JOURNAL OF ECONOMICS TEACHING

Interactive Web-based Simulations to Teach Econometrics: Making Abstract Concepts Tangible

We develop a series of publicly available interactive web-based modules and instructor materials to help students learn econometric concepts ranging from sampling distributions to omitted variable bias. The pedagogical literature supports using simulations to help students understand abstract statistical concepts but also cautions that ill-designed simulations may be ineffective. Most simulation tools currently available constrain users' parameter choices, require excessive programming infrastructure, or lack real-world economics examples. Our modules allow students to experiment with inputs like sample size, number of replications, and underlying population distributions; students then see the impacts of their choices displayed visually in tables, histograms, and scatter plots. The modules can be used in classroom demonstrations or group activities in computer labs but also have sufficient instructions to allow independent exploration.

Tanya Byker†, Amanda Gregg†, Dylan Mortimer‡

†Middlebury College, ‡ Atlanta Braves

1. Introduction

Increased access to data, statistical software, and computing power is allowing the current generation of undergraduate economics students to dive more deeply into the empirical side of economics than ever before. The goal of the quantitative sequence for economics majors at our institution is for our students to develop the statistical literacy that will enable them to tackle independent empirical research projects by the time they graduate. [1] However, we find it easier to teach the tools and procedures of statistical analysis than it is to help our students grasp the abstract theories that underpin their ability to draw valid conclusions about the results a powerful statistics package can easily "run." To address this difficulty, we have developed an interactive web-based application to help students learn statistics and econometric concepts ranging from sampling distributions to omitted variable bias to help students grasp the "concepts and ideas" that will make them effective users of the "tools and procedures" of empirical economic analysis.

The theoretical side of statistics and econometrics courses is notoriously difficult to teach, and students often find it difficult to feel confident and succeed in these courses (Kennedy, 2001; Tishkovskaya & Lancaster, 2010). This is often because students come from wide-ranging backgrounds that include very different past experiences with mathematics. Even undergraduates who have taken statistics in high school often lack hands-on experiences that unite statistical theory with real-world data generating processes. Much of the difficulty of teaching econometrics and statistics arises from the high level of abstraction of the material along with the need to develop a robust skill set including "logical reasoning, critical thinking, data analysis and interpretation and evaluation skills" (Boyle et al., 2014, p. 1). Finally, students often find statistics to be just plain boring (Boyle et al., 2014). Incorporating interactive technology offers the potential to develop intuition and invigorate the learning environment.

Our application provides a set of interactive web-based simulation modules. Simulation is a method using repeated draws or iterations of a procedure to demonstrate an abstract concept. Having students do activities that involve rolling dice or playing with cards are crude versions of simulation (and can be quite effective in the classroom). The goal of automating simulations on the computer is to draw large samples quickly and easily visualize the results of the simulation in graphs. Automation also greatly facilitates interactivity, making it possible for users to change parameters and quickly visualize (through animation) the impact of those choices.

A wide-ranging pedagogical literature supports using simulations to help students understand abstract statistical concepts (Briand & Hill, 2013; Kennedy, 2001; Marasinghe, Meeker, Cook, & Shin, 1996; Mills, 2005; 2002). In particular, a small number of studies have compared the effectiveness of computer-based simulations and animations to traditional static presentations such as textbooks or web-based lecture slides (Boyle et al., 2014). Mills (2005) randomly assigned students in university-level introductory statistics class to learn about the Central Limit Theorem using simulation in Excel or using traditional learning methods and found that simulation led to statistically significant improvements in test scores and attitudes towards the material. Liu, Lin, and Kinshuk (2010) found that high school students randomly assigned to use a dynamic simulation module rather than a lecture-based learning group had fewer misconceptions about the concept of correlation and improved understanding.

However, ill-designed simulations may be ineffective (Lane & Peres, 2006; Tversky, Morrison, & Betrancourt, 2002). Current best practice in designing simulations stresses visually

[1] Gal (2005, p. 70) defines statistical literacy as "the ability to interpret, critically evaluate, and communicate about statistical information and messages" (quoted in Boyle et al. 2014).

compelling, low-stress interactivity, though even the best-designed simulations have their limits. Interactive elements that allow students to focus on specific parts of an animation or simulation can help with comprehension (Tversky, Morrison, & Betrancourt, 2002). The visual reactivity that interactive applications provide also give students the ability to explore data in an engaging way, for example by dragging points around on a graph to affect the least squares sum of residuals (Chance, Ben-Zvi, Garfield, & Medina, 2007). The ability to rewind and restart modules or activities reduces stress, which may lead to improved comprehension (Tversky, Morrison, & Betrancourt, 2002; Wang, Vaughn, & Liu, 2011). Wang, Vaughn, and Liu (2011) compare the effectiveness of teaching hypothesis testing using static materials to animations with randomly assigned degrees of interactivity. Interestingly, they find that interactivity had no impact on lower-level learning (remembering) or on the highest level of learning (applying) compared to static methods, but that interactive modules did improve students' intermediate-level learning (understanding and low-level applying). Moreover, they find diminishing returns to interactivity: the most interactive treatments provide no significant improvements over simpler versions. Thus, simulations should be designed with care and presented as one component of a comprehensive course.

Other sites with elements of our application currently appear separately on the web or in posted code. Our assessment is that the simulation tools currently available either take too much of the choice in parameters away from the user, require an excessive amount of classroom time learning to set up the programming infrastructure, or lack any kind of economic context that ties them to the curriculum in our courses. The goal of our collection of modules is to allow students to explore abstract concepts and assumptions without becoming distracted by the complexity of programming simulations in statistical software. We also aim to incorporate real-world economic examples in our data choices to tie these modules to examples students will see in other economics classes.

The following short guide to our new applications proceeds as follows: first, we provide an outline of the simulations from which we took inspiration for this project, and how we hope to improve upon them. Rather than written descriptions, we provide links to video demonstrations of each module including the Law of Large Numbers, the Central Limit Theorem, joint distributions, ordinary least squares, and omitted variable bias. Next, we provide suggestions for integrating our modules into lectures and lab demonstrations, including a summary of each concept and selected examples of classroom demonstrations using the modules and links to an inventory of additional instructor materials we provide online. We conclude with a discussion of how our modules fit into a modern econometrics curriculum and outline modules we hope to develop in the future.

2. Simulation Tools Currently Available and How We Build on Them

We developed these modules because most simulation tools currently available constrain users' parameter choices, emphasize results rather than process, are too abstract in focus, or require excessive programming infrastructure. This section guides the reader through several other simulation exercises currently available and discusses the improvements we have made in developing ours.[2]

The Wolfram Demonstrations Project (https://demonstrations.wolfram.com/) provides an impressive array of simulation exercises, free to the public, for a wide variety of subjects including economics, statistics, mathematics, and the life sciences. We took considerable inspiration for our project from modules illustrating the Central Limit Theorem and Least Square

[2]This discussion is meant to be illustrative rather than exhaustive. There are, of course, other simulations available, but they share at least some of the limitations we outline here.

Regression.[3] Their Central Limit Theorem module allows the user to choose a sample size and take samples, but the user does not observe the sample process or the calculation of means from the sample. It may be clear, with an instructor's narration, *that* the Central Limit Theorem is at play, but it is not clear *why* the theorem should hold. We thus designed simulations that show as literally as possible how taking samples makes these two key results function how they do.[4]

The Wolfram Least Squares Regression simulation provides users with three options for datasets, sliders to choose a slope and intercept for a proposed line of best fit, and a button to show the least squares line. This module shows the user a visual representation of minimizing a sum of squared errors by drawing squares where each edge is the distance from a data point to the proposed line. We liked this visual representation, but ours has added the ability to choose the number of points to plot and then randomly generate the dataset, and our module makes the simulation into more of a "game," prompting the user to make a guess about the slope and intercept and then receive feedback in the form of the squares of errors.[5]

Instructors may also illustrate key ideas in statistics and econometrics by asking students to write simulations themselves, for example in Stata or R. Before developing these tools, one of us asked students to write simulations of the Law of Large Numbers and the Central Limit Theorem in Stata for a lab exercise. Doing so had advantages: students learned loops and how to generate random numbers in Stata. However, the exercise obscured a deeper understanding of the Law of Large Numbers or Central Limit Theorem. Students often did not realize the connection between the programming task and these concepts, because they were so focused on whether their program worked correctly or not. We thus developed exercises that refocused the students on the concepts rather than the code. Conceivably, with sufficient time, instructors could do both, i.e. illustrate the idea thoroughly through simulation and then ask students to write their own program as an exercise.

Finally, our application collects the concepts that instructors in econometrics would most want to emphasize on one site with a coordinated infrastructure that relies on economics examples throughout. Students can focus on the construction of new concepts rather than learning to navigate several new sites, each with its own formatting, internal vocabulary, and rules for interaction. A further advantage of our simulations is a set of supporting infrastructure for implementing these tools in the classroom, including video walkthroughs and instructor materials.

3. Access to Applications and Video Demonstrations

Our applications can be accessed at: go.middlebury.edu/econsims. Video demonstrations and additional instructor materials can be accessed at: go.middlebury.edu/econsims_materials

[3]Available: https://demonstrations.wolfram.com/TheCentralLimitTheorem/ and https://demonstrations.wolfram.com/LeastSquaresCriteriaForTheLeastSquaresRegressionLine/, respectively. Using this demonstration requires downloading the Wolfram Player.

[4]The Sampling Distributions simulation created by the Rice Virtual Lab in Statistics better approximates our objective, since it shows an underlying population distribution, samples taken from that distribution, and resulting sample means. However, choosing many samples simply results in a plotted result rather than an illustration of the process, and the user cannot see which observations were sampled from the original population. We hoped to improve on both aspects with our tools.

[5]The OLS App on the site Econometrics by Simulation provides a more sophisticated set of modules than what we currently offer. This app shows an underlying dataset and allows the user to explore many aspects of regression, including the choice of model relative to some true underlying data generating process. This app's versatility is attractive for instructing more advanced students, though our modules make more focused points as stand-alone exercises.

4. Guide to Module Content and Suggestions for Instructional Use

We encourage instructors to implement our simulations at all varieties of institutions and at varied moments within the econometrics curriculum. At our small liberal arts college, our econometrics sequence consists of two fundamental courses: Economic Statistics (introductory statistics at most colleges and universities, which we teach within the department) and Regression (a general introductory econometrics course that introduces multiple regression). We introduce the modules for the Law of Large Numbers and the Central Limit Theorem in Economic Statistics while the modules for Joint Distributions, Least Squares, and Omitted Variable Bias are best suited to the second course Regression. Instructors at other institutions may choose to use a similar sequencing, or may also find that upper-level students benefit from the refresher provided by the Law of Large Numbers and Central Limit Theorem simulations.

Our simulations have the greatest impact on learning when the instructor strikes the right balance between guiding students and allowing them to explore independently. Lane and Peres (2006) caution that demonstrating simulations to a passive classroom audience is likely no more effective than traditional instruction methods. However, they caution that "pure discovery learning" that has students interact with the application without any structure or direction has also not been shown to improve learning. Lane and Peres (2006) instead recommend "guided discovery," whereby interaction with the modules is based on a clear set of instructions and preferably asks students to answer questions about what they expect *before* they take the next step in the simulation. We thus encourage instructors who use our modules to find an equilibrium between allowing students to play freely with the modules, which can help make "dry" material fun, and guiding students with purpose towards revelatory moments. In the teaching materials that are linked to this paper, we describe strategies for structuring "guided" play or exploration.

<u>Law of Large Numbers</u>

The Law of Large Numbers is one of the first major results students learn that explains why we can use the sample mean as an estimator of the mean of the population from which that sample is drawn. The Law of Large Numbers states that, as the sample size *n* approaches infinity, the sample mean approaches the true population mean.[6] Students often struggle to differentiate the sample mean and population mean and to understand why the sample size would matter for the "accuracy" of the sample mean as an estimate of the population mean. Our simulation shows that the Law of Large Numbers really works in practice and why the sample size matters.

<u>Central Limit Theorem</u>

In our econometrics and statistics classes, the Central Limit Theorem is often taught alongside the Law of Large Numbers. Students often find the Central Limit Theorem difficult to grasp and, in particular, have trouble differentiating the thrust of the Central Limit Theorem from the Law of Large Numbers. The Central Limit Theorem is a statement about the *sampling distribution* of the sample mean, $\bar{x}$. The exact statement varies across textbooks, but generally, the Central Limit Theorem states that, for sufficiently large samples, the sample mean will be approximately normally distributed, no matter what shape the underlying population

[6] The standard deviation of the sampling distribution of the sample mean $\bar{x}$ is given by $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, where σ is the population standard deviation. As $n \to \infty, \sigma_{\bar{x}} \to 0$, and thus $\bar{x} \to \mu$. This theoretically important result is often omitted in introductory textbooks. Anderson et al. (2015), for example, do not discuss the Law of Large Numbers. Our treatment is similar to the one found in Stock & Watson (2015), p. 48. For a more technical discussion, see Miller & Miller (2004) p. 267 or Wooldridge (2010) pp. 41-2.

distribution has.[7] Our simulation here stresses the aspects of the Central Limit Theorem students find especially challenging: how the sample mean can have a distribution, what is implied by that sampling distribution being normal, and why the nature of the sample mean implies that the sample mean's distribution will be bell-shaped.

The following is a suggestion for teaching the Central Limit Theorem using this module, which we also illustrate using a video walkthrough:

> Probably the most powerful single exercise is the following: ask students to plot the means of many samples of size *n=1*. This should just give back the population distribution (check for understanding). Then, ask the students to plot the means of samples of size *n=2*. Already, the distribution of the sample means will begin to appear bell-shaped (Setting the Mystery Distribution as the population distribution produces especially stark results!) Allow them to continue to increase the sample size by 1 each iteration. This really drives home the central intuition: because the sample mean finds the center of the sample points in any sample, the sample means will be more likely to be located close to the population mean than far from the population mean. The bell-shaped curve implies that points are more likely to be located close to the center than far from the center.[8]

<u>Joint Distributions</u>

This module makes a transition from the focus on single random variables and their sampling distributions in the LLN and CLT modules to thinking about the joint distribution of two random variables. The joint probability distribution of two random variables describes their simultaneous behavior. The co-movement or relationship between two or more variables is the crux of most empirical economic analysis – what is the relationship between price and quantity demanded? How does one's level of education affect one's earnings? How does the corporate tax rate affect R&D expenditure? The key learning goal of this module is conceptualizing the relationship between the distribution of individual random variables and their co-movement as measured by covariance. Understanding covariance is an essential building block in learning about regression in the following modules. Our simulation allows students to input parameters (mean and standard deviation) for the distribution of two random variables and additionally input a value for their covariance. "Playing" with various combinations of these parameters allows students to visualize the resulting joint distribution on a scatter plot.

<u>Least Squares Regression</u>

Estimating an Ordinary Least Squares (OLS) regression line is a minimization problem.[9] We want to pick the best line to fit the data – where we usually start with the "data" displayed as a scatter plot conveying the relationship between two variables *X* and *Y*. Picking a line means picking an intercept and slope; and if there is anything students remember from high school,

[7]Anderson et al. (2015) states the Central Limit Theorem as: "In selecting random samples of size n from a population, the sampling distribution of the sample mean $\bar{x}$ can be approximated by a normal distribution as the sample size becomes large" (p. 313). Stock and Watson (2015), however, include a statement of the central limit theorem that builds in a rescaling of the values of the sample mean so that the resulting distribution is a standard normal distribution: "Suppose that $Y_1,\ldots,Y_n$ are i.i.d. with $E[Y_i]=\mu_Y$ and var $(Y_i) = \sigma_Y^2$, where $0 < \sigma_Y^2 < \infty$. As $n \to \infty$, the distribution of $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$ (where $\sigma_{\bar{Y}}^2 = \sigma^2/n$) becomes arbitrarily well approximated by the standard normal distribution" (p. 52).

[8]From our instructor materials for the Central Limit Theorem, p. 1. The Instructor Materials are available online with this publication and at the author's website at go.middlebury.edu/econsims_materials.

[9]Specifically, as Stock & Watson (2015) state, "The OLS estimator chooses the regression coefficients so that the estimated regression line is as close as possible to the observed data, where closeness is measured by the sum of the squared mistakes made in predicting Y given X" (p. 116).

it is that Y=mX+b, so they are happy to follow along here. In class, the discussion of why the squared distances between the points and the line is a good way to measure "fit" usually also goes smoothly. The confusion often arises when we talk about a way to solve the minimization problem, especially for students with rusty (or nonexistent) calculus backgrounds. Our module lets students visualize the square distances – they appear as actual squares that grow and shrink as the user guesses combinations of intercept and slope – and play a game of trying to minimize them. Students seem to find it fun to iteratively guess and check, with the calculated sum of squares updating in real time. They soon agree that having a convenient (closed form) mathematical solution to this problem is a beautiful thing.

Omitted Variable Bias

Omitted Variable Bias (OVB) is the bias in a regression estimator that arises when there is a variable (*V*) that is not included in the regression that is correlated with the regressor (*X*) and is a determinant of the outcome (*Y*). In this module, we focus on bias in the estimator of the OLS slope parameter. OVB is driven by two components: 1) The relationship between the omitted variable (*V*) and the regressor of interest (*X*), and 2) The relationship between the omitted variable and the outcome (*Y*).[10]

We rarely know the exact magnitudes of these two components (if we did, it must mean we have data on *V* and we could stop omitting it from the regression), but we often have (economic) intuition about their signs. Thus, we usually can (and should) think hard about the sign of the bias. Getting one's head around the sign (and direction) of the bias can be tricky because there are three signs floating around this problem – the sign of the two components above as well as the sign of the population slope on the relationship between *Y* and *X*. This module lets students experiment with all three components separately and visualize how they affect the regression estimates.

As a suggestion for teaching OVB using this module, ask students to consider the following relationship between time spent studying and score on a test:[11]

$$Test\ Score_i = \beta_0 + \beta_1\ Study_i + u_i.$$

In many real data sets, we see a zero or even negative *correlation* between hours of studying and grades. In the simulation, we explore the possibility that OVB explains this surprising finding. For example, imagine that sleep before a test also impacts scores *and* is correlated with how much a student studies:

$$Test\ Score_i = \beta_0 + \beta_1\ Study_i + \delta\ Sleep_i + \epsilon_i$$

[10]Stock & Watson (2015) state, "if an omitted variable is a determinant of Y_i, then it is in the error term, and if it is correlated with X_i then the error term is correlated with X_i. Because u_i and X_i are correlated, the conditional mean of u_i given X_i is nonzero. This correlation therefore violates the first least squares assumption and the consequence is serious: The OLS estimator is biased. This bias does not vanish even in very large samples, and the OLS estimator is inconsistent" (p 185).

[11] The following is largely excerpted from our instructor materials for OVB, p. 1, available online with this publication and at the author's website at go.middlebury.edu/econsims_materials.

Remind students that the equation for OVB in this case is given by:[12]

$$\beta_1 \xrightarrow{p} \hat{\beta}_1 + \frac{\delta \mathrm{Cov(Study,Sleep)}}{\mathrm{Var(Study)}}.$$

Ask students to think about the sign and relative magnitude of the population parameters β_1 and δ and enter these. Now, ask them to think about the sign of the correlation between hours of sleep and hours of studying – if you sleep more do you study more or less? Given these initial inputs, generate the scatter plot and the naïve regression line for a simple regression of test scores on hours of study. *Before* revealing the corrected line that re-estimates the slope on the study coefficient adding sleep as a regressor, ask students to think about the potential bias in the naïve regression based on the parameters for δ and Cov(Study,Sleep) we entered. Now we check their prediction by clicking the button to "show corrected regression line." Next, let students experiment with alternative hypotheses: If the correlation between sleep and study times are the opposite of the initial guess, how will the naïve and corrected line differ? If sleep has no impact on scores (even if it is correlated with study time), will the naïve and corrected lines differ? These experiments build an intuitive foundation for understanding cases of OVB they may encounter in articles they read or in their research projects.

We have posted online instructor materials with these and other suggestions for guiding students through our simulations in a traditional lecture class or interactive lab section (at: go.middlebury.edu/econsims_materials). These materials include suggestions for problem set questions to follow the demonstrations, which reinforce central concepts and allow the students to reflect on what they have learned by working through the simulations independently. These instruction guides and exercises could be adapted for use in various college-level statistics and econometrics classes.

5. Conclusion

As students become increasingly adept at using statistical packages, instructors should ensure that the big and beautiful ideas of statistics and econometrics do not get lost in the mix. Economists and educators, Angrist and Pischke recently called for a shift in focus away from "models and math" in econometrics teaching towards "causal questions and empirical examples" (Angrist & Pischke, 2017). A careful reading of their recommendations shows that this is not a suggestion that students no longer need to understand abstract concepts, but rather a call to focus less on "statistical technicalities" like tests for heteroscedasticity and instead to focus early and often on real-world examples of research strategies like Randomized Control Trials (RCT).

The idea behind starting with an RCT is that it provides a framework to understand causality because the OLS assumptions hold by design. Since many real-world research projects that our students study or conduct themselves will not involve random assignment, understanding how to move beyond that "ideal scenario" is the key learning goal in our courses. It makes sense to start with an RCT so that we can show what is different when we estimate a regression with observational data. There are several key (and abstract) concepts that students need to understand to make that leap successfully: What does it mean to have a sample of data (random or otherwise)? What is Stata or R actually doing when we type "regress"? What is the relationship between the distribution of the outcome and the distribution of our regressor of

[12]Assuming that $E(\epsilon_i|Study_i)=0$. This derivation of OVB builds on the way Byker teaches students about the consistency of the OLS slope estimator. Slightly different forms of an OVB equation are derived in Stock & Watson (2015) p. 214 and Wooldridge (2010) p. 65-7. The key concept, that the sign of the bias is determined by the relationship between the omitted variable and both the regressor of interest and the outcome, remains the same regardless of the specific exposition provided in each textbook.

interest? What happens when we do not have a random assignment to eliminate the problem of confounders (i.e., why does correlation not equal causation)? Is the result (including one from an RCT) statistically different from zero? We agree that each of these could be taught purely through equations on slides or the board; we also agree that many students find this strategy boring and difficult to grasp. The solution we propose in this paper is a set of interactive web-based tools that let students "play" with the concepts and gain intuition for these fundamental concepts.

This list of concepts – the keys to grasping how empirical analysis helps researchers address causal questions – was our motivation for building these tools and the outline of the menu page of our application. We began with the five included models because they represented our highest priorities in teaching statistics in econometrics. We plan to add modules in the future, including simulations of hypothesis testing, confidence intervals, multiple regression, and fixed effects regression. These will be available at go.middlebury.edu/econsims and go.middlebury.edu/econsims_materials as they are developed.

References

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. 2015. *Essentials of statistics for business and economics.* Stamford: Cengage Learning.

Angrist, J D., & Pischke, J-S. 2017. Undergraduate econometrics instruction: Through our classes, darkly. *Journal of Economic Perspectives, 31*(2): 125–144. https://doi.org/10.1257/jep.31.2.125.

Boyle, E. A., MacArthur, E. W., Connolly, T. M., Hainey, T., Manea, M., Kärki, A., & van Rosmalen, P. 2014. A narrative literature review of games, animations and simulations to teach research methods and statistics. *Computers & Education, 74*(May): 1–14. https://doi.org/10.1016/j.compedu.2014.01.004.

Briand, G., & Hill, R. C. 2013. Teaching basic econometric concepts using Monte Carlo simulations in Excel. *International Review of Economics Education, 12*(January): 60–79. https://doi.org/10.1016/j.iree.2013.04.001.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. 2007. The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, 1(1). https://sites.google.com/site/danibenzvi/TISETheRoleofTechnologyTISE2007.pdf.

Gal, I. "Statistical Reasoning: Meanings, Components, Responsibilities," in *The Challenge of Developing Statistical Literacy, Reasoning and Thinking,* ed. by Dani Ben-Zvi and Joan Garfield. Springer Netherlands: 2004: pp. 47-78.

Kennedy, P. E. 2001. Bootstrapping student understanding of what is going on in econometrics. *The Journal of Economic Education, 32*(2): 110–123. https://doi.org/10.1080/00220480109595177.

Lane, D. M, & Peres, S. C. 2006. Interactive simulations in the teaching of statistics: Promise and pitfalls. *Proceedings of the Seventh International Conference on Teaching Statistics*, Voorburg, The Netherlands: International Statistics Institute.

Liu, T.-C., Lin, Y.-C., & Kinshuk. 2010. The application of simulation-assisted learning statistics (SALS) for correcting misconceptions and improving understanding of correlation: Effects of SALS-based learning. *Journal of Computer Assisted Learning, 26*(2): 143–158. https://doi.org/10.1111/j.1365-2729.2009.00330.x.

Marasinghe, M. G., Meeker, W. Q., Cook, D., & Shin, T-s. 1996. Using graphics and simulation to teach statistical concepts. *The American Statistician, 50*(4): 342–351. https://doi.org/10.2307/2684932.

Miller, I., & Miller, M. 2004. *John E. Freund's Mathematical Statistics with Applications, Seventh Edition.* Prentice Hall.

Mills, J. D. 2002. Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1): 1-20. https://doi.org/10.1080/10691898.2002.11910548.

Mills, J. D. 2005. Learning abstract statistics concepts using simulation. *Educational Research Quarterly, 28*(4): 18–33.

Stock, J.H., & Watson, M.W. 2015. *Introduction to Econometrics, Update, 3rd Edition*. Pearson.

Tishkovskaya, S., and Lancaster, G. A. 2010. Teaching strategies to promote statistical literacy: Review and implementation. *Data and Context in Statistics Education: Towards an Evidence-Based Society. Proceedings of the Eighth International Conference on Teaching Statistics.* Voorburg, The

Netherlands: International Statistical Institute. http://iase-web.org/documents/papers/icots8/ICOTS8_C193_TISHKOVSKAY.pdf.

Tversky, B., Morrison, J. B., & Betrancourt, M. 2002. Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*(4): 247–262. https://doi.org/10.1006/ijhc.1017.

Wang, P-Y., Vaughn, B. K., & Liu, M. 2011. The Impact of animation interactivity on novices' learning of introductory statistics. *Computers and Education, 56*(1): 300–311.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*. *Second Edition.* Cambridge: MIT Press.