A Comparison of the Autocorrelation and Variance of NFL Team Strengths Over Time using a Bayesian State-Space Model

Joseph S. Koopmeiners

Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota

Corresponding author’s email address: koopm007@umn.edu

Abstract

Professional sports leagues are motivated to promote competitive balance in order to maintain fan interest. The National Football League (NFL) has taken several steps to promote competitive balance, most notably free agency and the salary cap, which were instituted prior to the 1994 season. Previous research into competitive balance in sports focused on the variability of team strengths but ignored the year-to-year autocorrelation in team strengths. We present a Bayesian state-space model for paired-comparisons that allows regression on the variance parameters. By modeling the variance parameters in a regression framework, we are able to simultaneously compare the variance and autocorrelation in team strengths over time. The autocorrelation in NFL team strengths has decreased over time while there has been little change in the variance in teams strengths since the 1970s.
1 Introduction

Professional sports leagues are motivated to promote competitive balance in order to maintain fan interest. A lack of competitive balance leads to lopsided games and predictable champions, which removes the primary reason why fans follow professional sports in the first place: we don’t know who is going to win. Professional sports have taken several approaches to promote competitive balance, including roster limits, the reverse order draft, salary caps, etc. Statistical approaches to summarizing competitive balance are needed to evaluate whether or not these approaches have the intended effect.

Competitive balance in the National Football League (NFL) has been a popular topic of conversation over the last fifteen years. Talk of increased competitive balance in the NFL began in 1999 when the St. Louis Rams won Super Bowl XXXIV a year after finishing 4-12. Two years later, the New England Patriots completed a similar worst-to-first turnaround winning Super Bowl XXXVI a year after finishing 5-11. These results led many to conclude that the NFL was experiencing increased parity due to the implementation of free agency and the salary cap before the 1994 season. Free agency allows players to play for the team of their choice when their contract expires, while the salary cap places a limit on the amount of money a team can pay their players in a given season. Both measures promote player movement and prevent wealthy clubs from stockpiling talent, which, in turn, promotes competitive balance. Of course, the Rams and Patriots turnarounds were not unprecedented. The San Francisco 49ers defeated the Cincinnati Bengals in Super Bowl XVI a year after both teams finished 6-10. Furthermore, the NFL has not been without dominating teams in the years since the Rams won the Super Bowl. The New England Patriots won three Super Bowls and were runner-ups in two more since 2001, while the Indianapolis Colts won at least 10 games in eleven of twelve seasons starting in 1999. While the Rams and Patriots Super Bowl victories are suggestive of increased parity in the NFL, a thorough analysis is needed to evaluate changes in competitive balance over time.

Competitive balance in sports has been studied extensively in the economics literature. Zimbalist provides an introduction to competitive balance in sports and discusses the use of the standard deviation of win percentages as a measure of competitive balance (Zimbalist, 2002). The win percentage (i.e. the percentage
of games won by a team for the entire season) is calculated for each team and is the simplest measure of
team strength available. A small standard deviation of win percentages would represent increased parity. If
all teams were of equal strength, the standard deviation of win percentages would be \( \frac{5}{\sqrt{N}} \), where \( N \) is the
number of games played by each team. Zimbalist refers to this as the idealized standard deviation of win
percentages. The NFL has the highest actual standard deviation of win percentages but the lowest ratio of
the actual standard deviation of win percentages to idealized standard deviation of win percentages of all
United States professional sports leagues.

Larsen et al. (2006) studied the impact of free agency and the salary cap on competitive balance in the
NFL. They used several measures of wealth distribution to summarize competitive balance in the NFL and
found a moderately significant increase in competitive balance after the implementation of free agency and
the salary cap.

The above measures summarize, at least indirectly, the variance in NFL team strengths. A large vari-
ance in team strengths would lead to a large number of very good and very bad teams, which would lead to
a large standard deviation of win percentages. Increased competitive balance would be characterized by a
decrease in the variance in team strengths.

An alternate approach is to study the year-to-year autocorrelation in team strengths. A decrease in au-
tocorrelation would allow larger changes in a team’s ability from year-to-year and increase the probability
of worst-to-first turnarounds like the Rams and Patriots. Furthermore, it is possible that fans will tolerate
large variability in team strengths (more accurately, that fans will tolerate their team being very bad; it’s
safe to say that fans do not mind if their team is very good!) if there is reason to believe that a team
can improve rapidly due to astute management. Ideally, an analysis of competitive balance in sports would
simultaneously consider the variance and autocorrelation in team strengths.

A second limitation of the analysis of Larsen et al. (2006) is that they do not consider results at the individual

In this paper, we evaluate changes in competitive balance in the NFL over time using a Bayesian state-space model for paired-comparisons that allows regression on the variance parameters. This analysis addresses the two limitations of previous studies of competitive balance in sports. First, by allowing regression on the variance parameters, we are able to simultaneously compare the variance and autocorrelation in team strengths over time. Second, we analyze results on the individual game level using a paired-comparison model.

The remainder of this paper proceeds as follows. In Section 2, we present a Bayesian state-space model that allows us to simultaneously compare the variance and autocorrelation of team strengths over time by estimating the variance parameters in a regression framework. In Section 3, we compare the variance and autocorrelation in team strengths over time in an unadjusted analysis and an analysis adjusted for potential confounders. Finally, we conclude with a discussion in Section 4.

2 Statistical Analysis

2.1 Data

Single game NFL regular season results from 1970 to 2011 were downloaded from Pro-Football-Reference.com. Teams played 14 regular season games in 14 weeks from 1970 to 1977. Starting in 1978, teams played 16 regular season games with the exception of the strike-shortened 1982 (9 games) and 1987 (15 games) seasons. All non-strike-shortened regular seasons were played over the course of 16 weeks from 1978 to 1989 and 17 weeks from 1990 to 2011 with the exception of 1993 when teams played 16 games in 18 weeks.

The NFL consisted of 26 franchises in 1970. Since then, the NFL has gone through four rounds of expansion, adding the Seattle Seahawks and Tampa Bay Buccaneers in 1976, the Carolina Panthers and Jacksonville
Jaguars in 1995, the “new” Cleveland Browns in 1999, who replaced the original Browns franchise that moved to Baltimore before the 1996 season and became the Ravens, and finally the Houston Texans in 2002 giving the NFL 32 franchises since 2002.

The outcome for each game will be a binary variable representing whether or not the home team won the game. In many cases, margin of victory is considered when modeling NFL game results and the difference in team scores is taken to be the outcome for each game (Glickman and Stern, 1998; Harville, 1977). Margin of victory is considered, by some, to be a better predictor of future performance but it may not be the best reflection of reality. The goal is to win the game, not to outscore the opposing team by as many points as possible. Therefore, a team’s strength should be measured in their ability to win games. Margin of victory only indirectly represents the true outcome of interest and, more importantly, the use of margin of victory does not adequately reflect differences in style across teams. For example, a defensive minded team will not score as many points as an offensive minded team, which in turn will decrease their margin of victory even if they are equally likely to win the game. For these reasons, margin of victory will not be considered for this analysis.

Prior to 1974, regular season games could end in a tie. Starting in 1974, the NFL instituted a “sudden-death” overtime for regular season games that end in a tie. In the event of a tie, teams play an extra 15 minute period and the first team to score wins. The game results in a tie if neither team scores during the overtime period. Only 46 of the 8931 regular season games, 0.5%, since 1970 and 17 since the start of the 1974 season have resulted in a tie. Therefore, tie games were ignored for the purposes of this analysis. We note, though, that the model proposed in Section 2.2 could easily be expanded to accommodate games that end in a tie.

### 2.2 Model

The data were modeled using a Bayesian state-space model. The outcome, $y_{j,j',k}$, is a binary variable indicating whether the home team, team $j$, defeated the visiting team, team $j'$, in year $k$. Therefore, $y_{j,j',k}$
is distributed with a Bernoulli distribution with probability $p_{j,j',k}$. The probability, $p_{j,j',k}$, is modeled using the Bradley-Terry model for paired comparisons (Agresti, 2002),

$$\logit(p_{j,j',k}) = \delta_{j,k} - \delta_{j',k} + \alpha_j,$$

where $\delta_{j,k}$ and $\delta_{j',k}$ are the team strengths for teams $j$ and $j'$ in year $k$ and $\alpha_j$ is the home-field advantage for team $j$. Home field advantage parameters were modeled hierarchically,

$$\alpha_j \sim N(\alpha, \sigma^2_\alpha),$$

to allow a sharing of information across teams when estimating the home field advantage parameters. Home field advantage is typically accounted for assuming a common home field advantage across teams. In reality, home-field advantage is dependent on many factors (stadium, crowd, weather, etc.), many of which are team dependent. For this reason, we allow home-field advantage to vary by team and note that this is similar to the approach taken by Gajewski (2006). The home-field advantage parameters, $\alpha_j$, could also be allowed to depend on year, $\alpha_{j,k}$, and, in fact, this may be more true to reality. Most teams played in multiple home stadiums over the last forty years and several relocated to different cities but, for simplicity, we hold the team-specific home-field advantages constant over time. Examples of paired-comparison models applied to NFL results can be found in the literature Harville (1977, 1980); Mease (2003).

Our primary interest lies in the variance and autocorrelation of the $\delta_{j,k}$s. The variance and autocorrelation are represented by the parameters $\sigma^2_k$ and $\rho_k$, respectively, which are indexed by $k$ to indicate that the variance and autocorrelation are allowed to change over time, and modeled using a state-space model. For $k = 1$, or, in the case of expansion franchises, a team’s first year of existence, team strengths are assumed to be normally distributed with mean 0,

$$\delta_{j,k} \sim N(0, \sigma^2_k),$$
otherwise, team strengths are assumed to follow a normal distribution that is a function of the team strength for the previous season and the autocorrelation parameter, $\rho_k$,

$$
\delta_{j,k} \sim N(\rho_k \delta_{j,k-1}, (1 - \rho_k^2) \sigma_k^2).
$$

The proposed distributional assumptions on the team strengths, although strong, are intuitively reasonable. Independent of knowledge from the previous season, we would expect team strengths to be approximately normally distributed with the average team winning approximately half of their games. An autocorrelation equal to 1 implies that after the initial season, a teams strength will be constant over time, while an autocorrelation of 0 corresponds to the situation where teams start over after each season. In reality, we expect the autocorrelation to be between 0 and 1, where a teams strength is shrunk back towards 0 but the team strength for the following year will randomly deviate from the shrunken strength from the previous year. Similar state-space models applied to NFL results can be found in the literature (Glickman, 2001; Glickman and Stern, 1998).

We model the variance and autocorrelation in a regression framework in order to facilitate comparisons over time. Covariates for $\sigma^2$ and $\rho$ are introduced into the model using the following transformations,

$$
\log(1 + \rho_k) = \beta_{\rho,0} + \sum_{i=1}^{n_\rho} \beta_{\rho,i} X_{\rho,k,i},
$$

and

$$
\log(\sigma_k^2) = \beta_{\sigma^2,0} + \sum_{i=1}^{n_{\sigma^2}} \beta_{\sigma^2,i} X_{\sigma^2,k,i},
$$

where the $X_{\rho,i}$’s and $X_{\sigma^2,i}$’s are covariates and the $\beta_{\rho,i}$’s and $\beta_{\sigma^2,i}$’s are regression parameters. This is the same formulation as the general linear model except that we are modeling the variance and autocorrelation rather than the mean.

Transforming the autocorrelation and variance parameters allows for the estimation of differences in $\rho$ and $\sigma^2$ on an unbounded scale. For a one unit difference between $X_{\sigma^2,k,i}$ and $X_{\sigma^2,k',i}$ and all other covariates
held equal,
\[ e^{\beta_{\sigma^2,k}} = \frac{\sigma_{k'}^2}{\sigma_k^2}, \]
which we will refer to as the variance ratio (VR). Similarly, for a one unit difference between \( X_{\rho,k,i} \) and \( X_{\rho,k',i} \) and all other covariates held equal,
\[ e^{\beta_{\rho,k}} = \frac{1 + \rho_{k'}}{1 - \rho_{k'}}/\frac{1 + \rho_k}{1 - \rho_k}. \]

For the remainder of this paper, we will refer to this quantity as the autocorrelation odds ratio (AOR) because the form and interpretation are similar to the odds ratio. An AOR of 1 implies that \( \rho_{k'} \) and \( \rho_k \) are equal, an AOR greater than 1 implies that \( \rho_{k'} \) is greater than \( \rho_k \) and an AOR less than 1 implies that \( \rho_{k'} \) is less than \( \rho_k \).

The regression framework provides great flexibility when comparing \( \sigma^2 \) and \( \rho \) over time. Our primary analysis will consider the simplest case, which only includes a linear term for year, but one could also consider other models for comparing the autocorrelation and variance over time. For example, time could be categorized by decade, for example, or a model could be fit with higher order polynomials for time. In a secondary analysis, we will consider an adjusted model that includes covariates for expansion and strike-shortened seasons. Finally, in our analysis we include the same set of covariates for \( \sigma^2 \) and \( \rho \) but a model could also be fit that includes a different set of covariates for each parameter.

### 2.3 Prior Distributions and Model Fitting

Normal priors with a mean of zero and a variance of ten were used for the regression parameters (\( \beta_{\sigma^2,i} \)’s and \( \beta_{\rho,i} \)’s). This represents a prior belief of an autocorrelation and variance in team strengths near zero and one, respectively, and against covariate effects on the variance and autocorrelation. That said, these priors are not strongly informative as all reasonable parameter values for the autocorrelation and variance are found within two standard deviations of the prior mean. For example, \( \pm 2 \times \sqrt{10} \) corresponds to autocorrelation values of \( \pm 0.996 \). Non-informative priors, \( \alpha \sim N(0,10^6) \) and \( 1/\sigma_\alpha^2 \sim Gamma(0.0001,0.0001) \), were used for the home field advantage hyperparameters.
Figure 1: Standard Deviation and Lag-1 Autocorrelation in Team Strengths Estimated From Individual Bradley-Terry models for each Season from 1970 to 2009

The posterior distributions are not available in closed form and must be estimated using Markov Chain Monte Carlo (MCMC) methods (Carlin and Louis, 2000). Gibbs sampling from the posterior distribution was implemented in WinBugs version 1.4.3. Convergence was monitored by visual inspection of trace plots from parallel chains. Parameter estimates were obtained from the first 1,000,000 iterations after burn-in with a thin of 100 to reduce autocorrelation within chains. This resulted in a total of 1000 samples from the posterior in each chain.

3 Results

3.1 Exploratory Data Analysis

Individual Bradley-Terry models with common home field advantage were fit using maximum likelihood for each season from 1970 to 2011 to serve as an exploratory analysis for the more formal regression analysis presented in Subsection 3.2. Using these results, we calculated a variance and autocorrelation in team strengths for each season. Autocorrelation of team strengths was calculated using Pearson’s correlation coefficient for each year with the previous season from 1971 to 2011. Five teams, the 1972 Miami Dolphins, the 1976
Table 1: Autocorrelation odds ratio and variance ratio estimates for time with and without adjusting for expansion seasons and player strikes

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted Model</th>
<th></th>
<th>Adjusted Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AOR</td>
<td>95% CI</td>
<td>P (AOR &gt; 1)</td>
</tr>
<tr>
<td>Time</td>
<td>0.83</td>
<td>(0.69,1.00)</td>
<td>0.028</td>
<td>0.83</td>
</tr>
<tr>
<td>Expansion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982 Strike</td>
<td>1.24</td>
<td>(0.38,22.36)</td>
<td>0.590</td>
<td></td>
</tr>
<tr>
<td>1987 Strike</td>
<td>1.16</td>
<td>(0.01,53.79)</td>
<td>0.453</td>
<td></td>
</tr>
<tr>
<td>VR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.94</td>
<td>(0.82,1.09)</td>
<td>0.218</td>
<td>0.94</td>
</tr>
<tr>
<td>Expansion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982 Strike</td>
<td>1.16</td>
<td>(0.01,53.79)</td>
<td>0.453</td>
<td></td>
</tr>
<tr>
<td>1987 Strike</td>
<td>0.60</td>
<td>(0.03,6.43)</td>
<td>0.453</td>
<td></td>
</tr>
</tbody>
</table>

1: Posterior Median  
2: Credible Interval  
3: Posterior Probability

Tampa Bay Buccaneers, the 1982 Baltimore Colts, the 2007 New England Patriots and the 2008 Detroit Lions, were excluded as outliers when calculating the standard deviations and autocorrelations because they either won all of their games (Dolphins and Patriots) or lost all of their games (Buccaneers, Colts and Lions).

Figure 1 presents plots of the standard deviation and lag-1 autocorrelation of the estimated team strengths by year. We see a clear linear trend in the autocorrelation with the autocorrelation decreasing steadily from 1971 to 2011. There is a hint of a quadratic trend in the standard deviation over time but, in general, the standard deviation has been relatively constant over time. We will initially consider a linear trend for time in our regression models for the variance and autocorrelation and evaluate the fit of the linear trend by examining a plot of the residuals.
Figure 2: Trace plots of the first 20,000 Gibbs Sampler iterations for the regression and home-field advantage parameters in the unadjusted model. From top to bottom: $\alpha$, $\sigma^2$, $\beta_{\rho,0}$, $\beta_{\rho,1}$, $\beta_{\sigma^2,0}$ and $\beta_{\sigma^2,1}$
3.2 Regression Results

For our initial analysis, we consider an unadjusted model with only a linear time effect for the variance and autocorrelation:

$$\log\left(\frac{1 + \rho_k}{1 - \rho_k}\right) = \beta_{\rho,0} + \beta_{\rho,1} \times year_k,$$

and

$$\log(\sigma_k^2) = \beta_{\sigma^2,0} + \beta_{\sigma^2,1} \times year_k.$$

Figure 2 presents trace-plots for the first 20,000 Gibbs sample iterations. We see that the model converged within the first 20,000 iterations and inference will be completed after excluding the first 20,000 iterations as burn-in.

Table 1 presents estimates of the autocorrelation odds ratio and variance ratio from the Bayesian state-space model. The posterior probability that the autocorrelation decreased over time was .972 and the AOR for a 10-year difference was 0.83 (95% credible interval (CI): 0.69, 1.00). This corresponds to a decrease in the autocorrelation from .754 (95% CI: .663, .832) in 1975 to .690 (95% CI: .624, .752) in 1990 and .607 (.491, .714) in 2005. We note that these estimates are considerably higher than the autocorrelation estimates observed in Figure 1. The state-space model will borrow information about a team’s strength from the surrounding seasons. In contrast, the exploratory analysis presented in Figure 1 only considers results from a single season when estimating a team’s strengths, resulting in a lower correlation in the estimates of a team’s strength. In contrast, there was very little change in the variance over time. The VR for a 10-year difference in time was 0.94 (95% CI: 0.82, 1.09) and the posterior probability that the variance decreased over time was 0.782.

We next completed an analysis adjusting for expansion, the 1982 and 1987 strike:

$$\log\left(\frac{1 + \rho_k}{1 - \rho_k}\right) = \beta_{\rho,0} + \beta_{\rho,1} \times year_k + \beta_{\rho,2} \times exp + \beta_{r_{\rho,3}} \times stk_{1982} + \beta_{r_{\rho,4}} \times stk_{1987},$$
Figure 3: Trace plots of the first 60,000 Gibbs Sampler iterations for the regression and home-field advantage parameters in the adjusted model. Left column (top to bottom): $\alpha$, $\beta_{\rho,0}$, $\beta_{\rho,2}$, $\beta_{\rho,4}$, $\beta_{\sigma^2,1}$, $\beta_{\sigma^2,3}$. Right column (top to bottom): $\sigma_{\alpha}^2$, $\beta_{\rho,1}$, $\beta_{\rho,3}$, $\beta_{\sigma^2,2}$ and $\beta_{\sigma^2,4}$
and

$$\log(\sigma_k^2) = \beta_{p,0} + \beta_{p,1} \times year_k + \beta_{p,2} \times exp + \beta_{rho,3} \times stk_{1982} + \beta_{rho,4} \times stk_{1987},$$

where \(exp\), \(stk_{1982}\) and \(stk_{1987}\) are dummy variables indicating that the season was an expansion season or shortened by strike. Separate covariates were used for the two strike-shortened seasons because we expect that the strike-shortened 1987 season (when replacement players were used) to have a different effect than the strike-shortened 1982 season. Figure 3 presents trace-plots for the first 60,000 Gibbs sample iterations. We see that the model converged within the first 60,000 iterations and inference will be completed after excluding the first 60,000 iterations as burn-in.

AOR and VR estimates for the adjusted analysis can also be found in Table 1. Adjusting for expansion and strike-shortened seasons had little effect on the AOR or VR for time. The AOR for a 10-year difference in time was 0.83 (95% CI: 0.68, 1.02) after adjusting for expansion and strike-shortened seasons and the posterior probability that the autocorrelation decreased over time was 0.966, while the VR for a 10-year difference in time was 0.94 (95% CI: 0.81, 1.10) in the adjusted analysis. The credible intervals for the expansion and strike-shortened seasons were much too wide to draw conclusions about the effect of expansion or player strikes on the autocorrelation or variance in team strengths.

Finally, Figure 4 presents the home-field advantage for each team for the unadjusted model. Results for the adjusted model are virtually identical and are not presented. Presented is the probability of defeating an average team at home for each team, along with a 95% credible interval. The solid line represents the average home-field advantage across the entire league. We see that the Steelers and Broncos had the best home-field advantage over the 42-year period from 1970 to 2011, while the Colts, Saints and Jets had the worst home-field advantage.
3.3 Model Fit

To this point, we have not discussed the fit of our model. Model fit is particularly important in this case because of the strong assumptions implicit in the underlying state-space and regression models. There are two aspects of model fit that need to be considered when evaluating this model: first, is the underlying state-space model described in Equation (1) appropriate for our data and, second, are the regression models presented in Table 1 adequate to explain changes in the autocorrelation and variance of team strengths over time?
Table 2: Deviance Information Criteria (DIC) for model with no autocorrelation process, state-space model with unadjusted regression model and state-space model with adjusted regression model

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>No autocorrelation process</td>
<td>11952.3</td>
</tr>
<tr>
<td>State-space model with unadjusted regression model</td>
<td>11462.2</td>
</tr>
<tr>
<td>State-space model with adjusted regression model</td>
<td>11474.3</td>
</tr>
</tbody>
</table>

To address the first aspect of model fit, we fit individual Bradley-Terry models for each season with no underlying state-space process and compare the fit of this model to the fit for our state-space model that allows regression on the variance and autocorrelation. We evaluate the fit of our models using the deviance information criteria (DIC) proposed by Spiegelhalter et al. (2002). DIC is a Bayesian measure of model complexity with smaller values representing a better fitting model. Table 2 presents DIC values for a model with no state-space process and for the state-space model with the unadjusted and adjusted regression models. We see that both state-space models have smaller DIC values than the model with no state-space process, indicating that they both represent a better fit to the data than the model with no state-space process. This suggests that, while the state-space model relies on strong assumptions about the underlying autocorrelation process, both models are reasonable fits to the data.

The adequacy of our regression models were evaluated by creating residual plots for the autocorrelation and variance regression models. Residuals were calculated by fitting a saturated model with a different variance and autocorrelation for each season and comparing these estimates to the fitted variance and autocorrelation for the unadjusted and adjusted regression models. Residual plots can be found in Figure 5. The residuals are scattered around zero and do not exhibit any consistent pattern. Large residuals were observed during the 1976 season for the variance, when the Buccaneers finished 0-14 as an expansion franchise, and in the 1981 (Super Bowl XVI) and 1999 (Super Bowl XXXIV) seasons for autocorrelation, both of which were discussed as motivating examples in the introduction. The analysis of the residuals suggest that our regression
models are an adequate fit to the data.

4 Discussion

This manuscript makes two major contributions. First, we highlight that both the variance and autocorrelation in team strengths should be considered when analyzing competitive balance in sports. Previous analyses of competitive balance in the NFL have focused only on the variance in team strengths. This represents the classical understanding of competitive balance (i.e. a professional sports league has good competitive balance
if there is little spread between the best and worst teams). Alternately, a league might also be interested in evaluating whether the same franchises field the best teams every year or if team strength is a more fluid process where a different set of teams are at the top of the league every year. This is measured by autocorrelation. To address both aspects of competitive balance, we propose a Bayesian state-space model that allows for a simultaneous comparison of the variance and autocorrelation over time. Our model estimates the variance and autocorrelation in a regression framework that allows for covariates to be included in the model.

The second major contribution is our analysis of the change in the variance and autocorrelation in NFL team strengths over time using our Bayesian state-space model. We show that there has, in fact, been a change in the competitive balance of the NFL over time but that the change manifested itself in a change in the autocorrelation in team strengths rather than a change in the variance in team strengths. This result is consistent with the worst-to-first turnarounds experienced by the Rams in 1999 and Patriots in 2001.

Our analysis confirms that there has been an increase in the competitive balance in the NFL but does not address why there has been a change in competitive balance in the NFL. One obvious explanation is the implementation of free agency and the salary cap before the 1994 season. This is a reasonable explanation given the temporal relationship between the implementation of the salary cap and free agency and Super Bowls XXXIV and XXXVI but it does not explain why the autocorrelation in team strengths has decreased consistently over the last forty years and has continued to decrease long after the implementation of the salary cap. Another explanation is the increased reliance on the passing game over the running game. An increased reliance on the passing game results in a larger percentage of the teams success being attributable to the success of a single player: the quarterback. In this case, an elite quarterback switching teams could have a dramatic impact on a teams fortunes (Kurt Warner to the Rams or Cardinals, for example). Of course, the increased emphasis on the passing game could also increase autocorrelation as elite franchises are likely to have an elite quarterback and their fortunes are not likely to change as long as their quarterback continues to play at an elite level. Future research is needed to identify factors that influence autocorrelation in the NFL.
One weakness of our analysis is the reliance on strong assumptions about the underlying autocorrelation process. We assume that team strengths are normally distributed for a given season and that there is an association between the variance and autocorrelation parameter. Model fit statistics were presented which indicate that the Bayesian state-space model is a better fit to the data than a model that does not include the underlying state-space process but it is possible that a different state-space process would provide a better fit to the data. A more flexible model would allow for more robust inference but, as can be seen from the parameter estimates for expansion and strike-shortened season, there is a limited amount of information available from which to estimate changes in the autocorrelation and variance over time.

A second weakness to this analysis is that game results are assumed independent. The extent to which game results are correlated is open for debate but it is unlikely that they are independent. The simplest example of correlated results within a season is the injury to a key player. If Aaron Rodgers were injured halfway through the season, the estimated team strength for the Packers would be between the team strength for the Packers when Aaron Rodgers is healthy and the team strength for the Packers when Aaron Rodgers is injured. The Packers would be more likely to win games with a healthy Aaron Rodgers than would be expected given their estimated team strengths and less likely to win games when Aaron Rodgers was injured than would be expected given their estimated strength. This could be addressed by adding another level to the model that allows for week-to-week autocorrelation but it is not clear how well we could estimate the regression parameters in this case given the limited amount of data.

Separate results were presented evaluating the change in the variance and autocorrelation over time. An alternate approach would be to develop a summary measure that combines both parameters. Combining the variance and autocorrelation into a single measure is appealing but might result in a loss of information compared to viewing the results separately. The NFL has experienced a decrease in the autocorrelation over time but, in contrast, there has been little to no change in the variance over time. Analyzing a composite measure would indicate that there has been a change in competitive balance over time but would
not indicate if the change were in the autocorrelation or variance. For this reason, we chose not to pursue a composite measure but acknowledge that a composite measure would be worth pursuing in future research.

Finally, this analysis focused on competitive balance in the NFL. While increased parity in the NFL has been a hot topic for the last fifteen years, competitive balance is of interest to all professional sports leagues. Small changes to the model would likely be needed to accommodate differences between the NFL and other professional sports leagues but, hopefully, this analysis will provide a framework for future analyses of competitive balance in sports that account for changes in both the autocorrelation and variance in team strengths over time.

Acknowledgments

The author would like to thank the Editor and two anonymous referees who’s helpful comments greatly improved this manuscript.

References


