Teaching Students Not to Dismiss the Outermost Observations in Regressions

Tomasz Kasprowicz
Academy of Business in Dabrowa Gornicza

Jim Musumeci
Bentley University

Journal of Statistics Education Volume 23, Number 3 (2015),
www.amstat.org/publications/jse/v23n3/kasprowicz.pdf

Key Words: Active learning; Outlier; Linear regression

Abstract

One econometric rule of thumb is that greater dispersion in observations of the independent variable improves estimates of regression coefficients and therefore produces better results, i.e., lower standard errors of the estimates. Nevertheless, students often seem to mistrust precisely the observations that contribute the most to this greater dispersion. This paper offers an assignment to help students discover for themselves the value of the observations that are farthest from the mean.

1. Introduction

Practitioners in several disciplines have come to appreciate that students' self-discovery leads to a clearer understanding than does a traditional lecture format. At the extreme, the Moore Method of teaching graduate mathematics (as described in Jones (1977)) features *no* lectures, but instead only a set of definitions, axioms, and propositions, with the students to use their own reasoning ability to either prove each proposition or give a counterexample. This is consistent with Moore's belief that "That student who is taught the best is told the least" (Parker 2005, p. vii). Cohen (1982) goes on to describe how this method can be adapted to the undergraduate teaching of mathematics, and indeed inquiry-based learning courses are often seen as an application of the Moore Method. Crouch and Mazur (2001, p. 970) report that "In recent years, physicists and physics educators have realized that many students learn very little physics from traditional lectures," and in two separate studies, White and Frederickson found that 11th and 12th graders in a traditional lecture-based classroom had a basic grasp of physics that was surpassed by that of sixth graders (1997) and seventh-ninth graders (1998) exposed to physics in an inquiry-based learning environment. Similarly, in chess, Soltis (2010, p. 23) quotes Mikhail Botvinnik as having told his students, "Chess cannot be taught. Chess can only be learned." Botvinnik, himself a world champion, founded a school that has itself produced three other world champions. Bransford, Brown, and Cocking (2000) also emphasize the importance of active learning, and Medina (2009, p. 74) observes "Before the first quarter-hour is over in a typical presentation, people have *usually* checked out. If keeping someone's interest in a lecture were a business, it would have an 80 percent failure rate."

Many statistics educators have also adopted the view that lectures often don't work well. Cobb (1992, p. 9), for example, observes that "Shorn of all subtlety and led naked out of the protective fold of educational research literature, there comes a sheepish little fact: lectures don't work nearly as well as many of us would like to think. Cobb (1992, pp. 15-18) recommends "active learning" as an alternative, with less of an emphasis on lectures and more on student learning. Snee (1993) also points out the need for experiential learning and emphasizes the importance of using real data to identify or solve real problems. Interestingly, both Snee and Moore refer to the same Chinese proverb to make this point: "I hear, I forget. I see, I remember. I do, I understand."

This approach is consistent with the GAISE (2005) guidelines, particularly the first and third of:

1. Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Foster active learning in the classroom.
4. Use technology for developing conceptual understanding and analyzing data.
5. Integrate assessments that are aligned with course goals to improve as well as evaluate student learning.
6. Use assessments to evaluate student learning.

Garfield, Hogg, Schau, and Whittinghill (2002) and Zieffler et al (2012) provide some preliminary evidence that teachers' beliefs and practices are consistent with the GAISE guidelines, and there is also evidence that the active-learning approach produces better results in statistics education. Keeler and Steinhorst (1995), for example, found that introductory statistics students in an active-learning environment had higher class averages and course-completion rates than students who took the same class in a traditional lecture format. Lesser and Kephart (2011) discuss the value of setting the tone for an inquiry-based format on the first day of class. Carter, Felton, and Schwertman (2014) and Vaughan (2015) provide two examples of assignments that use actual data (as per GAISE guideline #2) and that are sufficiently open-ended that students must decide for themselves what they are trying to find, and how to get there.

This paper provides another example that uses actual data and clears up another common misconception students have, namely, that the outermost observations are somehow untrustworthy because they are distant from the mean observation. We apply the active-learning approach to facilitate student understanding of the value of disperse observations in a linear regression analysis. Estimating parameters for the linear regression $Y = \alpha + \beta X + \varepsilon$ requires some dispersion in a sample's observations of the independent variable X. If all the observations x_i of the independent variable were identical, then any non-vertical line passing through the mean point $(\overline{x_i}, \overline{y_i})$ would produce identical residuals e_i, and consequently any estimate $\hat{\beta}$ would be just as valid as any other. At least one observation of the independent variable must be different from the others to help "anchor" the regression line.

Because random samples are, by definition, necessarily random, and ordinary least squares estimates $\hat{\alpha}$ and $\hat{\beta}$ rely on the sample, it follows that these estimates themselves (and their standard errors) are random variables. The Gauss-Markov theorem establishes that the ordinary least squares estimator is the best linear unbiased estimator (BLUE), where "best" is defined as having minimum variance (e.g., Studenmund (2006, p. 102)). An informal and admittedly less precise way of thinking about this minimum variance property is that, within the set of unbiased estimators, the best (unbiased) estimator is more likely to be closer to the true population parameter than is any other (unbiased) estimator.[1] In addition, according to Casella and Berger (1990, p. 556), when we apply a linear regression technique, "we are implicitly making the assumption that the regression of Y on X *is* linear" or that "the regression of Y on X can be adequately approximated by a linear function." In the rest of this paper, we are making the assumption that a linear relationship is appropriate. If there is doubt on this issue, then applying a linear model is premature.

After stating this requirement that there be some dispersion in observations of the independent variable, some econometrics textbooks go on to suggest the rule of thumb that greater variance in the observed x_i improves estimates of β. For example, Gujarti (2003, pp. 72-73) states "Looking at our family consumption expenditure example in Chapter 2, if there is very little variation in family income [the independent variable], we will not be able to explain much of the variation in the consumption expenditure [the dependent variable]." Kmenta (1997, p. 225) similarly observes "The more dispersed the values of the explanatory variable X, the smaller the variances of $\hat{\alpha}$ and $\hat{\beta}$… [thus] if we have an absolutely free choice of selecting a given number of values of X within some interval—say, from a to b (0 < a < b)—then *the optimal choice would be to choose one half of the Xs equal to* a *and the other half equal to* b." The reason is that the standard error of the estimate $\hat{\beta}$ features in its denominator the sum of the observed independent variable's squared deviations from the sample mean (e.g., see Kmenta (1997, p. 213)) so *ceteris paribus* a bigger dispersion in observations of the independent variable increases the denominator, thus reducing the standard error and producing better estimates.

Despite this, students often appear to believe that a sample's outermost observations are not to be trusted. For example, Dawson (2011, p. 2) defines a "mild outlier" as one that is more than 1.5 times the interquartile range (equal to the 75th percentile observation minus the 25th percentile observation), and observes that if the independent variable is normally distributed, about 0.8% of the observations qualify as mild outliers. He finds students often interpret such outliers "as evidence that the population is non-normal or that the sample is contaminated." As part of a project, we asked students a related question (question 1 of Part B of Appendix B) and found a

[1] It is less precise because, for example, another estimator that is not BLUE might be close to the actual parameter value almost all the time, but be extremely far away when it is not close. In such an instance, this alternate estimate is more likely to be close to the true parameter value than is the BLUE estimate, and yet this alternate estimate is not BLUE if its extreme misses cause it to have a higher variance.

similar mistrust of a sample's most extreme observations. For example, one student wrote that "extreme data points [top and bottom deciles] can be very far off from the average values…using the outliers can greatly skew the information," while another anticipated that "the calculated beta will be much more accurate when we look at the middle returns and not the extremes." These comments were fairly representative of the class' beliefs regarding the sample's most extreme observations. Cobb (1992, p. 10) points out that "As teachers, we consistently overestimate the amount of conceptual thinking that goes on in our courses, and under-estimate the extent to which misconceptions persist after the course is over." We believe it is important to address this misperception we observe about outliers.

In general, a larger sample size can be expected to produce better estimates. Absent a clear error in the data, it is never a good idea to throw away data, but are all the observations equally valuable, or are some more likely to produce better estimates than others? The Kmenta quote above suggests the peripheral observations will be most informative, but certainly a class assignment confirming this will be more persuasive to students than a sentence in a lecture or textbook. Alternately, as Dawson (2011, p. 3) expresses it, "Of course, this can be illustrated with a few well-chosen slides in a conventional lecture. However, my classroom experience has been that such a lesson is not well retained, and that students continue to take a very literal view of boxplot outliers as evidence either that the distribution is non-normal or that the flagged datum is somehow 'wrong.' This paper suggests an alternative approach based on experiential learning, via a simulation." Our goal is similar to Dawson's, and is very limited. We provide an exercise that is not designed to suggest students ignore any data they have collected, nor does it include a test for linearity; instead, its narrow focus is that students learn for themselves not to mistrust their sample's most extreme observations, but rather to welcome such observations because they increase the observed dispersion of the independent variable, and thus produce better regression parameter estimates.

In the remaining sections, we briefly present simulations showing that the most extreme values of the independent variable are the ones that are most useful in determining the regression parameter estimates $\hat{\beta}$ and $\hat{\alpha}$. We define the "most extreme values" as the outer deciles, although the same principle is broadly applicable to other definitions. We then describe an ungraded quiz we gave in different semesters to two sections of the same course to determine students' beliefs about what subset of a dataset would produce the best estimates of β and α. Finally, we include [in Appendix B] an out-of-class assignment that subjected those prior beliefs to empirical scrutiny. Most students were surprised at the results, and we report selected responses in Section III.

2. Example of Dispersion Improving Estimates

What is meant by the rule of thumb regarding the desirability of dispersion in the independent variable is that, provided the independent and dependent variables are not transformed in a way that changes dispersion, greater diffusion in observations of the independent variable will produce better estimates, i.e., estimates with lower standard errors. We assume that all the conditions of Ordinary Least Squares hold, specifically, that the relationship between the dependent and independent variable is linear, that the error terms ε are normally distributed with a mean of zero, are homoscedastic and not correlated with each other, and that the values of the independent variable X are uncorrelated with the error term[2] and do not have a sample variance of zero (e.g., see Kmenta (1997, p. 208)). We demonstrate the positive effects of greater dispersion of the independent variable by example, using the regression $Y = 0.5 + 12X + \varepsilon$, where ε is normally distributed with mean 0 and standard deviation 10. We consider three possibilities for the distribution of sample values of the independent variable, namely that observed x_i are drawn from a uniform distribution between 0 and 1, that they are normally distributed with mean 0 and standard deviation 10, and that they are normally distributed with mean 100 and standard deviation 10.

Suppose there exist 250 ordered pairs of data (x_i, y_i), but the data are very expensive and we can only afford 50 pairs out of the 250. However, a vendor can sort the ordered pairs in increasing order of the value of the independent variable, and sell us any 50 pairs. He gives us the following choices:

1. We can choose 50 randomly selected ordered pairs.
2. We can choose the first five ordered pairs in each decile.
3. We can choose the ordered pairs in any two deciles.

Given these choices, the following possibilities may seem plausible candidates for the best choice, with the “reasons” in parentheses:

A. Choose 20% of the pairs randomly (because random selection ensures unbiased tests).
B. Choose the first five ordered pairs in each decile (because this will give us representation across the entire spectrum of X).
C. Choose the fifth and sixth deciles’ ordered pairs (because these are closest to the median, and thus are most representative of the “typical” values for X).
D. Choose the first and tenth deciles’ ordered pairs (because these are the most extreme, and thus maximize the variance of the observations of X).

We simulated the specified dataset 2000 times and compared all four choices with the estimates obtained from the full dataset. Table 1 reports the results when the x_i are drawn from a uniform distribution between 0 and 1. When compared with the four limited-data choices, the full sample has the lowest standard errors for $\hat{\beta}$ (2.2115) and $\hat{\alpha}$ (1.2672). This is no surprise—in general, the reason larger samples are desirable is because they produce lower standard errors of the estimates. Using the full sample as a benchmark, we find that of the four limited-data choices, the top and bottom deciles (option D) had the largest average sample variance of the independent variable at 0.2058, resulting in the lowest standard errors (3.1708 for $\hat{\beta}$, or 43.4% higher than the full sample, and 2.1107 for $\hat{\alpha}$, or 66.3% higher than the full sample) as predicted by the Kmenta quote in the introduction. For the same reason, using the middle two deciles (option C), with their average sample variance of only 0.0035, produced the worst results, as implied by the earlier quote from Gujarti. Option C’s standard errors for $\hat{\beta}$ and $\hat{\alpha}$ are 25.6722 and 12.9232, both over ten times the corresponding values for the full sample.[3] The choice between options A and B is very close, but their standard errors are still substantially larger than D’s.

[2] This is the condition specified by Studenmund (2006, p. 89); it is often expressed in other econometric texts as “the observations X_i can be considered fixed in repeated samples” as in Kennedy (2008, p. 41) or Kmenta (1997, p. 208).
[3] While ignoring all but the two middle deciles of observations is an extreme example, it suggests that indiscriminate Winsorizing or trimming a dataset discards valuable information, and should be reserved for cases when there is strong reason to believe the data are erroneous. This point is also made by Studenmund (2006, p. 74).

Table 1. Regression Statistics for Four Different Methods for Choosing a Subset of the Population Independent-Variable Observations Uniformly Distributed Between 0 and 1

Selection of X from simulated dataset sorted on X	average independent variable sample standard deviation	average $\hat{\beta}$	average standard error for $\hat{\beta}$	average $\hat{\alpha}$	average standard error for $\hat{\alpha}$
Full Dataset	0.2887	11.9815	2.2115	0.4861	1.2672
Choice A: Random	0.2891	11.8282	5.0121	0.5631	2.9114
Choice B: First 5 from each decile	0.2897	12.1019	4.8784	0.4278	2.6713
Choice C: Middle two deciles	0.0592	12.8013	25.6722	0.0858	12.9232
Choice D: Top and bottom decile	0.4537	11.9282	3.1708	0.5242	2.1107

Summary statistics for different methods of selecting a sample of size 50 (from a total set of 250 ordered pairs). The regression line is $y_i = 0.5 + 12x_i + \varepsilon_i$, where ε_i is normally distributed with mean 0 and standard deviation 10. In this table, the x_i are uniformly distributed between 0 and 1.

Table 2 reports the results when the x_i are drawn from a normal distribution with a mean of 0 and a standard deviation of 10. Once again, the full sample produces the lowest standard errors of the estimates, and the outer deciles the second lowest. An interesting observation is that, when compared with the results of Table 1, choice D's relative advantage over the other limited-data options is larger when considering $\hat{\beta}$ (because the x_i have greater dispersion), but smaller when considering $\hat{\alpha}$. For example, in Table 1, the average standard error for choice A's $\hat{\beta}$ is, at 5.0121, about 58% larger than choice D's 3.1708. However, in Table 2, this percentage has increased to over 85% (= $\frac{.1442}{.0777} - 1$) because Table 2 features a greater dispersion in x_i.

Table 2. Regression Statistics for Four Different Methods for Choosing a Subset of the Population Independent-Variable Observations Normally Distributed with Mean 0 and Standard Deviation 10

Selection of X from simulated dataset sorted on X	average independent variable sample standard deviation	average $\hat{\beta}$	average standard error for $\hat{\beta}$	average $\hat{\alpha}$	average standard error for $\hat{\alpha}$
Full Dataset	10.0046	12.0016	0.0624	0.5203	0.6392
Choice A: Random	10.0038	11.9948	0.1442	0.5396	1.4327
Choice B: First 5 from each decile	10.5040	12.0056	0.1385	0.5752	1.4729
Choice C: Middle two deciles	1.4933	11.9818	1.0049	.5670	1.6084
Choice D: Top and bottom decile	18.1275	11.9996	0.0777	0.5451	1.4014

Summary statistics for different methods of selecting a sample of size 50 (from a total set of 250 ordered pairs). The regression line is $y_i = 0.5 + 12x_i + \varepsilon_i$, where ε_i is normally distributed with mean 0 and standard deviation 10. In this table, the x_i are normally distributed with mean 0 and standard deviation 10.

However, while choice D still dominates choice A as far as the standard error of $\hat{\alpha}$ is concerned, its advantage shrinks from Table 1's 38% ($= \frac{2.9114}{2.1107} - 1$) to only a bit over 2% ($\frac{1.4327}{1.4014} - 1$) in Table 2. The reason for this can be seen by examining the formulas for the variances of $\hat{\beta}$ and $\hat{\alpha}$. The expression for the variance of $\hat{\beta}$ is $\frac{\sigma^2}{\sum(X_i - \bar{X})^2}$ (e.g., Kmenta 1997, p. 224), which is inversely proportional to a denominator that measures the dispersion of the independent variable within the sample. The corresponding expression for the variance of $\hat{\alpha}$ is $\sigma^2(\frac{1}{N} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2})$. Here, ceteris paribus, increases in the dispersion of observations of the independent variable will decrease the variance of $\hat{\alpha}$, but the standard error of $\hat{\alpha}$ is also dependent on sample size and sample mean. For example, if the sample means all happened to be 0, the choices A–D would produce identical estimates of α, specifically, $\hat{\alpha} = \bar{y}_i$. Because the sample sizes of only 50 for the limited-data choices of Tables 1 and 2 are fairly small, and especially because the sample means, $\bar{X}$, are also fairly small, the relative benefit of the more dispersed samples from choice D is mitigated as far as estimating α is concerned. It is still best of the four limited-data choices, but by a narrower margin than in Table 1.

Things are quite different in Table 3, which reports the same simulations as Table 2, except this time the mean of the distribution from which the x_i are drawn has been increased to 100. This ensures a large value for $\bar{X}$, which in turn highlights the benefits of in-sample dispersion of the independent variable. The first three columns of Tables 2 and 3 are nearly identical—changing the mean of the distribution from which the x_i are drawn does not, except for a small sampling error, change the sample standard deviation of the x_i, the estimate $\hat{\beta}$, or the standard error of $\hat{\beta}$.

Table 3. Regression Statistics for Four Different Methods for Choosing a Subset of the Population Independent-Variable Observations Normally Distributed with Mean 100 and Standard Deviation 10

Selection of X from simulated dataset sorted on X	average independent variable sample standard deviation	average $\hat{\beta}$	average standard error for $\hat{\beta}$	average $\hat{\alpha}$	average standard error for $\hat{\alpha}$
Full Dataset	9.9951	11.9994	0.0631	0.5750	6.3694
Choice A: Random	9.9887	11.9967	0.1455	0.8076	14.6014
Choice B: First 5 from each decile	10.5014	11.9964	0.1381	0.8898	13.6284
Choice C: Middle two deciles	1.4904	12.0108	1.0168	-0.6221	101.6553
Choice D: Top and bottom decile	18.0973	11.9989	0.0779	0.6098	7.9383

Summary statistics for different methods of selecting a sample of size 50 (from a total set of 250 ordered pairs). The regression line is $y_i = 0.5 + 12x_i + \varepsilon_i$, where ε_i is normally distributed with mean 0 and standard deviation 10. In this table, the x_i are normally distributed with mean 100 and standard deviation 10.

However, $\hat{\alpha}$ is quite another matter. Instead of choices A, B, and D having similar standard errors for $\hat{\alpha}$, as they did in Table 2, now $\hat{\alpha}$'s standard error from choice D is about 25% larger than that for the full sample, while that of choices A and B is more than double that of the full sample. We stress that for purposes of estimating $\hat{\beta}$ and $\hat{\alpha}$ in both Tables 2 and 3, Choice D is best, but that the low mean of x_i in Table 2 makes its advantage small.

We emphasize that these simulations are in no way intended to suggest that observations near the mean should be discarded, or that only the outer deciles are informative. Among other things, considering only the outer deciles might well cause a researcher to conclude the relation is linear, when in fact Ramsey's (1969) RESET test or simply visual observation of the data would reveal a likely non-linear relationship. Rather, the simulations show that, if a researcher has a solid theory or another good reason to expect a linear dependence of Y on X, and if the most extreme observations are not so extreme as to imply an error (for example, a reported IQ of 900 is almost certainly an error, most likely an IQ of 90 with an extra "0" added), then the extreme observations are most valuable, just as the quote from Kmenta (1997) claims. Therefore, they are not observations that should be dismissed, but rather observations that improve the quality of the regression parameter estimates.

When there is good cause to believe the dependent variable is linear in X (in this case, we know it is because we generated the independent variable that way), then the best test will be a linear regression of Y on X, even if the variance of X is relatively small. When we are not sure of the correct form of the model, things become tricky. It is a bad idea to try alternatives and choose the one that appears to produce the best fit. Studenmund (2006, p. 212), for example, states "The choice of a functional form almost always should be based on the underlying theory..." rather than the data themselves. Instead of expending the effort to test alternatives, that effort would be better spent trying to determine *a priori* which choice of dependent variable best fits the theory being tested.

The next section suggests a take-home assignment that will lead students to learn the benefits of extreme observations for themselves. It is difficult to specify for exactly what courses the assignment might be appropriate, as this will vary substantially school by school and even class by class. If the students are required to do a take-home regression assignment, anyway, then we estimate the marginal in-class time required is under half an hour, and the extra out-of-class time is no more than an hour or two. This is likely not suitable in a beginning statistics class where the entire topic of linear regression may get only a week or two of class time; however, it is likely beneficial in any econometrics or other class that focuses on linear regression, or any subsequent class that applies linear regressions and has a linear-regression prerequisite. We made the assignment in a masters level finance class, not in a statistics, linear regression, or econometrics class, and we were limited to some extent in what we could require of the students. However, if we were making the assignment in a class for which linear regressions are a primary focus, we would also ask students to find and compare the standard errors of the slope coefficients for each of the full sample and the designated subsets of the full sample. After all, of the set of all unbiased linear estimates, it is the one with the lowest variance that is characterized as "best."

3. Student Quiz, Assignment, and Results

We tested how accurately students would speculate the ranking of these outcomes *a priori*, gave them an assignment that tested their conjectures, and asked for their explanations of why they believed they observed the results they did. The course in which this was implemented was a masters' level introductory corporate finance section.[4] A statistics course in which regressions

[4] Actually, prior to making the assignment in the masters' class, we did a trial run in an undergraduate investments class, but found that a sufficient number of students were confused at the concept of sorting data that the results were unusable. It is for this reason that the current assignment (Appendix B) spells out in detail how to sort an array by a specific column. Nevertheless, students still make such errors, as footnote 5 indicates. Because we changed the assignment from the trial run before spring 2013, and again from the spring, 2013 assignment to the spring, 2014 assignment, it is perhaps best to think of our results as those of multiple pilot studies.

are covered was a prerequisite, and in addition most students were also concurrently enrolled in another statistics class that studied regressions more extensively. We believe this assignment would be suitable in any class devoted primarily to linear regressions, or any class that subsequently analyzes and uses linear regressions; indeed, our results indicate it clarifies and reinforces a principle that is apparently often misunderstood by students. Depending on the class, some instructors may find it a suitable adjunct to the understanding of regressions in any but the most basic classes covering them.

Late in the spring, 2013 class, the Capital Asset Pricing Model (CAPM) was discussed. Given a set of assumptions, this model proves that stock I's expected return, $E[R_I]$, is linearly dependent on that of the market, $E[R_M]$, and a riskless rate of return, R_F, usually approximated by the T-Bill rate. Basically, the stock has two components of risk, systematic risk that is related to the market and unsystematic risk that is unique to the stock. Because the unsystematic risk is diversified away as the investor adds more stock to a portfolio, this type of risk is not rewarded with a higher expected return. However, risk that is related to the market as a whole cannot be diversified away, and so if a stock has greater sensitivity to the stock market, it earns a higher expected return. Specifically, the CAPM shows that this systematic risk is equal to

$$\beta = \frac{\sigma_{IM}}{\sigma_M^2} = \frac{covariance\ between\ the\ independent\ and\ the\ dependent\ variable}{variance\ of\ the\ independent\ variable}$$

and theorizes that $E[R_I] - R_F = \beta\{E[R_M] - R_F\}$, more commonly expressed as $E[R_I] = R_F + \beta\{E[R_M] - R_F\}$. The CAPM's β is identical to the slope of a regression line. Thus a stock's β is typically estimated from a linear regression of a stock's excess returns $(R_I - R_F)$ on those of the market $(R_M - R_F)$. Examples of the reporting of a stock's β as an integral piece of information about the stock can be seen in Exhibit A. A more thorough discussion of the CAPM can be found in virtually any corporate finance textbook, e.g., Brealey, Myers, and Allen (2014). After we discussed the fact that the parameter β was typically estimated by regression, we gave the students an unannounced, ungraded quiz asking which of the four options mentioned in the last section would produce the best estimate of β. (The quiz appears in Appendix A. It took only about 5 minutes to administer.) While the students were told the quiz would not be graded, they were required to put their names on them. Of those students in attendance, only about 10% (3 out of 28) thought the top and bottom deciles would produce the best estimate, while one student thought the middle two deciles would. The remaining 24 students were equally divided, with 12 thinking a pure random selection was best and 12 favoring the first five ordered pairs in each decile. After the quiz, several students asked for the correct answer, but we told them part of their subsequent written assignment was to discern this on their own.

Next, students were given an out-of-class assignment that, by requiring the use of real data to put their conjectures to the test, is consistent with GAISE recommendations 2 and 3. Each student was assigned one of six companies and asked to estimate its β based on weekly returns for the last 250 weeks of data. This part of the assignment is very similar to that given in Keller and Warrack (2003, pp. 627–629). Note that in this spring, 2013 class, students were not asked to compare the intercept term α's values resulting from the four choices. However, this was incorporated into the spring, 2014 assignment, and we report those results later in the paper. We selected the companies randomly and did not screen them to ensure they would produce the results we anticipated. Several students asked in class if it made a difference whether they estimated β from the market model (which uses raw returns, i.e., $Y = \alpha + \beta X$, where $Y = R_I$ and $X = R_M$) as per Peterson (2005), or from the CAPM itself (which uses excess returns, i.e., $Y = \alpha + \beta X$, where $Y = R_I - R_F$ and $X = R_M - R_F$). In response we used Excel to generate an example in which R_F was constant throughout the sample period and demonstrated in this case that as far as estimating β was concerned, there was no difference which model was used. When asked why, students didn't take long to realize that subtracting a constant R_F from either the dependent or independent variable essentially amounted to a translation of that variable along its axis, and thus did not affect the slope. We then asked what they thought would happen if R_F were not constant throughout the sample period, but had an extremely low variance when compared with the stock and market variances, and they correctly surmised it would not make a significant difference. At that point we told them they could complete their assignment using either raw returns (R_I and R_M) or excess returns ($R_I – R_F$ and $R_M – R_F$). Almost all elected to use raw returns.

After estimating β based on all 250 weeks of data, students were again asked as part of their written assignment which of the four specified subsets of 50 ordered pairs of data they thought would produce an estimate of β closest to the estimate based on all 250 observations. Twenty-one students both took the in-class quiz and explicitly answered this question in their submitted paper. Five of these 21 students changed their speculation (one in-class guess of A was changed to B for the submitted guess, two in-class guesses of A became submitted guesses of D, one C was changed to B, and one B was changed to D). It is possible that, given the extra time, three of the students decided D was best; however, given all three students who changed their answer to D also found D to be the best of the four options suggests they may have waited until they actually obtained their results, and then gone back and formed their "speculation" after the fact. Finally, they were asked to test all four options and explain their results (the complete assignment appears as Appendix B).

Of the 32 students submitting assignments, 25 found that the most extreme deciles produced the estimate of beta closest to the one they found from all 250 weeks of data.[5] Many of these students expressed surprise at the result, but offered good explanations of why it occurred. Specifically, the following were among the submitted observations and explanations:

1. "The one I thought would be the farthest resulted to be the closest estimate, which was D with a value of 2.16 (just 0.04 over)...Sample D might be the closest because it represents the maximum variation in the dataset by taking the most negative and most positive values in the sample."
2. "However, much to my surprise, I found that choosing the data points from the first and tenth deciles gave us a beta calculation that was closest to what was found when using all 250 data points!"
3. "1st and 10th Deciles—contrary to what I originally thought, the 1st and 10th deciles make sense to be the closest to the beta of the entire dataset. Beta is the responsiveness of a stock to movements in the market portfolio. By choosing the 1st and 10th deciles, I

[5] Of the other seven, two ranked the data by the dependent variable (stock return), not by the independent variable (market return) as the assignment specified. Four others made errors in Excel; one of these found that choice D was a close second to choice B for his dataset, and similarly one found D to be a close second to choice A. Finally, it was difficult to interpret from the seventh submission exactly what was done; it appears that the β from choice D was virtually identical to that from using all 250 pairs, but that the student thought a different question was being asked.

captured the most movement in the market, and thus captured the 'responsiveness' of the IBM stock to the movement of the market."

4. "Based on the beta calculations, I was way off base in thinking that the first and tenth deciles would be the furthest from the original beta calculation. These two actually provided a near identical beta from what was originally calculated."
5. "It is surprising to learn that from the four choices listed above the most accurate was the 1st & 10th decile and the least accurate was the 5th & 6th decile. I had assumed that using data points closer to the average would give us a more accurate calculation of Beta because of the relative closeness of these data points near the center. I believe that the fact that the 1st & 10th decile gave us a more accurate calculation of Beta tells us the data points that the most extreme data points in the first and 10th decile player [sic] a stronger role in determining beta than the ones near the mean. The fact that the 5th & 6th decile were extremely off from our calculated beta confirms this."
6. "I was surprised to find that this [B] was incorrect. Through statistical analysis, I was able to determine that the most accurate beta produced by 50 ordered pairs was found when using the information from the first and tenth decile."
7. "Out of my expectation, the accuracy of the four choices are D,A,B,C…The rationale behind this could be that data sets from first and tenth deciles largely determine the trend of the regression line, whereas data sets from fifth and sixth deciles are all squeezed together, which makes it hard to draw a trend line."

Being adherents of active learning, with its greater emphasis on student exercises than instructor lectures (e.g., Cobb 1992), we were quite gratified to see the responses above without any prompting on our part. However, not all students bought in to the idea that D was the best of the four alternatives. For example, one student who found D produced the best result seemed to attribute this to chance: "choosing the first and last groups of population…can result in time-period bias if a stock…[earns an]… abnormal return or loss in unstable market..." One other student also expressed similar skepticism.

At the request of a reviewer, the assignment was repeated in another section of the class two semesters later, with three major changes. First, the concept of "best" estimate was made more explicit by the insertion of the following sentence into the in-class quiz and the take-home assignment: "[By 'best' we mean most likely to be closest to the value you would get if you used the entire set of 250 ordered pairs.]" The results of the in-class quiz were fairly similar: of those students in attendance, just under 14% (5 out of 36) thought the top and bottom deciles would produce the best estimate, while three students thought the middle two deciles would. The remaining 24 students were nearly equally divided, with 10 thinking a pure random selection was best and 14 favoring the first five ordered pairs in each decile.

Second, while in the previous section we allowed students to assume the riskless rate, R_F, was zero, this time we required that they use the actual rate. The difference is very small—periodic weekly T-Bill rates averaged only 0.004% between June 15, 2009 and March 31, 2014, and the range of weekly rates was only 0.0081%, both of which are very small relative to individual stock returns and the S&P 500. For example, Apple's average weekly return was 0.6423%, with a range of 26.1425%, while the S&P 500 averaged a weekly return of 0.3106%, with a range of

14.5775%. Because R_F is so small and stable relative to individual stock and S&P returns, it makes little practical difference whether we use raw returns or excess returns.

The third major change was that this time, students were not only asked to estimate the stock's slope, β, but also its intercept, α. In the context of the Capital Asset Pricing Model, the α term is often called Jensen's α and is seen as a measure of whether a stock did better or worse than it should have given the market's movements (e.g., Bodie, Kane, and Marcus (2013, p. 603)). A positive Jensen's α is seen as a sign of superior performance, while a negative α is seen as an indication that the stock underperformed its expectations. Unfortunately, average weekly stock and stock market returns are very small compared with their respective standard deviations. As we found in Table 2, this implies that the standard errors of choice D's estimates of α, while still lowest of the four, are not so much lower that all students found it to be the best choice. Indeed, a summary of what students found to be the "best" estimate is shown in Table 4.

Table 4. Frequency of Estimates found to be closest to the Full-Dataset Estimates

Limited-Data Choice	Best Estimate of β	Best Estimate of α	Worst Estimate of β	Worst Estimate of α
A (Random Selection)	3	6	4	9
B (First Five in Each Decile)	5	6	2	6
C (5th and 6th Deciles)	0	6	25	12
D (1st and 10th Deciles)	24	14	1	5

This table reports the frequency with which students found the limited-data estimate to be closest (for best) or farthest away (for worst) to the full-sample values. While for both $\hat{\beta}$ and $\hat{\alpha}$, Choice D is characterized as best more frequently than any other choice, and is characterized as worst less frequently than any other choice, it is considerably less advantageous when estimating $\hat{\alpha}$ than when estimating $\hat{\beta}$ for the reasons pointed out in Section II. This gap could be mitigated if (a) sample size were larger and especially (b) $\bar{X}$ is large relative to σ_X. Unfortunately, the nature of the assignment dictated that (b) did not apply.

As the table shows, choice D was found by 75% of the students to be best for estimating β, but while choice D was the plurality choice for providing the best estimate of α (with 14 finding it best, as compared with the other competitors, which were tied with six each), only about 44% of the students found it best. The reason for this is identified in Section II, namely, that by the nature of the assignment, $\bar{X}$ is not sufficiently large relative to σ_X. Moreover, in contrast to the previous semester's two students who thought choice D's better estimate of β was just due to luck, this time four student's did. Their comments included:

1. Although the observed ranking proved slightly different then my predictions I still stand by my hypotheses. I still believe that the random sample is the most accurate.
2. While I maintain that my logic in ranking the choices a-d [with D ranked third] was sound, it shows that different sets of data behave differently.

3. The other two remaining data set purchase options do seem like the best alternatives but contrary to the results I would suggest for investors to use the first five values from each decile over the 1st and 10th decile data.
4. This [Choice D] is the best option based on beta….I don't think this option will always provide the best outcome though.

All four of these students found D's estimate of β to be best, but not its estimate of α.

However, there were still a number of students who concluded D was best overall:

1. Choice (d) by using the extreme datapoints (farthest from the mean) was more suitable since it provided the true representation of the sample and hence the best fit for the regression line as shown above.
2. Moreover, from Figure 2, the "first and tenth decile" has the closest Jensen's Alpha to the original data. In conclusion, the best choice is the first and tenth deciles.
3. … by choosing 1st and 10th deciles, D maximizes the variance of S&P 500 excess returns, which is the closest to the variance of 250 data points. Thus, β from Choice D is the closest to that obtained from the entire data set; β from Choice C is the farthest one.
4. … that's why when we chose the first and last deciles, we got the best estimation of beta, because it maximizes the variance of the market excess return.
5. For example, assume that there are two points, (1,1) and (2,2). The slope of the straight line passing through these two points is 1. Then, (2,2) moves two steps towards X-axis to (2,0), making the slope -1. Assume that there are two other points (1,1) and (100,100). The slope of the straight line passing through these two points is also 1. Then the (100,100) moves two steps towards X-axis to (100,98). Then the slope becomes 0.98, not too much different from 1. Therefore, the more distant the two sets of points, the more they reflect the whole trend than the movement of individual points.
6. … because the first deciles and last deciles include the most extreme numbers, which reflect the most volatility of the data.
7. To my surprise, the beta for choice (d) is almost equal the beta calculated from original 250 data points, and the Jensen's α is also the closest result. I guess the reason is that choice (d) covers the entire spread of returns, which are enough to represent the fluctuations of stock.

An example for Apple (ticker symbol AAPL) is provided in the linked spreadsheet, and summarized in Table 5. For the full sample period (June 15, 2009 through March 31, 2014), Apple's $\hat{\beta}$ is 1.0460 and its $\hat{\alpha}$ is 0.3177%. Of the four limited-data choices, the first and tenth deciles provide the value for $\hat{\beta}$ (1.0714) that is closest to the full sample estimate, but (for reasons pointed out in the discussion regarding Table 3) it does not feature the closest value for $\hat{\alpha}$, its 0.3170% distance from the full-sample value being nosed out by the first five datapoints in each decile's distance of 0.3070%.

Table 5. Summary of Regression Parameter Estimates for Apple (AAPL)

	Estimated $\hat{\beta}$	Estimated $\hat{\alpha}$	Absolute Deviation from Full-Sample $\hat{\beta}$	Absolute Deviation from Full-Sample $\hat{\alpha}$	Standard Error of $\hat{\beta}$	Standard Error of $\hat{\alpha}$
Full Sample	1.0460	0.3177%	--	--	0.09844	0.21625%
Random selection of 50 (Choice A)	0.7695	-0.3819%	0.2765	0.6995%	0.26343	0.49281%
First five each decile (Choice B)	0.9630	0.0107%	0.0831	0.3070%	0.20357	0.48825%
5th and 6th deciles (Choice C)	-0.7928	0.7270%	1.8388	0.4093%	1.69291	0.76636%
1st and 10th deciles (Choice D)	1.0714	0.6347%	0.0254	0.3170%	0.08975	0.37495%

Consistent with all four previous tables, Choice D's estimate of $\hat{\beta}$ (1.0714) is closest to the full-sample estimate of 1.0460. While Choice D's estimate of $\hat{\alpha}$ (0.6347%) was not closest to the full-sample estimate of 0.3177%, it was edged out only by one-hundredth of a percent. While Table 2 suggests Choice D will generally be best when the independent variable has a mean near zero, its average advantage over Choices A and B is very small. If there is flexibility in the assignment, it is better to make the point that greater dispersion improves estimates by selecting a dataset whose independent variable has a mean that is several standard deviations from 0, as in Table 3.

As mentioned earlier, what makes a linear unbiased estimate the "best" one is that it has the lowest standard error. Thus comparing the standard errors of the four limited-choice estimates $\hat{\beta}$ should give us a good idea how reliable the estimates are. Consistent with Tables 1–3, this example with Apple finds that of Choices A–D, D (1st and 10th deciles) has the lowest standard error for $\hat{\beta}$. This provides further evidence that its being the estimate closest to the full-sample estimate is not just a random accident.

We note in passing that while in this case the standard error for Choice D is lower than that for the full sample, this is not common (as suggested by Tables 1–3). Nevertheless, that D's estimate has a lower standard error than that of the full sample can occur, as it has here, and an instructor needs to be prepared to explain why. Probably the best way to do that is to give different students different stocks, and summarize in class what everyone found. As Tables 1–3 indicate, while it is common that D will have a lower standard error than A–C, it is not common that it will have a lower standard error than the full sample. Providing a summary that makes clear out this example with Apple (with the standard error of D lower than that of the full sample) is not the norm.

Overall, we believe that estimates of β and α may both matter for many purposes. Tables 1–4 all show that when the data are restricted, the top and bottom deciles are the best choice on average. Nevertheless, a student who just sees one test is likely to miss this point if his or her estimates produced mixed results for β and α. Again, while a solid majority of students will find Choice D produces the best estimate of β, only a plurality of students will find it produces the best estimate of α. Our experience is that the assignment, while effective when the intercept α is included, is considerably more effective when it focuses only on β. Moreover, when the assignment focuses only on β, most students will discover the principle involved on their own; when it includes an analysis of α, additional classroom discussion is required. If an instructor wants to emphasize the intercept, we recommend a pure simulation assignment along the lines of Tables 2 and 3. Once again, however, we emphasize that the exercise is not suggesting that use x values only at the extreme deciles; in practice it is never a good idea to discard data.

Similarly, the top and bottom deciles will generally produce an estimate of β with a larger standard error than that for the full sample (as in Tables 1–3). However, occasionally the top and bottom standard error for beta will be lower than that of the full sample (as it is in Table 5's Apple example). Once again, if several different datasets were assigned across the classroom, it is easier to show all the results and explain one result as an aberration, so that some students don't leave with the impression that it is better to discard all but the most extreme observations to get lower standard errors.

4. Conclusions

Some econometrics textbooks point out that greater dispersion in an independent variable produces better estimates, but many students appear to remain suspicious of extreme observations. We found that an assignment using actual data and comparing estimates from various subsets of the data convinced a large number of students that the inclusion extreme observations can improve the quality of estimates of slope. However, this does not mean that non-extreme observations should be ignored; rather, the only implication is that extreme observations should be welcomed, not shunned.

APPENDIX A
IN-CLASS QUIZ

Suppose that you have a solid reason to believe that the relationship between X_i and Y_i is linear,[6] and you would ideally like to buy a vendor's complete dataset of 250 ordered pairs (X_i, Y_i) for a regression analysis. However, the dataset is very expensive, and you can afford only 50 (20%) of the ordered pairs. The vendor can sort the pairs (X_i, Y_i) in increasing order of X_i and then will sell you any two deciles* if you want. Your goal is to get the "best" estimates of α and β in the linear regression $Y_i = \alpha + \beta X_i$. *[By "best" we mean most likely to be closest to the value you would get if you used the entire set of 250 ordered pairs.]*[7] Which of the following do you speculate is true?

a. It is best to let the vendor choose 50 of the ordered pairs randomly (because random selection produces unbiased tests).
b. It is best to choose the first 5 ordered pairs in each decile (because this will give us representation across the entire spectrum of X_i).
c. It is best to choose the fifth and sixth deciles (because they are closest to the median, and thus are most representative of "typical" values of X).
d. It is best to choose the first and tenth deciles (because these are the most extreme, and thus maximize the variance of X).

* A decile is a tenth of the data, i.e., the first decile of 250 ordered pairs contains the ordered pairs (X_i, Y_i) with the lowest 25 values of X_i, the second decile contains the ordered pairs with the 26^{th}—50^{th} lowest values of X_i,…and the 10^{th} (=last) decile contains all the ordered pairs with the 25 largest values of X_i.

[6] This first sentence was not part of the actual text of either the spring, 2013 or the spring, 2014 assignment, but was added so that students would not think examining *only* extreme observations was a good idea.
[7] This sentence in italics was omitted in the spring, 2013 assignment, but included in the spring, 2014 assignment.

Appendix B
Actual Spring, 2014 Assignment

I—Guidelines for FI 625 Written Case Analysis (150 points)

The following provides a description of the format for the document you will hand in. No more than four written pages plus a cover page and attachments. Keep the attachments to a reasonable number but be sure to include the Excel document containing the data series and regression results. Use Times New Roman 11-point font and 1.15 line spacing (this document is done with those specifications). Be sure to provide a title, the author, FI 625 and the date on the cover page. Please note that your grammar and format will be graded as well as the findings themselves. This should be a professional-looking document that could be handed to a potential employer. **This assignment is due before class Wednesday, April 30, 2014; electronic submissions only. Please note that this is also the day of the second exam. Please pace yourself and try to get the project done early so you do not have too many things to do that last weekend before the exam.**

INTRODUCTION
Describe the assignment in one paragraph thus providing the reader a synopsis of what he/she is about to read. Identify the company you have been assigned and its ticker symbol. Provide your results in one to two sentences.

DETERMINATION OF BETA
Provide one to two paragraphs describing the steps taken to determine beta. Compare your assigned company's computed beta to that from one other source (e.g., Bloomberg, FactSet, Yahoo! Finance). Explain the most likely cause(s) of the disparity between your beta and that cited.

DETERMINATION OF COST OF EQUITY
Describe the Capital Asset Pricing Model (CAPM) and solve for the required rate of equity for your company. Indicate your choice/derivation of the risk-free rate and the market risk premium.

DETERMINATION OF COST OF EQUITY
Suppose that, instead of being free, the data were expensive, and instead of buying all 250 ordered pairs of risk premia, you could only afford 50 pairs. Examine the possibilities suggested on the next page and speculate which is generally best, and why.

CONCLUSION
Conclude with what you have done (in the order of accomplishment) and the results for your company.

II—CALCULATING BETA COEFFICIENTS

A. Download the last 250 weeks of adjusted weekly stock prices for the company you have been assigned from Yahoo Finance (or any other reliable source) into an Excel spreadsheet and calculate a rate of return series. Do the same for the S&P 500 stock index (^GSPC in Yahoo). Finally, weekly T-Bill rates over the same period can be found at the Federal Reserve Economic Database (FRED) site at http://research.stlouisfed.org/fred2/. You may need to register at FRED, but once you do so, you can search for "weekly T-Bill rates" to get what you need. Note that the site gives you annual rates, and you will need to convert these to weekly rates. Note that the Yahoo site will give you weekly prices as of Monday, while the FRED site will give you rates as of Monday. Theoretically, these dates should match up exactly, but given T-Bill rates are pretty stable, you may assume the Friday rates you observe are the same on Monday.

Follow the steps in the associated document *Beta Estimation Computation Example* to download the data. Perform the following calculations and analyses (or use Excel's regression toolset to provide the equivalent):

1. Calculate the return series (adjusted for dividends and stock splits, if any) for the most recent 251 weeks of weekly prices for your stock and the S&P 500, and use these to calculate 250 weekly returns.
2. For each of your stock and the S&P Index, calculate the weekly excess return = amount by which each series exceeds the T-Bill return during that week.
3. Calculate the mean and standard deviation for each stock's return series and the index (use the Excel functions AVERAGE and STDEV). Note that these numbers may seem small because they are for weekly data instead of the more commonly seen annual data.
4. Calculate the correlation between your company's returns and the S&P 500 (use the Excel function CORREL).
5. Calculate the beta coefficient for your stock (use the Excel function SLOPE or regression function found in the Excel analysis toolpak). Obtain a beta estimate from a professional source (FactSet, Value Line, Bloomberg, etc.) and compare this to your estimate. Why might they differ?
6. Calculate Jensen's alpha for your stock over the five-year period in question. How do you interpret this number?
7. Using an appropriate estimate for the market return (consult Chapter 10 of your textbook for historical data), estimate the cost of equity for your stock using the Capital Asset Pricing Model. Be careful—your estimates are for weekly returns, but the cost of equity is typically expressed in annual terms, and so you will need to make the appropriate adjustment.

B. In part A, the data required to find β were free, but suppose they were very costly so that you could only afford to buy data for only a fifth of what is available (i.e., only 50 of the 250 weeks). The vendor has ranked the data by the independent variable (S&P 500 returns) and offers to let you choose

I. 50 random weeks of data
II. Any two deciles (a decile is one-tenth of the data, so the first decile would be the data corresponding to the lowest 25 S&P 500 returns, the second decile would be the data corresponding to the 26th lowest through 50nd lowest S&P 500 returns, and similarly

to the tenth decile, which would be the data corresponding to the 25 highest S&P 500 returns.

III. The first five datapoints in each decile, e.g., the data with lowest through fifth lowest S&P 500 returns, the 26th-30th lowest, etc., all the way up to the 226th-230th lowest.

Given these choices, and given you are certain the relationship between stock returns and market returns is linear,[8] we are considering each of the following data purchases, with the reasons in parentheses:

a. Choose 50 of the datapoints randomly (because random selection ensures unbiased tests).
b. Choose the first five datapoints in each decile (because this will give us representation across the entire spectrum of the independent variable, S&P 500 returns).
c. Choose the fifth and sixth deciles (because these are closest to the median, and thus are most representative of "typical" values of S&P 500 return).
d. Choose the first and tenth deciles (because these are the most extreme, and thus maximize the variance of S&P 500 returns).

1. Rank the four choices a—d from the one you suspect will produce the values for β and Jensen's α[9] that are closest to what you obtained using the entire set of 250 datapoints to the one you believe will be farthest away.

Before you continue, it is important to copy the returns and then use "Paste Special" to "Paste Values."

2. The easiest way to select 50 random datapoints is to add an extra column "Random," and enter "=rand()" in each cell of the column [Excel's "rand" function picks a randomly selected number between 0 and 1]. Copy this column, and then paste in the same position using "Paste Special and "Paste Values." Now sort the data from lowest to highest values of Random, and use the first 50 datapoints to estimate α. (Be sure that you keep the same stock returns associated with the same S&P 500 returns. For example, the lowest S&P 500 return is –7.19% for the week ending August 1, 2011; make sure this is paired with your stock's returns for the week ending August 1, 2011).

Now sort your entire dataset from lowest to highest S&P 500 return, once again being careful to make sure the S&P 500 returns remain paired with stock returns for the same days.

3. Next, test your speculation about the rankings of choices a—d above by estimating β and Jensen's α for each of the remaining three possible data purchases. [Even if your speculations in 1. and 2. were inconsistent with what you found, please do not revise those answers.]

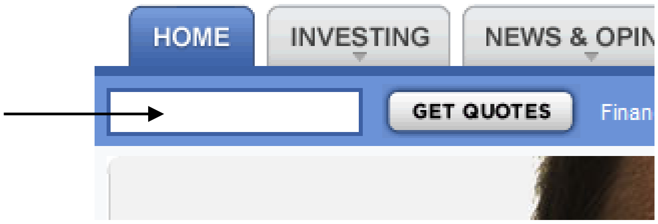
4. Were your speculations about the best choice from a—d consistent with what you actually found when you used all 250 datapoints in part A? Briefly discuss why you think that is. [If you speculated correctly, why do you think your choice is best? If your speculation was incorrect, do you think there is a reason, or that it was just due to the random nature of samples?]

[8] The expression "and given you are certain the relationship between stock returns and market returns is linear" was not part of either the spring, 2013 or the spring, 2014 assignment, but was added so that students would understand that, in general, focusing *only* on extreme observations is not an appropriate method.

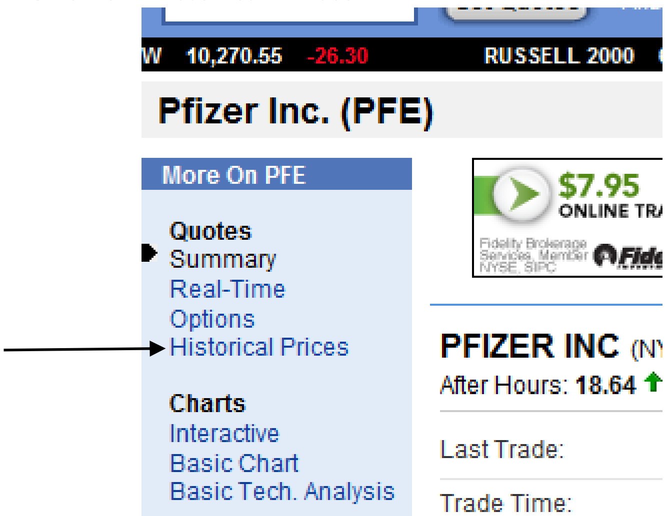
[9] The expression "and Jensen's alpha" was included only in the spring, 2014 assignment, not the spring 2013 assignment or a previous trial run.

III—BETA ESTIMATION COMPUTATION EXAMPLE

Go to Yahoo Finance and enter a company name or ticker in the Get Quotes box. Here, I am doing Pfizer (PFE).



Click on Historical Prices



In this example, we download five years (February 11, 2008 through November 26, 2012) of adjusted weekly stock prices from Yahoo Finance onto an Excel spreadsheet and calculate a rate of return series.

First | Previous | Next | Last

Prices

Date	Open	High	Low	Close	Avg Vol	Adj Close*
Nov 26, 2012	24.34	24.52	24.31	24.49	35,772,100	24.49
Nov 19, 2012	24.03	24.53	23.87	24.53	29,033,900	24.53
Nov 12, 2012	24.15	24.38	23.55	23.86	30,065,800	23.86
Nov 7, 2012	0.22 Dividend					
Nov 5, 2012	24.42	24.96	23.99	24.17	33,559,800	24.17
Oct 31, 2012	25.65	25.75	24.39	24.55	41,278,500	24.33
Oct 22, 2012	25.68	25.74	25.02	25.43	25,346,600	25.20

This shows the last few weeks of the series. Click on "download to spreadsheet" at the bottom of the page to get (for these same last few weeks)

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Volume	Adj Close
2	11/26/2012	24.34	24.52	24.31	24.49	35772100	24.49
3	11/19/2012	24.03	24.53	23.87	24.53	29033900	24.53
4	11/12/2012	24.15	24.38	23.55	23.86	30065800	23.86
5	11/5/2012	24.42	24.96	23.99	24.17	33559800	24.17
6	10/31/2012	25.65	25.75	24.39	24.55	41278500	24.33
7	10/22/2012	25.68	25.74	25.02	25.43	25346600	25.2
8	10/15/2012	25.17	26.09	25.14	25.76	29192800	25.53
9	10/8/2012	25.46	25.52	25	25.12	21475700	24.9
10	10/1/2012	24.8	25.59	24.8	25.52	25046100	25.29
11	9/24/2012	24.45	25.15	24.45	24.85	27623900	24.63
12	9/17/2012	23.68	24.66	23.68	24.52	30724800	24.3
13	9/10/2012	24.2	24.26	23.63	23.8	29427300	23.59

The only columns you really need are the date and adjusted close ("Adj Close") price series. Calculate weekly rates of returns using Return = $(P_1 - P_0)/P_0$. You may delete the other columns:

	A	B	C
1	Date	Adj Close	Return
2	11/26/2012	24.49	-0.00163
3	11/19/2012	24.53	0.02808
4	11/12/2012	23.86	-0.01283
5	11/5/2012	24.17	-0.00658
6	10/31/2012	24.33	-0.03452
7	10/22/2012	25.2	-0.01293
8	10/15/2012	25.53	0.025301
9	10/8/2012	24.9	-0.01542
10	10/1/2012	25.29	0.026797

Get the return series for the S&P 500 index for the same time period. Enter S&P in the "Get Historical Prices for" box and click GO.

S&P 500 (^GSPC) - SNP

1,416.18 0.23(0.02%) Nov 30

Historical Prices

Get Historical Prices f

Set Date Range

Start Date: Feb 8 2008 Eg. Jan 1, 2010

End Date: Nov 26 2012

Daily
Weekly
Monthly
Dividends Only

Get Prices

First | Previous | Next | Last

Prices

Date	Open	High	Low	Close	Avg Vol	Adj Close*
Nov 26, 2012	1,409.15	1,409.15	1,397.68	1,406.29	2,948,960,000	1,406.29
Nov 19, 2012	1,359.88	1,409.16	1,359.88	1,409.15	2,688,502,500	1,409.15
Nov 12, 2012	1,379.86	1,388.81	1,343.35	1,359.88	3,621,476,000	1,359.88
Nov 5, 2012	1,414.02	1,433.38	1,373.03	1,379.85	3,602,274,000	1,379.85
Oct 31, 2012	1,410.99	1,434.27	1,405.95	1,414.20	3,746,493,300	1,414.20
Oct 22, 2012	1,433.21	1,435.46	1,403.28	1,411.94	3,397,482,000	1,411.94
Oct 15, 2012	1,428.75	1,464.02	1,427.24	1,433.19	3,692,620,000	1,433.19
Oct 8, 2012	1,460.93	1,460.93	1,425.53	1,428.59	3,115,478,000	1,428.59

Appendix C
A Summary of the Meaning of the Information Obtained from the Yahoo Website

This Appendix outlines the meaning of the data that can be downloaded from the Yahoo Finance site discussed in part III of the student assignment presented in Appendix B.

All stocks are identified by a three or four letter ticker symbol (e.g., Apple is AAPL). Any publicly traded company can be selected by entering the company name or ticker symbol in the text box immediately after "Get Historical Prices for:"

There are options for daily, weekly, and monthly data. Daily data will report characteristics of the stock for any trading day (which excludes weekends and holidays during which stock is not traded). Weekly data will report the same values for any week, as measured by Monday to Monday (or, if Monday is a holiday, then Tuesday). Finally, monthly data will give the results measured from the first trading day of one month to the first of the next month.

The "open" price is the price at which the shares traded for the first transaction of the specified trading period, and similarly the "close" price is the last transaction of the period. Similarly, "high" denotes the highest price at which the stock was traded, and "low" the lowest price at which it was traded. "Avg volume" reports the number of shares that were traded during the trading period.

The only column that is required for the assigned project is the last one, "adj close" or adjusted close. The most common need for an adjustment is that many stocks pay quarterly dividends, and any dividend should be considered part of the return the shareholder earns. For example, if a stock closed at $20 at the end of the previous trading period and at $22 at the end of this trading period, and if the stock paid no dividend or required any other adjustment, then the return during that period would be $\frac{22-20}{20}$ = 10%. However, if the stock also paid a $1 dividend during the period, then the total return would be $\frac{1+22-20}{20}$ = 15%.

To keep the user from having to manually adjust for dividends every time they are paid, Yahoo and many other sites adjust the previous price so that you would earn the same return. In this case, for example, Yahoo would continue to report the $22 as "close" and as "adj close," and would also report the previous close as "$20." However, for the user's convenience they would report as "adj close" the price that would still produce a 15% return, or $\frac{22}{1.15}$ = 19.13. Thus to calculate returns (as required by the assignment in Appendix B), the user simply needs to divide the "adj close" by the previous "adj close;" there is no need to manually take dividends into account because the "adj close" has already done so.

The adjusted close also takes into account three other events that are substantially less common than cash dividends. Companies occasionally choose to have a stock split, in which, for example, the owner of one old share has it replaced by two new shares (as occurred for Apple on February 28, 2005). In a case like this, share price will typically fall by about 50%, but the adjusted price for will take this into account. Even rarer, companies sometimes choose to have a reverse stock split; for example, on May 9, 2011, Citicorp had a 10 for 1 reverse stock split, in which the owner of 10 old shares found them exchanged for one new share. In a case like this, share price will typically increase by a factor of about 10. [Indeed, this was the reason Citicorp chose the reverse stock split; the New York Stock Exchange requires that all shares traded there have an average price greater than $5/share, and Citicorp had fallen below this value.] Finally, companies occasionally issue stock dividends (as opposed to cash dividends), in which case the owner of one old share gets one plus some fraction of a new share. Typically if the fraction is more than 25%, it is treated as a stock split, and if it is less than 25%, it is treated as a stock dividend.

Because the variable of interest is typically returns (whether daily, weekly, or monthly), use of adjusted closing prices allow for the correct calculation of return as simply

$$\frac{adjusted\ close_{end\ of\ period} - adjusted\ close_{beginning\ of\ period}}{adjusted\ close_{beginning\ of\ period}}$$

without the need to manually adjust for dividends, stock splits, or stock dividends.

Exhibit A
Examples of a Stock's Beta as Publicly Reported

From Yahoo:

Apple Inc. (AAPL) - NasdaqGS Follow

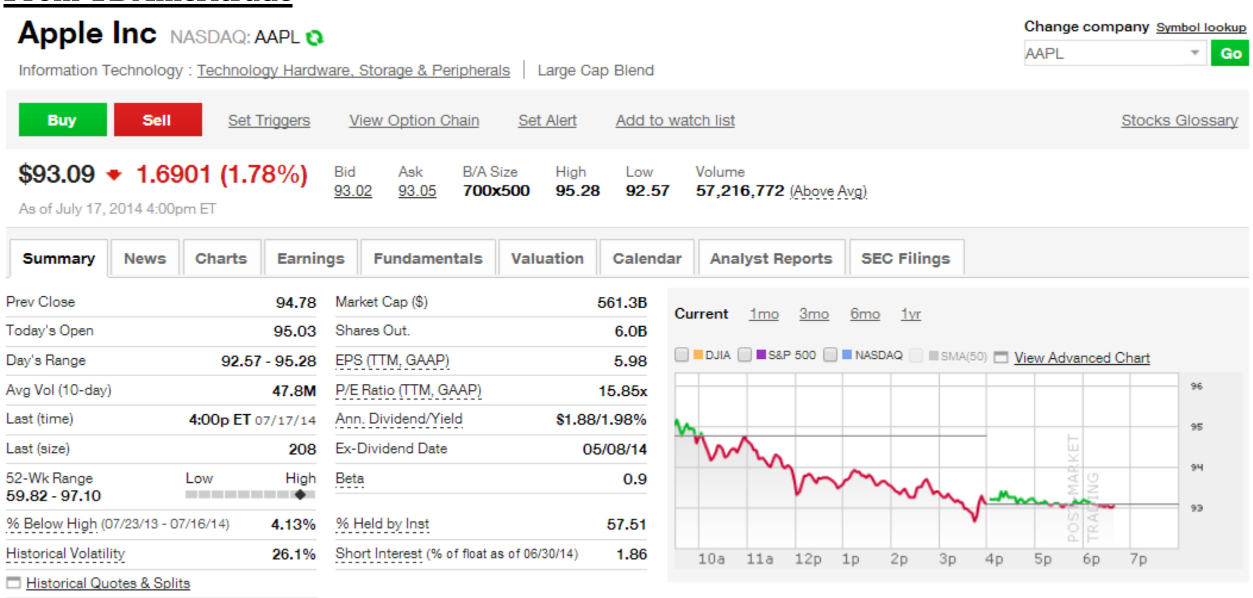
93.09 1.69(1.78%) 4:00PM EDT

After Hours : 93.01 0.08 (0.09%) 6:40PM EDT - Nasdaq Real Time Price

Prev Close:	94.78	Day's Range:	92.57 - 95.28
Open:	95.04	52wk Range:	59.82 - 97.10
Bid:	93.02 x 800	Volume:	53,685,081
Ask:	93.05 x 1000	Avg Vol (3m):	64,718,500
1y Target Est:	98.95	Market Cap:	561.30B
Beta:	0.74	P/E (ttm):	15.62
Next Earnings Date:	22-Jul-14	EPS (ttm):	5.96
		Div & Yield:	1.88 (2.10%)

People viewing AAPL also viewed:
APPL PCLN GOOG AMZN MA CMG

From TDAmeritrade



Apple Inc NASDAQ: AAPL

Information Technology : Technology Hardware, Storage & Peripherals | Large Cap Blend

Change company Symbol lookup
AAPL Go

Buy Sell Set Triggers View Option Chain Set Alert Add to watch list Stocks Glossary

$93.09 1.6901 (1.78%)

As of July 17, 2014 4:00pm ET

Bid	Ask	B/A Size	High	Low	Volume
93.02	93.05	700x500	95.28	92.57	57,216,772 (Above Avg)

Summary News Charts Earnings Fundamentals Valuation Calendar Analyst Reports SEC Filings

Prev Close	94.78	Market Cap ($)	561.3B
Today's Open	95.03	Shares Out.	6.0B
Day's Range	92.57 - 95.28	EPS (TTM, GAAP)	5.98
Avg Vol (10-day)	47.8M	P/E Ratio (TTM, GAAP)	15.85x
Last (time)	4:00p ET 07/17/14	Ann. Dividend/Yield	$1.88/1.98%
Last (size)	208	Ex-Dividend Date	05/08/14
52-Wk Range 59.82 - 97.10	Low High	Beta	0.9
% Below High (07/23/13 - 07/16/14)	4.13%	% Held by Inst	57.51
Historical Volatility	26.1%	Short Interest (% of float as of 06/30/14)	1.86
Historical Quotes & Splits			

Acknowledgements

Without implicating them, the authors thank Marcia Cornett, Dhaval Dave, Scott Gilbert, Atul Gupta, Charlie Hadlock, Lalatendu Misra, Kevin Sylwester, David Rakowski, and Woody Studenmund for their helpful comments.

References

American Statistical Association (2005), *GAISE College Report*. Available at http://www.amstat.org/education/gaise/GaiseCollege_Full.pdf.

Bodie, Z., Kane, A., and Marcus, A. (2013), *Essentials of Investments*, McGraw-Hill Irwin, New York.

Bransford, J. D., Brown, A. L. & Cocking, R. R. (eds.) (2000), *How people learn: Brain, mind, experience, and school*, Washington, DC: National Academy Press.

Brealey, R., Myers, C. and Allen, F. (2014), *Principles of Corporate Finance, Eleventh Edition*, McGraw-Hill Irwin, New York.

Carter, N., Felton, N., and Schwertman, N. (2014), "A Classroom Investigation of the Effect of Population Size and Income on Success in the London 2012 Olympics," *Journal of Statistics Education*, [online], 22, 2. Available at http://www.amstat.org/publications/jse/v22n2/carter.pdf.

Casella, G., and Berger, R. (1990), *Statistical Inference*, Duxbury Press.

Cobb, G. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L.A. Steen, Washington, DC: Mathematical Association of America, 3-43.

Cohen, D. (1982), "A Modified Moore Method for Teaching Undergraduate Mathematics," *The American Mathematical Monthly*, 89(7): 473-474 and 487-490.

Crouch, C. H., and Mazur, E. (2001), "Peer instruction: Ten years of experience and results," *American journal of physics* 69.9: 970-977.

Dawson, R. (2011), "How Significant is a Boxplot Outlier?," *Journal of Statistics Education* 19(2). Available at http://www.amstat.org/publications/jse/v19n2/dawson.pdf.

Garfield, J., Hogg, B., Schau, C., and Whittinghill, D. (2002), "First Courses in Statistical Science: The Status of Educational Reform Efforts," *Journal of Statistics Education* [Online], 10(2). Available at www.amstat.org/publications/jse/v10n2/garfield.html.

Gujarati, D. (2003), *Basic Econometrics, Fourth Edition*, McGraw-Hill Higher Education.

Jones, F. B. (1977), "The Moore Method," *The American Mathematical Monthly*, 84(4): 273-278.

Keeler, C. M. and Steinhorst, R. K. (1995), "Using small groups to promote active learning in the introductory statistics course: A report from the field," *Journal of Statistics Education*, 3(2). Available at www.amstat.org/publications/jse/v3n2/keeler.html

Keller, G., and Warrack, B. (2003), *Statistics for Management and Economics, Sixth Edition*, Thompson Educational Publishing.

Kennedy, P. (2008), *A Guide to Econometrics, 6th Edition*, Blackwell Publishing.

Kmenta, J. (1997), *Elements of Econometrics, Second Edition*, The University of Michigan Press.

Lesser, L., and Kephart, K. (2011), "Setting the Tone: A Discursive Case Study of Problem-Based Inquiry Learning to Start a Graduate Statistics Course for In-Service Teachers," *Journal of Statistics Education* [online]. Available at http://www.amstat.org/publications/jse/v19n3/lesser.pdf.

Medina, J. (2009), *Brain Rules*, Pear Press.

Parker, J. (2005), *R. L. Moore, Mathematician and Teacher*, Mathematical Association of America.

Peterson, P. (2005), "Estimating a Market Model Step by Step," Available at http://educ.jmu.edu/~drakepp/general/market_model.pdf.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society*, Series B 31(2): 350-371.

Snee, R. (1993), "What's Missing in Statistics Education?," *The American Statistician*, 47:2. 149-154.

Soltis, A. (2010), *Chess Studying Made Easy*, Batsford Chess.

Studenmund, A.H. (2006), *Using Econometrics, A Practical Guide, Sixth Edition*, Addison-Wesley.

Vaughan, T. S. (2015), "Analysis of Thursday Night NFL Winning Margins," Journal of Statistics Education 23(1). Available at http://www.amstat.org/publications/jse/v23n1/vaughan.pdf.

White, B.Y., and Frederickson, J.R. (1997), *The ThinkerTools Inquiry Project: Making Scientific Inquiry Accessible to Students*, Princeton, New Jersey: Center for Performance Assessment, Educational Testing Service.

White, B.Y., and Frederickson, J.R. (1998), “Inquiry, modeling, and metacognition: Making science accessible to all students,” *Cognition and Science* 16: 90-91

Zieffler, A., Park J., Garfield, J, delMas R., and Bjornsdottir, A. (2012), “The Statistics Teaching Inventory: A Survey on Statistics Teachers’ Classroom Practices and Beliefs,” Journal of Statistics Education 20(1). Available at http://www.amstat.org/publications/jse/v20n1/zieffler.pdf.

Tomasz Kasprowicz
Wyzsza Szkola Biznesu Dabrowa Gornicza (Academy of Business in Dabrowa Gornicza)

Jim Musumeci
Finance Department
Adamian Academic Center
Bentley University
175 Forest Street
Waltham, MA 02452
jmusumeci@bentley.edu
781-891-2235
