Is individual job performance distributed according to a power law? A review of methods for comparing heavy-tailed distributions

Seth M. Spain[a]

State University of New York at Binghamton

Marcus Credé

George Washington University

P.D. Harms

University of Nebraska – Lincoln

Bradley J. Brummel

University of Tulsa

Abstract

It has recently been proposed that individual job performance follows a power law distribution (O'Boyle & Aguinis, 2012). We present an argument and evidence for why the conclusion does not follow from the premises. We discuss the nature of generating mechanisms of statistical distributions, and compare the normal, lognormal, and Pareto distributions. We review statistically principled methods of testing power-law distributions, and point out how it is necessary to compare them to a *plausible alternative* distribution (Clauset, Shalizi, & Newman, 2009). We reiterate the importance of explicitly examining the assumptions and consequences of statistical models, and review the methods that are available to organizational researchers when the *norm of normality* is violated.

Keywords: Individual job performance, Power laws, Probability distributions

---

[a] Address: School of Management, PO Box 6000, Binghamton, NY 13902-6000, Email: sspain@binghamton.edu

Electronic copy available at: http://ssrn.com/abstract=2238126

Is individual job performance distributed according to a power law? A review of methods for comparing heavy-tailed distributions

As difficulties associated with the complexity of organizational phenomena become more prevalent (e.g., Marion & Uhl-Bien, 2001; Lichtenstein & Plowman, 2009), there is increasing interest in the applications of analytical tools from statistical physics in organizational science (e.g., Andriani & McKelvey, 2009; O'Boyle & Aguinis, 2012), in particular the use of power laws to describe distributions (or, more pointedly, the *tails of distributions*; Newman, 2005; Clauset, Shalizi, & Newman, 2009). The recent paper by O'Boyle and Aguinis shows some of the promise of using these tools—and the potential dangers of using them when they are not fully applicable. O'Boyle and Aguinis raise one very important concern: the distributions of the variables we study in the organizational sciences are given little attention and often entirely ignored—normality is often assumed, and seldom checked.

O'Boyle and Aguinis examine the *norm of normality*—that normality is generally assumed in organizational sciences, contrast it with the Pareto power law distribution, and marshal considerable empirical evidence that individual job performance is *not* normally distributed, concluding that it is, instead, distributed according to a power law. It is clear to visual inspection that the performance variables included in their paper are not normally distributed; there is, however, little evidence that those variables are distributed according to a power law as suggested by O'Boyle and Aguinis—that is, the tests provided are significant, but not critical.

The current paper will attempt to address four issues: (1) provide organizational researchers with an understanding of power law distributions, and distinguish them from similar, less exotic but heavy-tailed distributions, (2) discuss the evidence for power law distributions in

Electronic copy available at: http://ssrn.com/abstract=2238126

individual performance data in O'Boyle and Aguinis (2012), (3) illustrate rigorous methods for testing the claim of fit to a power law distribution, and (4) discuss statistical methods for fitting models to data drawn from distributions more exotic than the commonly known normal, log-normal, binomial, and Poisson distributions.

The literature in statistical physics has recently gone through a period of exuberance for power laws, along with strong claims about generating mechanisms and universality, only to have more sober findings tone the original claims down (e.g., Barabási, 2005; Clauset, Shalizi, & Newman, 2009; Malmgren, Stouffer, Motter, & Amaral, 2008; Stumpf & Porter, 2012; Vázquez, Oliveira, Dezső, Goh, Kondor, & Barabási, 2006). When we say *generating mechanisms*, we mean theoretical models for the processes that produce an outcome that has a power law distribution (e.g., some forms of growth, preferential attachment, see below). By *universality* we mean a property that belongs to all members of a class of models, such that a highly simplified *toy model* can be studied and the results will apply to any member of that class, including much more complicated phenomena in the real world (as is the case with phase transitions in statistical physics). We hope to provide a cautionary tale to help organizational researchers learn from this history, while also providing generally available tools to examine their assumptions.

Why the concern with power law distributions?

The search for power laws has become *de rigeur* in several sciences, including physics (Albert & Barabási, 2002; Vázquez et al., 2006) and computer science (Mitzenmacher, 2004a, 2004b). For instance, consider the networks of scientific collaborations in mathematics and neuroscience examined by several statistical physicists (Barabási, Jeong, Néda, Ravasz, Schubert, & Vicsek, 2002, cf., Newman, 2001). The degree distribution of researchers was found to follow a power law distribution (that is the number of collaborators that an individual researcher is connected with via coauthorship), and the authors argued this was due to a mechanism of *preferential attachment*—researchers with more collaborators are more likely to gain additional collaborators in any given time-step; the rich get richer. In general, the papers claiming a power law distribution for some phenomenon in the physics literature provide a theoretical *generating mechanism*—they propose a theoretical model and rigorously derive the power law as a consequence, and then fit the data. One concern in the physics literature is that the data-fitting component has often been based on heuristic methods that are demonstrably unreliable (Clauset et al., 2009).

This search for power laws in some fields has been driven by concerns over "black swan" events (e.g., Talib, 2007)—rare but important events that may have extraordinary consequences, such as fluctuations in financial markets (Gabaix, Gopikrishnan, Plerou, & Stanley, 2003). If the true distribution is Gaussian (normal), then very, very large fluctuations should be rare in the extreme. If, on the other hand, the true distribution of these events follows a heavy-tailed distribution, such as a power law, handling them with normal-theory based statistical models will severely underestimate their prevalence and give biased estimates of model parameters. This is an important consideration as most of the *standard* toolkit of statistical inference is built on normal theory and may fail spectacularly when the (conditional[1]) distribution is wildly non-normal. It is worth noting, though, that entire branches of statistical theory are devoted to understanding these types of events.

Reconsidering the *norm of normality*

Why is the normal distribution considered so fundamental? O'Boyle and Aguinis do not address the functional generation of normal distributions—or Pareto distributions. As a result, we would like to take some time to consider how the distributions of our variables come about. When an observed measurement is an additive linear function of many “small” random variables, its distribution will generally conform to a normal, or Gaussian, curve (unless the contributing random variables themselves have pathological distributions). This is a consequence of the central limit theorem (Cramér, 1946, p. 213). This may be considered a self-evident result, since it is taught in any introductory statistics course; however, understanding what it means is critical.

Many important variables have Gaussian distributions—at least when treated properly. For instance, consider height. The distribution of the heights of all adults in the US is very non-Gaussian—it is bimodal and exceptionally wide, unless it is conditioned on biological sex: within-sex, the distribution of heights is very nearly perfectly Gaussian. The central limit theorem drags distributions towards normality when they are the sum of many small causes, but when there is a single large (possibly non-normal) cause—like biological sex, the central limit theorem no longer applies (e.g., Gelman & Hill, 2007, p. 14).

But what if the contributing variables do not sum up, but multiply instead? The resulting distribution is the *lognormal,* that distribution whose log-transformation is normal. As Gaussian distributions arise from the previously mentioned additive mechanisms, the lognormal arises when the random variable under consideration is the *product* of many “small” random variables:

$$\log(xy) = \log(x) + \log(y) \tag{1}$$

For the same reasons (i.e., the properties of the central limit theorem) that Gaussian distribution is common, so is the lognormal.

The distribution of job performance

How is individual performance generated? The consensus in industrial/organizational psychology is that performance is some function of ability and motivation, and well predicted by measures of general cognitive ability and conscientiousness, among other individual differences (e.g., Schmidt & Hunter, 1998). If this were true, it would imply that the generating mechanism for individual performance has the form of the following regression model (under the assumption that the function is linear and additive):

$$\text{Perf}_i = B_0 + B_1\text{*ability}_i + B_2\text{*conscientiousness}_i + B_3\text{*skills} + B_4\text{*job-knowledge} + e_i \qquad (2)$$

where the conditional mean of performance depends linearly on the values of ability and conscientiousness and its variance is a function of the variance of those quantities and the error term (which is independent and identically distributed $N(0,\sigma^2)$; cf., Berk, 2004). Of course, if none of the generating variables—cognitive ability, conscientiousness, and error—is too pathological (i.e., very non-normal), then Performance will also be normally distributed.

If performance is not normally distributed, our consensus model has not answered the question: what is the generating mechanism for performance (cf., Mitzenmacher, 2004b)? In their model of human dynamics, Vázquez and colleagues (2005) proposed a priority-queuing model. That is, individuals effectively have a to-do list (a *priority queue*) in their heads. For performance, this model would imply that each employee has a list of tasks in their minds and executes the first in list. The list updates each time a completed task is pushed off of it.

This model leads to a power-law distribution for inter-event times—typically, tasks are completed in rapid succession with some very long intervals in between some tasks. Think about it in this overly simplified way: a person goes to work, checks their email for things they need to do during the day, the person executes a handful of basic tasks quickly and easily, then spend an hour working on a presentation, then execute a few more easy tasks and take a lunch break, next he or she works on several things during the afternoon and then heads home for the night, and comes in to start the cycle again the next day. Some gaps between tasks are seconds to minutes, some an hour or so, and one long break between days. Analogically, human movements have a similar pattern, in that most of our daily movements are small, that is we tend to move around close to our homes and work, but we occasionally take trips to other parts of the world.

The priority-queuing model is simple, and conceptually fits with certain theories of job performance (e.g., Beal, Weiss, Barros, & MacDermid, 2005). The only problem is that it does not apply to individual performance. It implies that the temporal distribution of, say, academic papers or home runs would be bursty and heavy-tailed, but not what the distribution of individual performance looks like. Specifically, papers should come out in lumps, for instance, an individual academic might be able to get a lot more writing done during winter and summer breaks than during semesters, so the pattern of paper submissions would clump in these times and be comparatively rarer during teaching times. The queuing model implies this sort of a temporal distribution of performance, with the lags between paper submissions following a power law, but the distribution of individual job performance in the form of total number of papers following a power law does not follow without imposing additional assumptions (e.g., time aggregation).

We could perhaps apply some sort of aggregating function to this model to determine a distribution for individual performance. Such a model might involve individual differences in queue length, speed of task execution, plasticity in task priority, etc. But that is just one option. Other models may be more appropriate. For instance, a variety of growth models conform to either a power-law or a lognormal distribution (Mitzenmacher, 2004b). If we take a view of individual performance as a repertoire of learned behaviors that grows over time (e.g., Spain, 2010), such models may make sense. That is, if individual job performance is a behavioral repertoire that grows according to, say, a logistic map, then the *size of the individual repertoires* should have a log-normal distribution, akin to the distribution of sizes of animal populations in ecology. Analogously to cognitive abilities (e.g., Humphreys, 1992), larger repertoires should generally equate with better performance.

So, you think you've got a power-law?

One important property of power law distributions is that they are *scale invariant*. But, what does it means for a phenomenon to be "scale invariant"? That is, a phenomenon whose frequency scales as an inverse power of its extremity (e.g., Clauset, Young, & Gleditsch, 2007). This is the definition of a power law distribution:

$$\Pr(x) \propto x^{-\alpha} \tag{3}$$

where α is a constant called the *scaling exponent*. Scaling parameters typically obtain values between 2 and 3, but there may be exceptions (Vázquez et al., 2005). Additionally, it is considered rare for empirical phenomena to obey power laws throughout the range of x. It is more common for the power law to apply only for values greater than some minimum value, $x_{\min}$. The *tail* of the distribution is said to follow a power law (Clauset et al., 2005, p. 662).

Many quantities of interest to OBHR researchers have a characteristic *scale*—in the manner that physicists use the word. While psychometricians will usually refer to scale as the variance of the characteristic, a physicist is likely to mean a typical size that measurements cluster around, such as the mean (e.g., Newman, 2005). For instance, males in the US have an average height of about 180cm, and the ratio of tallest ever male to the shortest ever is about 5—272cm versus 57cm, respectively. The heights of males cover two orders of magnitude and have a small dynamic range (cf., Newman, 2005, pg. 262). In contrast, we can characterize an arithmetic average for US cities, but it tells us little about the size of the typical US city—such a distribution leaves us with the question, is there even such a thing as a typical size for a US city?

O'Boyle and Aguinis make the interesting point that if performance is *scale-free* the distribution should be the same across levels of analysis. Specifically, they suggest that individual job performance and firm performance should both follow power-law distributions. This statement may be plausible, but it does not reflect the meaning of scale invariance as used in describing power laws. Instead, let us consider a hypothetical power-law regime, that is, some distribution that is distributed according to a power-law in the tail above the critical x_{min}. Scale-invariance means that the same power-law describes the entire tail as describes any more extreme subset. That is, if we compare $p(X > x_{\mathrm{min}})$ to $p(X > x_{\mathrm{k}})$ where $x_{\mathrm{k}} > x_{\mathrm{min}}$, the same power-law distribution will hold for both cases.

The origins of power laws. Above we discussed the typical central limit theorem, wherein the normal distribution is the limiting distribution. Instead suppose drawing a sample of independent random variables from heavy-tailed distributions. Under a version of the central limit theorem (Willinger, Alderson, Doyle, & Li, 2004), the sum of these variables is generically a power law (Stumpf & Porter, 2012). That is, under this regime, the power law emerges as the limiting distribution, analogously to the normal distribution in the more standard central limit theorem. Because of this, power law distributions may arise in empirical data for non-specific reasons (i.e., not due to a particular generating mechanism; cf., Stumpf & Porter, 2012), just because the phenomenon is a result of the mixing of various heavy-tailed distributions, none of which themselves need to follow a power law.

Fitting the power law model: Naïve approach. Andriani and McKelvey (2009) point out a fact well-known to physicists, but probably not among organizational scientists: when extremity and frequency are both log-transformed and plotted against each other (a *log-log plot*), variables following a power law distribution will decay linearly, while Gaussian variables will show an exponential decay. Generations of physicists have followed Pareto's lead and concluded that a power law underlies their data with such plots as confirmation. *Many* distributions decay approximately linearly on a log-log plot, including the log-normal.

A power-law distribution appears as a straight, negatively sloped line on a log-log plot (see Figure 2). Therefore, one basic approach is to regress log-frequency on log-rank and assess the R^2 (Gilmore, 2006, p. 21). The slope of the regression line can serve as an estimate of the scaling exponent. One major problem with this approach is that, over any sufficiently short range of values, a straight line can approximate any smooth function. This problem is general, as it can take a very large range of data in order to fully distinguish a power-law from other heavy-tailed distributions, such as the lognormal or Weibull.

--Insert Figure 2 about here--

Fitting the power law model: Principled approach. At this point, we provide a review of the statistical procedures for assessing whether a dataset is power-law distributed given by Clauset et al. (2009)[2].

Step 1: Use maximum likelihood methods to estimate the scaling exponent α from Equation 3.

Step 2: Use goodness-of-fit measures to estimate the critical value x_{min} where the scaling regime begins—that is, where the power law distribution takes over from whichever distribution rules the bulk of the data. As noted above, it is usually the tails of a distribution that are expected to follow a power-law distribution, not the entire distribution. Clauset et al. provide statistical tools for using the Kolmogorov-Smirnov statistic to estimate the location of x_{min}. In general, this approach attempts to find the *best* scaling region for the power law fit. That is to say, if a power law were fitted to the *entire* distribution it would generally fit more poorly than if fitted to only the region identified by this step.

Step 3: Use goodness-of-fit tests to assess the overall fit of the power-law model. Clauset et al. (2009) provide guidelines on using the Kolmogorov-Smirnov statistic to assess the fit of the distribution to data.

Step 4: Compare the fit of the power-law to other right-skewed distributions, such as the lognormal and the Weibull. These will probably fit equally well or better (e.g., Clauset et al. 2009; Stouffer, Malmgren, & Amaral, 2005; cf., Malmgren, Stouffer, Motter, & Amaral, 2008). Clauset et al. (2009) suggest using Vuong's (1989) likelihood ratio test to do this.

Is performance power-law distributed?

O'Boyle and Aguinis (2012) present a massive amount of data in their attempt to demonstrate that the distribution of individual performance follows a power law. We do not believe that their evidence demonstrates that performance is power-law distributed. There are several reasons we believe this to be the case.

First, the data themselves may not represent performance itself, but the outcomes of performance (Campbell, 1990; cf., Beck, Beatty, & Sackett, in press). According to the Campbell model, performance is behavior. For instance, the acts of reviewing the literature, engaging in theoretical reasoning, collecting and analyzing data, writing, and submitting this manuscript are performance; whether it is accepted for presentation or publication is a measure of effectiveness—an evaluation of the outcome of the earlier performance behaviors. Publications, awards, electoral wins, and even points scored/home runs are effectiveness measures, not performance. Of course, effectiveness is, itself, of considerable practical interest for real organizations. Effective performance behaviors are naturally more valuable to the organization than ineffective performance behaviors. So, the data presented are still relevant, but they do not directly inform questions regarding individual job performance. Since at least one other mechanism (evaluation) is involved in transforming performance to effectiveness, it is possible for performance and effectiveness to have decidedly different distributions. That is, it is possible that performance could be normally distributed and effectiveness could have some more exotic distribution, if the function mapping performance to effectiveness is nonlinear.

The data presented by O'Boyle and Aguinis cover a wide spectrum of professions: published researchers, entertainers of various stripes, elected officials, and professional and college athletes. However, these are all fairly elite professions. Even the publications track only publications appearing in the top 5 journals in each field. This condition brings with it two concerns: one mostly substantive and one purely methodological. First, even if we accept that the measures considered are performance rather than effectiveness, how general are the results? Do these findings apply to a random sample of mid-level managers—or would their performance be normally distributed? Second, it is impossible to be sure how the truncation of the sample due to talent (or previous success) affects the power law scaling, and it is effectively impossible to properly estimate the value of $x_{\min}$ at which the power law scaling takes over (a point made by

Beck et al., in press). For instance, there are presumably many researchers who never publish in the top five journals and many working actors who never appear in a venue to even potentially be nominated for an award.

Third, the vanilla Gaussian distribution is an inappropriate null model for comparison to the power-law distribution. Figure 1 in O'Boyle & Aguinis (2012) shows the extraordinary difference between a power-law distribution and the full range of a Gaussian—in general, it is the tail of the Gaussian that invites comparison with the power-law, as discussed above. If the data are substantially skewed, the power law will almost certainly fit better than the Gaussian, even if the power-law model is itself a fairly bad fit. Step 4 above is a principled approach to comparing the power law to other highly skewed distributions, albeit ones substantially less exotic than power laws. Additionally (see next section), it is clear by inspection that the data examined by O'Boyle and Aguinis are non-normal. Specifically, all data considered are discrete and all are counts of success. Therefore, the Poisson distribution is probably a better null model, in that it assumes a constant rate for occurrences of independent events (e.g., journal article publications; cf., Gelman & Hill, p. 110 - 111). The fit of the Pareto distribution was superior to the Gaussian, but since the Gaussian is *a priori* a poor model for the data-generating mechanism, this test is not severe—it does little to probe the hypothesis that the data truly follow the Pareto distribution (cf., Mayo, 1996, p. 7).

Finally, even if a more appropriate null model was used, the fits of the Pareto distribution appear fairly poor in many cases. The weighted average of the χ^2 statistics reported by O'Boyle and Aguinis for each of their 5 studies were: 27897, 2024, 8692, 1076, and 7975[3]—substantially bad fits. The fits for the Gaussian models were immensely worse in almost all cases, but this comparison does little to support the contention that the data *were* drawn from Pareto distributions. Before we conclude that individual performance follows a power law distribution, its fit to data should be assessed with a method designed for distributions – the Kolmogorov-Smirnov statistic (Clauset et al., 2009).

The previous point bears emphasizing: even if the power law was found to fit the data well, it should then be compared to *plausible* alternative models, such as the lognormal (or even the Poisson). For instance, the lognormal was not ruled out for any of the two-dozen datasets reanalyzed by Clauset and colleagues, except for HTTP connections (2009, p. 689), each of which had been previously identified in the literature as following a power law distribution. To be clear, the tests performed by O'Boyle and Aguinis are appropriate for testing the norm of normality, and that the test of that norm is one of their central concerns. The tests clearly demonstrate that the normal distribution produces extremely poor fits to all of their data sets. It is only in the claim that the data are distributed according to a power law that these tests are weak.

Recall, also, that it can take a very wide range of data values to distinguish a power law from other skewed, heavy-tailed distributions. O'Boyle and Aguinis (2012) do not provide min-max ranges, but given their reported means and standard deviations, it is unlikely that these data cover more than a couple of orders of magnitude. Without a wider range of data it may not even be possible to distinguish a power-law distribution from one of these other, plausible distributions. If this were the case, then it would take *strong* theory about the generating mechanisms for individual job performance to choose between different distributions, theory that has not been adequately provided by the literature on individual job performance (Campbell, 1990; Borman, 1991).

Empirical Illustration[4]

Example I: Earnings of professional golfers in tournaments

We begin with somewhat exaggerated example: data on the total earnings in 2012 for the 146 professional golfers on the PGA tour is displayed in a histogram in Figure 3. We use this as an *effectiveness* metric, and we will present a more plausible measure of golfer performance shortly. It is clear to visual inspection that these data are non-normal—they are strongly right-skewed and not even approximately symmetric about their mode. To demonstrate that the distribution is most definitely not Gaussian, we conducted a Kolmogorov-Smirnov (K-S) test comparing the data to a normal distribution with the sample mean and standard deviation (in R; R Development core team, 2012):

--Insert Figure 3 about here--

```
> ks.test(sm.money,"pnorm",mean=mean(sm.money),sd=sd(sm.money))

One-sample Kolmogorov-Smirnov test

data:  sm.money

D = 0.2117, p-value = 2.282e-05

alternative hypothesis: two-sided
```

The results show that there is *strong* evidence that the data do not support the hypothesis that these earnings are normally distributed. This clearly demonstrates that the data we are dealing with should not be approached directly using the norm of normality, consistent with the claims of O'Boyle and Aguinis. We can provide some initial evidence that these data are log-normally distributed by log-transforming them and seeing whether that distribution is normal using the K-S test:

```
> log.money <- log(sm.money)

> ks.test(log.money,"pnorm",mean=mean(log.money), sd=sd(log.money))

        One-sample Kolmogorov-Smirnov test

data:  log.money

D = 0.0901, p-value = 0.2534

alternative hypothesis: two-sided
```

--Insert Figure 4 about here--

Based on the results of this K-S test, we cannot rule out the claim that the money earned in tournaments within a given year by professional golfers is log-normally distributed.

We next fit a Pareto distribution to the data in two ways. We first used the naïve power law fit (regression method), which produced an exponent estimate of 1.49 (the slope of the regression line) and an R^2 of .66 (i.e. 66% of the variance in log-frequency was explained by log-rank). We then used the methods described in Clauset and colleagues' (2009) paper. Figure 4 displays the empirical upper cumulative distribution of the golfers' tournament earnings. The red line is the maximum likelihood estimate of the power law model and the blue line is the fit derived from the lognormal model. As can be plainly seen in Figure 4, the earnings of professional golfers can be approximated using a Pareto distribution with an exponent of -1.09, but the fit is poor. The earnings data appear to show approximately quadratic decay, and in fact appear to decline more quickly than implied by even the lognormal model. The log-likelihood for the Pareto model is -1922.65, the log-likelihood for the lognormal is -1815.72 (smaller is better), and the Vuong test comparing the two models is -9.29 (one-sided p-value 7.5e-21—or << .000001)[5].

In this case, we have set $x_{\min}$ to the minimum of the distribution of money earned. That is, we are testing whether the entire (observed) distribution follows a power law or a lognormal distribution. This approach appears to be consistent with that used by O'Boyle and Aguinis. Further, we attempted to use the Kolmogorov-Smirnov approach described in step 2 above, but doing so set $x_{\min}$ so high that only 31 data points remained in the sample. Under this model, the Pareto distribution and the lognormal fit the data nearly equally well (the log-normal fit trivially and non-significantly better). Essentially, using the fully principled approach left too little data to be informative about the research question.

Average score over tournaments. One problem with total earnings as a metric for player performance is that earnings per tournament are distinct, potentially nonlinear transformations of the players' scores. Another useful performance metric for professional golfers is the actual score they obtain in a given tournament. Here, we use each player's mean (averaged over both days of the tournament and across tournaments) to rank players. The histogram (with an accompanying normal approximation to the histogram) is displayed in figure 5. Since lower scores are better in golf, the raw scores were reflected around their mean prior to these analyses.

-- Insert Figure 5 about here --

Again, this distribution is not particularly normal, but it is much more symmetric than the earnings. A KS test of normality provides a test statistic of 0.14 (p = 0.013), indicating that the data are non-normal. In this case, KS test of the log-transformation of the average scores produced a test statistic of .14 (p = .017), indicating that the data are not log-normally distributed, either. The naïve fit of the power law model provides an estimated x_{min} of 70.57 and the unlikely exponent value of 97.85. Similarly to the above example, using the KS approach to detect x_{min} leaves only 68 data points above the threshold, leading to an inconclusive model comparison. When using all data points, the power law model has a log-likelihood of -250.73, and the lognormal model has a log-likelihood of -183.99, with Vuong test of -4.81 (one-sided *p*-value of 7.69e-07, or << 0.0001). However, we should make it very clear that *none* of the normal, lognormal, or Poisson distributions fit these data very well.

-- Insert Figure 6 about here --

Example II: Congressional fundraising data

Our second example analysis is fundraising by US congress members in 2012. The Kolmogorov-Smirnov test of normality produced a test statistic of 0.235 ($p < .000001$), indicating substantial non-normality. A histogram of the data is presented in Figure 5. As in the previous example, we then ran a K-S test on the log-transformed fundraising data, yielding a test statistic of 0.04 ($p = 0.432$). As with the previous example, this transformation and test provides some evidence that these data are distributed log-normally.

--Insert Figure 7 about here--

The naïve fit of the power law model provided an estimate of 2.5 for the scaling exponent and an R^2 of .89. The principled fitting provided an estimate of the scaling exponent of 1.5 and a log-likelihood of -5907.22. The log-normal model had a log-likelihood of -5634.57. The Vuong test produced a test statistic of -14.35 ($p < .000001$). Figure 6 displays the empirical upper cumulative distribution in black, the power law fit in red, and the lognormal fit in blue. When estimating x_{min} from data, the estimate was so high that only 110 data points remained in the sample, and the Vuong test could not distinguish between the power law and the log-normal models. Figure 6 shows the two models plotted against the data. Like the previous example, the power law model severely overpredicts high levels of funds raised; however, in the case of this data set, the lognormal *does* underpredict extremely high amounts of funds raised. We should also note that there are two Congress members who raised substantially more funds than the others. These individuals could be considered outliers, but we retained them in the analysis because their presence should *increase* the likelihood that the power law would fit the data better than the log-normal (i.e. these extreme values are substantially more likely under the power law than under the log-normal).

--Insert Figure 8 about here--

Example III: Professional football players

Our next example involves several productivity measures for 235 NFL wide receivers. We examine total receptions, total yardage of receptions, and total touchdowns for the 2012 season.

Touchdowns. We first modeled the total touchdowns by NFL wide receivers (histogram presented in Figure 9). The mean for this performance indicator was 3.22 touchdowns with standard deviation 2.72 touchdowns. The KS test for normality was strongly rejected (.21, *p*-value = 3.77e-9). The test for log-normality was also strongly rejected (.24, *p*-value = 4.55e-12). We also attempted to fit an exponential distribution (.27, *p*-value = 5.77e-15) and a Poisson distribution (.21, *p*-value = 1.89e-9).

-- Insert Figure 9 about here--

The naïve fit for the power law distribution provided an estimated scaling exponent of 1.15 with R^2 of .86. The principled fit of Clauset et al. provided a scaling exponent estimate of 3.55 and log-likelihood of -147.01. The log-normal model produced a log-likelihood of -146.54. With x_{min} set to 4 touchdowns, these models were indistinguishable. The exponential distribution was also indistinguishable (log-likelihood of -147.07). There were only 80 data points over this threshold.

When we used the minimum of touchdowns (1 touchdown), we found that the Pareto had an estimated scaling exponent of 2.17 and log-likelihood of -398.16. The log-normal model produced a log-likelihood of -397.35 and the exponential distribution produced a log-likelihood of -422.54. By the Vuong test, the Pareto and the log-normal were indistinguishable (-0.68, one-sided *p*-value = 0.25), and the Pareto fit the data significantly better than the exponential (2.23, two-sided *p*-value = 0.03).

-- Insert Figure 10 about here --

Given the low maximum of touchdowns (14 touchdowns), it is possible that the discreteness of the data leads to the generally bad fits above. So, we fit two discrete alternative models, as well, the zeta distribution for the power law and the Poisson distribution for the less-exotic right-skewed, asymmetric distribution. The zeta distribution produced a scaling exponent of 1.76 and a log-likelihood of -509.24. The Poisson distribution produced a log-likelihood of -556.82. The zeta did not fit the data significantly better than the Poisson, according to the Vuong test (1.83, one-sided *p*-value = 0.97, two-sided *p*-value = 0.07). Since none of these fits prove conclusively better, we also tried fitting a discretized exponential distribution (see Clauset et al., 2009 and accompanying source code). This model produced a log-likelihood of -543.09. The discrete exponential model *did* fit the data significantly better than the zeta distribution, according to the Vuong test (-2.58, one-sided *p*-value = 0.005, two-side *p*-value = 0.01). The data and all fitted models are presented in Figure 10.

Total yards. We also modeled the total yards gained by the 235 NFL wide receivers (histogram presented in Figure 11). The mean for this performance indicator was 269.2 yards and the standard deviation was 329.6 yards. The KS test was not consistent with the data being normally distributed (.20, p = 2.22e-16), while the test for the log-normal was better, it still did not fit the data well (.07, p = .03).

-- Insert Figure 11 about here --

The Pareto distribution (using the naïve approach produced an estimated scaling coefficient of 1.59 with an R^2 of .69 and) the Clauset et al approach produced an estimated scaling coefficient of 3.93 and estimated threshold value of 626 yards and a log-likelihood of -482.66. The log-likelihood for the log-normal model was -479.98. We also fit an exponential distribution, which had a log-likelihood of -479.80. With 72 data points above the threshold, the Vuong test was not fully conclusive (-2.79 against the exponential distribution, one-side *p*-value = .04, two-sided *p*-value = .07; -2.68 against the log-normal, one-sided *p*-value = .07, two-side *p*-value = .14). When we set the minimum to 1 yard (the minimum positive value, yardage can be negative), the estimated Pareto scaling exponent was 1.21 with log-likelihood of -3331.67. The exponential distribution had a log-likelihood of -3029.16 (Voung test comparing it to the Pareto: -12.03, two-sided *p*-value = 2.45e-33). The log-normal model had a log-likelihood of -3016.58 (Vuong test comparing it to the Pareto: -18.07, two-sided *p*-value = 2.53e-73).

So, it is clear that for total yardage, both the log-normal and the exponential distribution, neither of which appear to fit the data perfectly, by the KS test, fit substantially better than does the Pareto distribution. The fits of these models compared to the upper cumulative distribution is shown in Figure 12.

-- Insert Figure 12 about here --

Total receptions. Finally, we modeled the number of total receptions made by 235 wide receivers in the National Football League in the 2012 season (histogram presented in Figure 13). The mean for this performance indicator was 23.16 and its standard deviation was 24.85. The data robustly fail a KS test for normality (.19, p = 1.58e-14, or << .0001). A KS test of the log transformation is cleaner, but still fails (.09, p = .002).

-- Insert Figure 13 about here --

The Pareto distribution (using the naïve approach produced an estimated scaling coefficient of 1.55 with an R^2 of .72 and) fitted using the Clauset et al techniques produced a scaling coefficient of 3.95 with a log-likelihood of -387.84. The log-normal fit had a log-likelihood of -384.08 and so fit the data weakly better than the Pareto (LR = -3.75, one-sided p-value = .04, two-sided p-value = .08). We also fitted an exponential distribution to the data (se Figure 13), which fit the data significantly better than the Pareto (LR = -4.00, one-sided p-value = .01, two-sided p-value = .02). Figure 14 shows the fit of each of these models to the data, with the power law in red, the log-normal in blue, and the exponential in green.

-- Insert Figure 14 about here --

Discussion

In each of the empirical examples, the data were distinctly non-normal. The norm of normality was definitely violated for all cases, consistent with O'Boyle and Aguinis's (2012) main thesis. Both performance outcomes had qualitative features similar to those presented in O'Boyle and Aguinis (2012). Further, these are qualitative features that are germane to the Pareto distribution: right-skew and heavy tails. On the other hand, for both outcomes, we demonstrated that the normal distribution fit well enough to a reasonable transformation of the data. Further, we conducted reasonably principled statistical fittings of both the Pareto and the lognormal distributions. When fit to all of the data, the lognormal fit significantly better in both cases. When we attempted to find the best-fitting power law regime, as recommended by Clauset and colleagues (2009), we were left with too little data to distinguish the power law from the lognormal. This does not mean that variables similar to ours' *never* follow a Pareto distribution. It seems clear that to suggest such a distribution, a researcher should (a) have a good theoretical rationale for proposing the Pareto, and (b) take care to compare the fit of the Pareto to plausible alternative models.

The forgotten Step 5: Does the power-law matter?

The histograms presented by O'Boyle and Aguinis (2012) in their Figure 2 are clearly non-normal. Simple visual inspection shows them to be a) all-positive and b) highly right-skewed (they are clearly non-symmetrical and bounded on the left by zero). Regardless of whether the data conform to a power law or not, they are decidedly heavy-tailed. This is a point that goes back to Tukey (1977; Mosteller & Tukey, 1977)—maybe further—that empirical data often have wider, heavier tails than that predicted by normal-theory methods. Even if data are not properly Gaussian, there are a range of options for statistical inference.

If data are lognormal or similarly shaped, transformations to normality can be used (Gelman & Hill, 2009, p. 59 – 68). If data are symmetrical, but more variable than predicted by the Gaussian distribution, robust regression techniques based on the *t*-distribution (Gelman & Hill, 2009, pp. 124 – 125; Kruschke, 2011, pp. 430 – 433) or *M*-estimators (Huber, 1964; cf., Fox, 2008, pp. 530 – 539) can be used. Such robust approaches also have the added advantage of being more resistant to outliers than OLS models. Generalized linear models (Nelder & Wedderbrun, 1972) can be used to fit regression models to a variety of discrete, non-normal outcomes (including logistic, ordered logistic, and Poisson regression; Gelman & Hill, 2009, p. 109 – 124; given the count-based nature of most of the effectiveness measures considered by O'Boyle and Aguinis, some form of Poisson regression could work if one wanted to predict these outcomes). Other forms of non-normality may arise from mixtures of distributions (of which the combined heights of males and females described above is a simple example), which can be handled with mixture models (e.g., Benaglia, Chaveau, Hunter, & Young, 2009).

In many situations, the qualitative phenomenon of a right-skewed, heavy-tailed distribution is going to be far more important than the exact mathematical form of the distribution (Stumpf & Porter, 2012). An obvious situation occurs in market crashes: on one hand, it is important to note that these occur considerably more frequently than predicted by a Gaussian distribution (Gabaix et al., 2003), regardless of the form of the distribution. On the other hand, a thorough understanding of the form of the distribution may be essential to developing good theory for the underlying mechanisms that drive these fluctuations—understanding of which is important for designing policy interventions.

Implications of non-normal distributions for HRM practices

If it has been established that a particular performance measure is not normally distributed, how should personnel psychologists or HR practitioners deal with problems such as selection? The methods of classical linear regression are deeply entrenched in this area. A measure of linear association is the traditional standard of validity and incremental validity. Any sort of non-normality in performance raises the questions of how valuable such indices for validity are.

It is our opinion that statistics in this area can no longer be considered cut-and-dried. That is, the traditional toolbox of classical statistics needs upgrading to be useful to practicing organizational scientists. It is increasingly clear that plugging an outcome and a set of predictors into a linear regression model yields results that are not meaningful (Achen, 2002; 2005; we discuss only linear regression models here, but the argument applies to more sophisticated models, such as structural equations, with or without latent variables; Berk, 2004). One avenue is provided by Bayesian methods—which should also not be thoughtlessly applied. The Bayesian formalism encourages careful practitioners to build explicit probability models, which should simultaneously encourage some care in selecting the distributions used. It should be noted that the Bayesian formalism allows considerable freedom to researchers in specifying both likelihoods (data models) and prior distributions. This freedom can be misused or abused, so we are not suggesting that Bayesian methods will solve all scientific problems in organizational research. We merely suggest that these tools should become part of the regular toolkit of organizational scientists (Kruschke, Aguinis, & Joo, 2012).

Another route is to use non-parametric statistical models (e.g., Fox, 2009; Shalizi, 2012; Spain, Sotak, Tsai, & Harms, 2013). Kernel density methods can be used to estimate the shape of the distribution, and various non-parametric techniques can be used to fit essentially arbitrarily complex models linking predictors and criteria. Like more traditional approaches, these models too can be thoughtlessly applied, in which case they may be even more damaging than simple linear schemes. Non-parametric models have effectively infinite parameter spaces (in practice the parameter space *is* bounded, but it grows with the sample size). Without care, they can optimally fit noise easily. Non-parametric models need good out-of-sample tests of their predictions (so do linear models, but cross-validation is even more of a concern when the fitted model is allowed to be as *rough* as non-parametric models allow).

Furthermore, if the performance variable of interest really is distributed according to a power law, it can be regressed on predictor variables using *nonlinear least squares* (NLS, e.g., Fox, 2009, pp. 463 – 469; Shalizi, 2012). NLS approaches work by finding α that minimizes a loss function of the form $\sum_{n=1}^{N}(y_i - y_0 x_i^{-\alpha})^2$, where y is an outcome of interest, x is a predictor, and $-\alpha$ is the scaling parameter. This is a more difficult function to optimize than the linear least squares loss function, and usually must be fit using an algorithm like gradient descent—an iterative approach that is guaranteed to converge to a(n at least local) minimum. Methods like this—power law regression fit using NLS—are not implemented in general-purpose software. Organizational scientists need to become more familiar with applying modern, computational statistics (e.g., programming methods like the above in R or some similar language) if we are going to handle these sorts of non-standard models[6].

Checking assumptions

It is important that researchers and practitioners routinely perform checks of a variety of assumptions before applying rote statistical models to data. For instance, the Kolmogorov-

Smirnov (KS) test and the Shapiro-Wilk test can be used to assess whether a distribution matches a target distribution (e.g., Normal, Poisson, etc.) or is distributed normally, respectively. In particular, the KS test is useful for a wide variety of applied problems, and should probably be used more broadly (see, e.g., approaches to comparing entire distributions for differences, Handcock & Morris, 1999).

The distribution of the outcome (and more importantly, the error terms) is an important aspect of the *specification* of a statistical model. In general, the assumptions underlying statistical models are seldom checked in many fields (e.g., Achen, 2005; Antonakis, Bendahan, Jacquart, & Lalive, 2010; Berk, 2004; Shalizi, 2012; Spain et al., 2013). It is important to address these assumptions and to use methods consistent with the data at hand. Otherwise theoretically driven tests will not be *severe*; that is, the tests will do little to demonstrate whether or not the theory makes errors in predicting reality (Gelman & Shalizi, 2013; Mayo, 1996). The methods reviewed in this paper can help address assumptions about the distribution of data, and understanding the distribution of particular data can help to build better scientific models.

This question of specification is a broad concern for social sciences, in general, and the organizational sciences, in particular. Too often, traditional data models, like OLS regression or structural equation models, are applied in situations where any conclusions are suspect (e.g., Achen, 2005; Berk, 2004; Shalizi, 2012). Tools that can provide more robust inferences about associations are available (Wilcox & Keselman, 2012), as are methods for allowing flexible estimation of the functional form of relationships between predictor and criterion variables (e.g., Spain et al., 2013). A lot of modern approaches based in statistical learning theory make far weaker assumptions about distributions and functional forms than do traditional methods (see Hastie, Tibshirani, & Friedman, 2009 for an in-depth introduction, and Berk, 2008 for an overview using regression-type models). The emphasis of these new methods is on *learning from data*, with less emphasis on traditional null hypothesis significance testing (though such tests can generally be formulated, e.g., Fox, 2009; Shalizi, 2012; Spain et al., 2013). Greater emphasis on approaches based on statistical learning theory could help organizational scientists to formulate and test more complex, realistic models of organizational behavior and rely less on traditional toolboxes and heuristics like the norm of normality.

Conclusion

We believe that Andriani and McKelvey (2009) and O'Boyle and Aguinis (2012) have opened an important and interesting dialogue in the organization sciences. It is becoming nearly impossible to ignore the influence of complex systems thinking in our fields, even such traditional areas as individual job performance. We do not believe that O'Boyle and Aguinis have convincingly demonstrated that job performance follows a power law, but they have convincingly demonstrated that some indices of performance follow very badly non-normal distributions. Their findings should be especially important for anyone studying constructs like creative performance, where outcomes like published papers or entertainment awards are germane to the research questions. Papers such as theirs' should cause our research enterprise to pause and consider our methods—and how well they match the phenomena we wish to study.

References

Achen, C.H. (2002). Toward a new political methodology: Microfoundations and ART. *Annual Review of Political Science, 5,* 423 – 450.

Achen, C.H. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science, 22,* 327 - 339.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74,* 47 – 97.

Andriani, P., & McKelvey, B. (2009). From Gaussian to Paretian thinking: Causes and implications of power laws in organizations. *Organization Science, 20,* 1053 – 1071.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*, 1086 – 1120.

Barabási, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature, 435,* 207 – 211.

Barabási, A-.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaboration. *Physica A, 311,* 590 – 614.

Beal, D.J., Weiss, H.M., Barros, E., & MacDermid, S.M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology, 90*, 1054 – 1068.

Beck, J.W., Beatty, A.S., & Sackett, P.R. (In press). On the distribution of job performance: The role of measurement characteristics in observed departures from normality. Manuscript in press at *Personnel Psychology.*

Benaglia, T., Chaveau, D., Hunter, D.R., & Young, D.S. (2009). Mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software, 32,* 1 – 29.

Berk, R.A. (2004). *Regression analysis: A constructive critique.* Thousand Oaks, CA: Sage.

Berk, R.A. (2008). *Statistical learning from a regression perspective.* New York, NY: Springer.

Borman, W.C. (1991). Job behavior, performance, and effectiveness. In M. Dunnette & L. Hough (Eds.) *Handbook of Industrial/Organizational Psychology*, (Vol. 2, 271-326). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette, & L. M. Hough (Eds). *Handbook of Industrial and Organizational Psychology* (Vol. 1, 2nd ed.; pp. 687-732). Palo Alto, CA: Consulting Psychologists Press, Inc.

Clauset, A., Young, M., & Gleditsch, K.S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution, 51,* 58 – 88.

Clauset, A., Shalizi, C.R., & Newman, M.E.J. (2009). Power-law distributions in empirical data. *SIAM Review, 51,* 661 – 703.

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton, NJ: Princeton University Press.

Fox, J. (2008). *Applied regression analysis and generalized linear models.* (2nd Ed.) New York: Sage.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H.E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature, 423,* 267 – 270.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

Gelman, A., & Shalizi, C.R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology, 66,* 8 – 38.

Gilmore, A.L. (2006). *Social networks and social movements*. Unpublished master's theisis, University of Oxford.

Handcock, M.S., & Morris, M. (1999). *Relative distribution methods in the social sciences.* New York, NY: Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd Ed.). New York, NY: Springer.

Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics, 35,* 73 – 101.

Kruschke, J.K. (2011). *Doing Bayesian data analysis.* New York, NY: Elsevier/Academic Press.

Kruschke, J.K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods, 15*, 722 – 752.

Lichtenstein, B.B., & Plowman, D.A. (2009). The leadership of emergence: A complex systems leadership theory of emergence at successive organizational levels. *The Leadership Quarterly, 20,* 617 – 640.

Malmgren, R.D., Stouffer, D.B., Motter, A.E., & Amaral, L.A.N. (2008). A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences, 105,* 18153 – 18158.

Marion, R. & Uhl-Bien, M. (2001). Leadership in complex organizations. *The Leadership Quarterly, 12,* 389 – 418.

Mayo, D.G. (1996). *Error and the growth of experimental knowledge.* Chicago, IL: University of Chicago Press.

Mitzenmacher, M. (2004a). Dynamic models for file sizes and double Pareto distributions. *Internet Mathematics, 1*, 305 – 334.

Mitzenmacher, M. (2004b). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics, 2,* 226 – 251.

Mosteller,F., & Tukey, J. (1977). *Data analysis and regression: A second course in applied statistics.* Reading, MA: Addison-Wesley.

Nelder, J.A., & Wedderbrun, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A, 135*, 370 – 384.

Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences, 98,* 404 – 409.

Newman, M.E.J. (2005). Power laws, Pareto's law, and Zipf's law. *Contemporary Physics,* 323 – 351.

O'Boyle, E., Jr., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology, 65*, 79 – 115.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical findings from 85 years of research. *Psychological Bulletin, 124,* 262 – 274.

Shalizi, C.R. (2012). *Advanced data analysis from an elementary point of view.* Unpublished lecture notes, Carnegie Mellon University, [accessed June 27, 2012], URL: http://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ADAfaEPoV.pdf

Spain, S.M. (2010). *Multivariate dynamic criteria: A process model of job performance.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Spain, S.M., Sotak, K.L., Tsai, C.Y., & Harms, P.D. (2013). Testing the form of theoretical models by relaxing assumptions: Comparing parametric and additive models. Paper presented at the annual meeting of the Society for Personality and Social Psychology, New Orleans, LA. [accessed March 22, 2013] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2164297

Stouffer, D.B., Malmgren, R.D., & Amaral, L.A.N. (2005). Comment on *the origins of bursts and heavy tails in human dynamics.* arXiv:physics/0510216v1

Stumpf, M.P.H., & Porter, M.A. (2012). Critical truths about power laws. *Science, 335,* 665 – 666. doi: 10.1126/science.1216142

Taleb, N.N. (2007). *The black swan: The impact of the highly improbable.* New York: Random House.

Tukey, J. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Vázquez, A., Oliveira, J.G., Dezső, Z., Goh, K.-I., Kondor, I., & Barabási, A.-L. (2006). Modeling bursts and heavy tails in human dynamics. *Physical Review E, 73,* 036127-1 – 19.

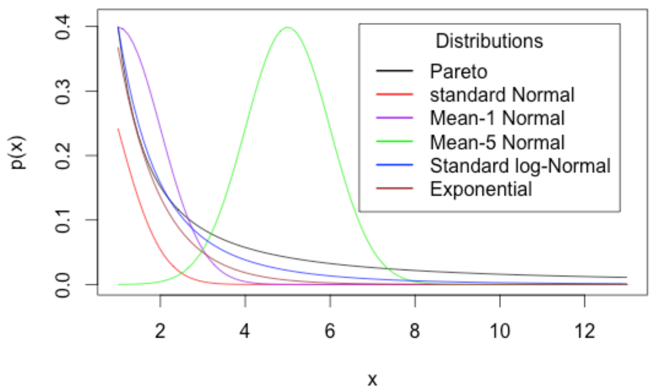
Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57,* 307 – 333.

Wilcox, R.R., & Keselman, H.J. (2012). Modern regression methods that can substantially increase power and provide more accurate understanding of associations. *European Journal of Personality, 26,* 165 – 174.

Figures

Figure 1a. Illustration of power law, several normal, log-normal, and exponential distributions

Figure 1b. Illustration of how the scaling exponent changes the shape of a Pareto distribution



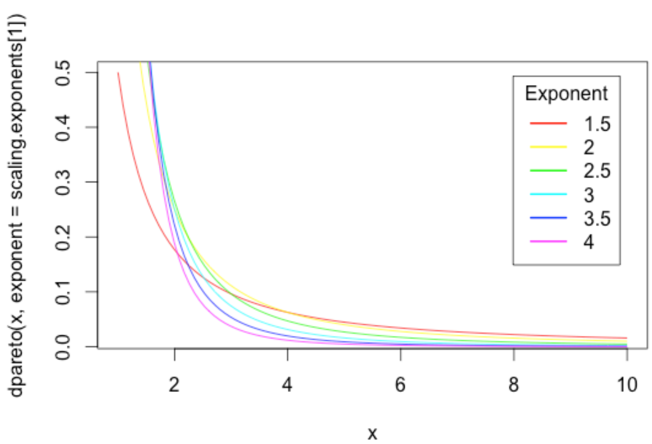
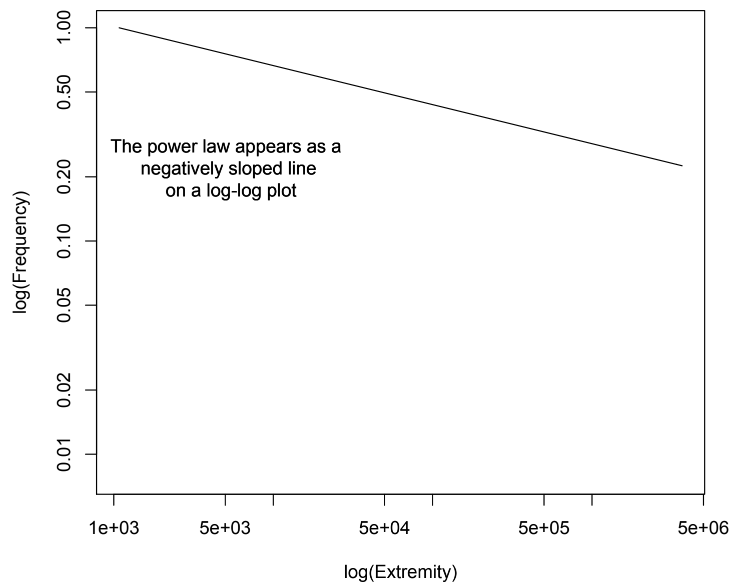


Figure 2a. Example of a power law fitted to a log-log plot.

Figure 2b. Illustration of scale invariance. The horizontal brackets span different ranges, but would be described by the same power law.



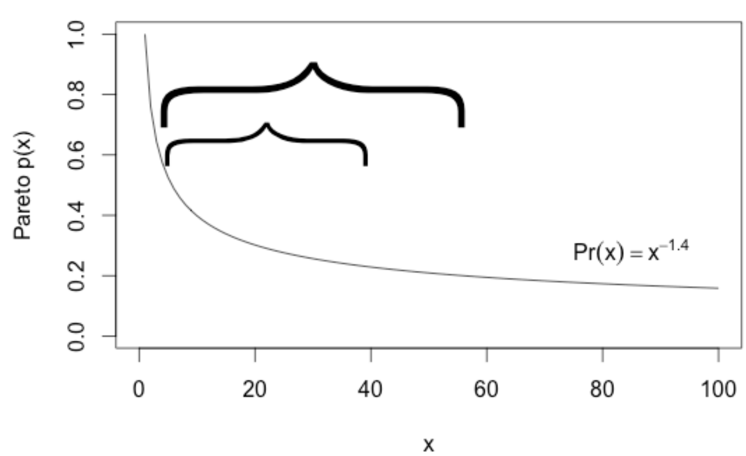


Figure 3. Histogram of total money earned in tournaments.

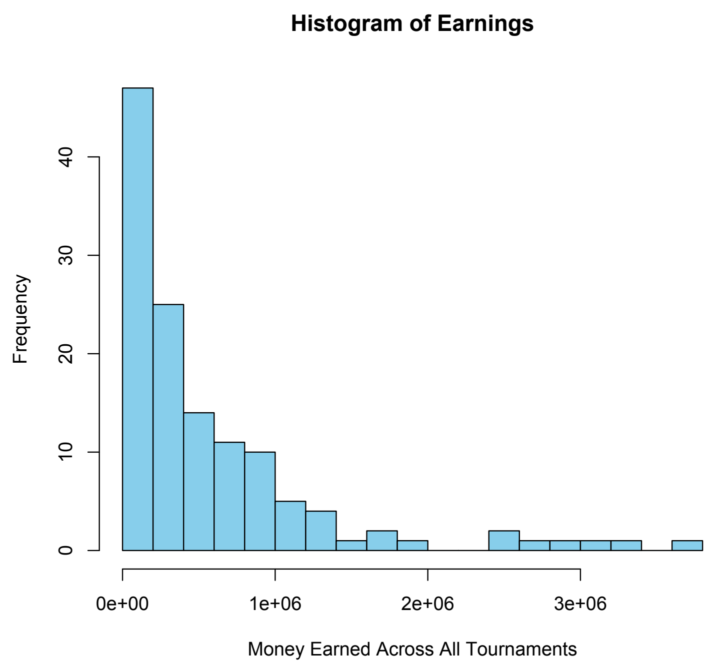
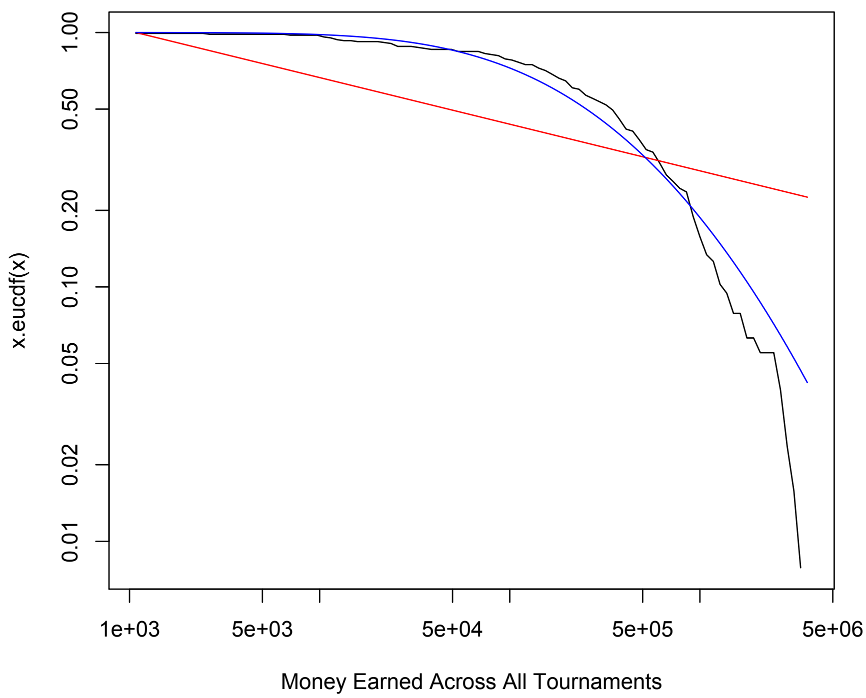


Figure 4. Data on total money earned by PGA players and empirical fits of the Pareto and log-normal distributions.



Note. Black line is the empirical upper cumulative distribution function (on the log-log scale, this is sometimes known as the "survival function"). The red line is the fit from the power law model. The blue line is the fit from a log-normal model.

Figure 5. Histogram of average PGA player scores for 2012 with normal density overlaid.

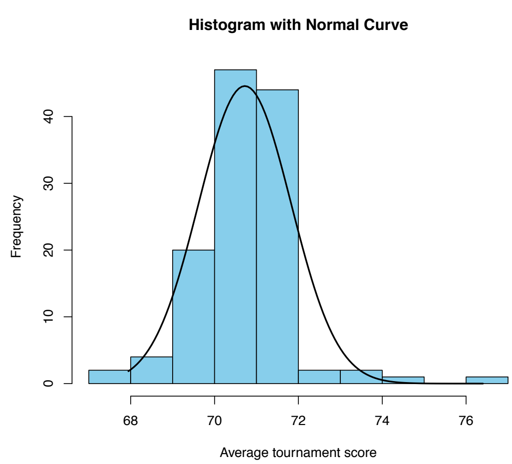
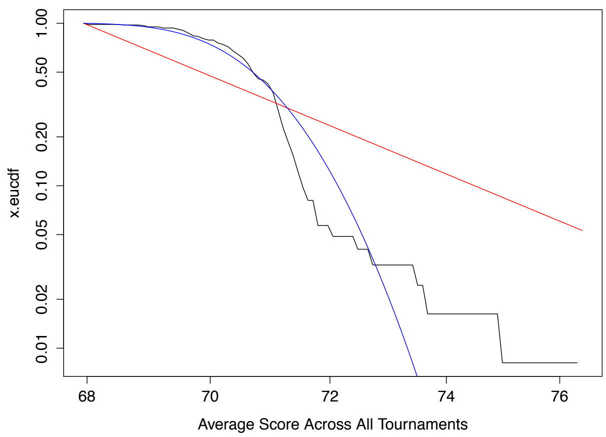


Figure 6. Data on average scores of PGA players and empirical fits of the Pareto (red) and log-normal (blue) distributions.



Note.

Figure 7. Histogram of fundraising by incumbent congress members in 2012.

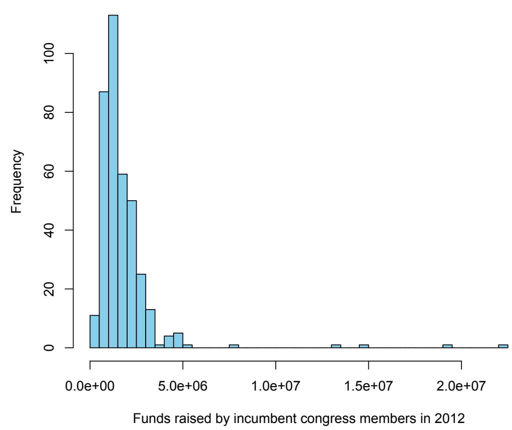


Figure 8. Data on fundraising by congress members and empirical fits of the power law and log-normal models.

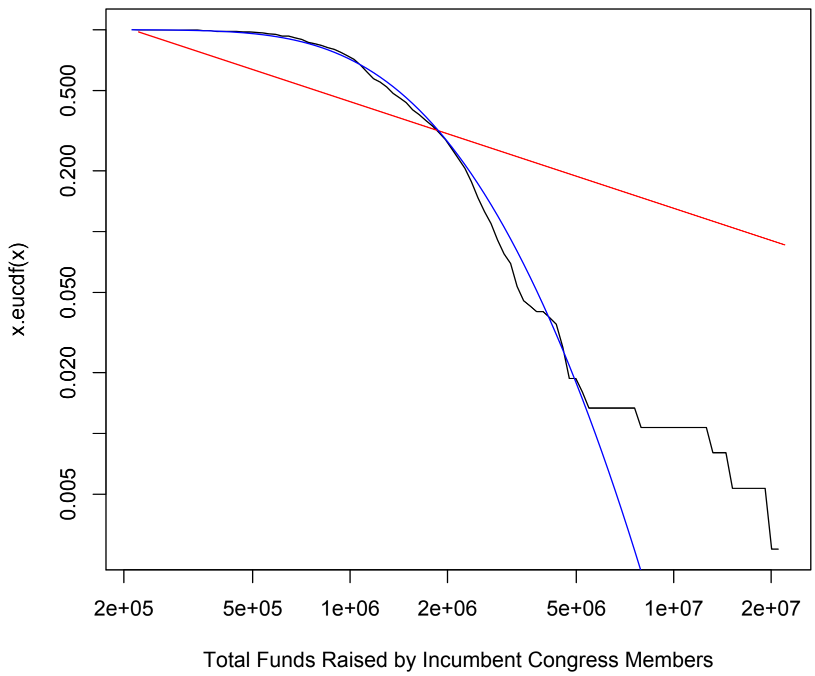


Figure 9. Histogram of touchdowns by NFL wide receivers with distribution functions for the log-normal (red) and exponential (black) superimposed.

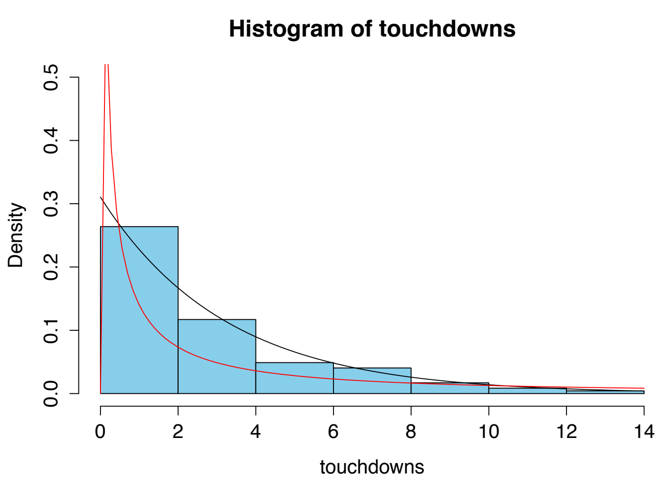


Figure 10. Fit of Pareto (red), log-normal (blue), and exponential (green) distributions to the empirical upper cumulative distribution of total touchdowns.

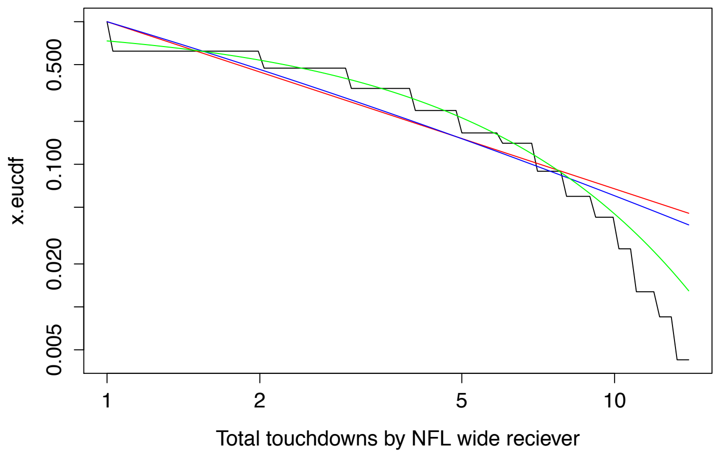


Figure 11. Histogram of total receptions with exponential distribution (black) and log-normal distribution (red) imposed.

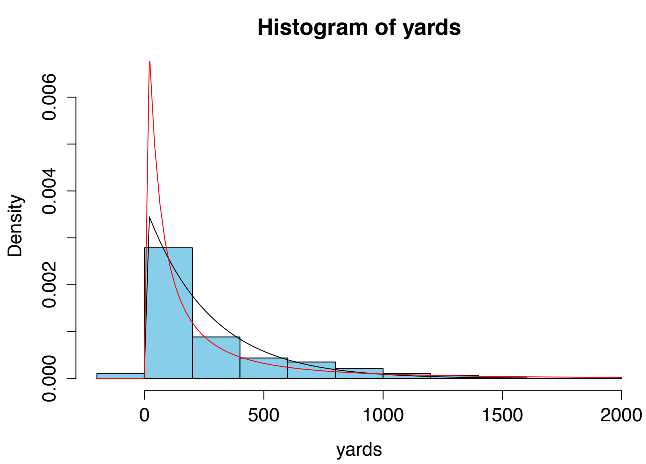


Figure 12. Fit of Pareto (red), log-normal (blue), and exponential (green) distributions to the empirical upper cumulative distribution of total yards.

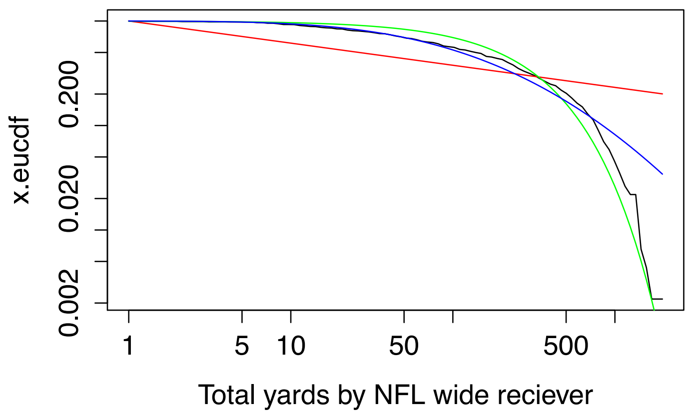


Figure 13. Histogram of total receptions with exponential distribution (black) and log-normal distribution (red) imposed.

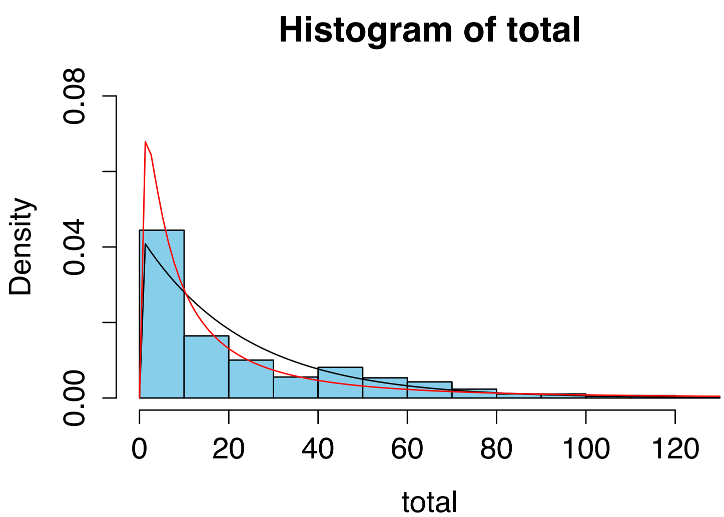
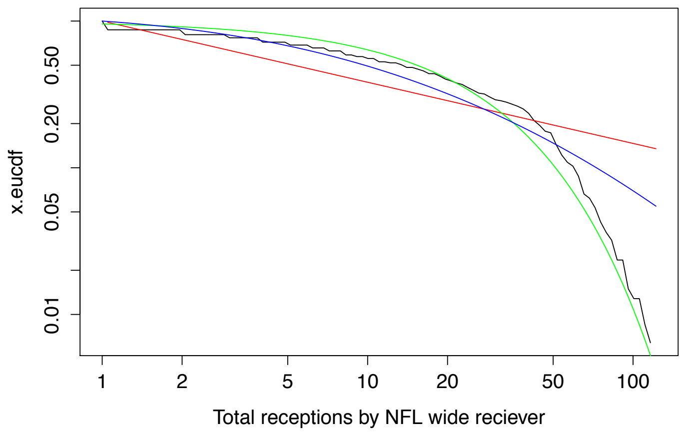


Figure 14. Fit of Pareto (red), log-normal (blue), and exponential (green) distributions to the empirical upper cumulative distribution of total receptions.



---

[1] It should be noted that standard statistical theory does not actually require a dependent variable to be normally distributed, but that it is *conditionally* normal. Consider the linear regression equation, $y_i = B_0 + B_1X_i + e_i$. Depending on the distribution of the predictors, X_i (say, binary), then the distribution of the y_is may be decidedly non-normal. The error term, e, is what is normally distributed in classical regression.

[2] In addition to the Clauset et al. (2009) paper, this discussion draws on C.R. Shalizi's comments on the paper at http://cscs.umich.edu/~crshalizi/weblog/491.html

[3] The algorithms used by the @Risk used by O'Boyle and Aguinis software to fit these models are not given in the company's publicly available user's manuals.

[4] The data for all empirical examples included in this paper are available for download at: http://bingweb.binghamton.edu/~sspain/adv-stat/gauss.html

[5] R functions for fitting these models are available from Aaron Clauset at http://tuvalu.santafe.edu/~aaronc/powerlaws/. Tutorial information on using the code is available at http://intersci.ss.uci.edu/wiki/index.php/Power-law_distributions. Some functions that accompany their paper require compiling C code, but none of the functions needed for the analyses in the current manuscript do.