The University Rankings Game:
Modeling the Competition among Universities for Ranking

APPENDIX
Adjacent Category Logit Model

Let:

- \( u = 1 \ldots U \) as the total number of universities that were ranked for every year in the observation period.
- \( t = 1 \ldots T \) as the number of the time period for which we observe the university ranks (\( T = 8 \) in our sample).
- \( r = 1 \ldots R \) as indexing ranks, where for top 50 universities we study \( R = 50 \).
- \( X_{ut} \) = the matrix of explanatory variables for university \( u \) at time \( t \). (These variables are the ones used by USNews)
- \( \pi_r \) = the probability of observing rank \( r \), such that \( \pi_r \geq 0 \), \( \forall r = 1 \ldots R \) and \( \sum_{r=1}^{R} \pi_r = 1 \).

For adjacent categories \( r \) and \( r+1 \), we define the adjacent-categories log-odds unit (LOGIT) as (Goodman 1983; Simon 1974):

\[
\text{logit}\left[ \frac{P(Y_{ut} = r \mid X_{ut}, Y_{u(t-1)})}{P(Y_{ut} = (r+1) \mid X_{ut}, Y_{u(t-1)})} \right] = \log\left( \frac{\pi_r(X_{ut}, Y_{u(t-1)})}{\pi_{r+1}(X_{ut}, Y_{u(t-1)})} \right),
\]

where, \( Y_{ut} \) is the rank for the university \( u \) at a given year \( t \) and

\[ \pi_r(X_{ut}, Y_{u(t-1)}) = P(Y_{ut} = r \mid X_{ut}, Y_{u(t-1)}) . \]

To incorporate explanatory variables, the logarithm is specified as a linear function of the explanatory variables (e.g., McCullagh 1980):

\[
\log\left( \frac{\pi_r(X_{ut}, Y_{u(t-1)})}{\pi_R(X_{ut}, Y_{u(t-1)})} \right) = \alpha_r + Y_{u(t-1)}\beta_r + X_{ut}\gamma_r + Y_{u(t-1)}X_{ut}\delta_r.
\]

As is the case in logit models with multiple categories, \( \pi_r(X_{ut}, Y_{u(t-1)}) \) is defined as:

\[
\pi_r(X_{ut}, Y_{u(t-1)}) = P(Y_{ut} = r \mid X_{ut}, Y_{u(t-1)}) = \frac{\exp(\alpha_r + Y_{u(t-1)}\beta_r + X_{ut}\gamma_r + Y_{u(t-1)}X_{ut}\delta_r)}{1 + \sum_{r=1}^{R-1} \exp(\alpha_r + Y_{u(t-1)}\beta_r + X_{ut}\gamma_r + Y_{u(t-1)}X_{ut}\delta_r)}.
\]

The probability specification in Equation 4 leads to the familiar likelihood function (\( L \)):
where, $I_{utr}$ is an indicator function that equals 1 if $Y_{ut} = r$, else it equals 0.

As we model lag of rank as an explanatory variable in Equation 4, $t = 2, \ldots, T$. To recognize the hierarchy inherent in the ranking we specify $\beta_r = (R - r)\beta$ and $\gamma_r = (R - r)\gamma$. Once parameter estimates are available by maximizing the logarithm of Equation 4, we calculate the probability of change in rank as:

$$
\frac{\pi_r(X_{ut}, Y_{ut(t-1)})}{\pi_{r+p}(X_{ut}, Y_{ut(t-1)})} = \exp(\alpha_r + pY_{ut(t-1)}\beta + pX_{ut}\gamma + pY_{u(t-1)}X_{ut}\delta).
$$

With $Y_{u(t-1)} = r$ modeled as an explanatory variable, we can use Equation 8 to calculate the probability of $Y_{ut} = r + p$, $\forall \ p = 0, 1, 2, \ldots$, $\forall r = 1, \ldots, R-1$. 

$$
L = \prod_{u=1}^{U} \prod_{t=2}^{T} \prod_{r=1}^{R} [\pi_r(X_{ut}, Y_{u(t-1)})]^{I_{utr}},
$$