

1 Introduction

Coaches are a high commodity in college football. The head coach for a division I college football team is many times the highest paid position in the respective state. blah blah blah.

2 Methods

Data were scraped from various web sources (primarily cfldatawarehouse.com) to build a large database. This database consisted of information at the team level (conference affiliation), coach level (number of wins, tenure length), and game by game level (game score, winning team, game location).

To get an estimate of the coaching ability, game by game win/loss status (dichotomously scored where a win was scored as 1, a loss or tie was scored as 0) was compiled within a given year and coach. This large matrix consisting of 1's and 0's was used in an item response theory (IRT) framework. The coaches were treated as the items and the games were treated as the subjects. A rasch model was fitted of the following form:

$$p(x_j = 1 | \theta, b_j) = \frac{e^{(\theta - b_j)}}{1 + e^{(\theta - b_j)}}. \quad (1)$$

In the above equation, the probability that a coach wins a given game is a function of the ability (θ) and the specific coaches ability (b_j). Treating coaches as items in an IRT framework gives an estimate of the coaches ability level (b_j). This procedure was done for all years for all coaches that had more than 2 games for a team in a given season.

2.1 Inferential Model

One of the strengths of IRT is that the resulting item parameters can be treated as a continuous scale of measurement (as opposed to the count metric of the number of wins in a given season). This allows the modeling of the coach ability parameter estimates over time to be simpler and easier to estimate. As a result, the following continuous linear mixed model was fitted to the coach ability IRT estimates:

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j + \mathbf{e}_j \quad (2)$$

This matrix equation is a linear representation of the coaches ability estimate (\mathbf{Y}_j) as a function of fixed effects ($\mathbf{X}_j\boldsymbol{\beta}$), random effects ($\mathbf{Z}_j\mathbf{b}_j$), and random error (\mathbf{e}_j).

The fixed effects are analogous to simple regression coefficients and represents the aggregate growth curve for all coaches. The random effects represent coach specific deviations from the aggregate or average growth curve. Lastly, the random error would represent deviations from the individual coach growth curve.

Covariates were mean centered when they did not have a meaningful zero by default. In addition, if a covariate had a significantly greater variance than others, they were scaled into z-score like terms before being entered into the model. Lastly, time (i.e. years coaching division I) was rescaled so that 0 represents their first year coaching and incrementing by 1 for subsequent years of coaching at the division I level.

3 Results

3.1 Inferential

Table 1: Model estimates

	Estimate	Std. Error	t value
(Intercept)	0.296	0.132	2.240
Year2	-0.054	0.017	-3.183
overWinmc	0.023	0.002	9.533
alltimemc	0.063	0.096	0.654
numAA	1.374	0.109	12.643
numGamesmc	0.124	0.046	2.707
power5conf	-0.590	0.194	-3.041
Year2:overWinmc	-0.001	0.000	-6.280
Year2:alltimemc	0.016	0.010	1.671
Year2:numAA	-0.018	0.009	-1.999
Year2:numGamesmc	0.028	0.004	6.297
Year2:power5conf	-0.030	0.019	-1.534

Notes: Variables above reflect: Year2 is the slope for time discussed briefly above, overWinmc is the total number of wins for a given coach all time – this variable makes more sense as a time varying covariate, but the way I merged this variable was best for now, alltimemc is the all time rankings as posted on cfb-datawarehouse – I would argue this represents prestige, numAA is the number of all americans for every year, numGamesmc is the number of games in a given year (hopes to account for the helping of those coaches with more chances to win), and

power5conf is an indicator where it equals 1 if the team/coach is a member of a power 5 conference and 0 otherwise (I wonder how far back this variable makes sense).

4 Discussion