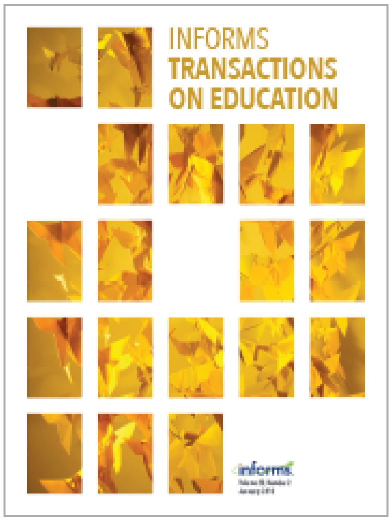

INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Risk Economies of Scale in the Finance and Insurance Industries: Placing the Right Emphasis in Introductory Business Statistics

Octavio Richetta,

Risk Economies of Scale in the Finance and Insurance Industries: Placing the Right Emphasis in Introductory Business Statistics.

Octavio Richetta
Department of Management Science and Information Systems
University of Massachusetts Boston
100 Morrissey Boulevard
Boston, MA 02125

octavio.richetta@umb.edu

ABSTRACT

Most undergraduate and graduate business programs include a required introductory course in statistics with statistical inference as the primary focus. Developing the proper understanding of statistical inference techniques requires an ambitious ramp-up effort devoted to the study of probability theory, discrete and continuous probability distributions, and the central limit theorem. Mastering this large amount of demanding material in a short time poses a significant challenge to most students. As a result, students resort to formula lists they apply without proper understanding. Frequently, even the best performing students in the class are incapable of applying simple statistical concepts soon after the course is over. Acquiring lifelong knowledge requires understanding. This paper illustrates how simple properties of the sum of independent identically distributed random variables account for risk economies of scale which are at the foundations of the finance and insurance industries. By highlighting this simple, yet extremely important result, students stand a better chance of achieving lifelong recollection and understanding of one of the most significant quantitative results in business.

Editor's note: This is a pdf copy of an html document which resides at
http://ite.pubs.informs.org/Vol3No2/Richetta/

INTRODUCTION

Even though they vary significantly in scope, most undergraduate and graduate business programs include a required introductory course in statistics with statistical inference as the primary focus. For perspective, we will focus on examples at both ends of the spectrum: an undergraduate course, Managerial Statistics, I teach in our business program; and a graduate course, Data Models and Decisions, in the MIT Sloan Fellows Program, which I teach as a visiting lecturer. Even though these two examples represent as wide a range in content and audience one may find, both audiences embrace the results that will be presented in this paper with great interest.

The undergraduate one-semester course currently uses a textbook by Black and Eldredge (2002). The objective of the course is to provide students with the basic tools in probability and statistics required by courses in the functional areas of finance, marketing and operations. Over half the course is devoted to a significant ramp-up effort (Chapters 1-7) that provides basic knowledge of probability theory and probability distributions, and the background needed for appropriate coverage of single population statistical inference (Chapters 8 and 9) and simple regression (Chapter 12). The nature of the subject matter, coupled with the amount of material covered, makes the introductory statistics course one of the most feared requirements among the less quantitatively inclined students. Simplifying the course content is not an option as the knowledge of statistical inference and regression required by other courses in the program is already at a minimum, and the ramp-up effort required for proper understanding of statistical inference and regression is also at a minimum. As a result, students have difficulty discerning the fundamental ideas in the course from the details and many surrender to long formula lists they do not understand as the only way of coping with the course.

The graduate course currently uses a textbook by Bertsimas and Freund (2001), which is covered in its entirety. The objective of the course is to present a unified treatment of probabilistic modeling, statistics, and optimization. This objective is significantly more ambitious than that of the undergraduate course. In the area of probability and statistics, the course comprises basic probability theory, decision analysis, statistical inference, simulation, and multiple regression. The optimization portion of the course covers linear pro-

 ©INFORMS ISSN: 1532-0545

gramming in significant depth, and non-linear and integer programming at an introductory level. Even though a large number of examples, including numerous cases, are used to reinforce the relevance of the material, most students view the course as one of the most challenging requirements in the program. Despite being a more mature audience, the more ambitious content results in difficulties in mastering the course material that are similar to those of the undergraduate students.

The problems encountered in teaching introductory business statistics are common to introductory statistics courses in all disciplines. According to Hogg (1991), who advocates radical changes in the way elementary statistics courses are taught, ”students frequently view statistics as the worst course taken in college.” The most dramatic findings on the difficulties students have grasping concepts in statistics are probably those of Savage (1998). Using simple classroom experiments, Savage showed that even masters level Operations Research students at Stanford University not only fail to understand the Central Limit Theorem, they even have difficulty understanding elementary probability distributions. Ingolfsson (1999), and Zalkind (1999), report similar results using experiments on large student audiences at business schools.

There has been significant debate on the causes and possible ways of fixing the problem. Romero et al. (1995), see an academic approach that places excessive focus on theory at the expense of covering practical problems in industry and business as the main cause of the problem. Hoerl et al. (1993), advocate a greater focus on managerial decision making as a possible solution. Apparently, some progress has been made. A broad-based survey of instructors conducted by Garfield et al. (2002), indicates changes in introductory statistics courses are being made; mainly in the form of greater emphasis on data analysis, big concepts and ideas and less on the number of topics covered, computation formulas and theory. However, the authors also report that a consensus has not been reached on what constitutes the right approach to fix the problem. Clearly, the outcome should be long-term learning of important concepts and results in statistics; however, how to teach a course so that this objective is achieved is not obvious.

In a recent paper, Sowey (2001), explores the introduction of striking demonstrations as one of the most important tools that can be used to reinforce the learning process, ”A striking demonstration is any proposition, exposition, proof, analogy, illustration, or application that (a) is sufficiently clear and self-contained to be immediately grasped, (b) is immediately enlightening, though it may be surprising, (c) arouses curiosity and/or provokes reflection, and (d) is so presented as to enhance the impact of the foregoing three characteristics.”

The objective of this paper is to present what I consider to be one of the most relevant striking demonstrations that can be presented in business statistics courses, illustrating how time and space economies of scale that are at the foundations of the finance and insurance industries can yield dramatic reductions in risk. The results that will be presented are suitable for both, undergraduate and graduate audiences; and, most importantly, conform to Sowey's definition of a striking demonstration.

The idea behind our striking demonstration is simple, additive processes involving random variables bring about economies of scale in risk. Specifically, we will focus on economies of scale that result when adding independent identically distributed (i.i.d.) random variables with finite mean and variance, for which the mean increases linearly with the number of variables while the standard deviation (SD) increases proportionally to the square root of the number of variables. We start by presenting this result, and extend it to the sample mean as a special case of the sum of i.i.d. random variables. We then use the result to present our first demonstration, showing how time economies of scale in the financial markets can yield a substantial reduction in the risk long term investors face, explaining the famous adage, ”invest for the long term”. We conclude the paper with our second demonstration, showing how space economies of scale can yield dramatic reductions in risk, making the business of selling insurance an attractive proposition.

The demonstrations are to be presented right after discussing theoretical results for the sample mean and its distribution (e.g., see Chapter 7, *Sampling and Sampling Distributions*, in Black and Eldredge, (2002), and Chapter 4, *Statistical Sampling*, in Bertsimas and Fre-

  ISSN: 1532-0545

und, (2001)). In most introductory business statistics courses, this material precedes the start of statistical inference coverage. Overall, I devote two lectures to this theme. The first demonstration comes at the conclusion of the first lecture. The second demonstration is presented at the beginning of the second lecture, motivating review of the knowledge acquired in the previous lecture; and, most importantly, reemphasizing the theoretical results obtained by presenting a second example that is also highly relevant in business.

RISK ECONOMIES OF SCALE FOR ADDITIVE PROCESSES

This section presents well-known simple results. The objective here is to exemplify methodologies that avoid the use of obscure results taken on faith to the extent possible, as I firmly believe we cannot learn what we do not fully understand.

We let random variable Y_n be the sum of n, i.i.d. random variables, X_i,with mean, μ_x, and SD, σ_x:

$$Y_n = \sum_{i=1}^{n} X_i \qquad (1)$$

Since the expected value of a sum of random variables is the sum of the expected values of the random variables in the sum, a result covered earlier in most introductory statistics courses, we see that:

$$\mu_{Y_n} = E(Y_n) = \sum_{i=1}^{n} E(X_i) = nE(X) = n\mu_X \qquad (2)$$

Deriving the variance of Y_n requires some more work. Using the definitions of expected value and variance of a random variable, we have that:

$$\sigma^2_{Y_n} = E[(Y_n - E[Y_n])^2] = \sum_{i=1}^{n} \sigma^2_{X_i} + 2\sum\sum_{i<j}(E[X_iX_j] - E[X_i]E[X_j]) \qquad (3)$$

Obtaining the right hand side of (3) involves some algebra work and a rearrangement of terms. I usually leave this exercise to a homework assignment to be handed in at the beginning of the next lecture instead of trying to have students follow the details in real time.

From the independence assumption[1], we see that in the last term of (3) $E[X_iX_j] = E[X_i]E[X_j]$, and we obtain:

$$\sigma^2_{Y_n} = \sum_{i=1}^{n} \sigma^2_{X_i}, \quad (3a)$$

(3a) is a well known result for the sum of independent random variables that provides an alternate way of presenting the variance of Y_n, when the audience lacks the background to follow the steps necessary to obtain the result in (3), and/or time is a constraint.

Since the X variables are identically distributed, we let $\sigma^2{}_{X_1} = \sigma^2{}_X$, for i=1,2, ... , n, so that (3a) becomes $\sigma^2{}_{Y_n}$=$n\sigma^2{}_X$, and:

$$\sigma_{Y_n} = \sqrt{n}\sigma_X \qquad (4)$$

(2) and (4) above show the result on risk economies of scale that is central to our demonstration, namely that the mean of Y_n grows linearly with n while its SD grows with the square root of n. A graph showing $\sigma_{Y\,n}$ vs $\mu_{Y\,n}$ as n increases (Figure 1), reinforces the result. An alternate way of presenting the risk economies of

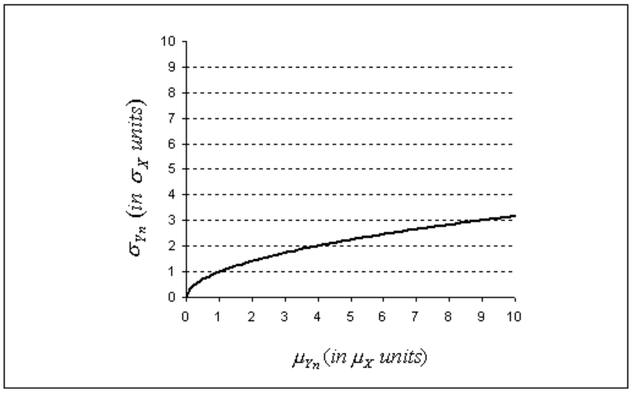

Figure 1: $\sigma_{Y\,n}$ vs $\mu_{Y\,n}$ for increasing n.

scale discussed above is to look at the average of i.i.d. random variables (i.e., in a sampling context, the sample mean). This is easily achieved by replacing X_i with

1. Most introductory business statistics courses would have introduced the concept of independence earlier in the course, in the context of independent events A and B, defined in probabilistic experiment, for which P(AB)=P(A)P(B). This result extended [informally] to include the case of independent random variables X_i and X_j, for which the joint probability density function is equal to the product of the marginal densities, yielding $E[X_iX_j] = E[X_i]E[X_j]$. The latter result is covered explicitly in courses such as the Sloan Fellows course mentioned at the beginning of the paper. Of course, there is no need to resort to any "informalities" when this is the case.

$(1/n)X_i$ in (1) and letting $Y_n = \overline{X}$; yielding the more familiar result:

$$\mu_{\overline{X}} = \mu_X; \quad and \ \ \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} \qquad (5)$$

My preference is to present the result for the sum of i.i.d. random variables so that one can present economies of scale in the form that is most familiar to students (Figure 1), and then just use the "one liner" described above to extend the result to the average. A graph showing $\sigma_{\overline{X}}$ vs. n (Figure 2) reinforces the result shown in (5). To show that both results yield equivalent

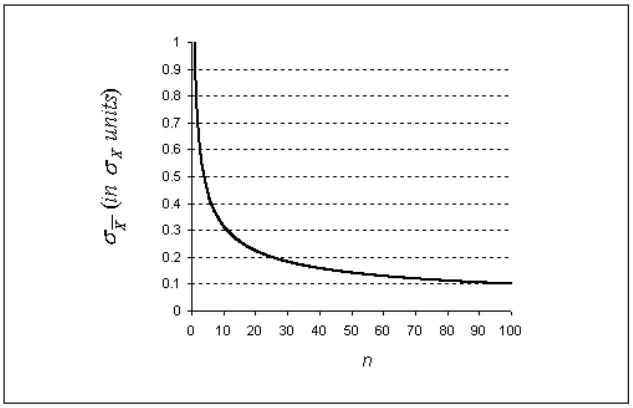

Figure 2: $\sigma_{\overline{X}}$ vs. n

economies of scale, one could calculate the coefficient of variation (CV) to demonstrate that both cases yield the same result, variability that is inversely proportional to the square root of n:

$$CV_{\overline{X}} = CV_{Y_n} = \frac{1}{\sqrt{n}} CV_X \qquad (6)$$

We are now ready to apply the result.

TIME-INDUCED RISK ECONOMIES OF SCALE IN THE FINANCIAL MARKETS

We let random variable X represent the nominal annual return on a stock market index. For the S&P 500 index, the mean of X is approximately 10% while its SD is in the order of 20%[2], indicating short term investing involves significant risks. When asked, most students venture reasonable guesses at the mean of X; however, most of them suggest a SD at or below the mean, and are quite surprised to find out the SD is approximately twice the expected return[3].

Lets consider instead the average return over a large number of years. If we let the random variable $\overline{X}$ represent the arithmetic average of the annual returns over a 25 year time span; by (5), we see that the mean of $\overline{X}$ is 10% but the SD of $\overline{X}$ is only 4%. This result shows that expanding the investment time horizon can bring about a significant reduction in risk. If the annual stock market return is assumed to be normally distributed over time, then we can see that 95% of the annual returns should be between 30% and 50% but the average return over 25 years should be between 2% and 18%, 95% of the time[4].

The analysis above assumes that annual stock market index returns are independent. Fama (1970), showed stock market prices follow a random walk and hence stock price returns are independent over time. When presenting the result, I make sure students clearly understand the difference between the random variables under consideration: X represents the return obtained in any given year, while $\overline{X}$ represents the average return over a period of several years.

Depending on the audience, I point out that there is a bit of a "wrinkle" in the analysis and ask what it may be. Occasionally, students come up with the right answer: the total return over a number of years is actually a multiplicative function of annual returns; thus, the average return is a geometric average, not

2. For the January 01, 1926- December 31, 2001 time period, Standard and Poor's reports arithmetic average annual returns for the S&P 500 index of 12.6% with a standard deviation of 20%.
http://fc.standardandpoors.com/htdocs/pdf/wac/5008.pdf

3. The "SD surprise" illustrates the conjecture that most investors do not have a good understanding of the risks involved in investing in the stock market, and sets the stage for demonstrating that there is a tangible advantage behind the "investing for the long term" proverb that is supported by the mathematical results obtained in the previous section.

4. The probability calculation for $\overline{X}$ requires knowing that $\overline{X}$ is normally distributed. This follows from the version of the central limit theorem (CLT) that is presented in the classroom as part of the theoretical results for the sample mean. The CLT is not explicitly discussed in this paper as we have purposefully limited our discussion to those results for the sample mean students can understand (i.e., those results that can be proven as opposed to those that need to be "taken on faith").

an arithmetic one[5]. The wrinkle in the analysis can be "ironed" by noticing that the result above can be applied to the log of the geometric return instead.

The learning is that in the long term, investors experience less risk. However, long-term investors must be mindful of the fact that the random variable that should really matter to them is average return over a large time span, not the return in any given year. This does not mean that short-term risk should not be an important consideration when deciding what fraction of a portfolio should be invested in the stock market. Investors should always be mindful of the fact that the SD of one-year returns is 20%, twice the expected return. In the short term, even long-term investors can feel significant pain when the relatively frequent negative return events materialize.

SPACE-INDUCED RISK ECONOMIES OF SCALE IN THE INSURANCE INDUSTRY

Let random variable X represent the total annual monetary loss to either an individual or a corporation using self-insurance. The expected monetary loss under self-insurance may be acceptable but the relative risks (i.e., coefficient of variation) involved are greater than those found in the stock market.

For simplicity, let's consider the case of collision/comprehensive car insurance coverage with no deductible for a vehicle with a $10,000.00 market value, and a 1% probability of total vehicle loss as the only unfortunate event possible. For this example, the expected value of the loss under self-insurance is $100 and the SD of the loss is $990 (for simplicity, we will use $1000 for the SD). The high level of risk involved is reflected in a SD that is one order of magnitude above the mean, resulting in a CV value of 10.

While most individuals would choose to carry liability insurance, the decision to purchase collision/comprehensive coverage would depend on how painful the loss experience might be in any given year. A wealthy individual with a car of low market value would tend not to carry collision/comprehensive insurance while the opposite would be most likely for a poor individual with an expensive car. The value of risk reduction for the latter case is such that these individuals are willing to purchase collision/comprehensive insurance coverage at a premium that is significantly higher than the expected value of the loss under self-insurance. For simplicity, we will use $200, twice the expected loss, for the policy premium.

From the perspective of an insurance company holding n i.i.d. policies; by (5), the expected average gain for random variable $\overline{X}$, the average net profit from writing n policies on any given year, is $100 (i.e., the policy premium minus the expected claim value), with a SD that declines proportionally to the square root of n. For n=10,000 policies, we see that the SD of $\overline{X}$ is only $10. Resulting in a CV of 0.1.

The result shows that by holding a large number of policies insurance companies can take advantage of significant economies of scale in risk. Holding a large number of policies is an important risk reduction strategy that can also provide a competitive advantage, as a larger number of policies results in a lower SD, which in turns allows more aggressive pricing.

For simplicity, in the analysis above, we have assumed that policies have identical claim probability distributions and that claim amounts are independent random variables. Both assumptions are reasonable for car insurance. The collision insurance example can then be used to provide a rationale for the risk reduction strategies insurance companies exercise when the independence assumption does not hold and/or economies of scale cannot be achieved, such as policy swaps across geographical areas and reinsurance.

We see that the time-induced economies of scale discussed in the previous section also apply to the insurance industry. For the insurance company example discussed above, holding 10,000 policies over a twenty-five year time span yields an expected profit per policy of $100, while the SD is reduced to only $2, for a CV of 0.02.

5. The difference between geometric and arithmetic annual returns can be illustrated with a relevant example: For the period 1802-1997, Siegel, 2002, reports arithmetic average annual stock market returns of 8.5% and geometric average returns of 7.0%.

CONCLUSION

We have seen how simple results for the distribution of the sample mean can explain time and space economies of scale that are at the foundations of the finance and insurance industries. We have shown that having a long-term time horizon in investing can reduce risks significantly, and demonstrated how holding a large number of policies is a crucial risk reduction strategy in the insurance industry. In addition, using the insurance industry example, we have seen that space and time economies of scale can apply concurrently.

My hope is that the prominence and relevance of the demonstrations presented in this paper will help students achieve life-long recollection and understanding of one of the most important results in statistics.

REFERENCES

Bertsimas, D., and R. Freund (2001), *Data Models and Decisions: The Fundamentals of Management Science*, South-Western, Cincinnati, Ohio.
Overview:
http://snapshot.swcollege.com/ss-overview.cfm?isbn=0-538-85906-7&disccode=4&v=2
Website:
http://www.swcollege.com/quant/bertsimas/bertsimas.html

Black, K., and D.L. Eldredge (2002), *Business and Economics Statisitcs Using Microsoft©Excel*, South-Western, Cincinnati, Ohio.
Overview:
http://snapshot.swcollege.com/ss-overview.cfm?isbn=032401726X&disccode=4&v=1
Website:
http://www.swcollege.com/quant/black/2002_first_edition/black.html

Fama, E. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work", *Journal of Finance*, Vol. 25, pp. 383-417.

Garfield, J., Hogg, R., Schau, C., and Whittinghill, D., (2002), "First Courses in Statistical Science: The Status of Educational Reform Efforts," *Journal of Statistical Education*, Vol. 10, No. 2.
www.amstat.org/publications/jse/v10n2/garfield.html

Hoerl, R. W., Hooper, J. H., Jacobs, P. J., and Lucas, J. M. (1993), "Skills for Industrial Statisticians to Survive and Prosper in the Emerging Quality Environment", *The American Statistician*, Vol. 47, pp. 280-291.

Hogg, R. V. (1991), "Statistical Education: Improvements Are Badly Needed", *The American Statistician*, Vol. 45, pp. 342-343.

Ingolfsson, A., Zalkind D. (1999), "The Teachers' Forum: Two Looks at the Spinner Experiment", *Interfaces*, Vol. 29, No. 6, pp. 112-122.

Romero, R., Ferrer, A., Capilla, C., Zunica, L., Balasch, S., Serra, V., and Alcover, R. (1995) "Teaching Statistics to Engineers: An Innovative Pedagogical Experience", *Journal of Statistics Education*, Vol. 3, No.1.
http://www.amstat.org/publications/jse/v3n1/romero.html

Savage, S.L. (1998), "Statistical Analysis For The Masses", *Statistics in Public Policy*, Ed.: Bruce Spencer, Oxford University Press, Oxford
http://www.stanford.edu/~savage/stat.pdf

Siegel, J. J. (2002), *Stocks for the Long Run*, McGraw Hill, NY.

Sowey, E. R. (2001), "Striking Demonstrations in Teaching Statistics", *Journal of Statistics Education*, Vol. 9, No. 1.
http://www.amstat.org/publications/jse/v9n1/sowey.html