A METRIC TO QUANTIFY THE TOPSYTURVINESS OF A COLLEGE FOOTBALL SEASON

Autar K Kaw and Ali Yalcin

Citation: "A Metric to Quantify the Topsyturviness of a College Football Season", Autar Kaw and Ali Yalcin, Chance, Vol. 22, No 3, pp. 18-26, 2009.

To garner attention of their audience, during every college football season, news media, sports commentators, and bloggers alike hope to have something to hype about. Luckily, for them, the 2007 season did give them something to talk about. One would be hard-pressed to recall a more topsy-turvy season where highly ranked teams lost regularly to low-ranked and unranked teams.

In just Week#1 of the 2007 season, Associated Press (AP) No. 5 team University of Michigan lost to an unranked Division-II team - Appalachian State. The Associated Press wasted no time in booting Michigan out of the Top AP 25. Two weeks later, No. 11 UCLA lost to unranked Utah by a wide margin of 44-6. UCLA also met the same fate as Michigan; UCLA was dropped from the AP Top 25.

The topsy-turvyness continued in the season, especially for No. 2 ranked teams. The University of South Florida, where I work, was ranked No. 2 when they lost to unranked Rutgers 30-27 in Week#8. This was the same week when three other teams (South Carolina, Kentucky, and California) ranked in the Top 10 of the AP poll also lost their games.

To top off the season, for the first time in history of the Bowl Championship Series (BCS), the title bowl game had a team (Louisiana State University (LSU)) with two regular season losses, and LSU ended up winning the national championship.

Although many ranted and raved about the anecdotal evidence of a topsy-turvy season, is it possible that the media and fans over-exaggerated the topsy-turvyness of the 2007 college football season. Were there other seasons that were more topsy-turvy than 2007?

To answer this question scientifically, this article proposes a metric to *quantify* the topsy-turvyness of the college football season. The authors are not aware of any

previous literature that has attempted to develop a metric that quantifies the topsy-turvyness of any sport that is ranked regularly during its season.

Two different topsy-turvy (TT) factors are calculated: one for each of week of the season, referred to as the *Week TT factor*, and one for the cumulative topsy-turvyness at the end of each week of the season, referred to as the *Season TT factor*.

Week TT factor

At the end of each college football week, the Associated Press (AP) poll rankings are calculated by polling 65 sportswriters and broadcasters across the nation. Each voter supplies his or her ranking of the top 25 teams. The individual votes are added by giving 25 points to the first place vote, 24 points to the second place vote, etc. The addition of the points then produces the list of the AP top 25 teams of the week.

The method to find the *Week TT Factor* is based on comparing the AP Top 25 poll rankings of schools from the previous week to that of the current week. The difference in the rankings of each school in the AP Top 25 from the previous week to the current week is squared, which hence allocates proportionately higher importance on bigger week-to-week changes in rankings for a given team.

The formula for the Week TT factor is given by

$$Week\ TT\ factor = \frac{S_k}{44.16} \times 100, \qquad (1)$$

where

S_k is the square root of the sum of square of the differences in rankings, given by

$$S_k = \sqrt{\sum_{i=1}^{25} (i - c_i)^2} \quad , \qquad (2)$$

and c_i= current week ranking of the previous week's i^{th} ranked AP Top 25 team.

In Equation (2), how do we account for teams that fall out of the AP Top 25 rankings? A team that gets unranked from the previous week is assumed as having become the No. 26 team in the current week, in other words c_i=26 for any unranked team *i* that gets unranked..

In Equation (1), where does the number 44.16 come from? It is a normalization number which is the mean of the lowest and highest possible value of S_k. The lowest possible value for S_k is for the case where all the rankings stay unchanged from the previous week. Since in this case, the numerical difference in the rankings between the current and previous week would be zero for all teams, the lowest possible value of S_k =0.

The highest possible value for S_k is obtained when the top 17 teams fall out of the top 25 ranks and the 25th to 18th ranked teams are ranked 1st to 8th, respectively. In this case, S_k =88.32.

Figure 1 shows the plot of the week TT factors for seasons between 2002 and 2007. Clearly, 2003 and 2007 seasons emerge as the two most topsy-turvy seasons, while 2004 season materializes as a very stable season.

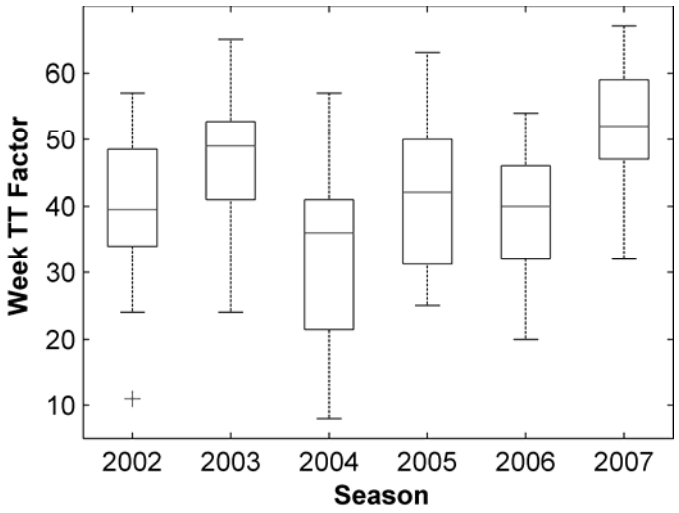
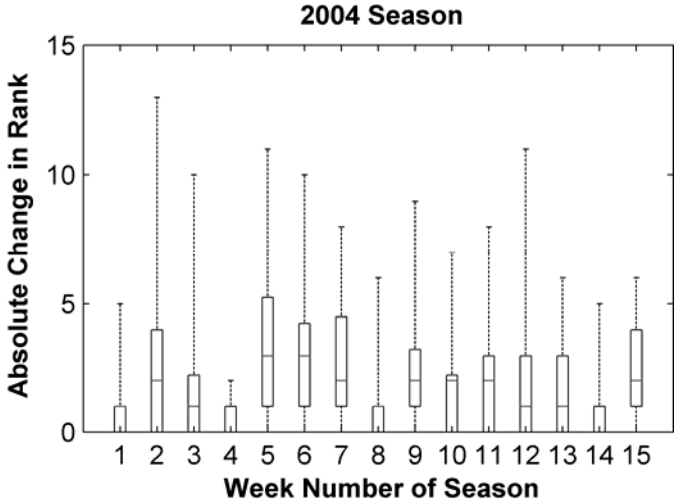
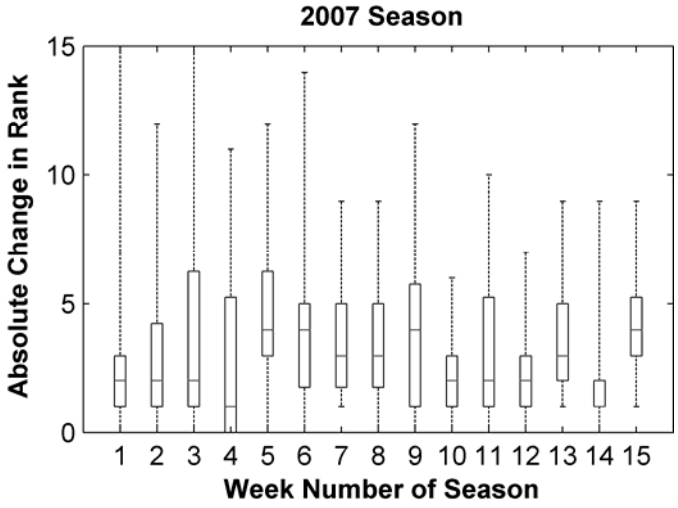


Figure 1. Box plot of Week TT factors of six seasons (2002-2007).

Each week, there are 25 teams in the AP Top 25 and there 25 changes in rank (some being zero). Figure 2 shows the box plots of the absolute change in the 25 rankings of the 2004 and 2007 season, respectively. This is a further illustration of the topsy-turvyness of the 2007 season and the stability of the 2004 season.





Figures 2. Box plot of changes in rank for seasons 2004 and 2007

Season TT factor

The *Season TT factor* is also calculated at the end of each week to gauge how topsy-turvy the season has been so far. The Season TT factor is calculated using weighted averages of the Week TT factors. As the season progresses, the Week TT factors are given more weight in the calculation of the Season TT factor because toward the end of the season, an upset of a ranked team is more topsy-turvy than an upset in the beginning of the season when the strength of a ranked team is less established.

The weight given to each Week TT factor in the Season TT factor formula is equal to 1+ (Week Number of the Season /Number of Seasons in a Week). For example, the weight given to the Week TT factor in 2007 of fifth week is 1+5/15=1.3333. The formula for the calculation of the Season TT factor at the end of the i^{th} week is

$$(Season\ TT\ \ factor)_i = \frac{\sum_{j=1}^{i}(1+\frac{j}{n})\times(Week\ TT\ \ factor)_j}{i(2n+i+1)/(2n)} \qquad (3)$$

where

n= number of weeks in the full season.

Based on the Season TT factor formula, Figure 3 shows a box-plot of all the season TT factors. Note that season 2004 was mostly a very stable season as compared to seasons 2007 and 2003. On the other hand, season 2005 that was mostly a "middle-of-the-way" season, exhibited high variability in weekly topsy-turvyness.

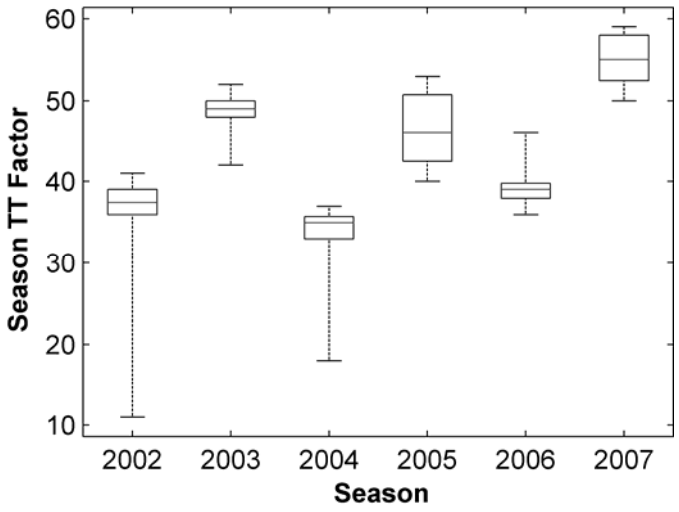


Figure 3. Box plot of season TT factors for six seasons (2002-2007)

The higher weights given to the later weeks in the Season TT factor do not result in a bias in the calculation of the Season TT factor. End-of-season TT factors calculated with the above weightage and equal weightage differ by less than 3%.

Effect of fall-out-of-rank number

In the calculation of the TT factors, for teams falling out of rankings, we used a rank number of 26. The reader may question that using some number other than 26 for the fall-out-of-ranking number may result in a different conclusion about the topsy-turvyness. The argument for choosing 26 as the rank of teams that become unranked is as follows.

The rankings of the teams could be extended beyond 25 by using the votes received by the unranked teams. However, this approach would suffer from several drawbacks.

1. Not all teams that get unranked get votes in the current week.
2. A team getting one or two votes is not a measure of a true ranking.
3. Low ranked teams falling out of the rankings do not warrant same weightage as the high ranked teams getting unranked because it is the fall of the high ranked teams that determines the topsy-turvyness.

We conducted a sensitivity analysis of the fall-out-of-ranking number. First, a suitable range for fall-out-of-ranking number needs to be found. Consider the votes

received by teams that fall out of ranking, and use those to give a ranking[1] of over 25 to them. For a topsy-turvy season such as that of 2007, the average fall in the ranking of the teams falling out of the Top 25 was 12.1 (standard deviation 5.2). For the same season, the average fall in rankings by using the rank number of 26 for teams falling out of the Top 25 is 6.7 (standard deviation 6.3). Based on this we chose a range of 26 to 42 (26 + difference in average fall in rank + two times the standard deviations $= 26 + (12.1 - 6.7) + 2 \times 5.2 \approx 42$) for the fall-out-of-rank number. The end-of-season TT factors show the same trend across the seasons (Figure 4). Note that a direct comparison cannot be made between the values of the TT factors obtained for each fall-out-of-rank number as both the numerator and the denominator of Equation (1) change accordingly.

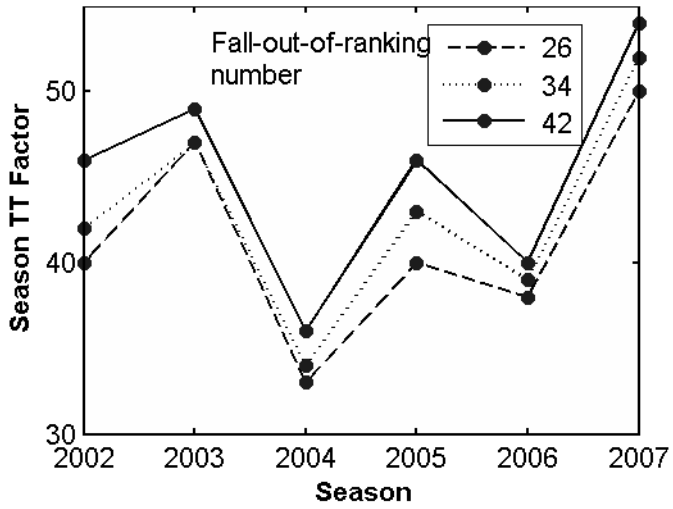


Figure 4. Effect of fall-out-of-ranking number on end-of-season TT factor

Other metrics

Another way of quantifying the topsy-turvyness of a college football season is to find the percentage of weeks of a season for which the Week TT factor is high. To do so, we calculated the average and the standard deviation of all the Week TT factors for the past six seasons (2002-07). For the Week TT factor, the average turns out to be 42.1, while the standard deviation is 12.4.

[1] In each weekly AP Poll, in addition to the Top 25 teams, other teams that also get votes are listed. We use the votes received by the teams to rank them beyond 25. If the previously ranked team received no votes in the next week, we ranked them as the team after the last team that received a vote.

If the Week TT factor is the average plus one standard deviation (that is 42.1+12.4=53.5) or more, we consider it as a measure of a topsy-turvy week. And, if the Week TT factor is the average Week TT factor less one standard deviation (that is 42.1-12.4=29.7) or less, it is a measure of a stable week.

Figure 5 shows the percentage of weeks for each of the last six seasons that were topsy-turvy and stable. Also shown are the end-of-season TT factors. These results agree with the previous assessment where seasons 2003 and 2007 are topsy-turvy, and the season 2004 is stable. In season 2007, no week fell in the category of a stable week, while 33% of the weeks were topsy-turvy. In contrast, in season 2004, 33% of the weeks were stable, and only 7% of the weeks were topsy-turvy.

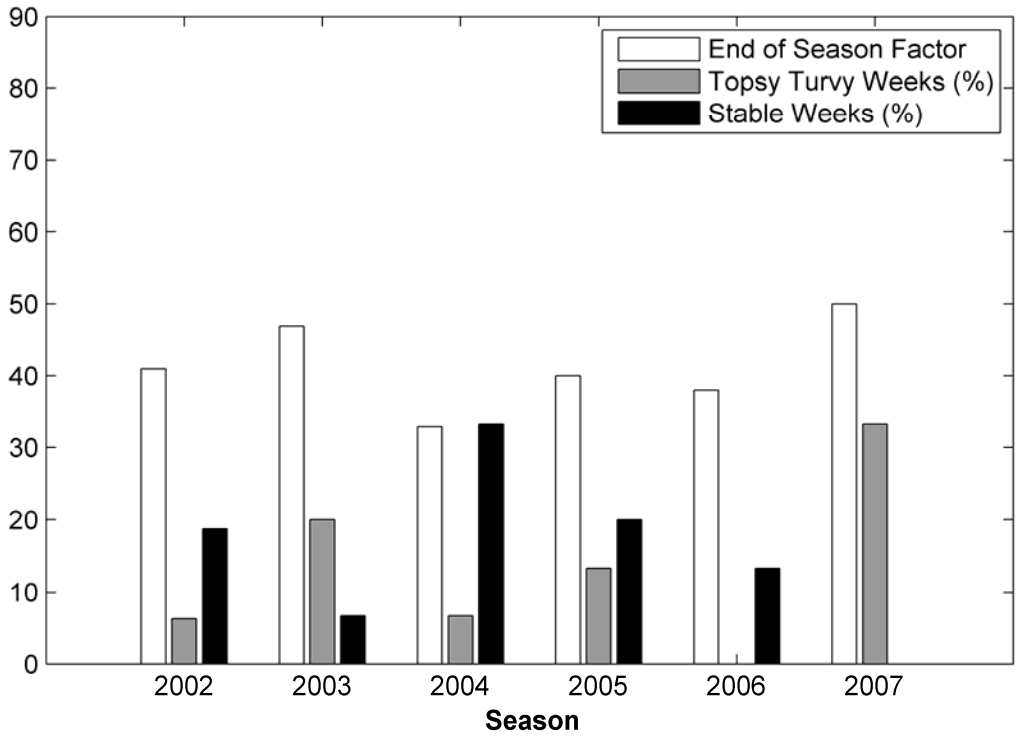


Figure 5. Percentage of stable and topsy-turvy weeks

TT factor based on other Polls

Would using ranking polls other than the AP Top 25 give different results? For this, we considered the USA Today poll rankings that are calculated by polling the USA Today board of 63 Division 1-A head coaches. Each voter supplies his or her ranking of the top 25 teams. The individual votes are added by giving 25 points to the first place vote, 24

points to the second place vote, etc. The addition of the points then produces the list of the USA Today top 25 teams of the week.

Figure 6 compares the week TT factors obtained from the AP and USA Today polls for the two seasons of 2004 and 2007. Although for a few weeks, the Week TT factors based on the AP and the USA Today polls differ slightly, both polls give very similar trends. Table 1 shows the end-of-season TT factors obtained using the AP and USA Today polls for all the six seasons. The maximum difference between the two is less than 5%.

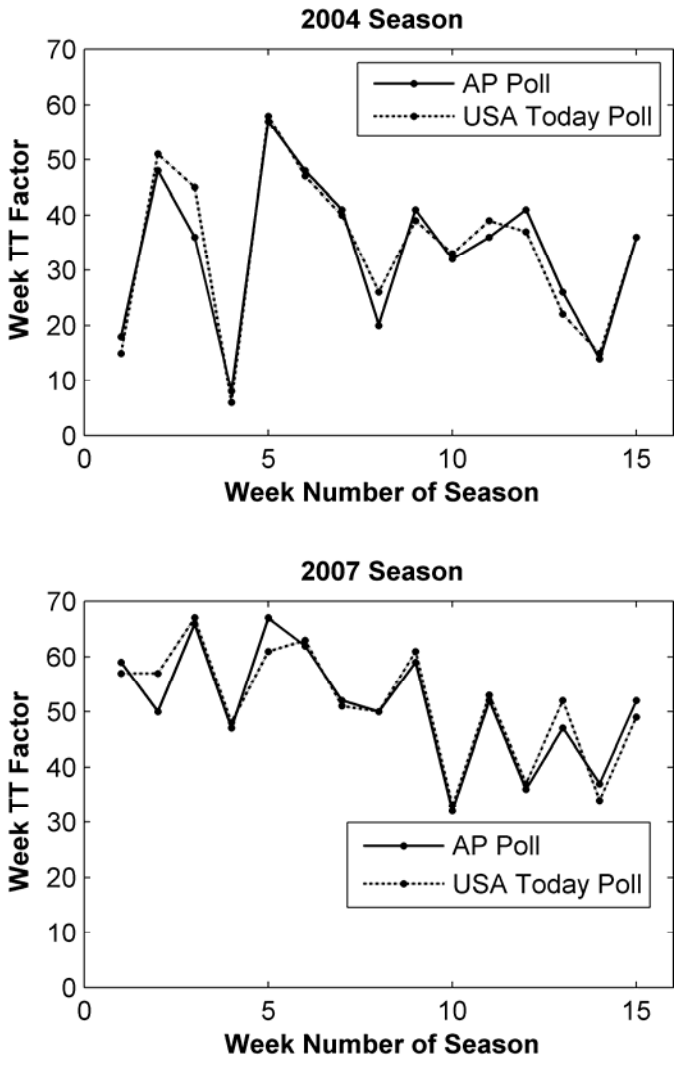


Figure 6. Comparing the Week TT factors from AP and USA Today poll for seasons 2004 and 2007.

Table 1. Comparing the end-of-season TT factors from AP and USA Today poll.

Season	End-of-season TT Factor	
	AP Poll	USA Today Poll
2002	41	41
2003	47	45
2004	33	34
2005	40	40
2006	38	39
2007	50	50

Other measures of disarray

How does the TT factor or the percentage weeks of high TT factor compare with other common measures of disarray such as the normalized Kendall's tau distance or the Spearman's rank correlation coefficient?

The normalized Kendall tau distance, *K* is a measure of discordant pairs between two sets, the sets in our case being the rankings from two consecutive weeks. The distance, *K* varies between 0 and 1, where 0 represents identical and 1 represents total disagreement in rankings. The trend of the number (1-*K*) for each week through the season is similar to the Week TT factors but the distinctness between the seasons is not as clear to differentiate between a topsy-turvy and a stable season.

The Spearman's rank correlation coefficient, ρ, is a measure based on the square of the difference between the rankings of the two sets. The coefficient, ρ varies from -1 to 1, where -1 represents total disagreement and 1 represents identical ranking. The trend of the number (1- ρ) for each week through the season is similar to the week TT factors. However, the TT factor presents a more appropriate measure of topsy-turvyness because of the following.

1) The teams that get unranked still get a rank of 25 or less in the formula for ρ, and hence introduces a bias which becomes larger in weeks where a significant number of teams fall out of rankings. For example, if four teams get unranked in a particular week, they all are assigned a rank of 23.5 [= (22+23+24+25)/4].
2) If a low ranked team loses and are out of the Top 25 ranking, they may get a higher rank in the formula for ρ. For example, if four teams get unranked in a

particular week, and one of the teams was ranked 25 in the previous week, it will be assigned a higher rank of 23.5 [= (22+23+24+25)/4].

Is topsy-turvyness random?

To determine the degree of randomness in weekly topsy-turvyness measured by the weekly TT factor, we calculated the lag one-autocorrelation coefficient r for the each season between 2002 and 2007. The results are shown in Figure 7. The autocorrelation factors for all seasons were ranging between ± 0.4, which does not indicate the presence of non-randomness in weekly topsy-turvyness throughout a season.

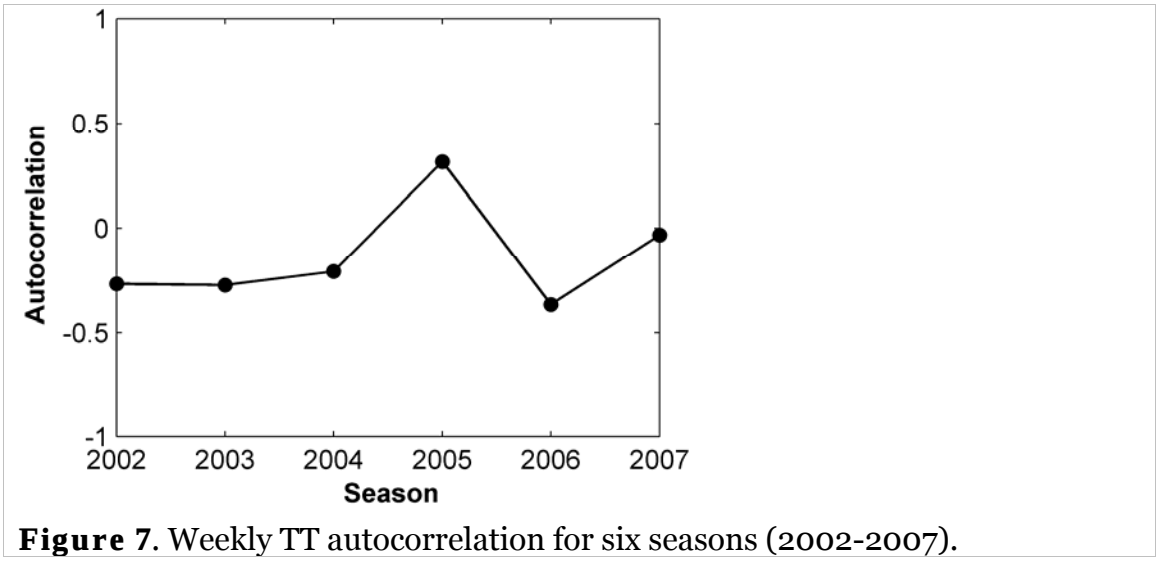


Figure 7. Weekly TT autocorrelation for six seasons (2002-2007).

Are the final week (postseason) TT factors statistically different from those of the regular season? Inspection of the results did not reveal any significant difference of final week TT factors from those of other weeks. This can be attributed to the sheer number of college bowl games in the postseason. In 2007 season alone, 64 teams played in the college bowls, which included 39 pre-bowl unranked teams. Sixteen bowls were played between unranked teams; 7 matched a ranked and an unranked team; and only 9 had ranked teams face each other. With match-ups like that, the bowl games are seemingly like any other regular season week except that more highly and closely ranked teams play each other.

To answer conclusively the question of whether any of the weeks in the season tend to be more or less topsy-turvy than the other weeks, we conducted an analysis of variance of topsy-turvyness based on a randomized complete block design where the weeks were treatments, and the seasons were blocks. The resulting ANOVA table is

shown in Table 2. As expected, the seasons showed a very significant difference between mean weekly topsy-turvyness across seasons. The analysis also indicated that there was a significant difference between the mean topsy-turvyness across weeks. Comparison of pairs of treatment means using Tukey's test did not reveal any significant trends but did indicate that week 14 was significantly less topsy-turvy than week 6. Though it was not statistically significant as compared with other weeks, the last week of the season (week 14), persistently had much lower Week TT factor scores than the rest of the weeks in the season. This is attributed to the fact that pre-bowl week involves conference championships (mostly a match-up between high ranked teams) and that many top ranked teams (44% in 2007) have already finished their regular season the week before.

Table 2. Analysis of variance for the weekly TT factors[2].

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-value
Weeks	3562.3	14	254.4	2.6	0.0044
Seasons	2936.5	5	587.3	5.9	0.0001
Errors	6914.4	70	98.8		
Total	13413.1	89			

Conclusions

Based on ranking change of college football teams from week to week in AP polls, a metric to measure the topsy-turvyness of college football weeks and seasons has been developed. Six recent seasons (2002-07) were used in the analysis. The 2007 season turned out to be the most topsy-turvy, while the 2004 season was the most stable. These findings were confirmed with other measurements such as change in ranking from week to week, and number of weeks that would be deemed topsy-turvy or stable based on the average and standard deviation of all the TT factor numbers. Other polls such as the USA Today poll resulted in similar trends of the TT factor as obtained using the AP polls.

In depth-statistical analysis of the weekly TT factors did not indicate presence of non-randomness in weekly topsy-turvyness through the season. Using a randomized

[2] For season 2002, of the 16 weeks, the first week results were not included in the analysis. All other seasons have data for 15 weeks.

complete block design statistically significant differences were detected in the mean TT factors across weeks and across seasons. However, no significant trends were found in the TT factors except in the last week before the bowls are played.

The Week TT factors and Season TT factors will continue to be calculated in every week of the future college football seasons so that media and fans alike can quantitatively judge the topsy-turvyness of a college football week.

The methodology to determine topsy-turvyness described here can be used for other sports that are ranked throughout the season such as college basketball and baseball.

Further Reading

BCS coordinator says increased parity won't encourage BCS change, November 28, 2007, Associated Press, http://sports.espn.go.com/ncf/news/story?id=3132578, Last accessed January 9, 2008.

Bagnato, Andrew, In topsy-turvy season, 6 teams remain perfect - so far, USA Today, October 16, 2007, http://www.usatoday.com/sports/college/football/2007-10-16-90477001_x.htm, Last accessed January 9, 2008.

Box, G. E. P. and Jenkins, G. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.

ESPN - NCAA College Football Polls, College Football Rankings, NCAA Football Poll, http://sports.espn.go.com/ncf/rankingsindex, Last accessed January 9, 2008.

Montgomery, D.C. (2001). Design and Analysis of Experiments, 5th Edition. John Wiley and Sons Inc., New York.

Thamel, Pete, Topsy-Turvy Season Faces More Flips, New York Times, October 15, 2007, http://www.nytimes.com/2007/10/15/sports/ncaafootball/15colleges.html, Last accessed January 9, 2008.

TT Factor for College Football, http://www.eng.usf.edu/~kaw/ttfactor

Contact Information

Autar K. Kaw
Professor, Department of Mechanical Engineering,
University of South Florida
4202 E Fowler Ave ENB118, Tampa FL 33620-5350.
Office: (813) 974-5626 | Fax: (813) 974-3539 |E-mail: kaw@eng.usf.edu

Ali Yalcin
Associate Professor, Department of Industrial and Engineering Management
University of South Florida
4202 E Fowler Ave ENB118, Tampa FL 33620-5350.
Office: (813) 974-5590 |E-mail: ayalcin@eng.usf.edu