

A commentary on socioeconomic data

Interpreting Economic and Social Data: A Foundation of Descriptive Statistics. By Othmar W. Winkler, New York, NY, Springer Publishing, 2009, 265 pp., \$119.00/hardback.

Natural scientists apply the fundamental assumptions of statistics to experimental data to draw conclusions about natural phenomena. Social scientists use the same methodology with socioeconomic data to create dynamic models of human behavior. Othmar Winkler's *Interpreting Economic and Social Data* calls into question the tendency of social scientists to treat quantitative summary data as objective measurements, as occurs in the natural sciences. Winkler's observations on the subject are both thought-provoking and insightful.

Measurement in natural science is performed using uniform building blocks. By contrast, socioeconomic data emerges from "real-life-objects," which are the projecting agents of socioeconomic phenomena: households, firms, contracts, and sales, to name just a few examples. Relying primarily on the statistical survey, these objects are reduced to "statistical-counting-units" or "still-pictures...somewhat like a photographic snapshot—except that less detail is retained." Since these counting units are typically self-reported and the samples are usually subjectively chosen, it follows that they are hardly truly random. Furthermore, socioeconomic data can be influenced by a number of factors that are of little concern in natural science, such as the place and time

period the data are collected. As Winkler says, "the assumption that [socioeconomic] data are only random deviations from some 'true value' is a carryover from the thinking developed in the natural sciences," and "to analyze them with statistical methods based on inference and on the concept of random sampling is pseudoscience."

This indictment serves as the basis for much of the book. Winkler, a professor of business and economic statistics, is devoted to rehabilitating the proper treatment and interpretation of socioeconomic data. He laments that introductory statistical texts increasingly emphasize inferential statistics to the exclusion of descriptive statistics, the traditional domain of social scientists. This book aims to reverse the trend, and would serve nicely as a complement to the typical formula-driven undergraduate or early graduate-level text. Although the author strives to be straightforward, the book requires an understanding of concepts such as time series, frequency distributions, probability, and linear regression.

The early chapters focus on the structure and nature of socioeconomic data. Through aggregation, statistical-counting-units can be organized into feasible units of analysis; ratios allow these aggregates to be put in context with each other. After some conceptual framework, Winkler cautions on the "loss of meaning in aggregation;" in other words, that it is easy to lose sight of the phenomenon of interest as the level of aggregation grows. He also warns that specification is important because there is a temptation to produce ratios with entirely unrelated aggregates. An example of this would be producing a measure of

accidents per hours worked; a better ratio would be accidents per hours worked in a given industry, since many industries are not especially dangerous.

The middle, and largest, portions of the text focuses on longitudinal analysis. Winkler notes time series data seldom reveal universal economic laws and instead tend to be a product of the historical condition and landscape in which the data were recorded. Because of this fact, socioeconomic time series data utilized in forecasting models will eventually become obsolete as broader societal changes take place. Winkler cautions forecasters against the temptation to treat socioeconomic time series data as "random samples from some hypothetical timeless populations," which can lead to a sense of complacency as the number of observations increases. This word of warning seems pertinent, especially in the aftermath of the financial crisis of 2008.

Winkler's discussion of longitudinal analysis drifts into the realm of price statistics. He asserts that transactions of money for goods should be thought of as the real-life-object and the price paid as the statistical-count-unit, which can vary widely depending on the place and time of purchase. Winkler suggests utilizing scanner data to produce a measure of average currency paid per transaction. This would negate the need to alter the basket of goods used in a price index as products enter or exit the market. Winkler also expresses concern that current measures of labor productivity focused on total hours of labor and output may fail to account for capital improvements over time.

Only late in the book does Winkler

approach cross-sectional analysis, beginning with a chapter on the interpretation of frequency distributions, central tendency, and dispersion. Turning to regression analysis, he notes that linear regression models originated in natural science are poorly suited to socioeconomic data, which is typically plagued by problems such as heteroskedasticity and low R-values. As a result, the line of best fit produced in regression equations is often fraught with misspecification issues and is likely to miss more complex relationships underlying the data. To remedy this, Winkler suggests using disaggregated data to the extent possible, and cautions against interpreting slope coefficients literally. A later chapter on the intersection of socioeconomic statistics and probability carries on

in this vein. The most common mistake is the application of statistical inference to populations or deliberately selected samples deemed “representative.” Modern social scientists tend to “view every situation as a random process or a random experiment, regardless of whether randomization was involved.” Winkler describes these misuses of inference as a trend, reinforced by statistics textbooks and journal editors. However, it seems likely the expanded use of mathematical statistics in social science is here to stay.

The book contains numerous diagrams to visually illustrate and reinforce the concepts described. The endnotes of each chapter contain detailed asides and citations should the reader be interested in pursuing any particular subject at length. Several

chapters also contain appendices that cover topics that may require a refresher, or need additional space to explore concepts mathematically. There also are two short chapters focused on the use of statistics in accounting and geography.

The writing in this book is easy to digest, although in the later portions it can seem repetitive, likely because it is intended for use as a reference. Overall, this is a very practical book; it would serve an aspiring social scientist or an experienced practitioner well to work through its lessons.

—Thomas Luke Spreen
Economist

Bureau of Labor Statistics
Division of Labor Force Statistics

Book review interest?

Interested in reviewing a book for the *Monthly Labor Review*? We have a number of books by distinguished authors on economics, industrial relations, other social sciences, and related issues waiting to be reviewed. Please contact us via email at mlr@bls.gov for more information.