full effort as the game nears its end. Thus estimation of the basic model would yield biased estimates of teams’ true abilities (overestimating the winners’ abilities). We correct for this possible systematic effect by estimating a function that discounts the actual pointspreads used to estimate team abilities.

The extended model wins more than 50% of the games against the line, but we find no evidence of strong inefficiency. Nevertheless, we find that the market has more difficulty in processing information that contains a great deal of noise. The model outperforms the line more frequently in games in which the two teams have recently been involved, in games in which noise benefited one team much more than the other. Also, by comparing our model (which forecasts game outcomes) to a comparable model which attempts to forecast the line, we show that the market tends to discount noisy games somewhat less than is optimal and tends to under estimate the importance of time-varying abilities.

The next section of the paper contains a description of the market for pointspread gambles and a brief review of some of the previous investigations of efficiency in the NFL pointspread market. Section 3 describes the stationary and time-varying parameters models and our estimation techniques. Section 4 discusses the results of estimation and simulation for both models. In § 5 we describe two extensions to the time-varying parameters model and discuss the results of estimation and simulation of the extended models. In § 6 we look for sources of bias in the way that the market sets the line by comparing our model to one which uses maximum likelihood estimation to forecast the Las Vegas Line. Section 7 concludes the paper.

2. Market Efficiency Tests in the Pointspread Wagering Market

In this section we describe football gambling markets, define the efficient markets hypothesis in the context of these markets, and briefly survey the existing literature. The typical pointspread wager is based on the “11-for-10” rule, which requires that the bettor risk $11 in order for the chance to win $10. The “line” on a game specifies the favored team and the pointspread. A bettor placing a wager on the favorite wins the bet if the favorite wins by a margin of victory greater than the pointspread. A bettor placing a wager on the underdog wins the bet if the underdog loses by less than the pointspread or wins the game outright.

Locking in a certain return is taken to be the goal of the house, or sports book, which operates the market. By setting the line to evenly divide the amount bet between the two teams, the sports book earns slightly less than 5% of the total amount bet with certainty. As a result, the line can be thought of as the best forecast of bettor behavior, rather than the best forecast of the game outcome. However, market efficiency requires that there

---

4 The terms spread and pointspread will be used to refer to the number of points by which team A is favored over team B (possibly negative), and the terms game outcome and point differential will be used to refer to the actual difference between team A and team B’s scores. For consistency we will always normalize A to be the home team.

5 In the event the outcome is identical to the line, known as a “push,” the gambler’s wager is refunded.
are no systematic deviations between expected game outcomes and the line great enough to allow profitable arbitrage. Under the 11-for-10 rule, this implies that no betting strategy should win more than 52.4% of the time.

Let $VL$ denote the pointspread in the Las Vegas Line and let $PD$ denote the actual point differential between the two teams (defined in a way that is consistent with the definition of $VL$). Let $i$ and $j$ denote two different teams, and let $t$ denote the week. Then, in its strongest form, the efficient markets hypothesis states that

$$\text{Median } [PD_{ij} | \Omega_{t-1}] = VL_{ij}, \quad (1a)$$

where the set $\Omega_{t-1}$ is the set of all information available to the bettor prior to the game. If the distribution of game outcomes is symmetric (supported by Stern 1991) then (1a) implies:

$$E_{t-1}[PD_{ij} | \Omega_{t-1}] = VL_{ij}. \quad (1b)$$

A variety of efficiency tests have been performed based on (1a) and (1b). One natural taxonomy of these tests is based on the information contained in the set $\Omega_{t-1}$. In general $\Omega_{t-1}$ will contain the current lines, past lines, past outcomes, known game conditions (e.g., type of playing surface, home team), past game conditions, past game statistics (e.g., fumbles, yards rushing), other public information (e.g. injuries, referees), and private information. Stronger tests of market efficiency generally use more information in $\Omega_{t-1}$.

In addition to variation in the information set, previous work can be divided into three distinct approaches: (i) conducting simple statistical tests of the hypothesis that the line is an unbiased estimate of the game outcome, (ii) simulating behavioral models based on specific theories about betting behavior, and (iii) estimating forecasting models of game outcomes to generate a betting rule based on model forecasts.

The simplest statistical test of efficiency in the literature involves estimating the following model using OLS:

$$PD_{ij} = a_0 + a_1 VL_{ij} + \epsilon_{ij}, \quad (2)$$

where market efficiency corresponds to the joint hypothesis that $a_0 = 0$ and $a_1 = 1$. Pankoff (1968) was the first to implement this test, and Zuber et al. (1985), Gandar et al. (1988), and Sauer et al. (1988) have more recently reported the results of such tests for a variety of NFL seasons. Gandar et al. (1988) concluded that statistical tests are not powerful enough to detect inefficiencies. A serious problem with these tests (see Golec and Tamarkin 1991a) is that the way in which the line and the outcome are defined dictates the null and alternative hypotheses. For example, if the point differential and the line are defined as home score less visitor score, then the alternative hypothesis is that the line provides a biased estimate of the home field advantage. On the other hand, if the point differential and the line are defined as the score of the favorite less the score of the underdog, then the alternative hypothesis is that bettors' over- or under-value the favorite. Because the alternative hypotheses are so restrictive, these statistical tests are essentially equivalent to simple tests of bettor behavior.

Direct behavioral tests of market efficiency begin with specific hypotheses about the likely nature of bettors' biases. Vergin and Scriabin (1978) tested a number of rules that attempt to exploit potential forms of bettor irrationality. A good example is the following. Suppose fans have a bias towards betting on their local team. Then betting against teams with the largest fan following (e.g., New York Jets, New York Giants, or Dallas Cowboys) may be a profitable strategy. Vergin and Scriabin test this on data from 1969-1974 for Giants and Jets games and find the strategy to be profitable.

\[7\] Gandar et al. (1988) indicate their conclusion is consistent with Summers' (1986) conclusion regarding statistical tests of efficiency in stock markets. He simulates a model of stock prices in which market participants behave irrationally and demonstrates that statistical tests are not powerful enough to detect irrationality. In principle wagering markets ought to have a better chance of detecting irrationality, if it exists, since one can compare bettor behavior with objective information available about fundamentals. It is the absence of this type of feedback that may reduce the power of efficiency tests in other markets.

\[8\] Golec and Tamarkin (1991a) examine both NFL and NCAA college football games and find evidence of home and favorite bias in the line. Remarkably, the bias is smaller in college football games.
However, this rule does not hold up in other data samples (see Tryfos et al. 1984). Another approach to testing market efficiency is to specify a forecasting model, and simulate betting rules based on the forecasts of the model. Earlier work by Zuber et al. (1985) and Gandar et al. (1988) estimated a structural model of the relationship between game statistics and game outcomes which they used to forecast future game outcomes: season averages are used as forecasts of game statistics. Zuber et al. (1985) present evidence of market inefficiency using this approach. However, Sauer et al. (1988) were unable to produce similar results in subsequent seasons. Two more substantive problems with this method are first, that season averages may be poor forecasts of future game performance (especially if teams' abilities are changing over the course of the season) and second, the structural model may be misspecified since it is unlikely that game statistics are exogenous. Especially because of the latter problem, we feel that a better approach is one which uses game outcomes rather than game statistics to forecast team abilities.

3. The Model

We assume that the outcome of a game is determined by the following model:

\[ PD_{it} = c + \theta_i - \theta_j + \epsilon_{it}, \]  

Lacey (1990) also tests a number of behavioral betting rules and concludes the NFL gambling market is efficient. Coile and Tamarkin (1991b) introduce evidence that equivalent bets have unequal returns, which can be construed as inefficiency, although it is difficult to generate positive returns using their betting rules. Sauer et al. (1988) also demonstrate that the variables used by Zuber et al. (1985) in their structural model add virtually no information to that already imbedded in the Vegas line. In more recent work, Ruzzo et al. (1995) find that the Zuber model is profitable for several seasons. (They also show that no evidence of inefficiency is detected using the cross-equation restriction methods advocated by Abel and Mishkin (1983) in spite of the existence of the profitable betting rule. They interpret this finding as evidence of the low power of cross-equation tests.)

This is because a team's strategy is a function of the evolution of the game. For example, once a lead is established, teams have more incentive to select running plays. Establishing the lead may not be a consequence of the running game, however, so the direction of causality between yards rushing and game outcomes is unclear.

where \( PD_{it} \) is the actual point differential, i.e. outcome, of the game in week \( t \) between team \( i \) (home) and team \( j \) (visitor), \( c \) is a constant, \( \theta_i \) and \( \theta_j \) are the ability indices for the teams \( i \) and \( j \), and the errors, \( \epsilon_{it} \), are independently and identically distributed random errors. The outcome is defined as the difference between the home team's score and the visitor's score; if \( PD_{it} > 0 \) then the home team won. Ability indices are interpreted as a composite index of the offensive and defensive strength of each team measured in points. The difference between two teams' ability indices, \( \theta_i - \theta_j \), measures the number of points that team \( i \) is likely to win by if it plays team \( j \) on a neutral field. The constant, \( c \), represents the home field advantage.

We estimate a transformation of equation (3) (see the appendix for a description) by ordinary least squares for each week \( t \), of the season beginning with week 4, using all of the observations from the current season through week \( t \). The forecast for a game in week \( t + 1 \) is calculated as the difference in the two teams' ability indices as estimated in week \( t \) plus an allowance for the home field advantage, \( c \).

Next we consider a more general model in which teams' abilities may vary over the course of the season. This may arise, for example, because of injuries or other staff changes. The time-varying parameters model is given by the following two equations:

\[ PD_{it} = c + \theta_{it} - \theta_{jt} + \epsilon_{it}, \]

\[ \theta_{it} = \theta_{it-1} + \omega_{it}, \]

where \( k \) indexes teams and the errors \( \epsilon_{it} \) and \( \omega_{it} \) are assumed to be i.i.d. normal random variables with variance \( \sigma^2_t \) and \( \sigma^2_t \) respectively. Equation (5) implies that the \( \theta_{it} \) follow independently distributed random walks. A transformed version of the model is estimated using the Kalman filter given an estimate of \( \sigma^2_t \).

The variance, \( \sigma^2_t \), must be estimated by maximum likelihood. Under our assumptions game outcomes are normally distributed. The predicted values are given by

\[ \tilde{PD}_{it|t-1} = \tilde{\theta}_{it-1} - \tilde{\theta}_{jt-1}. \]

Because of data limitations we were not able to consider the possibility that the home field advantage or game variance varied by team.

Four or five weeks of data are generally required to econometrically identify the 28 team abilities.
where the hat is used to denote the Kalman Filter model estimates, and \( \hat{\theta}_{i,t-1} = \hat{\theta}_{t-1} \) are the forecasts of team abilities. The prediction errors are defined as \( \pi_{it} = \tilde{\pi}_{it} - \tilde{\pi}_{it,t-1} \). The transformation and exact specification of the model are discussed in the Appendix.

The assumption that the errors are normally distributed is clearly only an approximation. Note that the Kalman Filter estimation is consistent as long as the errors are i.i.d. However, we assume normality because it implies that the prediction errors are normally distributed and that we can use standard maximum likelihood techniques to estimate the variance of \( u \). If the distribution is not normal, then maximum likelihood estimation (under the assumption of normal errors) may be biased. The lumpiness of the scoring technology (scores are restricted to be integer combinations of 2, 3, 6, and 7) may result in nonnormal errors (of course the errors are also not continuous). Stern (1986) shows that outcomes appear to be symmetric around the Las Vegas line, which encourages us to think that our specification is a good approximation. Also, for some of our results (in particular, the efficiency tests) this is not a serious limitation.

4. Results of Estimation and Simulations

We use game outcomes and closing lines from Jim Feist's Football Workbook and game statistics from The Sporting News Pro Football Guide. We use data from the 1980 to 1989 seasons excluding 1982 and 1987, the two years in which the NFL players were on strike. Because of the long gap between years in the 1982 strike and the use of "scab" teams during the 1987 strike, we felt that our model, which assumes weekly shocks to team ability from a fixed distribution, would be misspecified. Lacking sufficient observations to get good estimates of the variance of separate shocks for these unique periods, we felt it best to ignore these seasons. Within each season, estimation begins in week 5 and simulations cover weeks 6 through 16. So simulations are performed on 154 games per season, or a total of 1232 games in 8 seasons.

Maximum likelihood estimation of the time-varying parameters model is performed on the 1983–1986 samples, and the results are reported in column (i) of Table 1. For comparison, the results of the stationary parameters model (estimated by OLS) are reported in column (0). The estimate of \( \sigma^2 \) is 1.57, which suggests that team abilities exhibit substantial temporal variation. Although the null hypothesis of no temporal variation is rejected at conventional levels of significance (the p-value of 0.82 implies a significance level of 18%), we find estimates of the variance that are significant at the 5% level in the alternative specifications reported below. Also, the variance of the prediction errors is significantly lower in the time-varying parameters model, indicating that this version of the model forecasts game outcomes more accurately.

The betting simulations are performed by tabulating the number of games in which the model forecasts correctly whether the game outcome will fall above or below the line. The forecasted ability is \( \hat{\theta}_{i,t-1} = \hat{\theta}_{t-1} \), where the hat is used to denote the model estimate, and the forecasted point differential for a game between home team \( i \) and visiting team \( j \) is given by Equation (6). In the simulations we place hypothetical wagers on the home team when \( \tilde{\pi}_{it} > \pi_{it} \) and on the visiting team when \( \tilde{\pi}_{it} < \pi_{it} \).

The simulation results for the stationary parameters model are presented in column (i) of Table 2. They are somewhat unsatisfactory. The model wins 580 games, loses 626, and 26 games are ties (or pushes), so the winning percentage is 48.1% excluding ties.

The results of the time-varying parameters model are reported in column (v) of Table 2. Because the variance parameter is estimated by maximum likelihood using four years of the sample, 1983–1986, betting simulations for this period are not genuinely out-of-sample simulations. We use the variance estimates for the 1983–1986 period to perform out-of-sample simulations for 1980–1981 and 1988–1989. The winning percentage is 49.9% of the 609 out-of-sample games, excluding ties. Neither set of simulation results provides evidence of inefficiency. Although the simulation results using the time-varying parameters model are slightly better than the stationary parameters model, the likelihood value

---

14 However, our criterion in estimating the variances was not choosing winning bets, but maximizing the likelihood of the game outcomes as a function of the previous week’s estimates of team ability.
Table 1 Maximum Likelihood Estimation of Time-varying Parameters Model: 1983-1986

<table>
<thead>
<tr>
<th>Stationary Parameters Model*</th>
<th>(i) Basic Model</th>
<th>(ii) Discounting</th>
<th>(iii) Noise Variables</th>
<th>(iv) Discounting and Noise Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated Parameters*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_4 / \phi_1$</td>
<td>0.018</td>
<td>0.064</td>
<td>0.035</td>
<td>0.077</td>
</tr>
<tr>
<td>$\theta$</td>
<td>8.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>0.25</td>
<td></td>
<td></td>
<td>9.12</td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td>0.39</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-174.78</td>
<td>165.25</td>
<td>48.58</td>
<td>99.24</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>3.03</td>
<td>3.11</td>
<td>3.47</td>
<td>2.96</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>1.57</td>
<td>1.56</td>
<td>1.74</td>
<td>1.48</td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>224.9</td>
<td>224.7</td>
<td>204.3</td>
<td>216.8</td>
</tr>
<tr>
<td>Log-Likelihood Value</td>
<td>-2772.11</td>
<td>-2771.19</td>
<td>-2741.22</td>
<td>-2751.93</td>
</tr>
<tr>
<td>Likelihood-Ratio Tests:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0: \phi_2 / \phi_1 = 0$</td>
<td>x² = 1.80</td>
<td>(p-value = 0.82)</td>
<td>x² = 6.00</td>
<td>(p-value = 0.22)</td>
</tr>
<tr>
<td>$H_1: [\phi_2 = \phi_1]$</td>
<td>(p-value = 0.97)</td>
<td></td>
<td>(p-value = 0.99)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>x² = 59.94</td>
<td>(p-value = 0.99)</td>
<td>x² = 26.92</td>
<td>(p-value = 0.09)</td>
</tr>
</tbody>
</table>

Notes:
* The stationary parameters model is shown for comparison only. It is not estimated using maximum likelihood.
* $\phi_2 / \phi_1$ is the ratio of the variance of the error in the transition equation (random walk error) to the variance of the measurement equation (game outcome error). See Appendix for details. $\phi_2$ and $\phi_1$ are the discounting parameters defined in equation (8).
* $\phi$ is calculated as the sum of squared prediction errors divided by $n - 1$.

is only slightly smaller. However, when we consider two extensions of the model below, we find that the time-varying parameters specification statistically outperforms the stationary parameters model.

5. Extensions and Additional Results

In this section we consider two extensions to the time-varying parameters model that are aimed at detecting the possibility of inefficiency arising from two complications for learning: lopsided outcomes and mistake-riddled games. In each case, the game outcome is likely to be a biased or less efficient indicator of the relative strengths of the teams. We propose methods to account for the effect of each of these factors on the ability estimates and then evaluate whether the revised estimates improve our forecasts.

5.1. Discounting

The method which we use to control for lopsided games is to estimate a discounting function, $\gamma$, which describes the relationship between the actual score, and the predicted score plus error. In the Kalman Filter estimation, equation (4) is replaced with the following equation:

$$\gamma(PD_{ij}) = c + \theta_{ij} - \beta_{ij} + \epsilon_{ij}. \tag{7}$$

We assume that $\gamma(PD_{ij})$ is symmetric about the origin, that is, $\gamma(PD_{ij}) = -\gamma(-PD_{ij})$.

There are several justifications for treating lopsided games differently. First, if the losing team simply gives up early in a game, then there may be a systematic component of the error that is correlated with the expected game outcome. Second, if both teams let up once the game has been effectively decided then there may be a great deal more noise introduced into the final game.
outcome. In each case the informational value of an additional point is smaller the greater is the margin of victory, which suggests that we should put less weight on games with lopsided outcomes. However our methodology may be more consistent with the former explanation.

For simplicity of estimation, we selected a somewhat restrictive class of piecewise-linear discounting functions of the form:

$$
\gamma(PD) = \begin{cases} 
PD & \text{if } PD \in (-a, a), \\
 a + b(PD - a) & \text{if } PD > a, \\
-a + b(PD - a) & \text{if } PD < -a.
\end{cases}
$$

(8)

with slope one on the interval $(-a, a)$ and slope $b$ elsewhere. Although we would have preferred to estimate continuous discounting function, this function captures the desired nonlinearity while introducing the minimum number of new parameters.15

15 An anonymous referee suggested we generalize the discounting function to allow for other than unit slope in close games. The estimated slope for close games using the general model was 0.88. A likelihood ratio test indicated that the data accept the restriction of unit slope imposed by our simpler function, so we chose to proceed with the more restrictive model.

A problem with this approach is that the assumptions that the prediction errors are normal and that the errors in equations (4) and (5) are normal are no longer consistent. If the prediction errors are normally distributed then the errors in equation (4) may not even be i.i.d. This raises the possibility that the Kalman Filter estimates may be biased. This problem is difficult to rectify, since we are restricted to estimating models that are linear in the parameters that are estimated using the Kalman Filter. Note, however, that biased estimates are not a problem for statistical inference in any of our simulations.
umn (ii) of Table 1. The kink in the discounting function is estimated to occur at 8.3 points, implying no discounting of scores for games with margins of victory within a touchdown. The slope is estimated to be one-quarter of a point (0.25), a substantial reduction in the importance attached to margins in excess of one touchdown. Note that the log-likelihood value is significantly higher than in the time-varying parameters model without discounting. The hypothesis that there is no discounting function is rejected at the 99% confidence level. This improvement substantiates our claim that the information value of an additional point scored is smaller in lopsided outcomes. The estimated variance of the random walk error changes very little.

5.2. Noise Variables
The second extension of our model is to include additional explanatory variables in the model which are best described as “noise variables.” We hypothesize that fumbles lost, interceptions, and the number of penalties are variables which significantly affect the outcome of the game, but result largely from randomness. If this is true, then including them in the model will yield more precise estimates of team abilities.

This approach relies on two characteristics of the noise variables. First, these variables are highly correlated with game outcomes, and can statistically explain nearly half of the variance in game outcomes. Second, these variables are only weakly correlated with team abilities. The first characteristic implies that including the noise variables will significantly increase the efficiency of our estimation. The second implies that the risk of introducing bias is very small.

In the regression model, we include the net number of fumbles lost (by the home team), the net number of penalties, and the net number of interceptions as exogenous right-hand-side variables. However, for forecasting and in the maximum likelihood calculations, we assume that the predicted value of each of these variables is zero.

The maximum likelihood estimation of this model is reported in column (iii) of Table 2. Perhaps the most significant observation is that 40% of game variance is explained by these three variables: the estimate of the game variance in the time-varying parameters falls from 165.3 to 98.2 when fumbles, interceptions, and penalties are included. However, it is the improvement in the log-likelihood value which validates our treating these variables as noise in our model. Clearly, the increase in the accuracy of our estimates outweighs any possible bias that is introduced.

In column (iv) of Table 1 we report the results of maximum likelihood estimation for the model with both discounting and the noise variables included. Some additional improvement in the log-likelihood value is obtained. Interestingly, the noise variables explain only about 20% of the variance when the discounting procedure is used. This suggests that noise is greater in lopsided games, so that discounting for runaway games already corrects for some systematic noise.

5.3. Simulation Results of the Extended Models
In Table 2 we report the simulation results for the time-varying parameters model and the stationary parameters model with each of the extensions described above. Although Table 1 demonstrates that the time-varying parameters model is a better specification, it is clear from Table 2 that neither model is capable of systematically beating the line without further conditioning information. In the time-varying parameters model, adding discounting increases the out-of-sample performance of the simulations from 49.9% to 52.0%, while the inclusion of noise variables actually reduces the out-of-sample performance to 48.5%. Similar results are obtained when these extensions are applied to the stationary parameters model. The qualitative implication is that lopsided games pose more difficulty for revising assessments of team ability than do games with a large imbalance in mistakes. Although we find some evidence that the discounting model wins more than 50%, this difference is not statistically significant. More importantly, none of the models have an overall winning percentage that exceeds 52.4%, which is the requirement for earning arbitrage profits.

To better test whether market participants have difficulty extracting the appropriate signal from games with
a large imbalance in mistakes we consider a filtering mechanism to select games in which the betting strategy is most likely to be profitable. The objective is to select games in week \( t \) which involve teams that had relatively different experiences with turnovers and penalties in the previous week. For each game we use the parameters of the model and the actual number of fumbles lost, interceptions, and penalties to calculate the net number of points that can be attributed to randomness (NPAR). For games in week \( t \), we calculate the difference, DPAR, between the home and visiting teams' values of NPAR from week \( t - 1 \), and use it as our filter. Large values of DPAR occur when one of the two teams experienced much better "luck" than the other in the previous week (i.e. one team benefited much more from randomness).

The results of filtered sample simulations are reported in Table 3 for each of the models which includes the noise variables. The winning percentages tend to increase, although not monotonically, as the sample is restricted to games for which DPAR takes on higher values. This general pattern holds for all specifications. This result suggests that bettors have more difficulty extracting the appropriate signal from particularly noisy observations. In some cases the winning percentage exceeds 52.4%, the margin required for profitability. Note that the significance levels reported in the table are not independent since each row contains progressively fewer observations.

### 6. Sources of Learning Inefficiency

In this section we explore the implications of our model for market learning. In particular we compare our maximum likelihood estimates to those that are implied by the Las Vegas line. The procedure we use is to estimate the maximum likelihood model by maximizing the likelihood of the Vegas Line as opposed to the actual point differentials. That is, we look for estimates of the variance and the discounting parameters that enable the time-varying parameters model to best predict the lines.

<table>
<thead>
<tr>
<th>Time-Varying-Parameters Model</th>
<th>Stationary-Parameters Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Filter Cutoff:  
- DPAR ≥ 0  
- DPAR ≥ 3  
- DPAR ≥ 6  
- DPAR ≥ 9  
- DPAR ≥ 12

<table>
<thead>
<tr>
<th>Filter Cutoff</th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPAR ≥ 0</td>
<td>48.4</td>
<td>51.9</td>
<td>50.2</td>
<td>50.9</td>
</tr>
<tr>
<td>(n = 609)</td>
<td>(n = 609)</td>
<td>(n = 1232)</td>
<td>(n = 609)</td>
<td></td>
</tr>
<tr>
<td>DPAR ≥ 3</td>
<td>47.3</td>
<td>52.2</td>
<td>50.5</td>
<td>51.1</td>
</tr>
<tr>
<td>(n = 476)</td>
<td>(n = 435)</td>
<td>(n = 958)</td>
<td>(n = 421)</td>
<td></td>
</tr>
<tr>
<td>DPAR ≥ 6</td>
<td>48.5</td>
<td>53.2</td>
<td>52.2</td>
<td>50.0</td>
</tr>
<tr>
<td>(n = 373)</td>
<td>(n = 295)</td>
<td>(n = 736)</td>
<td>(n = 300)</td>
<td></td>
</tr>
<tr>
<td>DPAR ≥ 9</td>
<td>48.3</td>
<td>54.6</td>
<td>51.2</td>
<td>47.6</td>
</tr>
<tr>
<td>(n = 292)</td>
<td>(n = 185)</td>
<td>(n = 539)</td>
<td>(n = 185)</td>
<td></td>
</tr>
<tr>
<td>DPAR ≥ 12</td>
<td>49.5</td>
<td>53.3</td>
<td>53.9</td>
<td>54.0</td>
</tr>
<tr>
<td>(n = 212)</td>
<td>(n = 92)</td>
<td>(n = 386)</td>
<td>(n = 100)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(α = .26)</td>
<td>(α = .06)</td>
<td>(α = .21)</td>
<td></td>
</tr>
</tbody>
</table>

* \( \alpha \) denotes the marginal significance levels for a one-tailed test using the normal approximation to the binomial distribution and a binomial parameter of 50%. \( \alpha \) is reported only for simulation statistics greater than 50%.
The results of this estimation are reported in Table 4. The estimate of the variance of the random walk error is extremely small in every model. For example, in the basic time-varying parameters model the estimate is equal to 0.15, an order of magnitude smaller than the estimate obtained using the data on outcomes. We conclude that instead of putting greater weight on recent outcomes than on past outcomes, the market seems to weight all outcomes in a season equally (to a close approximation). This suggests that bettors tend to use season averages when determining how to place their bets, and do not attach particular significance to teams' most recent performances. It also suggests that the market does not believe in the "hot hand" hypothesis (see Camerer 1990). The finding is somewhat surprising in light of the large literature on excess volatility of prices in stock and foreign exchange markets. The short holding period and the absence of "macro" shocks in gambling markets may eliminate the instability that many believe exists in other asset markets.

The results of the estimation of the discounting model suggest that the market does discount lopsided outcomes, but in a different way. Although the intercept is lower, 5.4 versus 8.3, the slope is much greater, 0.37 versus 0.25. This implies that the market discounts outcomes between 5.4 and 24.1 points less than our model, but discounts outcomes over 24.1 points more than our model.

A third factor in which we are interested is the noise variables. At this time we do not know whether the market puts more or less weight on the noise variables than our model. However, the filtered sample simulations suggest that the market is not incorporating the noise information efficiently, and it appears likely that the inefficiency arises because the market treats the noise as a true indicator of ability, and not as randomness.

To evaluate the importance of these differences in the discounting function estimated using lines rather than outcomes, we performed a likelihood ratio test. In the model estimated using outcomes, we rejected the hypothesis that the kink in the discounting function was 5.4 and the slope was 0.37, the estimates obtained from estimation using the lines. This suggests that the differences in behavior implied by the estimates are significant.
7. Conclusion

We present a statistical model which uses data on National Football League games and betting lines to study how agents learn from past outcomes and to test market efficiency. Using Kalman Filter estimation, we show that teams' abilities exhibit substantial week-to-week variation during the season. This provides an ideal environment in which to study how agents learn from past information.

Contrary to much published research on this topic, we find scant evidence that agents are incapable of formulating good forecasts, as summarized by the Las Vegas line. However, we find some indication that learning is inhibited both by runaway games and by the presence of an unusually large number of turnovers or penalties. When betting simulations are restricted to the subset of games which involve teams that had unusually noisy observations in the previous week, the model earns positive profits over a large sample. The approach also enables us to study how agents learn over time in comparison with how the model updates estimates of team ability. The behavior of market forecasts over time suggests that the market's assessment of team abilities exhibits less week-to-week variation than we find is optimal.

The lessons drawn from this research might carry over to research on other asset markets. Indications that bettors overreact to noisy outcomes and put too much weight on past relative to recent performance when forming expectations about the future suggests directions for efficiency tests on stock price data. For example, investors may overreact to random events that affect the current balance sheet of firms which do not affect the firms' future earnings potential.19

19 The authors are grateful to Carsten Kowalczyk, Mark Hooker, Skip Sauer, Steve Venti, Frank Wolak, Alex Zanello, and two anonymous referees for helpful comments and discussion. We would especially like to thank Jon Ezrow for valuable insights and expert research assistance. Remaining errors are our responsibility.

Appendix

Stationary Parameters Model

We estimate the following equivalent transformation of equation (3):

\[ y_t = c + D \theta + \epsilon_t, \]  
(A1)

where \( y_t \) is a \( 14 \times 1 \) vector of the outcomes of the games played in week \( t \), defined as the difference between the home team's score and the visiting team's score, \( c \) is a constant equal to the home field advantage, \( \theta \) is a \( 27 \times 1 \) vector of team ability indices, \( D \) is a \( 14 \times 27 \) matrix of dummy variables where \( d_{ij} = 1 \) if team \( j \) is the home team in the \( i \)th game, \( d_{ij} = -1 \) if team \( j \) is the visiting team in the \( i \)th game, and \( d_{ij} = 0 \) otherwise, and \( \epsilon_t \) is a \( 14 \times 1 \) vector of i.i.d. random variables with variance \( \sigma^2_\epsilon \). Because we have included a common intercept term, reflecting the home field advantage, one of the 28 teams is dropped from the matrix \( D \). Thus, \( \theta \) measures the ability of a team \( i \) relative to the ability of the omitted team (we chose to eliminate the Minnesota Vikings). This approach is equivalent to estimating all 28 team abilities while imposing the restriction that the average of the team ability estimates must be equal to zero.

Time-Varying Parameters Model

We transform (4) and (5) into the following measurement and transition equations:

\[ y_t = Z_t \theta_t + \epsilon_t, \]  
(A2)

\[ \theta_t = \theta_{t-1} + \eta_t, \]  
(A3)

where \( Z_t \) is a \( 14 \times (r + 27) \) matrix equal to \([X_t, D_t]\), the state variables, \( \beta_t = (c, \alpha, \theta_t)' \), are composed of \( c \), an \( r \times 1 \) vector of stationary parameters, \( \alpha \), a vector of stationary parameters associated with \( X_t \), and \( \theta_t \), a vector of time-varying ability indices associated with \( D_t \), \( X_t \) is a \( 14 \times r \) matrix of exogenous variables (noise variables) including a constant, \( D_t \) is defined as above, \( r \) is the number of exogenous variables including the constant (1 in the basic time-varying parameters model and 4 in the noise variables model), \( \epsilon_t \) is a \( 14 \times 1 \) vector of i.i.d. random variables with variance \( \sigma^2_\epsilon \), and \( \eta_t \) is an \((r + 27) \times 1 \) vector of errors with covariance \( \sigma^2_\eta Q \), where

\[ Q = \begin{bmatrix} 0 & 0 \\ 0 & W \end{bmatrix} \]  
(A4)

is an \((r + 27) \times (r + 27)\) and \( W \) is a \( 27 \times 27 \) matrix with elements \( w_{ii} = 1 \) if \( i = j \), and \( w_{ij} = 1/2 \) otherwise. Note that the errors in equation (A3) are related to the errors in equation (5) by the equation, \( \eta_t = \eta_{t-1} - \eta_{t-1} \), for all \( i > r \), where \( \eta_{t-1} \) is the error associated with the omitted team. Therefore, \( \sigma^2_\epsilon = 2 \sigma^2_\eta \).

The estimation of (A2) and (A3) has three components. First, we estimate initial parameters or starting values. We use the method proposed by Cooley and Prescott (1976), suggested as a means of initialization by Harvey (1981), in which the initial values are estimated using GLS. This method yields the most efficient estimates since it incorporates all of the information about the error structure. We estimate the GLS model on the first \( k = 4 \) weeks of data in each season, which gives us estimates of the parameters, \( \beta_t \), and their covariance matrix, \( P_t \).

Second, we estimate (A2) and (A3) iteratively using the Kalman Filter, assuming that the relative value of the variances \( \sigma^2_\epsilon \) and \( \sigma^2_\eta \) is known. The parameter estimates, \( \beta_t \), and their variance-covariance matrix, \( P_t \), are estimated using the following iterative update equations:
\[ \hat{\beta} = \hat{\beta}_{oner} + P_{oner} Z_{oner} \gamma \]

\[ P_t = P_{oner} + (\chi_t^2 / \nu_t^2) Q_t + P_{oner} Z_{oner}^2 P_{oner} \]

where

\[ F_t = Z_{oner} Z_{oner}^2 + I_{oner}. \]

The prediction errors, \( \nu_t \), are defined as the residuals, \( \nu_t = y_t - Z_{oner} \hat{\beta} \), and the covariance matrix of the prediction errors is given by \( E(\nu_t \nu_t^\prime) = \sigma F_t \). By construction, the update equations are independent of the variance \( \sigma^2 \) and depend only on the ratio of the variances, \( \sigma^2 / \sigma^2_t \).

Finally, we estimate the ratio of the variances, \( \sigma^2_t / \sigma^2 \), using maximum likelihood. The prediction errors are normally distributed since, by assumption, \( \epsilon_t \) and \( \nu_t \) are normally distributed. The multi-

\[ L(y_{t1}, \ldots, y_T) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{T/2}(\sigma^2)^{1/2}} \exp \left( -\frac{1}{2\sigma^2} y_t^2 \right). \]

Differentiating with respect to \( \sigma^2 \) yields the first-order condition

\[ \frac{\partial L}{\partial \sigma^2} = \frac{1}{N(T-k)} \sum_{t=1}^{T} y_t^2 / \sigma^2 \]

which is our estimate of \( \sigma^2 \). The ratio \( \sigma^2 / \sigma^2_t \) is found by numerically maximizing the log of the likelihood function, given by

\[ \log L(y_{t1}, \ldots, y_T) = -\frac{1}{2} N(T-k) \log 2\pi - \frac{1}{2} N(T-k) \]

\[ \times \log \left( \frac{1}{N(T-k)} \sum_{t=1}^{T} y_t^2 / \sigma^2 \right) \]

\[ -\frac{1}{2} \sum_{t=1}^{T} \log |F_t| - \frac{1}{2} N(T-k) \]

(10)

where \( \sigma^2 / \sigma^2_t \) is incorporated into the definition of \( F_t \), and equation (9) has been used to substitute for \( \sigma^2_t \). Since the first and fourth terms of (10) are constants, only the 2nd and 3rd terms need to be calculated.

References


Accepted by William T. Ziemba, former Departmental Editor; received April 1992. This paper has been with the authors 6 months for 3 revisions.