Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Points Gained in Football: Using Markov Process-Based Value Functions to Assess Team Performance

Timothy C. Y. Chan , Craig Fernandes , Martin L. Puterman

informs
http://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 69, No. 3, May–June 2021, pp. 877–894
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Methods

Points Gained in Football: Using Markov Process-Based Value Functions to Assess Team Performance

Timothy C. Y. Chan,[a] **Craig Fernandes,**[a] **Martin L. Puterman**[b]

[a] Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada; [b] Sauder School of Business, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada

Contact: tcychan@mie.utoronto.ca, https://orcid.org/0000-0002-4128-1692 (TCYC); craig.fernandes@mail.utoronto.ca, https://orcid.org/0000-0001-6508-6755 (CF); martin.puterman@sauder.ubc.ca, https://orcid.org/0000-0001-7866-4892 (MLP)

Received: October 18, 2018
Revised: August 4, 2019; February 21, 2020
Accepted: April 1, 2020
Published Online in Articles in Advance: January 20, 2021

Subject Classifications: probability: Markov processes; recreation and sports; dynamic programming: Markov
Area of review: Stochastic Models

https://doi.org/10.1287/opre.2020.2034

Abstract. To develop a novel approach for performance assessment, this paper considers the problem of computing value functions in professional American football. We provide a theoretical justification for using a dynamic programming approach to estimating value functions in sports by formulating the problem as a Markov chain for two asymmetric teams. We show that the Bellman equation has a unique solution equal to the bias of the underlying infinite horizon Markov reward process. This result provides, for the first time in the sports analytics literature, a precise interpretation of the value function as the expected number of points gained or lost over and above the steady state points gained or lost. We derive a martingale representation for the value function that provides justification, in addition to the analysis of our empirical transition probabilities, for using an infinite horizon approximation of a finite horizon game. Using more than 160,000 plays from the 2013–2016 National Football League seasons, we derive an empirical value function that forms our points gained performance assessment metric, which quantifies the value created or destroyed on any play relative to expected performance. We show how this metric provides new insight into factors that distinguish performance. For example, passing plays generate the most points gained, whereas running plays tend to generate negative value. Short passes account for the majority of the top teams' success and the worst teams' poor performance. Other insights include how teams differ by down, quarter, and field position. The paper concludes with a case study of the 2019 Super Bowl and suggests how the key concepts might apply outside of sports.

Keywords: Markov reward process • Bellman equation • value function • points gained • American football

1. Introduction

A major stream of sports analytics research seeks to attribute a value to game states in order to assess performance and develop strategy. In American football, the game state may correspond to the down and yard line; in golf, the distance from the hole and the terrain; or in baseball, the inning, number of runners on base, score differential, and number of outs. This paper develops a value function for professional American football using Markov process theory and National Football League (NFL) data. Although this paper should be of considerable interest to sports enthusiasts, it also provides new ways of looking at dynamic programming-based value functions that may lead to novel insights beyond sports.

The goal of this paper is to develop a novel metric for performance assessment in football, using a dynamic programming-based value function that has a rigorous theoretical underpinning. Our approach to modeling football can be seen as an infinite horizon discrete time dynamic Markov reward process (Puterman 1994). We begin by defining the state space to be all possible down and yard line combinations and the reward to be expected number of points scored from each state. Although more general state descriptors are typically used when the goal is tactical decision making, a restricted state space is appropriate when the goal is performance assessment and the determination of how value is generated in football, which has received limited attention in the literature. In addition, our smaller state space helps to ensure that we have enough data in each state to derive reasonable estimates of the transition probabilities.

Our data consists of the start and end state for 164,299 plays over the 2013–2016 regular seasons in the NFL. From this data we derive empirical distributions for the transition probabilities between states. We then compute the value function by directly solving a derived set of evaluation equations. These equations model the natural transitions between consecutive plays in a game, resulting in a Bellman equation characteristic of Markov reward processes.

Our modeling approach employs a Markov reward process, in which there is no action undertaken in any state, in contrast to a Markov decision process. Since our goal is performance assessment, we use the entire data set to generate a single transition probability matrix for the league, which can be thought of as representing average performance. A value function computed using these transition probabilities thus represents expected performance across the league, which facilitates the assessment of individual team performance relative to this league-wide baseline. Thus, instead of considering team-specific policies as part of a Markov decision process, we use a Markov reward process to develop a league-wide value function that serves as a benchmark against which individual plays, and teams by extension, can be measured.

Although the concept of a value function in football has been proposed previously (Winston 2009, Yurko et al. 2019), we are the first to provide a rigorous theoretical foundation that establishes that the system of evaluation equations has a unique solution equal to the bias of the underlying Markov reward process. In contrast, previous studies did not attempt to provide a theoretical underpinning for their value function, or even demonstrate uniqueness of their value function. Instead, they simply propose a model and generate numerical results in lieu of a deeper exploration of the value function itself. Our fundamental representation of the value function enables a rigorous interpretation of it in the football context as the expected number of points gained/lost over and above the steady state points gained/lost per play. Additionally, solving these equations enables us to obtain performance insights through the computed value functions. Finally, drawing inspiration from the work of Broadie (2012) in professional golf, we develop a performance assessment metric we refer to as *points gained* using value function differences to obtain insight into how different teams generate value from their plays.

This paper makes the following contributions:

1. We provide the first general theoretical justification for the dynamic programming approach to estimating value functions in sports. We prove that the standard approach to deriving the value function results in a unique solution and that this unique value function is equal to the bias of the underlying Markov reward process. We also provide a martingale interpretation of our value function, which connects it to literature computing value functions based on individual possessions, rather than entire games or seasons.

2. We define a metric, points gained, as the difference between the actual outcome of a play and the expected outcome using the Bellman equation. Applied to extensive NFL data, the points gained metric provides insight into which play types and game states generate the most value for each team as compared with the league average. Our findings include: (a) passes generate the most points gained by far, whereas running plays tend to generate negative value; (b) among passing plays, short passes account for the majority of the top teams' success and the worst teams' poor performance; and (c) the best teams distinguish themselves in the middle of the field, between the two red zones.

3. We present a case study using points gained to analyze the 2019 Super Bowl. Our analysis suggests that New England's advantage over Los Angeles was accumulated primarily on second downs, and attributable to both passing and running plays, as well as penalties.

2. Related Literature

This paper relates to two bodies of sports research: (1) dynamic programming and (2) football value function derivation and interpretation.

Dynamic programming has been widely applied in sports to solve tactical problems and compare teams or players. For instance, Washburn (1991) built on previous work (Morrison 1976, Morrison and Wheat 1986, Erkut 1987) and utilized a Poisson process model and dynamic programming to suggest that National Hockey League teams trailing by one goal late in the game should be pulling their goalies much sooner than standard practice. Boronico and Newbert (1994) utilized stochastic dynamic programming to evaluate optimal offensive/defensive strategies for 1st and goal scenarios in the NFL. Fry et al. (2007) employed a deterministic dynamic programming approach to model the sequential decisions a coach must make during a draft in order to maximize the total draft value. Chan and Singal (2016) used dynamic programming to determine an appropriate handicap between two tennis players of unequal ability in order to generate fair matches. Most closely related to this paper, Broadie (2008, 2012) used dynamic programming to generate a value function for every position on a golf course. Utilizing these values, he created a strokes gained metric that identifies strengths and weaknesses in a golfer's game and what factors contribute most to a golfer's success. Golf was modeled as a finite horizon game, in contrast to the infinite horizon model we develop in this paper. Also, the definition of our points gained metric flips the terms related to actual and expected points, since more points is better in football whereas fewer strokes is better in golf.

In football, several papers have constructed value functions using NFL data. Carter and Machol (1971) used data from the first half of 59 games in the 1969 regular season. Due to their limited data set, they only determined values for first and 10 situations at 10 different yard lines (5, 15, 25, . . . , 95). In an unpublished manuscript, Cabot et al. (1981) extended these results by determining values for every combination of yard line, down and yards to go. Bell (1982) used a similar approach to Carter and Machol (1971) to analyze strategy and performance in the Canadian Football League. Romer (2002) expanded Carter and Machol's work by using first quarter data from the 1998–2000 regular seasons and determined values for first and 10 situations on each yard line. This analysis suggested that NFL coaches should be more aggressive on fourth down, even if they are close to their own end zone. Winston (2009) used a value function to answer several tactical questions including when to "go for it" on fourth down, whether to accept penalties, and when to try for a field goal. He also created a point value added metric that measures the value of individual plays. Goldner (2012, 2017) utilized empirical transition probabilities to build an absorbing Markov chain model of a football drive. He computed the absorption probabilities for each scoring event and then determined the expected points for each state composed of the down and five-yard buckets for yard line and yards to go. Pelechrinis et al. (2019) utilized a value function approach to determine the value of plays adjusted for the offensive and defensive strengths of each team. Lastly, Yurko et al. (2019) created a multinomial logistic regression model to generate a value function to evaluate plays based on their impact on expected points and win probability. Their focus was to generate a wins above replacement metric to break down individual player contributions to team wins. Almost all research on value functions uses them for tactical decision making. In contrast, our goal is to understand how teams create value. The few papers that do focus on value creation do not complement their numerical results with a theoretical justification for their approach.

3. Value Function Derivation and Interpretation

In this section, we provide a rigorous foundation for computing and interpreting football value functions. We model a game between two teams, A and B. We assume an infinitely long game where the time index n corresponds to the play number (excluding points after touchdown). Let element s in the state space S represent the state of the game from the perspective of the team with possession of the ball. Although our development in this section is general, to help make things concrete, consider the state s being composed of the down d, yard line y, and team with possession t. Let P denote the matrix of transition probabilities between states and the vector r the expected reward received on a play starting in a designated state. As a convention, we assume the expected points are from the perspective of team A so that when team B possesses the ball, the expected rewards will be negative. The value function, which we interpret later, will be represented by a vector v. Subvectors and submatrices will be identified by subscripted versions of the corresponding vectors and matrices. Our results are based on the fundamental literature on undiscounted Markov reward processes, appearing in chapter 8 of Puterman (1994). Recall that in a Markov reward process, unlike a Markov decision process, there is no decision being undertaken in any state.

3.1. The General Case

We first consider the case where the two teams may have different transition probabilities between states. The transition probability matrix P can be written as a partitioned matrix in the following form:

$$P = \begin{bmatrix} P_{AA} & P_{AB} \\ P_{BA} & P_{BB} \end{bmatrix}. \tag{1}$$

Submatrices P_{AA} and P_{BB} correspond to transitions where team A and team B retain possession of the ball, respectively. The other two submatrices correspond to changes of possession. Note that this representation implicitly suggests that the team with possession is encoded into the state space. The expected reward vector r, assumed bounded, can be written as follows:

$$r = \begin{bmatrix} r_A \\ r_B \end{bmatrix}. \tag{2}$$

Although the actual number of points does not vary by team on a scoring play, the expectations above differ because transition probabilities vary by team. This can be seen in the next section when we explicitly write out the evaluation equations in component notation.

Since there is a positive probability of reaching any other state from any starting state, all states of the underlying Markov chain are recurrent. Thus, the limiting matrix $P^* := \lim_{n\to\infty} P^n$ exists, has positive entries for all states, and has equal rows. An entry $P^*_{s,t}$ represents the long run probability of occupying state t for any starting state s. When all states are recurrent, the *gain* of this Markov reward process, $g = P^* r$, is a constant vector that represents the average points gained (if component-wise positive) or lost (if component-wise negative) per play over the infinite horizon. Equivalently, g can be viewed as the steady state reward per play. Theorem 8.2.6 in Puterman (1994) establishes that $g = P^* r$ is the unique solution of the system of equations

$$v = r - g + Pv, \tag{3}$$

and the value function

$$v = \begin{bmatrix} v_A \\ v_B \end{bmatrix} \tag{4}$$

is unique up to a single additive constant. Equation (3) is also known as Poisson's equation. The extra condition $P^{*}v = 0$ uniquely determines v to be the bias or transient reward of the underlying Markov reward process. Note that even without this condition, the relative values $v(s) - v(t)$ are unique for any two states s and t. This fact is important because the empirical relative values are used in the definition of the points gained metric subsequently.

3.2. Interpretation of the Value Function

The bias of the Markov reward process may be written as follows:

$$v(s) = \mathbf{E}\left[\sum_{n=0}^{\infty}\big(r(X_n) - g\big) \mid X_0 = s\right], \tag{5}$$

where the random variable X_n denotes the state of the game at the nth play. Equation (5) provides a precise interpretation for the value function: $v(s)$ represents the expected number of points gained or lost over and above the gain (i.e., steady state points gained/lost per play) in an infinite number of plays starting in state s. In the language of Markov reward processes, the value function represents the expected total difference between the reward and the stationary reward, where the stationary reward is defined as the average reward per period in steady state. Since the steady state is reached quickly in a recurrent Markov chain, $v(s)$ may be regarded as the relative advantage or disadvantage of starting in state s. Equivalently, $v(s)$ is often referred to as the transient reward associated with starting in state s, which can be seen by noting (from equation (8.2.3) in Puterman (1994)) that

$$v = \sum_{n=0}^{\infty}(P^n - P^{*})r, \tag{6}$$

and that the terms in the sum converge to zero at the rate of the subdominant eigenvector of P.

Equation (8.2.5) in Puterman (1994) gives further insight into the meaning of $v(s)$. It shows that

$$v(s) = v_n(s) - ng + o(1), \tag{7}$$

where $v_n(s)$ denotes the expected total points over n plays starting in state s and $o(1)$ is an expression that converges to zero as $n \to \infty$. This means that after a sufficiently large number of plays, the value (or bias) is the expected number of points over or above the average gain or loss per play. Since convergence to steady state is fast (see Section 4.6), this approximation should be accurate after a relatively small number of plays.

3.3. A Martingale Perspective on the Value Function

Further insight into the meaning of the value function can be derived through martingale theory. From section 10.3 in Glynn (2013), it follows that

$$M_n := v(X_n) - v(X_0) + \sum_{j=1}^{n-1}\big(r(X_j) - g\big) \tag{8}$$

is a zero-mean martingale. Consequences of this result are the following:

1. The $\lim_{n\to\infty} M_n = 0$.
2. For any stopping time, τ, $\mathbf{E}[M_\tau] = 0$.
3. For any $s \in S$,

$$v(s) = \mathbf{E}[v(X_\tau)|X_0 = s] + \mathbf{E}\left[\sum_{j=1}^{\tau-1}\big(r(X_j) - g\big)|X_0 = s\right]. \tag{9}$$

Equation (9) shows that with a judicious choice of a stopping time, we can interpret $v(s)$ as an expected total excess reward until a specific event occurs plus the excess reward after the event. In particular, specifying τ to be the number of plays until the next kickoff by the team with possession, the expression $\mathbf{E}[v(X_\tau)|X_0 = s]$ is independent of s. In this case, $v(s)$ can be interpreted as the expected total excess reward until the next kickoff by the team with possession plus the value of a kickoff. Later we find that for the data under consideration and this specific choice of τ, $\mathbf{E}[v(X_\tau)|X_0 = s] = -0.63$ for all s, which is the expected value of a kickoff from the 35 yard line. Note that choosing a stopping time in which $\mathbf{E}[v(X_\tau)|X_0 = s]$ varies with s would not afford such a concise interpretation.

This martingale representation relates our choice of value function to research that computes expected values based on a single possession such as Yurko et al. (2019). This interpretation becomes even more meaningful when we use points gained to evaluate performance and the offsets of −0.63 cancel since we are computing the difference in value function at two states.

3.4. A Model with Identical Teams

We now consider the special case of two teams with identical transition probabilities. This is the model that will be the basis for data analysis in subsequent sections, where our objective is to derive a single, league-wide value function (see Section 4). This is the same approach taken by previous investigators, but without the formal justification we provide next. Since we consider identical teams, we assume that $P_{AA} = P_{BB}$ and $P_{AB} = P_{BA}$ and the expected reward vector $r_B = -r_A$, allowing us to omit the state variable that denotes the team with possession.

Lemma 1. *Let g be the gain in a model with identical teams. Then $g = 0$.*

Proof. Since the underlying Markov chain is recurrent, the limiting matrix P^* exists and can be written as follows:

$$P^* = \begin{bmatrix} Q & Q \\ Q & Q \end{bmatrix}, \tag{10}$$

where Q is a stochastic matrix. Since $g = P^*r$ and $r_A = -r_B$, $g = 0$. □

Applying Lemma 1 to Equations (5) and (7), we obtain:

$$v(s) = \mathbf{E}\left[\sum_{n=0}^{\infty} r(X_n) \mid X_0 = s\right] \tag{11}$$

and

$$v(s) = v_n(s) + o(1), \tag{12}$$

respectively. This means that $v(s)$ represents the total expected point advantage or disadvantage starting in state s. Since $g = 0$, it follows from Equation (3) that v defined in Equation (11) is the unique solution of $(I - P)v = r$ subject to $P^*v = 0$. Thus, the underlying stochastic system is a zero-gain Markov reward process and its bias equals its expected total reward (see section 10.4.1 in Puterman (1994)).

We next show that the equation

$$(I - P_{AA} + P_{AB})v_A = r_A \tag{13}$$

has a unique solution and this solution is a subvector of the bias defined in Equation (11). In our subsequent numerical results, this is the equation we solve to find empirical value functions. To begin, we write the system $(I - P)v = r$ in block form:

$$(I - P_{AA})v_A - P_{AB}v_B = r_A, \tag{14}$$
$$-P_{BA}v_A + (I - P_{BB})v_B = r_B, \tag{15}$$

which in the symmetric case is equivalent to

$$(I - P_{AA})v_A - P_{AB}v_B = r_A, \tag{16}$$
$$-P_{AB}v_A + (I - P_{AA})v_B = -r_A. \tag{17}$$

Lemma 2. *The system of Equations (16) and (17) has a solution (v_A, v_B) with $v_B = -v_A$.*

Proof. Adding Equations (16) and (17), we get

$$(I - P_{AA} - P_{AB})(v_A + v_B) = 0. \tag{18}$$

Hence, there exists a solution with $v_B = -v_A$. □

It is clear from Equation (18) in the proof of Lemma 2 that adding a constant vector to any solution of the system satisfying $v_B = -v_A$ will also result in a solution to the system. This result also follows from the previous subsection, since Equation (3) is equivalent to Equations (16) and (17), and it was shown previously that (3) has a unique solution up to an additive constant. In other words, any solution of the form $(v_A + c\mathbf{1}, -v_A + c\mathbf{1})$, where c is a scalar and $\mathbf{1}$ is a vector of 1's with the same dimension as P_{AA}, is a solution to this system.

Theorem 1. *The system of equations*

$$(I - P_{AA} + P_{AB})v_A = r_A \tag{19}$$

has a unique solution equal to the upper subvector of the bias, v, as defined by (11).

Proof. As noted previously, the bias is the unique solution of Equations (16) and (17) subject to $P^*v = 0$. Combining this result with the representation for P^* given by (10), it follows that for $(v_A + c\mathbf{1}, -v_A + c\mathbf{1})$ to be the bias, it must be that $c = 0$. Thus, since (16) is equivalent to (17), the upper subvector of the bias is the unique solution to Equation (19). □

What has been established is that in the identical team case, where the average reward per play is $g = 0$, the unique solution of Equation (13) has the representation given in (11). Hence, the value function is the total expected reward associated with starting in state s. In steady state, the team with possession will change infinitely often so the realized reward will alternate between offsetting positive and negative values. Thus, the value function is a combination of the transient reward associated with starting in state s until reaching steady state, plus the steady state reward, which should be zero in the case of identical teams.

4. Value Function Estimation

We now apply the results of the preceding section to compute empirical value functions. In particular, we consider a model with identical teams represented by average transition probabilities and expected reward vectors. Teams are not identical in practice, of course. But the identical teams model allows us to use all of the league-wide data to form a single transition probability matrix, which represents league-average performance. In other words, the value function is computed using average transitions over all plays for all teams during the time period being evaluated, and is not identified with any specific optimal policy. The resulting value function represents average performance, which facilitates performance assessment of individual teams relative to this baseline.

4.1. State Space

Recall that with identical teams, we represent the state by a pair (d, y) where d denotes the down and y

the yard line at the start of the play. We adopt the convention that a yard line is denoted by integers 1 to 99 relative to the goal line of the team with possession. A more complete state description is possible, including yards to go, score differential, time remaining, and other variables in addition to down and yard line. However, for our ultimate objective, which is to relate play outcomes to scoring, our simplified state representation suffices. The resulting state space S has just over 400 states. The combination of downs one to four and yard lines one to 99 constitute 396 possible states. Kickoffs (or punts) from the 20 (following a safety) and 35 (following a touchdown or field goal) yard lines, plus kickoffs at other nearby yard lines due to penalties, make up the remaining states. We do not explicitly model points after touchdown. Instead we include the average values in the expected reward functions corresponding to a touchdown.

4.2. Evaluation Equation

Let the vector V with components $V(d,y)$ denote the value of being in state (d,y). We use (13) to compute the value of starting a play at (d,y). Our identical team assumption means that the value of an offensive possession at state (d,y) equals the negative value of an opponent's possession at the same state. Thus, when possession changes through either a fumble, interception, punt, kickoff, missed field goal, safety, or a loss of possession on downs, the sign of V is negative. This can be seen by rewriting (13) as follows:

$$V = r_A + P_{AA}V + P_{AB}(-V). \tag{20}$$

This equation shows that V can be calculated by taking into account the values of all the different states to which the current play may transition, including changes of possession and the resulting points scored.

Before writing out the evaluation equation in detail, we require additional notation. Let $p_{(d,y),(d',y')}$ be the probability of transitioning from state (d,y) to state (d',y') on a single play. Let $p^{TO}_{(d,y),(d',y')}$ be the probability of a turnover for a play starting in state (d,y) and resulting in the defensive team receiving possession at state (d',y'). Let $p_{(d,y),TD}, p_{(d,y),FG}, p_{(d,y),SAF}, p_{(d,y),defTD}$ be the probabilities of scoring a touchdown, scoring a field goal, surrendering a safety, and having the defense score a touchdown on a turnover, respectively, when the play starts at down d and yard line y. We let the zeroth down denote a kickoff, meaning $V(0,20)$ and $V(0,35)$ are the value of kickoffs from the 20 and 35 yard lines, respectively. For scoring plays, field goals are worth three points and safeties are worth negative two points (two points for the defense). The value of a touchdown is six plus a possible extra zero, one or two points based on the type of extra points conversation play attempted and whether the attempt is successful. In our model, we use 6.97 as the value of a touchdown, since the average value of an extra points conversion attempt following a touchdown was 0.97 according to our data from 2013–2016, which is described later.

We now rewrite (20) in component notation as follows:

$$\begin{aligned} V(d,y) = &\sum_{(d',y')\in S} p_{(d,y),(d',y')} V(d',y') \\ &- \sum_{(d',y')\in S} p^{TO}_{(d,y),(d',y')} V(d',y') \\ &+ p_{(d,y),TD}(6.97 + V(0,35)) + p_{(d,y),FG} \\ &\times (3 + V(0,35)) + p_{(d,y),SAF}(-2 + V(0,20)) \\ &+ p_{(d,y),defTD}(-6.97 - V(0,35)), \\ &\qquad \forall (d,y) \in S. \end{aligned} \tag{21}$$

By Theorem 1, this system of equations has a unique solution, which equals the long run expected total reward starting in state (d,y).

The empirical probability of transitioning between two particular states serves as the basis for computing the transition probabilities. We use a bootstrapping approach by first identifying all plays that start in a particular state (d,y). Then we sample with replacement from these observations 1,000 times and record the proportion of times each next state, (d',y'), is drawn. These proportions serve as the transition probabilities from (d,y) to (d',y'). The (d',y') states can also be a result of penalties, where the starting and ending down may be the same or where the team could receive a new first down without advancing the ball 10 yards. This entire process is repeated for every down and yard line pair, resulting in a transition probability matrix between all possible states.

We use bootstrapping (Efron 1979) to obtain value function estimates and confidence intervals. We create 30 replicates of $V(d,y)$ by repeatedly generating transition probabilities and expected rewards by the process above and solving the system of Equations (21). We use the average of these replicates as our estimate of $V(d,y)$ and obtain empirical 95% confidence intervals for each $V(d,y)$.

4.3. Data

We obtained play-by-play data from www.NFLsavant.com for four NFL regular seasons (2013–2016). Our data set contained 1,034 games and 164,299 plays. For each play, we extracted the year, quarter, down, yards to go for first down, yard line, offensive team, defensive team, and play type. To prepare the data for modeling, we computed two new variables, *NextDown* and *NextYardLine*, which represent the down and yard line of the ensuing play. Note that similar data are now available through nflscrapR (Horowitz et al. 2019).

4.4. Value Function Results

Figure 1 shows the value function by down averaged over all four years of data. Not surprisingly, as the down increases, the value of that field position decreases. On average, going from first to second down with no gain in yardage results in a loss of roughly 0.4 points. Going from second to third down results in a loss of roughly 0.6 points. The biggest decrease is from third to fourth down, which results in a loss of 0.7 to 1.3 points, depending on field position. This larger decrease makes sense because the majority of fourth down plays are punts, which give the opposing team possession of the ball. The value function for fourth down increases sharply after the opponent's 40 yard line (offense's 60 yard line), which is when most teams enter field goal range. Since a missed field goal from this point places the opponent in an advantageous position, the gap between third and fourth down is largest around the opponent's 45 yard line.

Observe also that for plays starting near the opponent's goal line (the 100 yard line in our notation), the value is less than 6. In contrast, the value of a touchdown, which is the most likely outcome, is 6.97. This result is because in our model, the value of a touchdown is offset by the subsequent kickoff, which has a value of −0.63. This shows how using an infinite horizon dynamic model accounts for the evolution of the game beyond the current scoring play.

4.5. Comparison with Previously Published Values

As shown in Table 1, our value function appears most similar to that of Romer (2002), at least on first down. Compared with the other value functions, our values tended to be higher near a team's own end zone but lower toward the opponent's end zone. We believe this result is due to higher-scoring offenses in today's NFL. For example, average points per game increased from approximately 20 points in 2003 to 22 points in 2016. Similarly, the average points per drive increased from 1.68 to 1.97 over the same time period. Since teams are better at scoring in general, the value of being deep in one's own end is higher. By symmetry, the value of being deep in the red zone (yard lines 80 to 99 in our notation) is lower because after a team scores, the value of the other team having the ball at their end of the field is greater.

4.6. Quality of the Finite Horizon Approximation

With regard to using an infinite horizon value function to represent a finite horizon game, note that a typical NFL game has about 150 plays per game, or 75 plays per half. With our state description, the second half can be considered the start of a separate game, since the last play of the first half does not carry over to the first play of the second half. To validate our use of an infinite horizon value function empirically, we rewrote the transition probabilities used to evaluate

Figure 1. (Color online) Aggregate Value Function by Down

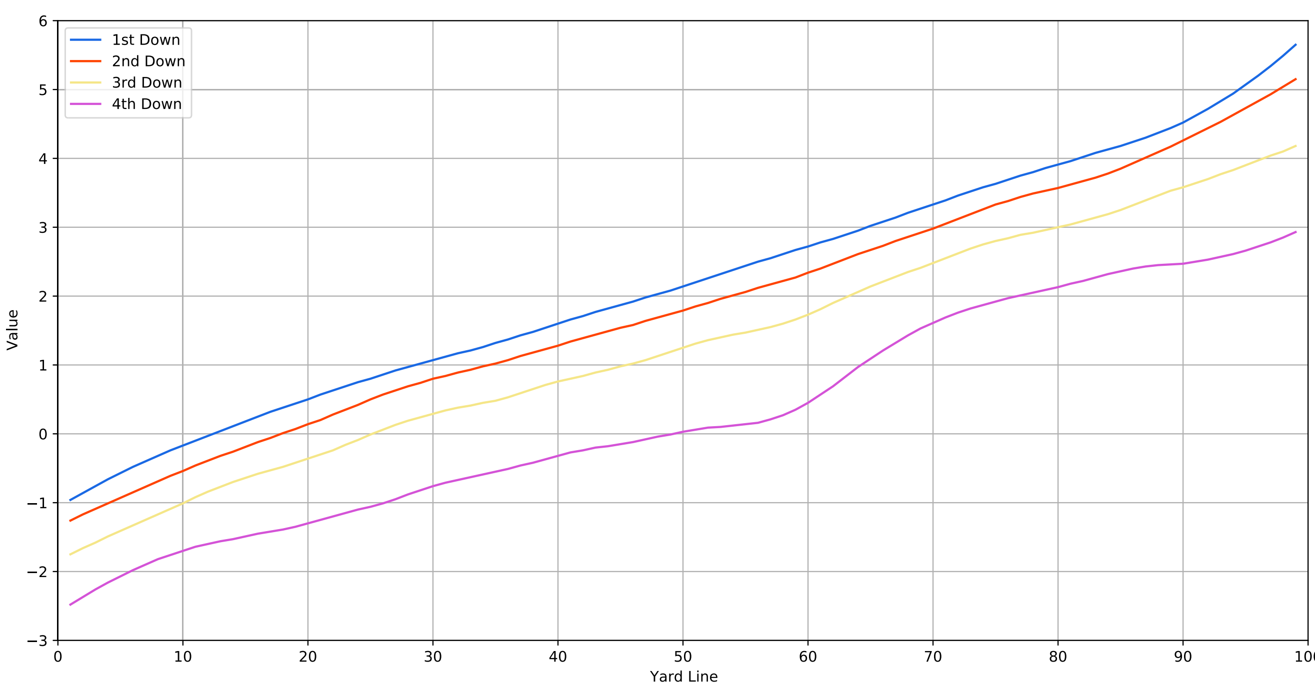


Table 1. Comparison Between the $V(d, y)$'s on First Down

Yard line	Carter and Machol 1969	Cabot et al. 1981	Romer 2002	Our model
5	−1.25	−1.33	−0.80	−0.57 (-0.59, -0.55)
15	−0.64	−0.58	0.00	0.18 (0.17, 0.20)
25	0.24	0.13	0.60	0.80 (0.79, 0.81)
35	0.92	0.84	1.15	1.32 (1.30, 1.33)
45	1.54	1.53	1.90	1.87 (1.86, 1.89)
55	2.39	2.24	2.20	2.44 (2.43, 2.45)
65	3.17	3.02	2.80	3.02 (3.00, 3.04)
75	3.68	3.88	3.30	3.63 (3.62, 3.65)
85	4.57	4.84	4.00	4.18 (4.17, 4.20)
95	6.04	5.84	4.90	5.07 (5.06, 5.08)

Note. Our values contain a 95% confidence interval for each $V(d, y)$.

(21) in matrix form and found that the second largest eigenvalue equalled 0.81. Consequently, P^n and P^* differed by at most 0.001 after 30 iterations. Thus, it follows from Equation (6) that V represents the expected advantage or disadvantage of starting in state (d, y) over a small, finite number of plays. We acknowledge that the infinite horizon interpretation will be less accurate toward the ends of the two halves. However, we believe this is a minor limitation since plays toward the end of the halves constitute a small number of plays in our total sample.

5. Points Gained

5.1. Derivation and Interpretation

In this section, we explore insights into football performance based on a metric we refer to as points gained (PG). Suppose the ith play starts at state (d_i, y_i) and ends at state (d_{i+1}, y_{i+1}) and r_i points are scored. We define the points gained PG_i to be the difference between the observed and expected number of points scored on play i. That is:

$$PG_i = r_i + V(d_{i+1}, y_{i+1}) - \mathbf{E}(R_i + V(D_{i+1}, Y_{i+1}) | (D_i, Y_i) = (d_i, y_i)). \quad (22)$$

By substituting the Bellman equation

$$V(d_i, y_i) = \mathbf{E}(R_i + V(D_{i+1}, Y_{i+1}) | (D_i, Y_i) = (d_i, y_i)) \quad (23)$$

into (22), it follows that

$$PG_i = r_i + V(d_{i+1}, y_{i+1}) - V(d_i, y_i). \quad (24)$$

As is evident from Equation (24), points gained represents the change in value from the beginning of a play to the end of the play, adjusted by the number of points scored during that play. Note that for nonscoring plays, $PG_i = V(d_{i+1}, y_{i+1}) - V(d_i, y_i)$. Viewed through Equation (22), points gained equivalently measures the value of a play relative to the league average. If PG_i is positive, the play exceeded expectation. If PG_i is negative, it fell below expectation. By analyzing points gained by game situation, we can precisely distinguish on what type of plays a particular team gains or loses value. Thus, using a value function based only on down and yard line suffices, especially if adjusted by how other teams performed in this situation. For example, if team A gained 2 points per game on short passes on third down, and the league average was 1, team A would have gained an advantage on this type of play in this situation.

The martingale representation in Section 3.3 also provides another interpretation of points gained. Based on Equation (9), the result that $g = 0$ when we assume identical teams, and the choice of stopping time to be the number of plays until the next kickoff by the team with possession, it follows that:

$$PG_i = r_i + \mathbf{E}\left[\sum_{j=1}^{\tau-1} r(X_j) \middle| X_0 = (d_{i+1}, y_{i+1})\right] - \mathbf{E}\left[\sum_{j=1}^{\tau-1} r(X_j) \middle| X_0 = (d_i, y_i)\right]. \quad (25)$$

Consequently, points gained gives the change in expected total points until the next kickoff by the team in possession of the ball as the result of the ith play.

By virtue of how it is defined, points gained has a convenient additive property: the points gained on a sequence of plays is the sum of the points gained on the individual plays. Suppose a team's possession lasts for n plays. Then the total points gained for this sequence of plays satisfies the following:

$$\sum_{i=1}^{n} PG_i = \sum_{i=1}^{n} \left(V(d_{i+1}, y_{i+1}) - V(d_i, y_i) + r_i\right) \quad (26)$$

$$= V(d_{n+1}, y_{n+1}) - V(d_1, y_1) + \sum_{i=1}^{n} r_i, \quad (27)$$

because of the telescoping sum. The interpretation of points gained for a drive is the change in the expected point differential from the first play to the last. Likewise, the points gained for a game is the change in the expected point differential from the opening kickoff until the last play. Since a team is equally likely to be the kicking or receiving team, and the *PG* calculations utilize a league-wide value function, the expected value of the starting game state is zero. Hence, the points gained for a finite game can be interpreted simply as a team's expected margin of victory for each game. This interpretation of the value function has been proposed previously, but without any formal justification (Winston 2009).

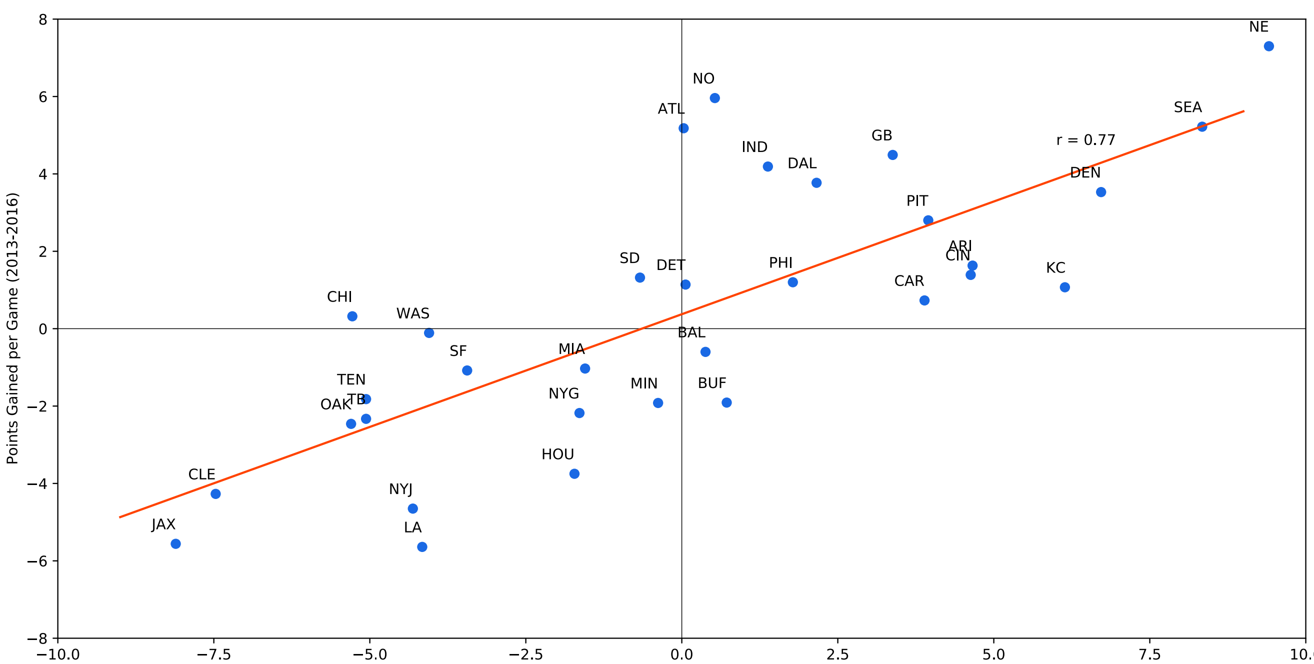
Figure 2 shows a strong positive correlation of 0.77 between the observed average margin of victory for each team over 2013–2016 and the average points gained per game for each team over the same time period. Moreover, points gained underestimates the observed margin of victory, perhaps because of scores near the end of the half and the game, and the fact that many field goals result in negative points gained. This is consistent with our findings in Section 5.3, where we examine a case study of the 2019 Super Bowl.

Finally, we provide a few illustrative points gained computations. An incomplete pass (zero yards gained) on first down at the 20 yard line results in a points gained of -0.36 ($V(2,20) - V(1,20) = 0.14 - 0.50$), whereas a five yard pass or run would result in a points gained of zero ($V(2,25) - V(1,20)$). Figure 1 shows that gaining about five yards on first down, about 10 yards on second down, and more than 15 yards on third down are required to break even. In other words, the loss of a down becomes more costly as the down progresses. For another example, scoring a touchdown from the 90 yard line (opponent's 10 yard line) corresponds to gaining 1.82 points. To see this, note that $V(1,90) = 4.52$, $r_i = 6.97$, and $V(0,35) = -0.63$. The key observation here is that the points gained on this single play are considerably less than the 6.97 points scored on the touchdown because the possibility of a score is already captured in $V(1,90)$ and the effect of the subsequent kickoff. Suppose instead that there were two incompletions before a touchdown from the 90 yard line. This would result in a points gained of $V(0,35) - V(3,90) + 6.97 = -0.63 - 3.58 + 6.97 = 2.76$, which is higher than before since there are fewer chances remaining to score a touchdown on third down compared with first down. Three incompletions at the 90 yard line followed by a field goal results in a points gained of $V(0,35) - V(4,90) + 3 = -0.63 - 2.47 + 3 = -0.10$, reflecting the expectation that a fourth down at that field position should typically result in a successful field goal. Finally, a turnover on downs would result in a points gained of $-V(1,10) - V(4,90) = 0.17 - 2.47 = -2.64$, a significant loss.

5.2. Using Points Gained to Assess Performance

We next discuss how to use the points gained metric to obtain insight into how football teams generate value, focusing on team offense across multiple seasons. Similar analyses are possible for team defense and individual players, and at the season, game or within game levels. As Winston (2009) notes, measures that can isolate the contribution of different

Figure 2. (Color online) Relationship Between the Average Number of Points Gained Per Game and the Observed Average Margin of Victory of Each Team, 2013–2016



plays to performance are far more meaningful than standard statistics that do not directly relate to scoring.

Our main observations are based on Table 2. This table provides the rank and points gained per game for each NFL team overall ("Total" column) and by play type for the 2013–2016 seasons. Although rule changes were implemented during this time frame, the effect of these changes on the overall rankings and points gained estimates are very small. Thus, we decided to use all of the data together, taking advantage of the larger pool of plays from which to derive transition probabilities. Summary measures include the rank correlation for each play type with the rank based on the overall points gained, and the average points gained across the league overall and by play type. Teams are ordered on the basis of their rank on overall points gained.

5.2.1. League-Wide Analysis. On average, points gained per game are positive for passing and scrambling only. Teams gain on average 5.19 points through passing and 0.41 points through scrambling. Average points gained are highly negative for sacks (−2.97) and negative for field goals (−0.73), runs (−0.68), and punts (−0.66). Clearly, teams gain most of their points through passing and lose the most through sacks. Most of the other average effects are smaller.

Figure 1 provides justification for these negative points gained values. Since sacks involve loss of yardage and loss of a down, they are quite costly. In our data set, each team averaged 2.46 sacks per game, meaning that each sack contributes to a loss of 1.21 points. This is roughly equivalent to a 10-yard sack on second down at the opponent's 30 yard line. Running plays tend to result in negative points gained since the average gain in yardage on a running play is not enough to offset the decrease in value due to the loss of a down.

Somewhat surprisingly, the average points gained for field goal attempts is −0.73 and all teams have negative values for this metric. We believe there are two reasons for this. First, successful field goals on fourth down inside the opponent's 14 yard line result in negative points gained (i.e., $3 - 0.63 - V(4, y) < 0$) for $y > 86$). Also, missed field goals on fourth down contribute to this effect. To illustrate, a successful field goal on fourth down at the 70 yard line gains $3 - 0.63 - 1.61 = 0.76$ points, whereas an unsuccessful field goal from this position gains $-1.07 - 1.61 = -2.68$ points. During the 2013–2016 seasons, there were 97 successful and 42 unsuccessful field goals from the 70 yard line. Hence, the expected points gained from this field position is $(0.7) \times 0.76 + (0.3) \times (-2.68) = -0.27$.

5.2.2. Team Comparisons. We now explore the insights gained with respect to individual team performance. Note that team rank on total points gained is highly correlated with passing rank (0.94) and moderately correlated with rank on running (0.58) and rank on sack avoidance (0.50). Teams gain most of their points through passing often by a significant margin. For example, New England's offense gains 7.30 points per game overall, but their passing accounts for 10.05 points per game of this total, which means that other types of plays are producing negative value. The top five teams in points gained passing are also the top five in points gained total, whereas the bottom three teams in points gained passing are also the bottom three teams in total points gained. The top three teams in points gained passing (New England, New Orleans, and Green Bay) gain at least four points passing more than an average team. Clearly, outstanding passing contributes significantly to outstanding performance.

Although points gained passing is positive for all teams, most teams lose value on running plays. In fact, teams that rank highly in points gained running are not necessarily those that rank highly in points gained total. From a confirmatory perspective, we found the points gained scrambling statistics to be in strong agreement with our football knowledge. San Francisco with quarterback Colin Kaepernick had a points gained scrambling of 1.41, exceeding the overall average by one point. Other teams (quarterbacks [QBs]) that excelled on scrambling were Green Bay (Rodgers), Indianapolis (Luck), and Seattle (Wilson); the worst team in scrambling was Denver (Manning up to 2016 and Osweiler for 2016–2017).

Other observations from Table 2 include:

- Dallas ranked best in points gained running, gaining almost a point more than the league average. Unfortunately, they were only slightly above average in terms of passing points gained. If they improved their points gained passing while maintaining their high quality running game, they could potentially improve their offensive production substantially.
- New England gained over half a point more than an average team on its kickoffs. We hypothesize that this corresponds to New England's ability to avoid touchbacks and leave opponents deep in their own end of the field.
- New Orleans was the lowest ranked team overall on points gained field goal attempts, losing almost a whole point to the best ranked team Indianapolis. This clearly identifies an area where improvement was necessary.
- Denver excelled on sack avoidance, losing 0.85 points less than the league average. Since this value clearly did not come from scrambling (recall Denver ranked last in points gained scrambling), we believe Denver is generating this value from having a quarterback who is noted to have a quick release.

Table 2. Average Points Gained Per Game, Categorized by Play Type

	Rank									Points gained per game								
Team	Total	Pass	Run	FGA	KO	Punt	Scr	Sack	Pen	Total	Pass	Run	FGA	KO	Punt	Scr	Sack	Pen
NE	1	1	13	6	1	10	20	3	2	7.30	10.05	−0.21	−0.37	0.52	−0.43	0.22	−2.49	0.00
NO	2	2	4	32	7	9	26	7	19	5.96	9.87	0.27	−1.33	0.18	−0.38	0.15	−2.59	−0.20
SEA	3	4	3	7	5	19	4	19	32	5.22	8.52	0.28	−0.43	0.24	−0.68	0.80	−3.05	−0.46
ATL	4	5	12	8	13	2	18	12	15	5.18	8.26	−0.19	−0.48	0.04	0.02	0.37	−2.69	−0.16
GB	5	3	9	12	32	21	2	27	9	4.49	9.19	0.01	−0.66	−0.76	−0.77	0.91	−3.32	−0.11
IND	6	9	22	1	2	3	3	6	1	4.19	6.91	−1.27	−0.04	0.43	−0.15	0.84	−2.57	0.04
DAL	7	14	1	3	6	5	12	8	28	3.77	5.73	0.67	−0.26	0.18	−0.19	0.50	−2.59	−0.27
DEN	8	7	16	11	18	23	32	1	3	3.53	7.67	−0.42	−0.64	−0.05	−0.85	−0.04	−2.12	−0.02
PIT	9	6	11	15	28	30	27	10	7	2.80	7.87	−0.18	−0.69	−0.38	−1.30	0.15	−2.62	−0.06
ARI	10	10	18	22	3	22	22	13	27	1.63	6.57	−0.87	−0.88	0.37	−0.82	0.19	−2.70	−0.25
CIN	11	13	15	25	12	20	29	5	21	1.39	6.03	−0.39	−0.94	0.08	−0.72	0.09	−2.55	−0.21
SD	12	8	28	19	29	17	25	15	5	1.32	7.52	−1.67	−0.81	−0.38	−0.66	0.17	−2.80	−0.05
PHI	13	15	2	24	19	16	16	14	12	1.20	4.94	0.39	−0.93	−0.09	−0.64	0.42	−2.74	−0.14
DET	14	11	26	30	20	11	10	9	4	1.14	6.43	−1.50	−1.13	−0.13	−0.44	0.57	−2.61	−0.05
KC	15	16	5	26	26	1	5	23	25	1.07	4.72	0.19	−0.94	−0.26	0.02	0.77	−3.20	−0.24
CAR	16	19	8	17	4	14	9	22	6	0.73	4.37	0.04	−0.79	0.34	−0.56	0.57	−3.19	−0.06
CHI	17	17	6	21	22	15	8	17	23	0.32	4.40	0.14	−0.83	−0.14	−0.63	0.61	−2.99	−0.24
WAS	18	12	14	23	31	27	31	26	13	−0.11	6.16	−0.36	−0.92	−0.42	−1.13	−0.01	−3.30	−0.14
BAL	19	26	21	5	10	8	17	4	18	−0.60	3.13	−1.02	−0.35	0.14	−0.23	0.41	−2.49	−0.20
MIA	20	18	17	28	23	4	23	29	26	−1.03	4.39	−0.42	−1.09	−0.15	−0.19	0.18	−3.49	−0.25
SF	21	20	24	9	8	28	1	28	17	−1.08	4.05	−1.38	−0.49	0.16	−1.16	1.41	−3.48	−0.18
TEN	22	24	27	4	30	13	6	21	22	−1.82	3.61	−1.51	−0.33	−0.38	−0.52	0.69	−3.15	−0.21
BUF	23	27	7	14	11	29	15	20	8	−1.91	2.59	0.13	−0.68	0.10	−1.26	0.42	−3.14	−0.07
MIN	24	25	10	16	27	25	11	30	14	−1.92	3.49	−0.12	−0.79	−0.35	−0.87	0.52	−3.64	−0.15
NYG	25	23	31	2	14	26	28	11	11	−2.18	3.77	−2.01	−0.25	0.03	−1.01	0.09	−2.65	−0.14
TB	26	21	29	29	9	12	19	25	24	−2.33	4.05	−1.67	−1.12	0.16	−0.48	0.28	−3.30	−0.24
OAK	27	29	19	20	21	6	24	2	30	−2.46	2.27	−0.94	−0.83	−0.13	−0.22	0.17	−2.47	−0.33
HOU	28	28	30	27	24	18	7	18	10	−3.75	2.28	−1.71	−0.99	−0.16	−0.67	0.66	−3.04	−0.11
CLE	29	22	23	31	17	31	21	32	31	−4.27	3.85	−1.35	−1.17	−0.05	−1.36	0.21	−4.04	−0.37
NYJ	30	31	20	18	15	32	14	16	29	−4.65	1.44	−0.97	−0.80	0.00	−1.62	0.45	−2.86	−0.29
JAX	31	30	32	10	25	24	13	31	20	−5.56	1.70	−2.09	−0.62	−0.17	−0.86	0.49	−3.80	−0.20
LA	32	32	25	13	16	7	30	24	16	−5.64	0.18	−1.45	−0.66	−0.01	−0.22	0.00	−3.29	−0.18
League average										0.37	5.19	−0.68	−0.73	−0.03	−0.66	0.41	−2.97	−0.17
Rank correlation		0.94	0.58	0.19	0.31	0.30	0.05	0.50	0.29									

Notes: Play types are passes, runs, field goal attempts (FGA), kickoffs (KO), punts, QB scrambling (Scr), sacks, and penalties (Pen). Summary measures include league averages and rank correlations between total points gained and play type.

Table 3. Breakdown of Points Gained Passing (PGP) by Depth and Direction

	Rank							Points gained per game						
		Short			Deep				Short			Deep		
Team	PGP	L	M	R	L	M	R	PGP	L	M	R	L	M	R
NE	1	1	1	9	16	1	18	10.05	2.48	2.68	1.17	1.01	1.59	1.13
NO	2	4	10	2	1	8	4	9.87	1.72	1.36	1.78	2.03	1.03	1.94
GB	3	16	2	7	6	4	1	9.19	0.50	2.54	1.21	1.34	1.12	2.47
SEA	4	7	17	8	2	13	3	8.52	1.34	1.06	1.17	1.99	0.93	2.04
ATL	5	3	7	5	26	3	12	8.26	1.77	1.61	1.50	0.65	1.40	1.34
PIT	6	18	25	1	10	7	2	7.87	0.44	0.84	2.23	1.21	1.04	2.10
DEN	7	5	13	4	5	18	13	7.67	1.51	1.17	1.61	1.35	0.75	1.30
SD	8	2	18	3	11	24	16	7.52	1.80	1.05	1.75	1.18	0.56	1.18
IND	9	21	5	13	4	21	6	6.91	0.32	2.13	0.56	1.38	0.68	1.84
ARI	10	17	14	6	7	14	8	6.57	0.45	1.09	1.49	1.25	0.89	1.41
DET	11	15	4	11	19	5	26	6.43	0.56	2.35	0.83	0.83	1.06	0.78
WAS	12	19	9	10	9	22	14	6.16	0.43	1.54	0.99	1.23	0.68	1.30
CIN	13	12	19	16	8	2	19	6.03	0.83	1.05	0.35	1.23	1.51	1.06
DAL	14	20	8	12	15	17	15	5.73	0.41	1.55	0.72	1.08	0.76	1.20
PHI	15	14	21	20	21	9	11	4.94	0.65	1.03	0.11	0.79	1.01	1.35
KC	16	8	12	19	20	25	21	4.72	1.21	1.19	0.16	0.79	0.48	0.89
CHI	17	6	29	22	23	12	23	4.40	1.34	0.48	0.00	0.76	0.94	0.88
MIA	18	9	30	15	3	29	28	4.39	1.18	0.44	0.45	1.42	0.33	0.56
CAR	19	10	22	23	22	23	24	4.37	1.14	0.99	−0.05	0.77	0.63	0.88
SF	20	13	15	17	29	20	27	4.05	0.74	1.09	0.33	0.40	0.73	0.76
TB	21	26	16	25	14	19	10	4.05	−0.08	1.08	−0.17	1.11	0.73	1.38
CLE	22	25	26	14	12	11	29	3.85	−0.06	0.74	0.51	1.15	1.00	0.49
NYG	23	28	3	30	31	10	20	3.77	−0.23	2.48	−0.75	0.29	1.01	0.96
TEN	24	11	11	26	27	16	31	3.58	0.92	1.35	−0.19	0.65	0.83	0.03
MIN	25	29	20	18	18	31	9	3.49	−0.25	1.03	0.22	0.85	0.23	1.41
BAL	26	31	6	24	17	15	30	3.13	−0.54	1.78	−0.12	0.87	0.83	0.32
BUF	27	27	32	21	25	27	7	2.59	−0.13	0.00	0.07	0.65	0.35	1.66
HOU	28	24	24	28	24	30	25	2.28	0.04	0.87	−0.38	0.66	0.32	0.79
OAK	29	30	28	31	28	6	5	2.27	−0.38	0.54	−1.26	0.49	1.05	1.85
JAX	30	32	23	27	30	28	17	1.70	−0.86	0.95	−0.22	0.34	0.34	1.15
NYJ	31	23	27	29	32	26	22	1.44	0.11	0.54	−0.61	0.13	0.38	0.88
LA	32	22	31	32	13	32	32	0.18	0.24	0.11	−1.53	1.15	0.22	−0.01
League average								5.19	0.61	1.21	0.44	0.97	0.79	1.17
Rank correlation		0.72	0.56	0.90	0.64	0.58	0.50							

Notes: Depth short (≤15 yards) versus deep, and direction is left (L), middle (M), and right (R). Summary measures include league averages and rank correlations between categories and totals.

Table 4. Points Gained by Down and Kickoffs (KOs)

	Rank						Points gained per game					
Team	Total	KO	1st	2nd	3rd	4th	Total	KO	1st	2nd	3rd	4th
NE	1	1	1	9	3	4	7.30	0.52	3.00	0.78	2.20	0.79
NO	2	6	5	3	2	11	5.96	0.18	1.56	1.54	2.30	0.38
SEA	3	5	2	8	5	17	5.22	0.24	2.29	0.86	1.72	0.11
ATL	4	13	10	11	1	2	5.18	0.04	0.98	0.66	2.41	1.09
GB	5	32	11	1	4	18	4.49	−0.76	0.90	2.47	1.81	0.07
IND	6	2	12	6	10	3	4.19	0.43	0.66	1.04	1.11	0.95
DAL	7	6	7	7	13	6	3.77	0.18	1.46	0.88	0.57	0.68
DEN	8	18	6	5	8	23	3.53	−0.05	1.52	1.07	1.21	−0.22
PIT	9	28	8	4	7	26	2.80	−0.38	1.11	1.21	1.29	−0.43
ARI	10	3	3	19	12	30	1.63	0.37	1.74	−0.42	0.61	−0.68
CIN	11	12	18	2	19	14	1.39	0.08	−0.04	1.56	−0.40	0.19
SD	12	29	24	10	6	22	1.32	−0.38	−0.33	0.70	1.52	−0.19
PHI	13	19	9	15	18	16	1.20	−0.09	1.00	0.39	−0.22	0.13
DET	14	20	25	13	9	21	1.14	−0.13	−0.36	0.60	1.13	−0.10
KC	15	26	4	17	21	8	1.07	−0.26	1.57	−0.31	−0.52	0.58
CAR	16	4	14	25	15	9	0.73	0.34	0.52	−0.95	0.33	0.49
CHI	17	22	15	21	11	19	0.32	−0.14	0.40	−0.58	0.63	0.01
WAS	18	31	20	12	14	31	−0.11	−0.42	−0.11	0.65	0.50	−0.73
BAL	19	10	17	29	23	1	−0.60	0.14	−0.02	−1.34	−0.62	1.24
MIA	20	23	13	14	30	24	−1.03	−0.15	0.63	0.54	−1.80	−0.25
SF	21	8	22	20	26	10	−1.08	0.16	−0.22	−0.53	−0.89	0.40
TEN	22	29	16	32	16	5	−1.82	−0.38	0.12	−2.16	−0.12	0.71
BUF	23	11	23	16	28	28	−1.91	0.10	−0.30	0.11	−1.32	−0.51
MIN	24	27	19	22	25	20	−1.92	−0.35	−0.06	−0.73	−0.76	−0.02
NYG	25	14	30	18	22	12	−2.18	0.03	−1.67	−0.33	−0.58	0.37
TB	26	9	26	23	20	27	−2.33	0.16	−0.84	−0.75	−0.41	−0.49
OAK	27	21	21	28	29	7	−2.46	−0.13	−0.16	−1.14	−1.66	0.62
HOU	28	24	28	27	24	25	−3.75	−0.16	−1.52	−1.13	−0.64	−0.31
CLE	29	17	27	30	27	29	−4.27	−0.05	−1.19	−1.41	−1.08	−0.54
NYJ	30	15	32	26	17	32	−4.65	0.00	−2.14	−0.95	−0.14	−1.41
JAX	31	25	31	24	32	13	−5.56	−0.17	−2.04	−0.85	−2.86	0.36
LA	32	16	29	31	31	15	−5.64	−0.01	−1.53	−2.03	−2.22	0.15
League average							0.37	−0.03	0.22	−0.02	0.10	0.11
Rank correlation		0.31	0.85	0.83	0.86	0.30						

Note: Summary measures include league averages and rank correlations between categories.

5.2.3. Points Gained Subcategories. Next, we drill down into subcategories of points gained by examining how points gained varies by pass type, down, and yard line. Additional results on points gained running and points gained by quarter are provided in the appendix. To understand what types of passes contributed most to points gained passing, Table 3 provides points gained values based on pass depth (short [15 yards or less] versus deep) and direction (left, middle, or right). Note that short passes include passes that were caught within 15 yards of the line of scrimmage but were then run for a net gain of more than 15 yards. Overall, teams gain 56% of their points gained passing on long passes, but the breakdown varies considerably from team to team. On short passes, more than half of the points gained are on passes over the middle. For the top four teams, short passes contribute to 50% (4.75/9.41) of total passing points gained. The bottom four teams lose 43% (−0.60/1.40) of their total passing points from their short passes. The top four teams are all ranked seventh or higher in at least one of the categories for short passes. The bottom four teams are ranked in the bottom third of all teams for every short passing category. Since points gained on passes is the largest contributor to overall points gained in Table 2, these results suggest that short passes are the most important factor that differentiates the best from the worst teams.

New England, which ranks first in points gained passing, generates over 60% of its value in this category with short passes, one of only two teams to do so. Most of the points gained are over the middle, which is a characteristic of other outstanding passing teams including the Green Bay Packers and the New York Giants. Similarly, New England generates the most value of any team on deep passes to the middle, but they are average when it comes to deep passes to the left or the right. Overall, New England generates 6.75 points from the three types of passes it

Table 5. Points Gained by Yard Line (from Offense's Perspective)

	Rank					Points gained per game				
Team	Total	1–20	21–50	51–79	80–99	Total	1–20	21–50	51–79	80–99
NE	1	8	2	1	2	7.30	0.50	3.09	2.28	1.43
NO	2	9	1	2	18	5.96	0.49	3.33	2.24	−0.10
SEA	3	6	3	6	16	5.22	0.74	2.45	1.87	0.16
ATL	4	10	4	5	9	5.18	0.33	2.29	1.98	0.57
GB	5	4	6	8	10	4.49	0.80	1.69	1.49	0.51
IND	6	18	5	7	5	4.19	−0.09	1.81	1.76	0.70
DAL	7	21	7	4	13	3.77	−0.30	1.67	2.02	0.38
DEN	8	16	19	3	1	3.53	0.06	−0.14	2.05	1.56
PIT	9	2	12	13	14	2.80	1.09	0.82	0.56	0.33
ARI	10	1	18	12	17	1.63	1.18	−0.09	0.59	−0.05
CIN	11	7	17	22	3	1.39	0.61	0.17	−0.57	1.18
SD	12	22	11	9	19	1.32	−0.38	0.83	0.98	−0.11
PHI	13	19	10	10	24	1.20	−0.15	0.90	0.77	−0.32
DET	14	15	8	27	6	1.14	0.13	1.45	−1.13	0.69
KC	15	5	14	24	8	1.07	0.75	0.55	−0.83	0.60
CAR	16	17	15	19	7	0.73	−0.09	0.54	−0.33	0.61
CHI	17	14	16	20	15	0.32	0.13	0.39	−0.49	0.29
WAS	18	32	9	14	23	−0.11	−1.43	1.38	0.22	−0.28
BAL	19	11	13	25	27	−0.60	0.22	0.58	−0.92	−0.49
MIA	20	3	27	17	20	−1.03	0.86	−1.43	−0.23	−0.23
SF	21	24	23	16	12	−1.08	−0.47	−0.94	−0.08	0.41
TEN	22	13	29	29	4	−1.82	0.16	−1.76	−1.39	1.17
BUF	23	25	22	15	26	−1.91	−0.51	−0.92	0.00	−0.48
MIN	24	12	25	21	25	−1.92	0.18	−1.07	−0.57	−0.46
NYG	25	20	28	11	32	−2.18	−0.25	−1.66	0.69	−0.95
TB	26	26	20	23	30	−2.33	−0.69	−0.22	−0.59	−0.83
OAK	27	23	26	26	11	−2.46	−0.43	−1.41	−1.12	0.50
HOU	28	29	21	31	21	−3.75	−1.03	−0.73	−1.75	−0.24
CLE	29	27	24	30	29	−4.27	−0.86	−1.01	−1.68	−0.71
NYJ	30	31	30	18	31	−4.65	−1.21	−2.37	−0.24	−0.83
JAX	31	30	32	28	28	−5.56	−1.05	−2.59	−1.22	−0.70
LA	32	28	31	32	22	−5.64	−0.93	−2.55	−1.89	−0.27
League average						0.37	−0.05	0.16	0.14	0.13
Rank correlation		0.67	0.89	0.81	0.62					

Note: Summary measures include league averages and rank correlations between categories.

ranks first in, which together comprise 67% of its overall points gained passing.

Table 4 summarizes points gained by down. Recall kickoffs are modeled as the zeroth down, so we include them here to keep the total points gained values consistent with Table 2. On average, teams gain the most points on first down but the variability between teams by down is high. For example, the top team, New England, gains points on all downs, ranking particularly high on first, third, and fourth downs. Other top teams, such as New Orleans, Seattle, and Atlanta, gain the most points on either first or third down. Overall, the top seven teams gain points on all downs (excluding kickoffs), whereas the bottom seven teams lose points on all downs. The rank correlations suggest that overall performance is highly correlated to performance on the first three downs, but not kickoffs or fourth down. Note also that points gained on second down is negative, in contrast to first, third, and fourth down, which all have positive points gained. The reason points gained on second down is negative is because the loss in value going from second down to third down is quite large (average 0.6 points on the same yard line). Since the percentage of first down conversions when starting on the second down is quite low, only 26.5%, teams usually get a negative points gained value on second down. The points gained for first down is positive because the drop in value from first to second is not as large. Third downs have the highest first down conversation rate, which supports a positive points gained. Finally, fourth downs have positive points gained values because they often end with punting the ball deep into the opposing team's end, which has low or even negative value for the opponent, or a field goal that gains points for the offense.

Table 5 shows that the best teams tend to distinguish themselves in the middle of the field, between

Table 6. Score by Quarter in the 2019 Super Bowl

Quarter	1	2	3	4	Total
New England	0	3	0	10	13
Los Angeles	0	0	3	0	3

the two 20 yard lines. However, among the top teams overall, performance in the opponent's red zone and their own red zone is more variable, with some of the top teams being worse than average in those parts of the field. Notably, the worst teams tend to lose the most points in their half of the field.

5.3. Using Points Gained to Evaluate a Single Game: A Case Study of the 2019 Super Bowl

On February 3, 2019, the New England Patriots defeated the Los Angeles Rams 13–3 in the 2019 NFL Super Bowl. In this section, we show how the points gained metric provides insight into which aspects of the game gave New England its advantage. This case study demonstrates how these metrics can be used for performance assessment at the game level, which is currently done postgame using fairly simple statistics (e.g., QB rating) that may not properly or fully ascribe value to the plays that mattered.

Table 7 shows that New England gained a total of −0.63 points and Los Angeles gained a total of −7.94 points, a difference in favor of New England of 7.31 points. This difference aligns with the observed 10 point margin of victory. Comparing the quarter-by-quarter score (Table 6) versus points gained (Table 7) shows that points gained differences are in close agreement to the actual score differences for the first two quarters, and are underestimates in the third and fourth quarters. Note also that total points gained was negative for both teams, supporting the fact that the game was a close defensive battle throughout. In fact, it was the lowest scoring Super Bowl in history, which is noteworthy since Los Angeles had the second best offense that season.

Table 8 shows that New England gained a significant advantage on plays on first through third down, with the biggest advantage at second down. Fourth down differences may be explained by a missed field goal by New England and the fact that Los Angeles ran more fourth down plays.

Table 9 provides the most insight into how New England won. It shows that New England excelled at both pass and run defense, and dominated through its 4.43 points gained advantage in running, again notable since Los Angeles had one of the best run defenses in the NFL that season. Another factor was New England's points gained advantage of 3.04 points in penalties, which puts a clear, quantifiable value to a couple of key penalties that went in New England's favor during the game. Los Angeles had an almost four point advantage in field goals; again, New England's missed field goal heavily contributed to this deficit.

Digging into specific plays in more detail, we find that New England gained its passing advantage through short passes over the middle, which is consistent with its performance during 2013–2016, as shown in Table 3. Also, New England's running advantage was gained primarily by runs off left tackle. Finally, New England gained 2.57 points on a single play (a pass from Tom Brady to Rob Gronkowski) near the end of the fourth quarter. This play has been consistently singled out as one that essentially won the game for New England.

Table 7. Points Gained by Quarter in 2019 Super Bowl

Quarter	1	2	3	4	Total
New England	−3.59	−3.68	0.01	6.80	−0.63
Los Angeles	−3.76	−6.54	0.87	1.32	−7.94

Table 8. Points Gained by Down and Team for 2019 Super Bowl

Down	New England	Los Angeles	Difference
1	1.35	−0.93	2.28
2	5.29	−2.08	7.37
3	−1.72	−5.71	3.99
4	−4.70	1.86	−6.56

6. Conclusion

In this paper, we developed a value function for the NFL using a Markov reward process model. Although value functions have been studied previously, this paper is the first to develop a rigorous theoretical foundation for this model and an interpretation of the value function as the bias of the underlying Markov reward process or expected excess reward up to a stopping time. We use the value function to derive a points gained metric that enables us to provide novel insights into team performance. For example, our analysis shows that passing plays generate the most points gained by far, whereas running plays tend to generate negative value. We further analyzed passes

Table 9. Total Points Gained by Play Type and Team for 2019 Super Bowl

Play type	New England	Los Angeles	Difference
Pass	−1.42	−5.16	3.74
Run	1.62	−2.81	4.43
Punt	0.23	0.75	−0.52
Penalty	2.12	−0.92	3.04
Field goal	−2.33	1.62	−3.95

by type of pass, and found that short passes account for the majority of the top teams' success and the worst teams' poor performance. Other insights include how teams differ in terms of points gained by down, quarter, and yard line, allowing us to identify teams that arguably perform better in important situations such as third down or the fourth quarter, and opportunities for teams to improve in various aspects of their game. Finally, we use the points gained metric to analyze the 2019 Super Bowl and show how it reveals aspects of the game that contributed to the outcome. We believe that this analysis sets a direction for future reporting of game outcomes.

Opportunities that remain are to apply this methodology to analyzing defensive performance in football and to other sports. Discrete event sports, such as golf, baseball, cricket, and curling, are amenable to such analysis. Even more fluid sports, such as basketball, hockey, soccer, and tennis, can be analyzed from a points gained perspective with appropriately defined state spaces that include possessions or geospatial coordinates of the players/ball; data on the latter is increasingly abundant as motion tracking increases in popularity. For example, all National Basketball Association team arenas are fitted with camera systems that track player and ball movement throughout each game.

More importantly, there are novel opportunities to explore how the value function difference method (points gained) can be used to assess performance in nonsports application. For example, in appointment scheduling, it could be used to assess the cost to the system of no-shows or of excess arrival of high priority customers. Or in a large organization with multiple units that are managed separately, a points gained approach may be used to determine a corporate average performance level, which is then used to assess performance of individual units.

Acknowledgments

The authors thank John Birge, Mark Broadie, Wayne Winston, Ed Kaplan, Peter Bell, the area and associate editor, and two anonymous reviewers for helpful suggestions and comments that improved the paper.

Appendix. Points Gained by Running Plays and Quarter

Table A.1 provides a breakdown of points gained by running plays. Note that the league average is negative and although a few teams are positive, 12 teams lose more than a point per game on running plays. We believe this occurs because running plays often result in a gain of only a few yards.

Table A.1. Points Gained Per Game Running, Categorized by Direction

			Rank						Points gained per game							
Team	Total	C	LE	LG	LT	RE	RG	RT	Total	C	LE	LG	LT	RE	RG	RT
DAL	1	3	12	10	9	22	1	3	0.67	0.22	0.16	–0.04	–0.08	–0.09	0.34	0.16
PHI	2	1	17	21	13	26	16	5	0.39	0.74	0.04	–0.16	–0.12	–0.14	–0.12	0.15
SEA	3	4	15	26	27	1	9	7	0.28	0.21	0.06	–0.23	–0.32	0.48	–0.03	0.11
NO	4	2	3	30	6	12	4	17	0.27	0.28	0.43	–0.53	–0.03	0.09	0.15	–0.12
KC	5	25	2	5	4	7	5	24	0.19	–0.57	0.52	0.08	0.06	0.25	0.10	–0.25
CHI	6	12	6	18	2	11	11	20	0.14	–0.06	0.35	–0.11	0.10	0.10	–0.05	–0.21
BUF	7	5	19	6	25	5	19	6	0.13	0.19	–0.05	0.07	–0.31	0.29	–0.20	0.14
CAR	8	11	1	20	26	13	28	4	0.04	–0.05	0.68	–0.14	–0.32	0.08	–0.36	0.15
GB	9	13	25	2	5	17	7	8	0.01	–0.07	–0.11	0.13	0.01	–0.02	–0.01	0.08
MIN	10	16	14	25	29	8	2	9	–0.12	–0.11	0.08	–0.22	–0.34	0.21	0.19	0.07
PIT	11	15	7	19	7	14	15	18	–0.18	–0.11	0.28	–0.12	–0.03	0.06	–0.12	–0.15
ATL	12	10	27	4	10	32	3	1	–0.19	–0.02	–0.15	0.09	–0.10	–0.48	0.19	0.29
NE	13	20	16	14	23	4	6	10	–0.21	–0.36	0.05	–0.07	–0.26	0.39	0.01	0.04
WAS	14	18	28	9	3	3	24	23	–0.36	–0.20	–0.19	0.00	0.09	0.41	–0.23	–0.25
CIN	15	14	13	11	22	21	20	2	–0.39	–0.07	0.10	–0.05	–0.26	–0.08	–0.21	0.18
DEN	16	6	9	17	24	10	14	31	–0.42	0.17	0.21	–0.10	–0.27	0.11	–0.09	–0.45
MIA	17	17	4	22	21	19	23	13	–0.42	–0.14	0.41	–0.16	–0.25	–0.02	–0.21	–0.06
ARI	18	28	5	8	31	9	8	22	–0.87	–0.68	0.38	0.02	–0.46	0.11	–0.02	–0.21
OAK	19	24	24	3	8	18	26	12	–0.94	–0.56	–0.11	0.13	–0.06	–0.02	–0.29	–0.02
NYJ	20	19	8	15	18	16	29	25	–0.97	–0.31	0.21	–0.08	–0.16	–0.01	–0.36	–0.26
BAL	21	8	10	32	20	6	31	21	–1.02	0.10	0.18	–0.66	–0.23	0.28	–0.48	–0.21
IND	22	26	32	1	14	27	13	19	–1.27	–0.59	–0.31	0.17	–0.12	–0.14	–0.07	–0.20
CLE	23	7	23	31	17	20	21	28	–1.35	0.16	–0.10	–0.64	–0.16	–0.06	–0.21	–0.34
SF	24	31	11	28	15	15	25	11	–1.38	–0.88	0.17	–0.30	–0.15	0.01	–0.27	0.03
LA	25	30	20	23	28	2	27	27	–1.45	–0.76	–0.05	–0.18	–0.33	0.45	–0.31	–0.27
DET	26	9	30	7	32	24	17	26	–1.50	0.04	–0.24	0.05	–0.82	–0.12	–0.13	–0.27

Table A.1. (Continued)

	Rank								Points gained per game							
Team	Total	C	LE	LG	LT	RE	RG	RT	Total	C	LE	LG	LT	RE	RG	RT
TEN	27	22	18	29	1	25	32	14	−1.51	−0.48	−0.03	−0.34	0.12	−0.13	−0.60	−0.06
SD	28	29	21	27	12	29	18	16	−1.67	−0.74	−0.06	−0.26	−0.10	−0.23	−0.16	−0.11
TB	29	27	26	13	16	28	12	30	−1.67	−0.65	−0.13	−0.07	−0.16	−0.16	−0.06	−0.44
HOU	30	23	22	24	11	31	10	29	−1.71	−0.54	−0.09	−0.18	−0.10	−0.34	−0.04	−0.42
NYG	31	21	29	16	19	30	30	32	−2.01	−0.42	−0.19	−0.09	−0.19	−0.26	−0.36	−0.50
JAX	32	32	31	12	30	23	22	15	−2.09	−0.99	−0.28	−0.05	−0.35	−0.11	−0.21	−0.09
League average									−0.68	−0.23	0.07	−0.13	−0.18	0.03	−0.13	−0.11
Rank correlation		0.67	0.52	0.14	0.21	0.46	0.52	0.57								

Notes: Directions are center (C), left end (LE), left guard (LG), left tackle (LT), right end (RE), right guard (RG), and right tackle (RT). Summary measures include league averages and rank correlations between categories.

The loss of value due to incrementing the down largely offsets the gain in value from a short advance of the yard line, which leads to an overall negative average points gained for running plays. Top running teams tend to run well down the center or left side; only these plays net a positive points gained for all of the top four running teams. In contrast, the

Table A.2. Points Gained by Quarter and Overtime (OT)

	Rank						Points gained per game					
Team	Total	1st	2nd	3rd	4th	OT	Total	1st	2nd	3rd	4th	OT
NE	1	1	2	4	1	19	7.30	2.21	1.42	1.53	2.17	−0.03
NO	2	10	5	2	3	11	5.96	1.03	1.16	1.96	1.79	0.02
SEA	3	13	1	12	2	14	5.22	0.33	2.31	0.66	1.91	0.01
ATL	4	2	14	1	7	20	5.18	2.12	0.38	2.08	0.62	−0.03
GB	5	3	10	14	6	18	4.49	2.07	0.70	0.53	1.22	−0.02
IND	6	16	4	3	5	28	4.19	0.08	1.22	1.66	1.34	−0.11
DAL	7	7	6	7	8	4	3.77	1.18	0.97	1.01	0.51	0.09
DEN	8	6	7	6	10	17	3.53	1.20	0.86	1.04	0.45	−0.01
PIT	9	17	3	21	4	5	2.80	−0.05	1.33	0.05	1.39	0.08
ARI	10	18	9	11	13	13	1.63	−0.05	0.73	0.74	0.19	0.02
CIN	11	5	15	13	19	31	1.39	1.22	0.35	0.55	−0.58	−0.15
SD	12	9	21	9	15	21	1.32	1.16	−0.32	0.80	−0.27	−0.05
PHI	13	11	11	8	22	15	1.20	0.50	0.53	0.96	−0.79	0.01
DET	14	4	23	18	12	12	1.14	1.24	−0.55	0.13	0.30	0.02
KC	15	8	8	26	17	3	1.07	1.18	0.77	−0.62	−0.36	0.09
CAR	16	25	12	5	18	10	0.73	−0.51	0.52	1.14	−0.45	0.03
CHI	17	28	13	17	9	30	0.32	−0.70	0.50	0.18	0.47	−0.13
WAS	18	21	20	15	14	22	−0.11	−0.15	−0.08	0.25	−0.07	−0.05
BAL	19	12	29	19	11	25	−0.60	0.47	−1.45	0.08	0.37	−0.08
MIA	20	29	18	10	27	8	−1.03	−0.81	0.16	0.75	−1.17	0.05
SF	21	15	16	30	21	16	−1.08	0.18	0.33	−0.93	−0.66	0.00
TEN	22	20	24	16	23	7	−1.82	−0.13	−0.96	0.20	−0.98	0.05
BUF	23	14	26	23	20	27	−1.91	0.22	−1.14	−0.23	−0.66	−0.10
MIN	24	22	28	22	16	1	−1.92	−0.26	−1.44	−0.08	−0.27	0.13
NYG	25	23	17	28	25	24	−2.18	−0.43	0.16	−0.78	−1.04	−0.08
TB	26	24	19	24	28	32	−2.33	−0.49	0.06	−0.33	−1.41	−0.17
OAK	27	27	22	25	24	6	−2.46	−0.64	−0.37	−0.53	−0.99	0.07
HOU	28	31	31	20	26	9	−3.75	−0.92	−1.84	0.08	−1.11	0.04
CLE	29	19	30	27	31	29	−4.27	−0.12	−1.68	−0.64	−1.71	−0.12
NYJ	30	26	25	31	32	2	−4.65	−0.60	−1.12	−1.07	−1.97	0.11
JAX	31	30	32	29	30	23	−5.56	−0.87	−2.14	−0.92	−1.56	−0.06
LA	32	32	27	32	29	26	−5.64	−1.02	−1.29	−1.72	−1.52	−0.09
League average							0.37	−0.02	0.21	−0.02	0.09	0.11
Rank correlation		0.78	0.85	0.83	0.90	0.06						

Note: Summary measures include league averages and rank correlations between categories.

bottom four running teams have a negative points gained in every running direction.

Table A.2 shows points gained by quarter. On average, points gained is positive in the second and fourth quarters, but like points gained by down, there is significant variability across teams by quarter. The top 10 teams gain points in all quarters whereas the lowest ranked four teams lose points in all quarters. The top team, New England, is ranked highly in all quarters. In particular, the team ranks first in the first and fourth quarters, which suggests New England tends to start out strong and finish strong. Other teams that are fairly consistent across quarters include Dallas and Denver, which are ranked seventh and eighth overall. On the other hand, some teams exhibit more variability. For example, Atlanta and Green Bay rank highly in the first quarter, but are closer to league average in the second quarter. On the flip side, Seattle and New Orleans tend to finish strong, but start games less strong, suggesting that there is room for improvement in their first quarter performance.

References

Bell P (1982) Analysis of Strategies in the Canadian Football League. *Inform. Systems Oper. Res.* 20(2):116–125.

Boronico JS, Newbert SL (1994) Play calling strategy in American football: A game-theoretic stochastic dynamic programming approach. *J. Sport Management* 13(2):103–113.

Broadie M (2008) Assessing golfer performance using golfmetrics. Crews D, Lutz R, eds., Sci. Golf V: Proc. World Sci. Congress Golf (Energy in Motion Inc., Mesa, AZ), 253–262.

Broadie M (2012) Assessing golfer performance on the PGA tour. *Interfaces* 42(2):116–146.

Cabot V, Sagarin J, Winston W (1981) A stochastic game model of football. Working paper, Indiana University, Bloomington.

Carter V, Machol RE (1971) Operations research on football. *Oper. Res.* 19(2):541–545.

Chan TCY, Singal R (2016) A Markov decision process-based handicap system for tennis. *J. Quant. Anal. Sports* 12(4):179–189.

Efron B (1979) Bootstrap methods: Another look at the jackknife. *Ann. Statist.* 7(1):1–26.

Erkut E (1987) More on Morrison and Wheat's 'Pulling the goalie revisited.' *Interfaces* 17(5):121–123.

Fry MJ, Lundberg AW, Ohlmann JW (2007) A player selection heuristic for a sports league draft. *J. Quant. Anal. Sports* 3(2): 121–123.

Glynn PW (2013) Lecture notes in stochastic systems for MS&E 321. Stanford University, Stanford, CA. https://stanford.edu/class/msande321/Handouts/10%20Martingales.pdf.

Goldner KA (2012) A Markov model of football: Using stochastic processes to model a football drive. *J. Quant. Anal. Sports* 8(1): 1–18.

Goldner KA (2017) Situational success: Evaluating decision-making in football. Albert J, Glickman ME, Swartz TB, Koning RH, eds., *Handbook of Statistical Methods and Analyses in Sports* (CRC Press, Boca Raton, FL), 183–198.

Horowitz M, Yurko R, Ventura S (2019) nflscrapR: Compiling the NFL play-by-play API for easy use in R. https://www.rdocumentation.org/packages/nflscrapR/versions/1.4.0.

Morrison DG (1976) On the optimal time to pull the goalie: A Poisson model applied to a common strategy used in ice hockey. *TIMS Studies in Management Science*, vol. 4 (North-Holland, New York), 137–144.

Morrison DG, Wheat RD (1986) Misapplications reviews: Pulling the goalie revisited. *Interfaces* 16(6):28–34.

Pelechrinis K, Winston W, Sagarin J (2019) Evaluating NFL plays: Expected points adjusted for schedule. Brefeld U, Davis J, Van Haaren J, Zimmermann A, eds. *Machine Learning and Data Mining for Sports Analytics (MLSA 2018)*, Lecture Notes in Computer Science, vol. 11330 (Springer, Cham).

Puterman ML (1994) *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (John Wiley & Sons, New York).

Romer D (2002) It's fourth down and what does the bellman equation say? A dynamic-programming analysis of football strategy. NBER Working Paper 9024, National Bureau of Economic Research, Cambridge, MA.

Washburn A (1991) Still more on pulling the goalie. *Interfaces* 21(2):59–64.

Winston W (2009) *Mathletics* (Princeton University Press, Princeton, NJ).

Yurko R, Ventura SL, Horowitz M (2019) nflWAR: a reproducible method for offensive player evaluation in football. *J. Quant. Anal. Sports* 15(3):163–183.

Timothy C. Y. Chan is the Canada Research Chair in Novel Optimization and Analytics in Health, a professor in the Department of Mechanical and Industrial Engineering, the director of the Centre for Healthcare Engineering, and the director of the Centre for Analytics and AI Engineering at the University of Toronto. His primary research interests are in optimization under uncertainty and the application of optimization methods to problems in healthcare, medicine, and sports.

Craig Fernandes is a second year MASc candidate in the Applied Optimization Laboratory supervised by Professor Timothy C. Y. Chan at the University of Toronto. His research focuses on optimization methods and applications of machine learning as they relate to sport decision problems.

Martin L. Puterman is professor emeritus at the University of British Columbia Sauder School of Business. His research has focused on Markov decision processes, healthcare operations, sports analytics, and applied statistics. He is an INFORMS Fellow and recipient of the INFORMS Lanchester Prize, the Canadian Operations Research Society (CORS) Award of Merit, the CORS Practice Prize, and the INFORMS case prize.