Research in Higher Education, Vol. 42, No. 3, 2001

NEWS OR NOISE?
Estimating the Noise in the *U.S. News* University Rankings

Ilia Dichev

This study investigates the quality of the *U.S. News* annual ranking of national universities and liberal arts colleges. The main finding is that current rankings changes have a strong tendency to revert over the following two rankings. Using a simple model, this study estimates that about 70 to 80 percent of the variation in rankings changes is transitory and reversible. Thus, most of the "news" in the annual rankings is essentially meaningless noise. An analysis of possible explanations suggests that the noise in annual ranking changes is most likely due to various measurement, estimation, and other information processing errors in the rankings' underlying components.

KEY WORDS: university; rankings; *U.S. News*.

INTRODUCTION

This study investigates the quality of the *U.S. News and World Report* (hereafter *USN*) annual ranking of U.S. colleges and universities. Although there are other ranking sources (e.g., the *Money* magazine ranking of best buys in college education), *USN* publishes by far the most important college rankings. The *USN* ranking has the longest history and is generally regarded as the most authoritative and least problematic of all college rankings (Webster, 1992). Additionally, it is the most commercially successful. *USN* estimates that it sells over 2.2 million copies of the college rankings issue, reaching an end audience of 11 million people, and an additional 1 million copies of the related college guidebook. Taken together, the *USN* ranking publications account for nearly half of the total market of 6.7 million copies of newsmagazine college rankings and guides (McDonough, Antonio, Walpole, and Perez, 1998). The success of the

Address correspondence to: Ilia Dichev, University of Michigan Business School, 701 Tappan Street, Ann Arbor, MI 48109-1234; dichev@umich.edu.

college rankings has prompted a substantial increase in the breadth and the depth of the *USN* coverage. Today, *USN* produces rankings for a variety of institutions of higher education, including national and regional universities and liberal arts colleges, specialty schools, and graduate and professional schools.

The influence and the importance of the rankings have also gradually grown for users of the rankings. Perhaps the most important reason for the success of the rankings is users' demand for objective and reliable information about college quality. College quality is not directly observable, and traditional college catalogues and viewbooks are often uninformative and self-serving. Thus, faced with the large cost of college education, applicants and their parents have been increasingly looking for independent verification of quality (Webster, 1992). Each year about 400,000 prospective students consider the rankings an important factor in deciding which school to attend (McDonough et al., 1998). In addition, McDonough and associates show that prospective students who use the rankings tend to have higher high school grades and in general are more sophisticated and better prepared for college. Thus, from the universities' perspective, college rankings influence more heavily the most desirable college applicants. Another sign of the importance of the rankings is that changes in the rankings cause perceptible ebbs and flows in the number and quality of applicants (e.g., Fombrun, 1996). As a result, albeit grudgingly, most top university administrations have accepted the importance of the rankings, and many have even attempted to manipulate their rankings. For example, schools have manipulated their admission policies and the reporting of their average SAT scores to affect the rankings (Hunter, 1995).

Despite the importance of college rankings, very little is actually known about the properties and the quality of these rankings. Existing sources (e.g., Webster, 1992) mainly argue whether *USN* uses "good" inputs into the rankings (e.g., SAT scores vs. high school grades), and what is the "best" way to combine the inputs to produce the rankings (e.g., how much weight to assign to each input). These issues are undoubtedly important but there are disagreements about what "best" should be, and it is difficult to address these issues directly in a rigorous manner. Instead, the present study uses a different approach. It takes the existing rankings as given and investigates their properties for quantitative evidence about their quality.

More specifically, this study investigates the predictability of rankings changes. One motivation is that finding little or no predictability in changes essentially means that current rankings are unbiased forecasts of future rankings. Forecasting ability is likely an important consideration for college applicants because, while they choose a college based on current rankings, what they really care about is the college rankings at graduation and beyond, when they will affect the value of their human capital. Since college applicants are the most important users of the rankings, it seems useful to investigate the predictability of rankings changes. A related motivation is that if current rankings summarize all relevant information efficiently, then changes should be triggered only by new information. Since new information is by definition unpredictable, "good" rankings should have little or no predictability in changes.

In any case, finding predictabilities in rankings changes can yield important insights into the rankings. For example, it is possible that current rankings are poor forecasts of future rankings but are a good measure of contemporaneous real performance—much like a bestseller ranking is often a poor predictor of future sales but accurately captures contemporaneous performance. To further illustrate the implications of potential predictabilities in changes, assume that rankings capture contemporaneous performance and that one finds a positive autocorrelation in rankings changes. This evidence could be indicative of feedback effects in real performance, where an improvement in the rankings translates into better applicants, and better applicants in turn lead to a further improvement in the rankings.

The empirical evidence is based on an investigation of time-series predictability in changes for the two most important rankings, *USN*'s Top 25 rankings of national universities and national liberal arts colleges. The main finding is that changes in the *USN* rankings have a strong tendency to revert in the next two rankings. The reversibility in rankings is strong not only in statistical terms but seems to account for a strikingly large part of the total variation in rankings changes. Using a simple model of two-period reversibility, it appears that between 70 to 80 percent of the variation in rankings changes is due to "noise," to transitory effects, which quickly disappear in later rankings. Thus, most of the "news" in *USN*'s annual college rankings is essentially meaningless noise.

The study also identifies and evaluates three possible sources of reversibilities in the rankings changes: reversibilities in methodology, "fight-back" effects in real performance, and that *USN* rankings are essentially mechanical aggregations of noisy and reversible components. A consideration of published ranking methodologies reveals that methodology reversals are unlikely to account for the quick reversals in the rankings. Changes in methodology are mostly gradual, fairly small, and do not show obvious signs of reversals. Thus, if anything, changes in methodology likely cause nonreversible changes in the rankings.

The observation that changes in the methodology likely cause nonreversible changes in the rankings is interesting because it has two important implications. First, if 70 to 80 percent of the rankings changes are reversible, and methodology-induced changes account for a material portion of the nonreversible changes, then it follows that real changes in fundamental school quality (which are nonreversible) account for possibly as little as 10 percent of the variation in published rankings changes. Second, critics of the *USN* rankings often point to methodology-induced changes as a major problem in interpreting and comparing the rankings (e.g., Webster, 1992; Karl, 1999). However, the quantitative estimates above suggest that in interpreting the rankings, predictable reversibilities in changes pose an even bigger problem than methodology-induced changes.

"Fight-back" effects in real performance are also unlikely to account for the large and quick reversibility in rankings changes. If schools that experience a deterioration in the rankings fight back extra hard to regain their status, then one would expect that fight-back effects would be especially pronounced for intermediate to long-term changes in the rankings. However, additional evidence reveals little or no reversibility in 5-year changes in the rankings. This evidence is the opposite of what one might expect if fight-back effects in real performance were to account for the reversibilities in the rankings.

The third possible explanation for the reversibilities in annual rankings changes is based on the observation that published *USN* rankings are essentially simple mechanical aggregations of noisy underlying components. Thus, the rankings mechanically incorporate the noise in the underlying components. Since the noise in the underlying components predictably reverses in the future, the aggregated rankings also exhibit large and quick reversals in changes. This explanation is supported by additional evidence that, unlike with methodology, ranking components exhibit the same large and quick reversals in changes. The statistical evidence is complemented by considering specific cases where evidently large reversals in ranking components produce large reversals in the overall rankings.

The evidence in this study has implications for both preparers and users of the rankings. It seems that, unless happy with producing meaningless "news," *USN* should employ some sort of explicit statistical analysis and filtering of the noise in the rankings components. Such procedures would result in a substantial reduction of the present variability of annual rankings changes. Rankings users themselves can also correct the deficiencies of *USN*'s ranking methodology. If *USN* continues its current practices, users should discount heavily the revisions in the annual rankings updates, and look to long-term rankings and trends for more reliable evidence on relative school quality.

INVESTIGATING THE PREDICTABILITY OF RANKINGS CHANGES

Searching for predictability in rankings changes necessarily involves specifying a model of possible predictabilities in changes. A logical place to start is the investigation of time-series predictability, that is, whether one can use past changes in the rankings to explain and predict current changes in the rankings. I use regression analysis to identify time-series patterns in predictability in rankings.

An investigation of the rankings data and preliminary analyses showed that it is useful to present more explicit evidence about the nature and the characteristics of the rankings data. This preliminary evidence illustrates the logic behind the test design and helps in the interpretation of the results. Table 1 presents the sample of *USN* Top 25 rankings of national universities and national liberal arts colleges for 1987 to 1998. *USN* produces a variety of college rankings, including Top 25 for national universities and national liberal arts colleges, second-, third-, and fourth-tier rankings for national-level schools, and rankings of regional schools and graduate and professional schools. However, to keep the analysis tractable, I focus on the two most important rankings—the Top 25 for national universities and national liberal arts colleges. Additionally, these two rankings are the oldest and most complete, which allows for the following fairly extensive statistical analyses.

USN issued the first rankings of national universities and liberal arts colleges in 1983, issued new releases in 1985 and 1987; the rankings have become an annual event since 1987. The 1988 ranking reflected a pronounced shift in methodology. The rankings before 1988 were based exclusively on surveys of college reputation. The rankings after 1988 are derived as summary measures from several underlying components, which reflect both surveys and objective measures like entering students' exam scores, graduation rates, and so on. Thus, the large shift in methodology from 1987 to 1988 caused a large and somewhat anomalous shift in the rankings for many schools. For example, Panel A in Table 1 shows that the California Institute of Technology was ranked 21 in 1987, 3 in 1988, and it was never ranked below 9 thereafter. The University of California at Berkeley was ranked 5 in 1987, 24 in 1988, and it was never ranked above 13 later. Similar large and mostly sustained drops occurred for the University of Michigan and the University of North Carolina. An examination of Panel B demonstrates similar large and anomalous shifts in the rankings between 1987 and 1988 for national liberal arts colleges. The rankings methodology has continued to evolve over the remaining years, but the changes have been comparatively small and incremental. An inspection of the rankings also does not reveal such numerous and drastic changes in the rankings for later years. Thus, to avoid the effects of regime shifts and nonhomogeneity, I restrict the test sample to observations from years 1988 to 1998.

An examination of the raw data in Table 1 also helps to develop a qualitative feel for the data, which is useful in interpreting the statistical results later. One striking feature that emerges from this examination is that the composition of the Top 25 schools is remarkably stable over time. For example, Panel A reveals that there are only 29 schools that have been ranked at least once in the top 25 over the 11 years spanning 1988 to 1998. Twenty schools have always been in the Top 25 schools, while schools that appear in and out of the rankings tend to be confined mostly to the lower tiers of the Top 25 ordering. Apparently, it is quite difficult for new entrants to break into this elite group of schools, and it is uncommon for current Top 25 members to leave (especially for schools from the upper echelons). A related observation is that relative standing in the

TABLE 1. The *USN* Top 25 Rankings for National Universities and National Liberal Arts Colleges for 1987 to 1998

	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Panel A: National Universities												
Brown University	10	13	15	12	17	18	12	11	9	8	9	10
California Institute of Tech.	21	3	4	5	5	5	5	7	7	9	9	9
Carnegie Mellon University	—	—	—	22	24	19	24	24	23	—	23	25
College of William and Mary	22	—	—	—	—	—	—	—	—	—	—	—
Columbia University, NY	18	8	11	10	9	10	11	9	15	11	9	10
Cornell University, NY	11	14	11	9	12	11	10	15	13	14	14	6
Dartmouth College, NH	6	7	8	8	8	7	8	8	7	7	7	10
Duke University, NC	7	12	5	7	7	8	7	6	6	4	3	6
Emory University	25	22	—	—	—	21	25	16	17	19	9	16
Georgetown University	—	17	25	19	19	17	17	25	21	23	21	20
Harvard University	2	4	3	1	1	1	1	1	1	3	1	1
Johns Hopkins University	16	11	14	15	11	15	15	22	10	15	14	14
MIT, MA	11	5	7	6	6	5	4	4	5	5	6	4
Northwestern University, IL	17	16	19	23	14	13	13	14	13	9	9	10
Princeton	4	2	2	4	4	2	2	2	2	2	1	1
Rice University	14	9	10	16	15	12	14	12	16	16	17	18
Stanford University, CA	1	6	6	2	3	4	6	5	4	6	5	4
Tufts University	—	—	—	—	—	—	—	—	25	22	23	25
UC–Berkeley	5	24	13	13	16	16	19	23	—	—	23	22
UCLA	—	21	16	17	23	23	22	—	—	—	—	25
University of Chicago	8	10	9	11	10	9	9	10	11	12	14	14
University of Illinois at Urbana	20	—	—	—	—	—	—	—	—	—	—	—
University of Michigan	8	25	17	21	22	24	23	21	24	24	23	25
University of North Carolina	11	23	18	20	25	—	—	25	—	24		
University of Notre Dame	—	18	23	—	—	—	25	19	18	17	19	18

University of Pennsylvania	19	15	20	13	13	14	16	12	11	13	7	6
University of Rochester	—	—	—	25	—	—	—	—	—	—	—	—
University of Texas at Austin	25	—	—	—	—	—	—	—	—	—	—	—
University of Virginia	15	20	21	18	21	22	21	17	19	21	21	22
Vanderbilt University	—	—	24	—	19	25	20	18	22	20	19	20
Washington University	23	19	22	24	18	20	18	20	20	17	17	16
Yale University	3	1	1	3	2	3	3	3	2	1	3	1
Panel B: National Liberal Arts Colleges												
Amherst College	4	2	2	1	3	2	1	1	1	2	2	1
Barnard College	—	19	25	25	19	—	—	—	—	23	—	24
Bates College	—	21	21	19		23	18	21	18	22	20	19
Bowdoin College	12	9	13	4	4	6	6	6	4	8	8	7
Bryn Mawr College	11	13	5	23	11	10	15	14	9	10	9	15
Bucknell University	—	—	—	—	—	—	—	25	—	—	—	—
Carleton College	3	12	14	10	12	12	13	10	11	9	7	9
Centre College of Kentucky	22	—	—	—	—	—	—	—	—	—	—	—
Claremont McKenna College	23	18	21	13	14	16	11	—	16	15	10	15
Colgate University	—	16	16	22	17	17	19	18	17	20	21	21
Colby College	—	22	23	18	20	15	17	23	18	18	18	17
College of Holy Cross	—	—	—	—	23	—	25	24	—	—	24	—
Colorado College	20	25	—	—	—	—	—	20	—	—	—	24
Connecticut College	—	—	—	—	—	—	—	—	25	—	—	24
Davison College	15	20	15	10	21	18	12	8	21	11	11	11
Earlham College	16	—	—	—	—	—	—	—	—	—	—	—
Grinnell College	10	8	10	16	16	14	16	17	14	16	14	11
Hamilton College	25	24	23	19	25	21	22	—	23	25	22	21
Haverford College	9	10	11	21	8	9	7	6	5	6	6	5
Lafayette College	—	—	—	—	—	25	—	—	—	—	—	—
Macalester College	—	—	—	—	—	—	—	—	—	—	24	24

TABLE 1. (*Continued*)

	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Middlebury College	17	14	12	8	9	8	8	11	7	7	12	7
Mount Holyoke College	17	17	17	17	18	19	23	12	19	19	19	19
Oberlin College	5	11	9	14	15	20	21	19	22	24	23	24
Occidental College	—	—	—	24	24	—	—	—	—	—	—	—
Pomona College	6	5	4	6	5	5	5	5	8	5	5	5
Reed College	12	—	—	—	—	—	—	—	—	—	—	—
Smith College	14	7	7	9	10	10	9	13	10	12	15	11
St. John’s College	19	—	—	—	—	—	—	—	—	—	—	—
St. Olaf College	20	—	—	—	—	—	—	—	—	—	—	—
Swarthmore College	2	1	1	2	2	3	3	3	2	1	1	2
Trinity College	—	—	20	—	—	24	24	22	23	21	24	23
University of the South	—	—	—	—	—	—	—	—	—	—	—	24
Vassar College	24	15	19	10	13	13	14	16	13	17	17	17
Washington and Lee Univ.	25	23	18	15	22	22	20	15	15	13	13	11
Wellesley College	8	4	5	5	6	4	4	4	5	4	3	4
Wesleyan University	6	6	8	7	7	7	10	9	12	14	16	9
Williams College	1	3	3	3	1	1	2	2	2	3	4	3

rankings is also quite stable. For example, Harvard, Princeton, and Yale are always in the top 4, Dartmouth is never above 6 or lower than 10, and Cornell and Brown oscillate around 10. An examination of Panel B reveals a similar stability in composition and relative standing over time for national liberal arts colleges (although this stability seems not as pronounced as that for the national universities).

Table 2 presents some evidence about the empirical distribution of rankings changes. Panel A reveals that the distribution of ranking changes for national universities is fairly bell-shaped and centered on zero. However, the distribution has somewhat heavy tails (kurtosis of 2.36), especially the left tail (skewness of –0.72), and a test for normality is easily rejected at the 0.001 level. Since most of the following tests rely on normality assumptions, this evidence implies that inferences based on the raw data might be problematic. I address this problem by deleting the extreme observations in the sample. To illustrate the effect of progressive deletions, Panel A presents a set of descriptive statistics for the sample after deleting all observations that have absolute value of changes greater than 8 and another set of statistics after deleting all changes greater than 6. As one might expect, the successive deletions leave the mean substantially unchanged, reduce the standard deviation, and most importantly they substantially reduce the skewness and kurtosis of the distribution. At the end, the specification that deletes all observations with absolute value of changes greater than 6 leaves a sample for which normality is not rejected at the 0.10 level. This trimmed sample is used for all statistical tests later in the study. Additional tests reveal that the results are qualitatively similar for the less stringent trimming and for the untrimmed sample (actual results not presented).

Panel B in Table 2 presents evidence about the empirical distribution of rankings changes for national liberal arts colleges. Again, the distribution is centered on zero and fairly symmetrical. However, the distribution has more extreme observations for both negative and positive changes and has discontinuities for large positive changes. The statistics confirm that the distribution is more spread and has heavier tails than the one for the national universities as evidenced by higher standard deviation and kurtosis. Not surprisingly, the test for normality rejects at the 0.001 level for the untrimmed sample. To assess the effects of deletions, I present descriptive statistics after deleting all changes with absolute value greater than 10 and another one for values greater than 8. Again, the successive deletions reduce the skewness and kurtosis, and the more stringent trimming leaves a sample for which normality is not rejected at the 0.10 level. This sample is used for all following tests. In any case, using a less stringent trimming or no trimming has only a moderate effect on the results.

Table 3 presents regression results for the time-series predictability in changes for *USN* rankings. Since there are no priors about the length of time-series predictability, I start by regressing current rankings changes on 1-year lagged

TABLE 2. Properties of the Empirical Distribution of Changes in the *USN* Top 25 Rankings for National Universities and National Liberal Arts Colleges

Panel A: National Universities

The empirical distribution of rankings changes

Ch	−12	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8
N	1	1	1	2	2	2	5	4	7	6	18	44	52	35	30	11	7	7	4	2	2

Descriptive statistics for the distribution of changes for selected trimmings of outliers

Trimming Rule	*N*	Mean	*SD*	Skewness	Kurtosis	Test for *f*(Ch) ~ Normal?
All available observations	243	–0.05	3.03	–0.72	2.36	$p < 0.001$
Delete all observations with –8 < Ch < 8	229	0.10	2.65	-0.14	1.25	$p < 0.005$
Delete all observations with –6 < Ch < 6	212	0.14	2.30	-0.02	0.68	$p > 0.100$

Panel B: National Liberal Arts Colleges

The empirical distribution of rankings changes

Ch	−13	−12	−11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	—	10	11	—	13	—	18
N	1	1	1	1	2	2	1	2	13	7	10	18	32	58	31	21	11	7	5	6	2	—	1	1	—	1	—	1

Descriptive statistics for the distribution of changes for selected trimmings of outliers

Trimming Rule	*N*	Mean	*SD*	Skewness	Kurtosis	Test for *f*(Ch) ~ Normal?
All available observations	236	–0.12	3.60	0.28	4.45	$p < 0.001$
Delete all observations with −10 < Ch < 10	220	–0.12	2.85	–0.16	1.31	$p < 0.005$
Delete all observations with –8 < Ch < 8	212	–0.11	2.67	–0.13	0.61	$p > 0.100$

Table 2 presents the empirical distribution of rankings changes. It also presents descriptive statistics for the distribution after successive trimmings of outliers. Ch denotes a change of a given magnitude. *N* denotes the number of available observations. *SD* denotes standard deviation. The test for normality of the distribution of rankings changes is the Jarque-Bera test. The *p* values are from chi-square scores of 77.39, 15.66, and 4.10 for national universities, and 197.81, 16.67, and 3.88 for national liberal arts colleges (chi-square test with 2df).

TABLE 3. Regression Results for the Relation Between Current and Lagged Changes for the *USN* Top 25 Rankings of National Universities and Liberal Arts Colleges (Regression: CurCh = b_0 + b_1*Lag1Ch + b_2 *Lag2Ch + b_3*Lag3Ch + b_4 *Lag4Ch + ε)

	Intercept	Lag1Ch	Lag2Ch	Lag3Ch	Lag4Ch	Adj. R^2	*N*
Panel A: National Universities							
Coefficients	0.10	–0.20				0.04	193
(*t* statistics)	(0.68)	(–3.15)					
Coefficients	0.08	–0.30	–0.23			0.12	162
(*t* statistics)	(0.55)	(–4.17)	(–3.63)				
Coefficients	–0.01	–0.24	–0.18	0.10		0.11	133
(*t* statistics)	(–0.07)	(–3.07)	(–2.36)	(1.40)			
Coefficients	–0.05	–0.20	–0.15	0.09	0.11	0.07	104
(*t* statistics)	(–0.30)	(2.08)	(–1.57)	(0.96)	(–1.23)		
Panel B: National Liberal Arts Colleges							
Coefficients	–0.09	–0.26				0.06	187
(*t* statistics)	(–0.50)	(–3.56)					
Coefficients	0.07	–0.38	–0.19			0.14	156
(*t* statistics)	(0.37)	(–5.17)	(–2.66)				
Coefficients	0.05	–0.40	–0.16	0.05		0.12	128
(*t* statistics)	(0.28)	(–4.25)	(–1.88)	(0.67)			
Coefficients	0.12	–0.42	–0.23	0.08	–0.02	0.16	104
(*t* statistics)	(0.55)	(–4.18)	(–2.22)	(0.83)	(–0.23)		

CurCh denotes current changes, Lag1Ch denotes 1-year lagged changes, and so on. All regressions are run after the trimming of outliers. For national universities, outliers are defined as all changes with an absolute value greater than 6. For liberal arts colleges, outliers are defined as all changes with an absolute value greater than 8.

changes and then proceed to include longer-lagged changes, up to four lags before current changes. An inspection and comparison of adjusted R^2, and magnitude and significance of the coefficients across regressions allows one to draw inferences about the nature and the length of the relation between current and lagged changes. The main conclusion seems to be that for both national universities and liberal arts colleges there is a significant negative relation between current changes and first- and second-lag changes, but no reliable evidence for higher-order lags. For both panels, it is the second regression (the one that includes first- and second-lagged changes only) that offers the best combination of explanatory power and parsimony. Therefore, this study adopts the model of current changes as a function of first- and second-lagged changes for the remaining analyses.

The results for national universities in Panel A of Table 3 also reveal that the negative relation between current and first- and second-lagged changes is quite strong in both statistical and commonsense terms. In statistical terms, both coefficients are significant at the 0.001 level. Perhaps more importantly, the sheer magnitude of the coefficients seems quite large at –0.30 and –0.23. Taken literally and somewhat oversimplifying the matter, these coefficients imply that about 30 percent of a given change in the rankings reverses in the very next ranking, and another 23 percent unravels in the 2-year ahead ranking. In other words, the regression results suggest that there is a strong and quick reversibility in the rankings changes for national universities. The results in Panel B are qualitatively similar to those in Panel A. The coefficients on both first- and second-lagged changes are negative and highly significant, with magnitudes suggesting a pronounced reversibility in rankings changes.[1]

The analysis and trimmings in Table 2 indicate that the results in Table 3 are unlikely to be due to outliers or deviations from normality. Nevertheless, I also present some simple probabilistic evidence about the relation between current and past changes that does not rely on any distributional assumptions. Given the evidence of Table 3, I investigate the probability of obtaining a positive or negative current rankings change, conditional on the signs of first- and second-lag changes. More specifically, to test for a negative relation between current and first-lag changes, I calculate the probability of a switch in the sign of a current rankings change, conditional on the sign of the most recent rankings change. Excluding observations with zero changes, if the sign of current changes does not depend on the sign of past changes, this probability should be about 0.50. However, the empirical probability of a switch in the rankings change is 0.57 for national universities and 0.66 for liberal arts colleges. I then calculate the probability of a switch in the sign of the current rankings change, given that the preceding two rankings changes have the same sign. This probability is 0.77 for national universities and 0.75 for liberal arts colleges.

Summarizing, both regressions and simple nondistributional evidence indicate a strong negative relation between current rankings changes and first- and second-lag changes. This evidence suggests that rankings changes include a lot of noise: They have a large transitory component, which quickly and predictably reverses in the next two rankings.

ESTIMATING THE NOISE IN RANKINGS CHANGES

The results in Table 3 imply that rankings changes have both a permanent and a transitory component. The permanent component captures unexpected and permanent changes in the rankings while the transitory component is largely meaningless noise that quickly reverts in the following two rankings. In turn, these findings raise an intriguing issue. It seems worthwhile to obtain a quantitative estimate of the relative importance of the permanent and transitory component in the rankings changes, which could serve as a gauge of the "fundamentals" vs. "noise" proportion in the those changes.

Indeed, the results in Table 3 already suggest that noise accounts for a large component of the rankings changes. However, it is difficult to derive precise estimates of the relative importance of noise from the results in Table 3 because of two considerations. First, the coefficients of the multiple regressions reflect the incremental explanatory effect of each lagged variable on current changes, while we are more interested in the combined explanatory effect of the lagged variables. The combined effect is likely to be different from merely summing the incremental effects because the two lagged variables are negatively correlated as well. Second, the standard measure of combined explanatory power, adjusted R^2, is also not quite appropriate. The reason is that the R^2 in the regressions in Table 3 provides an estimate of variation in current changes that is due to the reversal of *past* realizations of noise. However, R^2 does not capture the variation in current changes that is due to the *current* realizations of noise. These current realizations of noise will also predictably reverse in the future, and they need to be accounted for in deriving a quantitative estimate of the proportion of permanent and transitory changes in the rankings.

I derive quantitative estimates of this proportion by using a simple model of reversion in rankings changes and the empirical estimates of variances and covariances in the data. The model of rankings changes is defined as follows:

$$RC_t = p_t + n_t - k{*}n_{t-1} - (1-k){*}n_{t-2} \quad (1)$$

$$RC_{t-1} = p_{t-1} + n_{t-1} - k{*}n_{t-2} - (1-k){*}n_{t-3} \quad (2)$$

$$RC_{t-2} = p_{t-2} + n_{t-2} - k{*}n_{t-3} - (1-k){*}n_{t-4} \quad (3)$$

where:

RC_t denotes ranking change for year t

p_t denotes the permanent component of the change in the rankings for year t

n_t denotes the current realization of the noise component for year t

k is a coefficient that captures what part of the current realization of the noise component reverses in the next ranking. The entire current realization of noise is assumed to revert within the next 2 years, with a k proportion occurring in the very next ranking, and a $(1-k)$ proportion occurring in the 2-years ahead ranking.

Equations 1 to 3 simply state that current changes in the rankings are driven by contemporaneous changes in the permanent component of the rankings, contemporaneous changes in the noise component of the rankings, and a reversal of past realizations of the noise component of the rankings. Additionally, I make the following assumptions:

1. The permanent changes p_t are by definition unexpected, so current realizations of p are uncorrelated with past realizations of p or with current or past realization of noise.
2. Current realizations of noise are uncorrelated with past realizations of noise.
3. The realizations RC_t, p_t, and n_t over time are drawn from the same underlying distributions with constant variances Var(RC), Var(p), and Var(n).

The variables of interest in this model are p, n, and k. To be more precise, we are interested in the relative magnitude of variation of p and n because together with k they provide an estimate of the relative proportion of permanent changes to noise in the rankings changes. However, p, n, and k are unobservable and have to be estimated. There are different ways to estimate these variables, but the intuition about their estimation is straightforward. The relations between the unobservable variables on the right-hand side of equations 1 to 3 determine the relations between the observable left-hand side variables. Thus, one can use the relations between the observable left-hand side variables to derive estimates of the relative importance of the unobservable right-hand side variables.

Perhaps the easiest way to estimate the relative importance of the unobservable variables is by using the observable variance-covariance matrix of current and first- and second-lagged rankings changes. More specifically, one can use equations 1 to 3 to derive a system of equations for the variance of current rankings changes and its covariances with first- and second-lagged rankings changes and solve it for k, Var(n), and Var(p):

$$\mathrm{Var}(RC) = \mathrm{Var}(p) + 2*(1 - k + k^2)*\mathrm{Var}(n) \quad (4)$$

$$\mathrm{Cov}(RC_t, RC_{t-1}) = -k^2*\mathrm{Var}(n) \quad (5)$$

$$\mathrm{Cov}(RC_t, RC_{t-2}) = -(1 - k)*\mathrm{Var}(n) \quad (6)$$

This system of equations is well identified and has a unique (positive values) solution for given Var(RC), Cov(RC_t, RC_{t-1}), and Cov(RC_t, RC_{t-2}). The empirical estimates of Var(RC), Cov(RC_t, RC_{t-1}), and Cov(RC_t, RC_{t-2}) are calculated from the *USN* samples and are given in Table 4. Panel A presents the variance-covariance matrix for the rankings changes for national universities, and Panel B presents the results for the liberal arts colleges. Using the estimates for national universities, one obtains the following solution for k^U, Var$(n)^U$, and Var$(p)^U$ (where superscript U denotes university):

$$k^U = 0.65$$

$$\mathrm{Var}(n)^U = 2.68$$

$$\mathrm{Var}(p)^U = 1.18$$

TABLE 4. Variance—Covariance Matrices for the Relation Between Current and Lagged Changes in the *USN* Top 25 Rankings of National Universities and Liberal Arts Colleges

	CurCh	Lag1Ch	Lag2Ch
Panel A: National Universities			
CurCh	5.32		
	(212)		
Lag1Ch	–1.13	5.48	
	(185)	(190)	
Lag2Ch	–0.94	–1.36	6.04
	(163)	(167)	(172)
Panel B: National Liberal Arts Colleges			
CurCh	7.13		
	(212)		
Lag1Ch	–1.86	6.68	
	(181)	(187)	
Lag2Ch	–0.61	–6.76	6.93
	(157)	(161)	(167)

CurCh denotes current changes, Lag1Ch denotes 1-year lagged changes, and Lag2Ch denotes 2-year lagged changes. The numbers without parentheses are variances and covariances for the corresponding pair of variables. The numbers in parentheses are the available number of observations.

The estimate for k in this solution suggests that about 65 percent of current realizations of noise revert in the next ranking, and the rest unravels in the 2-years ahead ranking. It also suggests that the variation of the noise component is considerably larger than the variation of the permanent component of the rankings changes. Perhaps most importantly, this solution allows one to obtain a quantitative estimate of how much of the variation in rankings changes is due to permanent changes and how much to transitory and reversible noise. As a convenient summary statistic, I define a rankings change noise ratio as:[2]

$$\begin{aligned}\text{Noise ratio} &= (\text{variance in changes due to noise})/(\text{total variance in changes})\\ &= \{2*(1 - k + k^2)*\mathrm{Var}(n)\}/\mathrm{Var}(RC)\end{aligned}$$

Using the estimates of k^U, $\mathrm{Var}(n)^U$, and $\mathrm{Var}(p)^U$ yields a noise ratio of 0.78 for national universities. In other words, the data suggest that about 78 percent of the variation in rankings changes is due to transitory and reversible effects rather than to permanent innovations in the rankings.

Panel B in Table 4 presents the variance-covariance matrix for current and lagged changes for liberal arts colleges. Using the variance and covariance estimates allows one to solve for the unobservable variables k^{LAC}, $\text{Var}(n)^{LAC}$, and $\text{Var}(p)^{LAC}$ (where superscript *LAC* denotes liberal arts colleges):

$$k^{U} = 0.79$$

$$\text{Var}(n)^{U} = 2.95$$

$$\text{Var}(p)^{U} = 2.19$$

Further computations yield a noise ratio for liberal arts colleges of 0.69. Thus, the estimate of noise in the rankings changes of liberal arts colleges is on the same magnitude as that for national universities.

POSSIBLE EXPLANATIONS AND IMPLICATIONS OF THE NOISE IN RANKINGS CHANGES

The preceding sections provide strong evidence that *USN* rankings changes are largely transitory and have a strong tendency to revert over the next two rankings. This finding raises the natural question of what might be the possible causes for this phenomenon. To address this question, the present study identifies and investigates further three possible explanations. First, the reversibilities in changes might be due to reversibilities in the rankings methodology. Second, the reversibilities in rankings changes might be due to reversibilities in the real performance of schools. Third, the reversibilities might be due to the fact that *USN* rankings are essentially mechanical aggregations of noisy and reversible components.

Reversibilities in the Rankings Methodology

The reversibilities in the rankings methodology explanation seem intuitive and plausible. Holding the rankings information inputs constant, reversibilities in the rankings methodology will produce reversibilities in rankings changes. To illustrate, assume that *USN* has a fairly stable core model of how to produce university rankings. However, suppose that *USN* constantly tweaks the core model to respond to various fads. Bending the core model to accommodate the current fad produces some changes in the rankings. After the fad subsides, the return to the core model will produce predictable reversions of these changes. For example, a lunge toward rewarding "research excellence" followed by a return to "focus on teaching," will produce predictable reversibilities in the rankings. If such revisions in methodology are sufficiently prevalent and large, they can account for the documented reversibilities in rankings.

To evaluate the relative importance of this explanation, I examine the available methodology information. Each year, the *USN* rankings issue includes a fairly elaborate description of the rankings procedures—including the components and the subcomponents of the rankings, their relative weights, and a description of the way the components are combined in the rankings. The *USN* Web site and communications with the editors also helped to confirm or clarify some issues. As an illustration of how *USN* describes its rankings methodology, the appendix provides a snapshot from *USN*'s Web site. As the appendix demonstrates, the explanations are fairly straightforward and clear, and there is a good level of available detail. Overall, it seems that the available material provides a good understanding of the evolution of the *USN* rankings methodology over time.

The main conclusion from the available information is that the changes in methodology between 1988 and 1998 seem to be relatively incremental and gradual, more of a fine tuning than drastic changes. For example, the component "Financial resources" accounted for 20 percent of the rankings in 1991, 18 percent in 1992, and 15 percent in 1993. "Faculty resources" accounted for 25 percent in 1991 and 1992, and 20 percent in 1993. In 1993 the "Alumni satisfaction" component entered the rankings with a weight of 5 percent. Thus, the annual changes in the rankings methodology likely produce relatively minor changes in the rankings. More importantly, the examination of rankings methodologies over time revealed no pattern of repeated and predictable reversals. Thus, it seems unlikely that reversibilities in the rankings methodology could account for the reversibilities in rankings changes.

In fact, the examination of the methodology data leaves the impression that *ex ante* the changes in methodology appear largely unpredictable. Thus, changes in the rankings methodology produce unpredictable changes in the rankings. In other words, methodology-induced changes in the rankings account for at least some of the nonreversible changes in the rankings. This observation is interesting because it has two important implications.

First, the evidence in the preceding section reveals that about 70 to 80 percent of the variation in annual rankings changes is due to transitory and reversible noise. The remaining 20 to 30 percent variation is due to nonreversible changes. Also note that the nonreversible changes in the rankings can be due to two reasons: nonreversible changes in the rankings methodology and real changes in fundamental school quality. Thus, methodology-induced changes can account for, at most, 20 to 30 percent of the variation in rankings changes.[3] The comparative magnitude of these two sources of variation is intriguing because critics of the *USN* rankings have often brought up the problem of methodology-induced changes as a major, or the major, problem with the rankings (e.g., Webster, 1992; Gottlieb, 1999; Karl, 1999). Critics have rightly pointed out that tweaking the methodology produces meaningless changes in the rankings, creating excess volatility and obvious problems in the interpretation of the rankings. However, this study essentially finds that reversibility in rankings changes is a more subtle but far more dangerous source of meaningless changes in the rankings than changes in methodology. Thus, the public debate on the merits of the rankings can be enriched by a consideration of the importance of the reversibility in rankings changes.

A second and related implication is that the 20 to 30 percent estimate is also the upper limit on the proportion of changes in the rankings that can be due to real changes in fundamental school quality. Assuming a more realistic halfway split of the 20 to 30 percent estimate suggests that perhaps only as little as 10 percent of the variation in rankings changes is due to changes in fundamental quality. Admittedly, this 10 percent estimate is imprecise. However, based on a combined consideration of the preceding evidence, it seems that changes in fundamental school quality account for only a discomfortingly low proportion of the changes in published school rankings.

Reversibilities in Real Performance

The second explanation is that reversibilities in rankings changes could be a natural reflection of reversibilities in real performance. For example, such reversibilities in real performance could be due to fight-back and complacency effects where downgraded schools work extra hard to regain their lost positions, and upgraded schools become complacent and tend to lose their gains. However, the fight back and complacency explanation seems somewhat strained *ex ante* when one considers the speed of the reversibility in published rankings changes. Considering the possible speed of school reactions, one would expect that fight back and complacency effects will be felt only in medium- to long-term time horizons, possibly 3 to 5 to even 10 years. This expectation seems inconsistent with the preceding evidence that most of the reversion occurs in the very next ranking, and the reversions are pretty much complete by the second year ahead. It just seems improbable that the average university can identify a turnaround plan, implement it, and the effects are felt and are recorded within less than a year or two—in time for the new ranking.[4]

Table 5 provides some additional evidence on the importance of reversibilities in real performance as an explanation for reversibilities in published rankings changes by examining the properties of annual versus long-term changes. Panel A starts with a simple comparison of the descriptive statistics for annual versus 10-year changes for national universities. Ten-year changes are the longest horizon changes available for the sample period of 1988 to 1998. To preserve comparability between observations available for annual and 10-year changes, Panel A presents evidence only for the subset of schools that have rankings for both 1988 and 1998. However, these data restrictions turn out to be surprisingly mild. The number of available observations for 10-year changes in Panel A is 25, which means that all schools that had a Top 25 ranking in 1988 were still ranked

TABLE 5. A Comparison of Short-Term (annual) and Long-Term (10-year) Changes in the Rankings

N	Mean	*SD*	Minimum	Maximum
Panel A: National Universities				
Descriptive statistics for annual changes				
243	–0.05	3.03	–12	8
Descriptive statistics for 10-year (1988 to 1998) changes[2]				
25	–0.52	4.42	–9	9
Panel B: National Liberal Arts Colleges				
Descriptive statistics for annual changes[3]				
236	–0.12	3.60	–13	18
Descriptive statistics for 10-year (1988 to 1998) changes[4]				
25	–0.52	5.08	–12	13

[1]Correlation between current and first-lag changes (*p*-value): –0.38 (0.001).
[2]Correlation between 1988–1992 and 1993–1998 long-term changes (*p*-value): –0.14 (0.393).
[3]Correlation between current and first-lag changes (*p*-value): –0.46 (0.001).
[4]Correlation between 1988-1992 and 1993-1998 long-run changes (*p*-value): –0.05 (0.689).

in the Top 25 in 1998! The only caveat for this remarkable fact is that *USN* allowed ties in 1998 but not in 1988. Thus, there were exactly 25 schools in the Top 25 in 1988, while a four-way split of rank 25 resulted in 28 schools being ranked in the Top 25 in 1998. In any case, this statistic is another dramatic illustration that membership in the Top 25 universities is quite stable over time.

An examination of the descriptive statistics in Panel A reveals two other interesting findings. First, the standard deviation of 10-year changes (4.42) is only marginally larger than the standard deviation of annual changes (3.03). Second, the absolute values of the minimum and the maximum 10-year changes (−12 and 8) are quite similar to those for extreme annual changes (–9 and 9). Intuitively, one would expect that the standard deviation and the extremes in 10-year changes to be much larger than those for annual changes. In fact, if rankings revisions are only triggered by new and random information, one would expect that the standard deviation of long-term changes to be increasing linearly in time. Thus, if there was no predictability of rankings changes, the standard deviation of 10-year changes should be 10 times the standard deviation of annual changes. Instead, Panel A finds that there is only a 40 percent increase in standard deviation from annual to 10-year changes. In addition, even this 40 percent increase is probably largely accounted for by the fact that, as discussed earlier, the long-term changes in rankings methodology are almost by definition somewhat random and larger than the annual changes in rankings methodology.

In other words, the descriptive evidence in Panel A of Table 5 implies that controlling for changes in the rankings methodology, the standard deviation in 10-year changes is likely no larger than the standard deviation of annual changes. Recall that real changes in school fundamentals are likely unexpected, and thus the standard deviation of real changes should be increasing linearly in time. Therefore, this evidence implies that about 90 percent of annual changes is transitory noise that does not affect long-term rankings, and only perhaps 10 percent or so is due to real changes in the underlying fundamentals. Thus, the descriptive evidence in Panel A of Table 5 is consistent with earlier evidence that only a small portion of the variation in rankings changes is due to changes in real performance.

Panel B portrays pretty much the same picture for the rankings of national liberal arts colleges. Just like with national universities, all Top 25 schools in 1988 are in the Top 25 in 1998 as well (partly because of a six-way split of rank 24 in 1998), demonstrating that there is remarkably little long-term mobility at top schools. The standard deviation of 10-year changes (5.08) is only marginally larger than the standard deviation of annual changes (3.60), and the absolute values of the 10-year extreme changes (−12 and 13) are actually *lower* than those for annual changes (−13 and 18). Thus, a comparison of the descriptive statistics in Panel B also leaves the impression that annual changes are mostly noise that matters little for long-term changes in the rankings. It seems remarkable that the average Top 25 school faces a 10-year variation in its rankings that is not that different from what it has to endure in annual rankings changes.

Table 5 also offers more specific evidence about the relative importance of fight-back effects in real performance by comparing the first-order autocorrelations in annual and 5-year rankings changes. The 5-year rankings changes are computed over years 1988 to 1992 and 1993 to 1998 to allow for the longest possible horizons for computing a relation between current and past changes. As one would expect from the preceding evidence, the benchmark first-order autocorrelation in annual changes in Panel A is negative and large (–0.38), and highly significant ($p < 0.001$). If medium-to-long-term fight-back and complacency effects are to account for the high proportion of noise in annual rankings, then one would also expect a significant negative correlation between 5-year changes in the rankings. Indeed, the first-order autocorrelation in 5-year changes is negative (–0.14), but it is not nearly significant ($p = 0.393$).[5] The difference is even more clear-cut for the liberal arts colleges in Panel B. The first-order autocorrelation in annual changes is –0.46 ($p < 0.001$), while the autocorrelation in 5-year changes is –0.05 ($p = 0.689$).

Summarizing, the evidence in Table 5 confirms earlier findings that annual rankings changes have a large transitory component that reverses rather quickly. In addition, there is no reliable relation between long-term changes in the rankings. Thus, it seems unlikely that longer-range fight-back effects in real performance can account for the reversibilities in annual rankings changes.

Reversibilities in the Rankings Components

A third possible explanation for the reversibilities in rankings is that the rankings contain noisy information, and the predictable reversibilities in this noise lead to the predictable reversibilities in the rankings. A closer examination of the rankings methodology seems to be consistent with this explanation. *USN* rankings are essentially simple mechanical aggregations of several weighted components (e.g., Academic reputation, Faculty resources, Financial resources, etc.). Over the years, the components and the weights have evolved, but otherwise the method of deriving the rankings remains basically the same. Since the information in the components likely has some measurement errors or other sources of noise, the simple aggregation of components implies that this noise will be carried over into the rankings as well.[6] By its nature, such noise will tend to predictably reverse in the future, which implies that rankings changes will be reversible as well.

To explore this explanation further, Table 6 offers some more specific evidence about the properties of the rankings components. Panel A presents descriptive statistics for the untrimmed distributions of changes for the four most important and most stable components of the rankings for national universities.[7] These four components are available for most or all years, where Academic reputation, Faculty resources, and Financial resources typically account for 20 to 25 percent of the rankings each, and Student selectivity accounts for another 5 percent.[8]

The main message of Panel A is that the distributions of changes for Academic reputation, Financial resources, and Student selectivitiy are fairly similar with comparable means, standard deviations (2.30 to 3.21), and extremes. However, the Faculty resources component has a much greater variation in changes than any of the other components, with a standard deviation of 13.55, and much greater extremes. This evidence implies that the variation in changes for Faculty resources dominates the variation in changes from the other components and is likely the chief driver of the variation in changes for the overall rankings.

Panel B in Table 6 investigates whether rankings components are reversible in changes. Recall that earlier evidence suggests that the reversibilities in rankings are unlikely to be due to reversibilities in the rankings methodology. Thus, one would expect that there are likely reversibilities in the rankings components, which could account for the reversibilities in rankings. Panel B investigates this conjecture by regressing current changes in the rankings components on first- and second-lagged changes. Since the extremes in Panel A of Table 6 suggest that OLS estimation might be unduly influenced by outliers, for each regression I eliminate all observations that are more than three standard deviations away from the mean (the tenor of the results is the same for the untrimmed samples).

Overall, the evidence in Panel B demonstrates a strong negative relation be-

TABLE 6. Properties of the Changes in Selected Components of the *USN* Rankings of National Universities

Changes in Component	*N*	Mean	*SD*	Min	Max
Panel A: Descriptive Statistics for Changes in Selected Rankings Components					
Academic reputation	169	–0.49	2.30	–8	5
Faculty resources	221	–0.42	13.47	–82	60
Financial resources	199	0.13	2.84	–8	11
Student selectivity	222	–0.34	3.10	–17	10

Component	Intercept	Lag1Ch	Lag2Ch	Adj. R^2	*N*
Panel B: Regression Results for the Relation Between Current and Lagged Changes in Selected Rankings Components (Regression: CurCh = $b_0 + b_1$*Lag1Ch + b_2 *Lag2Ch + ε)					
Academic reputation	–0.62	–0.49	–0.21	0.18	113
	(–3.13)	(–5.24)	(–2.20)		
Faculty resources	–0.18	–0.43	–0.21	0.17	149
	(–0.29)	(–5.67)	(–3.33)		
Financial resources	0.26	–0.12	0.07	0.01	136
	(1.49)	(–1.36)	(0.89)		
Student selectivity	–0.36	–0.12	–0.09	0.01	157
	(–1.89)	(–1.60)	(–1.38)		

CurCh denotes current changes, Lag1Ch denotes 1-year lagged changes, and Lag2Ch denotes 2-year lagged changes. The numbers in the table for each corresponding variable are coefficient in the regression and *t* statistic (in parentheses).

Component	Academic Reputation	Faculty Resources	Financial Resources	Student Selectivity
Panel C: Correlations Between Concurrent Changes in Selected Rankings Components				
Academic reputation	1.0			
Faculty resources	–0.02	1.0		
	(0.82)			
Financial resources	–0.04	0.25	1.0	
	(0.62)	(0.001)		
Student selectivity	0.13	0.08	0.00	1.0
	(0.09)	(0.23)	(0.99)	

Each cell in the table contains the Pearson correlation between the two corresponding variables, and a P-value (in parentheses).

tween current and lagged changes for Academic reputation and Faculty resources and a weak negative to no reliable relation for Financial resources and Student selectivity. These results indicate that at least some components have a strong reversibility in changes, which could potentially account for the reversibility in overall rankings changes. It also seems telling that Faculty resources, which is by far the most important determinant of the overall rankings, also manifests the most pronounced reversibility in changes.

Panel C in Table 6 presents the correlations between concurrent changes in these four components of the rankings. The premise of Panel C is that if these components capture some common underlying fundamental like school quality or real performance, then there should be at least a moderate positive relation between concurrent changes in these components. However, all correlations in Panel C except one are insignificant. The only sizable and significant correlation is between Faculty resources and Financial resources, but this relation is also somewhat mechanical because of a common denominator (both Faculty resources and Financial resources are expressed per student). Thus, the variation in rankings components changes seems almost entirely idiosyncratic, rather than systematic, and possibly largely due to idiosyncratic noise. Of course, one might argue that a very strong positive relation between components is also undesirable because it will imply that the components are redundant to each other as sources of information. However, it still seems telling that there is virtually no relation between contemporaneous changes in the components of the rankings.

Table 7 presents additional and more specific evidence about the relation between reversibilities in changes in rankings components and overall rankings. The table illustrates specific cases where large reversions in changes for one component, Faculty resources, result in matching reversions for the overall rankings. The Faculty resources component is a natural choice for this illustration because it accounts for 25 percent of the overall rankings, has by far the largest variation in changes, and exhibits a strong reversibility in changes. In view of the preceding evidence, the tenor of the results in Table 7 is not entirely surprising. However, the consideration of specific cases brings the preceding statistical evidence to life and helps to further discriminate between possible alternative explanations about the reversibility in rankings changes.

The time-series of rankings of Johns Hopkins University is an interesting extreme case that vividly illustrates the major themes in this study. The shaded areas in Table 7 reveal two prominent dip-and-recovery periods in Faculty resources, 1989–1991 and 1993–1995, where the changes in the overall rankings clearly mirror the changes in the Faculty resources rankings. However, the really fascinating part of this evidence is the sheer magnitude of the changes. Taken literally, the rankings suggest that Johns Hopkins was 9 in Faculty resources in 1989, 30 in 1990, and 9 again in 1991, followed by a 37-97-15 sequence during 1993–1995. It is difficult to understand how a proper measure of the faculty

TABLE 7. A Comparison of the Faculty Resources Rankings and the Overall Rankings for Selected Universities

University		1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Johns Hopkins	Faculty Resources	9	30	9	24	37	97	15	19	14	17
University	Overall Ranking	14	15	11	15	15	22	10	15	14	14
Harvard University	Faculty Resources	11	2	1	4	3	3	1	11	2	2
	Overall Ranking	3	1	1	1	1	1	1	3	1	1
University of	Faculty Resources	18	21	18	20	50	13	8	14	5	6
Pennsylvania	Overall Ranking	20	13	13	14	16	12	11	13	7	6
Emory University	Faculty Resources	—	—	—	17	29	21	16	20	8	12
	Overall Ranking	—	—	—	21	25	16	17	19	9	16
Washington	Faculty Resources	6	46	15	29	23	30	30	9	8	9
University	Overall Ranking	22	24	18	20	18	20	20	17	17	16

This table illustrates the relation between the Faculty resources component of the rankings and the overall rankings for selected national universities. The shaded areas highlight particular instances where large and immediate reversibilities in the Faculty resources component clearly translate into matching reversibilities in the overall rankings.

resources of Johns Hopkins could ever manifest such a change. In fact, the magnitude and the pattern of these changes simply defy common sense about how true faculty resources might evolve for one of the most elite U.S. universities. The magnitude of these changes is such that it raises the possibility of measurement or classification irregularities in the source data, or even outright data-handling and processing errors. Additionally, note that these roller coasters in the Faculty resources component also cause material swings in the overall rankings. For example, the "miraculous" recovery in Faculty resources from 97 in 1994 to 15 in 1995 probably accounts for most or all of the 22-to-10 improvement in overall rankings. Such a swing in overall rankings moves Johns Hopkins from the bottom of the Top 25 to Top 10 and probably has very real effects on applications and enrollment decisions. At the same time, this swing appears to be mostly noise: Over the entire 10-year period Johns Hopkins has never been ranked above 10 or below 22. It is either that 1995 was truly exceptional in restoring Johns Hopkins' faculty resources or more likely that the *USN* component "Faculty resources" was a rather poor reflection of the underlying true faculty resources.

The pattern of changes for Harvard University is another interesting case. Throughout the 1990s, Harvard has always been ranked number 1 except for a number 3 ranking in 1996. Thus, the pattern of overall changes does not reveal any conspicuous irregularities. However, note that the dip to number 3 in 1996 is most probably related to the dip-and-recovery pattern in Faculty resources (1-11-2 over 1995–1997). Again, it is difficult to imagine that this pattern could occur in terms of Harvard's true faculty resources. The more likely explanations that come to mind are again classification, measurement, and various data-processing irregularities. The rest of the specific cases for Table 7 largely repeat the same theme and are presented for two reasons. First, they demonstrate that cases like Johns Hopkins and Harvard are not that uncommon. Second, they reveal that such pronounced reversals occur over all years and are not the result of some year-specific regime-shifting effects (like an unidentified change in methodology).

Summarizing the evidence about possible explanations, the reversibilities in *USN* rankings are unlikely to be due to reversals in methodology or real performance. However, the reversibilities in rankings seem to be at least partly due to reversibilities in the underlying rankings components. The reversibilities in rankings components are most likely due to classification, measurement, and various types of data-handling errors, which by nature quickly reverse in the future. On one hand, such errors are partly unavoidable. On the other hand, it is disturbing that the variation in rankings due to such errors exceeds greatly the variation due to real performance. The bottom line seems to be that *USN* rankings changes are highly inaccurate measures of changes in school quality, with all its related implications.

Potential Implications for Preparers and Users of the Rankings

This study identifies strong reversibilities in *USN* rankings changes and offers some possible explanations for these reversibilities. Assuming that *USN* and its audience view these reversibilities as undesirable, *USN* could use its own proprietary information to investigate further and potentially take corrective measures. An important consideration here is that *USN* can take successful corrective measures even with only an approximate or incomplete knowledge of the underlying causes for the reversibility in changes. Given that the nature of the phenomenon is known (negative autocorrelation in changes, up to second lag), *USN* could use various standard statistical smoothing techniques or Bayesian updating to correct the rankings before releasing them.

However, *USN* might be reluctant to implement such changes because the elimination of noise will leave precious little "news" in the rankings. In other words, ranking most top schools each and every year about the same might indeed be a better reflection of their true relative academic quality. However, that leaves little surprise to report in the annual rankings, and the absence of surprise questions and threatens the importance of the annual editions of the rankings. In effect, the elimination of noise could actually reduce the popularity, the impact, and the commercial success of the rankings. Thus, it will be interesting to observe whether *USN* implements some noise-elimination changes or simply adheres to the status quo where most of news in the rankings is actually reversible noise.

Note that ultimately it may not matter that much whether *USN* modifies its methodology. Rankings users can also make explicit or implicit adjustments that will remove the reversibilities in changes. Effectively, users need to greatly discount current changes in *USN* rankings and view annual changes as largely transitory noise. As a rule of thumb, long-term averages in the rankings or long-term changes in the rankings seem to be a much better indication of the underlying real levels and changes in fundamental school quality.

CONCLUSION

Changes in the *USN*'s annual college rankings have a strong tendency to revert in the following two rankings. Using a simple model, this study estimates that about 70 to 80 percent of the variation in rankings changes is due to transitory and reversible noise. In other words, most of the news in the annual rankings is essentially meaningless noise. After accounting for rankings changes due to changes in methodology, it seems that perhaps as little as 10 percent of the variation in annual rankings changes is due to real changes in school fundamentals. Thus, changes in *USN*'s annual rankings seem to be a highly inaccurate measure of real changes in fundamental school quality.

Acknowledgments. Richard Bagozzi, Eric Dey, Jim Hines, Joel Slemrod, and an anonymous referee have provided valuable comments and suggestions.

APPENDIX. Undergraduate Ranking Criteria and Weights

U.S. News bases its college rankings on up to 16 different measures of academic quality that fall into seven broad categories: academic reputation, student selectivity, faculty resources, student retention, financial resources, alumni giving, and, for national universities and national liberal arts colleges only, graduation rate performance. The following table lists each quality indicator, its component subfactors, and the weight or relative importance assigned to each category and subfactor. A brief definition of each indicator may be found at the bottom of the table. For a more detailed explanation of the rankings indicators and methods, please see How *U.S. News* Ranks Colleges.

Rankings Category	National Universities and Liberal Arts Colleges-Category Weight	Regional Universities and Regional Liberal Arts Colleges-Category Weight	Subfactor	National Universities and National Liberal Arts Colleges-Subfactor	Regional Universities and Regional Liberal Arts Colleges-Subfactor
Academic reputation	25%	25%	Academic reputation survey	100%	100%
			Acceptance rate	15%	15%
			Yield	10%	10%
Student selectivity Fall '97	15%	15%	High school class standing—top 10%	35%	0%
			High school class standing—top 25%	0%	35%
			SAT/ACT scores	40%	40%
Faculty resources '97	20%	20%	Faculty compensation	35%	35%
			Faculty with Ph.D.	15%	15%
			Full-time faculty	5%	5%
			Student/ faculty ratio	5%	5%
			Class size, 1–19 students	30%	30%
			Class Size, 50+ students	10%	10%
Retention rate	20%	25%	Graduation rate	80%	80%
			Freshman retention rate	20%	20%

APPENDIX (*Continued*)

Rankings Category	National Universities and Liberal Arts Colleges-Category Weight	Regional Universities and Regional Liberal Arts Colleges-Category Weight	Subfactor	National Universities and National Liberal Arts Colleges-Subfactor	Regional Universities and Regional Liberal Arts Colleges-Subfactor
Financial resources	10%	10%	Educational expenditures per student	100%	100%
Alumni giving	5%	5%	Alumni giving rate	100%	100%
Graduation rate performance	5%	0%	Graduation rate performance	100%	0%
Total	100%	100%	—	100%	100%

Definitions of ranking criteria:

Academic reputation. The average rating of the quality of a school's academic programs as evaluated by officials at similar institutions. The survey was conducted in the spring of 1998.

Acceptance rate. The ratio of the number of students admitted to the number of applicants for the fall 1997 admission.

Alumni giving. Percent of undergraduate alumni of record who donated money to the college or university, averaged over the 1996 and 1997 rates.

Class size, *1–19 students*. The percentage of undergraduate classes, excluding class subsections, with fewer than 20 students enrolled during the fall of 1997.

Class size, *50+ students*. The percentage of undergraduate classes, excluding class subsections, with 50 students or more enrolled during the fall of 1997.

Expenditures per student. Total educational expenditures per full-time-equivalent student.

Faculty compensation. Average faculty pay and benefits adjusted for regional differences in cost of living during the 1996 and 1997 academic years. Includes full-time assistant, associate, and full professors.

Faculty with Ph.D.'s. The proportion of full-time faculty members with a doctorate or the highest degree possible in their field or specialty during the 1997 academic year.

Freshman retention rate. Percent of first-year freshmen who returned to the same college or university the following fall, averaged over the classes entering between 1993 and 1996.

Full-time faculty. The proportion of total faculty employed on a full-time basis during the 1997 academic year.

Graduation rate. Percent of freshmen who graduated within a 6-year period, averaged over the classes entering between 1988 and 1991.

Graduation rate performance. The difference between the actual 6-year graduation rate for students entering in the fall of 1991 and the rate expected from entering test scores and education expenditures. Note: In past years, this indicator was referred to as "value added."

High school class standing. The proportion of students enrolled for the fall 1997 academic year who graduated in the top 10 percent (for national universities and liberal arts colleges) or 25 percent (for regional universities and liberal arts colleges) of their high school class.

SAT/ACT scores. Average test scores on the SAT or ACT of enrolled students, converted to percentile scores by using the distribution of all test takers.

Student/faculty ratio. The ratio of full-time-equivalent students to full-time-equivalent faculty members during the fall of 1997, excluding faculty and students of law, medical, and other stand-alone graduate or professional programs in which faculty teach virtually only graduate-level students. Faculty numbers also exclude graduate teaching assistants.

Yield. The ratio of students who enroll to those admitted to the fall 1997 freshman class.

NOTES

1. College rankings are ordinal numbers, which are bounded from above, that is, no ranking can be higher than 1. Thus, the negative relation between current and past changes might be due to a "ceiling effect," where especially at the top, ranking improvements tend to be followed by rankings decreases because there is little or no room for improvement while there is plenty of downside potential. I investigate the relative importance of this effect by recalculating the results after the deletion the top five universities for each year. The motivation is that the ceiling effect should be most pronounced for schools at the very top, and the negative relation between current and past changes should be markedly weaker if it was due to the ceiling effect. However, the results remain almost identical for this specification, which indicates that empirically the ceiling effect is not important.
2. Of course, 1 minus the noise ratio is equal to Var(p)/Var(RC), which is the permanent change ratio of the rankings.
3. Note that this upper limit is attained only under the rather extreme condition that there is no relation whatsoever between changes in published rankings and real changes in fundamental school quality.
4. Note that this argument assumes that schools fight back mostly by trying to improve real performance. However, as noted earlier, there is evidence that schools attempt to improve their rankings by manipulating the submission data. Such fight back data manipulations could show as immediate reversals of deteriorations in the rankings.
5. Autocorrelations in 5-year changes are corrected for the negative small sample bias (e.g., Campbell, Lo, and MacKinlay, 1997, p. 46). For first-order autocorrelations, if the computed autocorrelation is r, the bias-corrected estimate is given by $r + (1 - r^2)/(N - 1)$. The correction is negligible for annual changes.
6. For example, objective measures like SAT scores are likely an imperfect measure of student quality. Further, research in psychology indicates that people commonly use information-processing heuristics, which will manifest as noise in rankings survey responses.
7. The rankings components data are voluminous and fairly heterogeneous. Thus, for parsimony, I only present evidence about the rankings components for national universities, which is the more important ranking. Since all previous results are similar for both national universities and national liberal arts colleges, there is no reason to expect that the results for components will be different.
8. The components data has the following limitations. None of the components has available data for 1988. Faculty resources and Student selectivity have complete data for all remaining years, 1989–1998. Academic reputation has no data for 1997 and 1998 because for this component *USN* provided rankings before 1997 but switched to scores out of a maximum of 5 for 1997 and 1998. On examination of the data, I also found that there was a large and systematic shift in relative Financial resources rankings between 1988 and 1989, most probably due to a changing definition or some other methodological reason (similar to the one observed for overall rankings between 1987 and 1988). To illustrate the effect of the shift with respect to steady-state changes, the standard deviation of changes in Financial resources is about 11 for 1989–1998, and 2.78 for 1990–1998. Thus, to avoid the effects of regime shifts and outliers, I deleted 1989 for Financial resources.

REFERENCES

Campbell, J., Lo, A., and MacKinlay, A. (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Fombrun, C. J. (1996). *Reputation: Realizing Value from the Corporate Image*. Boston, Harvard Business School Press.

Gottlieb, Bruce (1999). Cooking the school books. *Slate* (a Web-based magazine), August 31.

Hunter, Bruce (1995). College guidebooks: background and development. In D. Walleri and M. K. Moss (eds.), *New Directions in Institutional Research: No. 88. Evaluating and Responding to College Guidebooks and Rankings*. San Francisco: Jossey-Bass.

Karl, Jonathan (1999). Do these ballyhooed college rankings mean anything at all? A tutorial. *The Wall Street Journal* (September 10): W13.

McDonough, Patricia M., Antonio, Anthony L., Walpole, MaryBeth, Perez, Leonor X. (1998). College rankings: Democratized college knowledge for whom? Forthcoming in *Review of Higher Education* 39.

Webster, David S. (1992). Rankings of undergraduate education in *U.S. News and World Report* and *Money*: Are they any good? *Change* (March/April): 19–31.

Received September 28, 1999.