

## **“Six Challenges for the Statistical Community”**

**As Prepared for Delivery at the Joint Statistical Meetings, July 30, 2012**

**Alan B. Krueger**

**Chairman, Council of Economic Advisers**

Thank you very much [Bob Rodriguez] for your introduction and for inviting me to give the President’s Invited Lecture at the Joint Statistical Meetings this year. I also want to commend you for the theme you have set for this year’s Meeting: using statistics to serve a data dependent society.

I thought I would do something a bit uncharacteristic, and tell you about my background in statistics and economics and how it comes into play in my current job. I also want to use this occasion to challenge the statistical research community – or some subset of you – to work on a set of statistical problems that I have encountered in my capacity as an economic adviser.

My current job, as Bob mentioned, is serving as Chairman of the President’s Council of Economic Advisers. When President Obama announced my nomination, he said, “I rely on the Council of Economic Advisers to provide unvarnished analysis and recommendations not based on politics, not based on narrow interests, but based on the best evidence, based on what’s going to do the most good for the most people in this country.” Obviously, determining the best available evidence and what’s going to do the most good for the country requires knowledge of economics and statistics.

When I started college I had no intention of studying economics or statistics. I planned to be a lawyer, a profession in which I thought I could represent clients in need of help. I only took statistics because two semesters were required. Like many others in my class at Cornell, I feared the introductory stats course. My course was taught by Paul Velleman, who was a student of John Tukey, and a firm believer in Exploratory Data Analysis. Fortunately, Paul appreciated that many of my classmates were math averse. He spent days describing basic ideas like the mean, median, box plot, and stem and leaf display. He taught the resistant line along with OLS. I discovered that these concepts were intuitive and central to understanding the world around me.

By the time he taught hypothesis testing and confidence intervals I was hooked. I remember sending my mother, a first-grade teacher, a long letter explaining the difference between Type I and Type II errors, and the tradeoff between them with a fixed sample size. The letter had neatly drawn diagrams, showing overlapping distributions under the null and alternative hypotheses. She had no idea what I was talking about. But to me the idea that one could use statistics to make scientific statements and test hypotheses about economic and social phenomenon like inequality was transformative.

After Velleman's introductory courses I took two courses from Phillip J. McCarthy on survey design and multivariate methods. (I later learned that P.J. McCarthy was the first to use Markov chain methods to model labor force flows.) These courses taught me that statistics could be used proactively, to design and carry out surveys when the appropriate data were not available.

Around the time I was taking these courses I came to the view that the field of economics had more potential to positively address society's problems than the field of law. Economics as taught in college, however, is mainly a set of principles. These principles, while very logical and intuitive, are by and large untested. It was then that I decided that my interests in statistics could be married with my excitement about economics.

Of course, there is a giant field of econometrics that does exactly this, but much of econometrics assumes the model is correct and goes about extracting the relevant parameters in the most efficient way possible. In fact, when I first took econometrics in graduate school it took me the longest time to figure out how one could specify a likelihood function after Paul Velleman drilled into me that we should use methods that are robust to underlying distributional assumptions. The methods I was learning didn't seem like the ideal way to test theoretical models.

Anyway, in much of my research as an academic economist I tried to identify natural experiments where the forces of nature have conspired to produce something like a randomized experiment to test economic theories and estimate key parameters. For example, Josh Angrist and I compared wages of students who had more or less schooling depending on where their birthday fell in relation to the school start age, which influenced the grade they were in when they were no longer bound by the compulsory schooling law. Unless one believes in astrology, it seemed to me that having one's birthday fall on one side or the other of school-start-age was as close to random assignment as we were going to get in variability in educational attainment.

The problems that micro economists are involved in mainly involve causal modeling in some way or another.

But I also discovered that producing descriptive statistics on important economic phenomenon could be of enormous value, both for economics and public policy. In one classic example, Richard Freeman and Brian Hall designed a survey in 1985 to measure the number of homeless people in the U.S. Their estimates, which were highly controversial at the time, were later found to be very close to more sophisticated Census Bureau estimates. At Princeton I founded the Survey Research Center in 1992 to help economists and other social scientists, as well as students, collect new data. In one recent project, the Survey Center helped Alan Blinder and me to measure the fraction of jobs that are susceptible to being offshored. And Morris Kleiner and I included questions on the same survey to measure the number of workers who are required to obtain a license to do their job.

My interest in surveys also intersected with another of my frustrations with econometrics – we take the data as given and try to use ever more sophisticated econometric techniques to overcome weaknesses of the data. Wouldn't it be better to collect more appropriate data? This outlook helped my colleague Orley Ashenfelter and me to address a longstanding puzzle in labor economics: Research on identical twins had found that, controlling for twin-pair effects, the payoff from education was much lower than that found in the rest of the literature. One problem with this approach, identified by Zvi Griliches and others, was that identical twins had very similar education levels, so any differences were likely to be dominated by measurement errors. Sophisticated econometric methods were thrown at this problem.

Ashenfelter and I decided to collect multiple measures of the twins' education. To do this, we went to the world's largest twins festival and conducted our own survey. Specifically, we separated the twins and asked them about their own and their twin's education level. So we had data from twin "i" and twin "j" on twin i's education. The correlation was about 0.9, high, but not very high considering that the correlation between i and j's true education was around 0.6. Almost half of the variance in the difference in education between pairs of identical twins was statistical noise. With the data we collected, we could do something very simple, like use the average education reported by i and j to measure i's education. When we did this, our results showed a high payoff to education, like the rest of the literature. My point is that by collecting our own data we were able to solve a longstanding estimation puzzle.

Now I should fast forward to the present. In my current role I am much more of a consumer of statistics than a producer. While I was Chief Economist of the Treasury Department I helped to develop a confidence interval for when the government would hit the debt ceiling, but that was a rare incidence of applying statistics to a new problem.

In many more situations, however, I come across problems where I say – wouldn't it be nice if someone had solved this.

Let me give you *six* challenges that I see as very important for the statistics and policy communities.

The **first challenge** involves developing appropriate methods and infrastructure to take full advantage of the commercial, administrative and use-related data that are being captured at an exponential rate, with adequate privacy protections. Larger and larger caches of data are becoming available in new and ingenious ways. These include data on credit card transactions, scanner data on retail sales, Google searches for various terms, eBay bids and sales, LinkedIn members' occupational transitions and Monster.com job listings. I'm sure this list just scratches the tip of the iceberg given the spread of information technology.

There is enormous potential to use these Big Data sets to cut survey costs and reduce respondent burden, to improve and expand our existing social and economic indicators and make them more timely, to assess the reliability of traditional survey data, to study network and GIS-related

issues, and to answer a myriad of previously unanswerable questions. But the risks of misusing the data are also high.

The sampling methods that I learned from P.J. McCarthy were invented in part because it was too expensive to analyze data on an entire universe, and in part because the only way to get data was to go out and collect it. This is why the BLS designed the Consumer Price Index and started surveying price data during World War I. But scanners now capture an enormous volume of retail trade transactions every day. And computing power makes it possible to analyze huge quantities of data.

While the goal of sampling – to produce representative and accurate statistics – remains paramount, that goal can often be achieved more cost effectively and with more precision by using data that are administratively collected, or, more likely, by *combining* such data with other information.

Now, I do not mean to suggest that this process is straightforward or easy. It is not. Indeed, much of the administrative data that is captured is proprietary, and nonrepresentative. There are few safeguards to ensure transparency and that the data are accurately reported and evenhandedly analyzed, and there often is little effort applied to adjust for the nonrepresentativeness of the data.

This does not stop people from using such data, however. Prominent examples include: the ICSC retail chain store sales index (which represents weekly sales from a fixed set of stores from major chains); the ADP/Macroadvisors employment projection; and various indices of help wanted ads.

To encourage more research on “Big Data” in March 2012 the Obama Administration launched a \$200 million “Big Data Initiative” spearheaded by the White House Office of Science and Technology Policy in concert with several other Federal departments. As OSTP Director John Holdren said at the time, “In the same way that past Federal investments in information-technology R&D led to dramatic advances in supercomputing and the creation of the Internet, the initiative we are launching today promises to transform our ability to use Big Data for scientific discovery, environmental and biomedical research, education and national security.”

The main statistical challenge for economic data and research, it seems to me, is to design ways to appropriately weight data that are collected administratively from a large but partial share of the relevant universe. In many situations, we could use statistical techniques to combine or blend the Big Data sets with more traditional survey data. In addition, noisy data will need to be filtered, and privacy must be protected. And efficient computational methods of processing data will likely be necessary.

My **second** challenge involves sample weights in traditional surveys. Techniques developed in this area can possibly be applied to nonsurvey data as well.

The typical private sector survey has a response rate of only around 5 to 20 percent, and even high quality surveys by research institutes can have response rates around 60 or 70 percent. If nonresponse was unrelated to the underlying phenomenon of interest, the response rate would not matter, and the sample of respondents could be used to represent the phenomenon of interest. Interestingly, a meta-analysis of research by Bob Groves and Emilia Peytcheva that compared sample estimates to population estimates of the same parameter found that, looking across many studies, the response rate is unrelated to the estimation bias in many situations.<sup>1</sup>

One conclusion I draw from this finding is that all surveys need to worry about sample weights. There is a large and well developed literature on sample weights that mainly involves adjusting for intentional sample design features. But an equally important problem is respondent nonresponse for behavioral reasons. In practice, sample weights are often derived by rules of thumb or stepwise regression procedures. The added variability due to the uncertainty surrounding sample weights is rarely, if ever, reported and accounted for in standard errors.

We tend to exaggerate the precision of our estimates by ignoring the estimation errors introduced by nonrepresentative samples even when we have weighted the sample by estimated sample weights.

So my suggestion is that more analytical work can be done on sample weights, and we can think about additional data that can be surveyed to assist with the development of nonresponse weights. At a minimum, the variability due to the estimation of sample weights should be routinely reported and incorporated into standard errors.

But a more ambitious agenda would be to develop methods for deriving robust sample weights. My challenge to you is to develop new techniques for deriving sample weights that are robust to various forms of nonresponse biases, or that bound the effect of nonresponse bias.

A **third** area in need of renewed research that I want to highlight involves seasonal adjustments. There is an extensive literature on seasonal adjustments from the 1970s and 1980s, but relatively little work has been done recently. Seasonal adjustment is a particularly challenging statistical problem because the estimand is not well defined and because seasonal patterns can change.

Another challenge in economic data is to separate the business cycle from seasonal factors. This problem is exacerbated when seasonal factors evolve.

Commonly used seasonal adjustment packages take very little external information into account, such as temperature or snowfall. Research could look into the optimal way of combining external information related to seasonal movements in relevant series, especially to help model seasonal adjustment factors that could be evolving.

---

<sup>1</sup> Groves, Robert M., and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72(2): 167-189.

Advances in nonparametric and semiparametric methods, as well as computational methods, could also help to improve the estimation of seasonal adjustment factors. In addition, in applied work attention is rarely paid to the extra variation introduced due to the fact that seasonal factors are estimated, rather than known with certainty.

A **fourth** area that I would like to challenge you to conduct research involves traditional evaluation methods as applied to budget-related issues. The Office of Management and Budget issued a landmark call for Federal agencies to use evidence and rigorous evaluation in budget, management and policy decisions to make government work more effectively. Acting Director Jeffrey Zients requested that agencies “demonstrate the use of evidence throughout their Fiscal Year 2014 budget submissions” and that Budget submissions “include a separate section on agencies’ most innovative uses of evidence and evaluation.” He specifically highlighted that agencies can often use administrative data to conduct rigorous evaluations, including evaluations that rely on random assignment.

Indeed, Zients’s memo stated explicitly: “One of the best ways to learn about a program is to test variations and subject them to evaluation, using some element of random assignment or a scientifically controlled design.” OMB is encouraging agencies to use waiver authority to evaluate different approaches to improving outcomes.

I want to underscore how sweeping a change this is in the operation of government. For the first time, agencies are expected to have a high-level official who is responsible for program evaluation, such as a chief evaluation officer, and agencies are invited to strengthen their evaluation capacity in the Budget submissions.

CEA, together with OMB has begun holding a series of discussions with senior policy officials and researchers in the relevant agencies to ensure that an increasing share of Federal policies follow evidence-based practices.

This unprecedented initiative should be music to the ears of statisticians. While RCT methods have been employed to great effect ever since R.A. Fisher accepted a job at Rothamsted Agricultural Station, issues unique to the Federal Budget and government programs provide some new challenges for statisticians. For example, as Paul Rosenfeld, Don Rubin and others have shown, interpreting RCTs when there is partial compliance is not always straightforward.<sup>2</sup> Furthermore, ingenuity could be used in the process of implementing waivers to produce unbiased estimates for the relevant population as a whole, as well as for specific subsets. Another question is whether insights gleaned from randomized designed research can be optimally combined with incoming observational data on participants in ongoing programs to adjust ongoing programs in real time.

---

<sup>2</sup> For a recent survey of the literature, see Guido Imbens and Jeffrey Woolridge, “Recent Developments in the Econometrics of Program Evaluation,” *Journal of Economic Literature* 47:1, 2009, pp. 5-86.

While existing statistical tools are sufficient to support this initiative, the potential for applying evaluation methods to an expansive list of programs and in new ways provides many opportunities for research and innovation.

The **fifth** challenge on my list concerns one of the greatest threats facing our Nation: the decades' long rise in inequality and decline of the middle class. By any measure, the income distribution has become more kurtotic. We have gone from 50 percent of households earning within 50 percent of the median income in 1970 to 44 percent in 2000 and 42 percent today. The hollowing out of the middle class is a fundamental threat to our economy.

In the past I have raised the concern that rising inequality is jeopardizing our tradition of equality of opportunity. One piece of evidence comes from comparing social mobility across generations and cross-sectional inequality across countries. This figure shows what I call the "Great Gatsby Curve": the relationship between income mobility across generations, as measured by the slope from a regression of son's log income on father's log income, and inequality in the parent's generation, as measured by the Gini coefficient for after-tax income in the mid-1980s. Countries that have a high degree of inequality also tend to have less economic mobility across generations.

The rise in inequality in the U.S. is a concern because it portends a decline in social mobility. Our nation benefits when gifted children from low-income families have the opportunity to fully develop and exploit their talents. Yet a child from a family making less than \$35,000 a year who scores around 1100 on the SATs has a lower chance of graduating college than a less gifted child who scores around 950 on the SATs if his or her parents earn over \$100,000 per year.<sup>3</sup>

New research by Robert Putnam finds that children born to more affluent parents are increasingly raised in ways that lead to more nurturing and supportive environments, while there is little if any improvement for children born to less affluent parents. Since the early 1980s, for example, participation in nonsports extracurricular activities has been rising for children born in the top quartile of the SES distribution and falling for children born in the bottom quartile. There is also a growing gap based on parental education in the amount of time that parents spend with their children. These changes forebode a growing opportunities gap between children born to affluent and poor parents.

It is easy to devise theoretical models that predict that a rise in income inequality leads to a decline in intergenerational mobility. For example, in the Becker and Tomes human capital model, education of parents and their children is positively correlated. A rise in the education-earnings gradient over time therefore causes a rise in the correlation in income between parents and children. But there may be other mechanisms at work. And if we understand the mechanisms, we can design public policy initiatives to support intergenerational opportunity

---

<sup>3</sup> Based on calculations from the Department of Education's 2009 Beginning Postsecondary Students (BPS) longitudinal survey. All individuals in the sample began post-secondary school, and the tabulations refer to the proportion who graduated with a Bachelor's degree within six years of starting. Family income is from 2002, the year before postsecondary-school enrollment.

despite the past rise in inequality. For example, by increasing Pell grant generosity and making student loans more affordable, children of less affluent families can overcome financial obstacles to completing college.

So my challenge is to study the mechanisms underlying the Great Gatsby Curve. If we can learn more about the channels that restrict opportunities for children from disadvantaged families, we can more effectively design and implement strategies that support economic mobility.

My **sixth** and final challenge involves our financial regulatory system. Our universities and business schools have done an outstanding job developing techniques and teaching methods for financial engineering and financial econometrics – all tools for making money. But we have not devoted nearly as much effort to developing tools for detecting financial fraud.<sup>4</sup>

There is a tremendous opportunity here. Financial regulators, notably the Commodities Futures Trading Commission (CFTC) and Securities and Exchange Commission (SEC), have reams of data flowing in on financial transactions. Their tools for financial surveillance and enforcement were recently strengthened by the Dodd-Frank Wall Street Reform Act, which, among many other things, requires new financial players such as swap dealers to register and report to the financial authorities.

We need appropriate statistical surveillance tools to analyze the massive amount of data that flows in to detect suspicious behavior, market manipulation, and collusion.

The CFTC alone receives from 400-600 million records on time and sales each day from over 200,000 customer accounts. With such massive amounts of data, the regulators need cluster techniques to reduce the data into traders who follow similar strategies. Moreover, some of the traders are not people, but machine trading algorithms. Machine learning techniques are needed to reverse engineer the algorithms that the machines are using to detect whether the algorithms change, and why. Enforcement actions often result when a trader brags that he rigged prices or cornered a market, but machine algorithms do not send emails or go to bars. We need new methods to detect suspicious behavior, and I can think of no better tool than statistics.

Hidden Markov Chain and nonlinear statistical models can be developed to detect discontinuities in market behavior. In addition, financial regulators would benefit from the development of network analytic tools that link traders, banks and payment system nodes, and then determine what happens when a particular link drops out.

---

<sup>4</sup> One notable exception is William Christie and Paul Schultz (1994), who found that trades of NASDAQ stocks were unlikely to occur at odd eighths-of-a-dollar. Buyers were more likely to quote shares at 17 1/2 or 17 3/4 than at 17 5/8, for example. Their statistical finding resulted in 30 securities firms eventually paying a \$910 million settlement to end a class-action suit alleging price fixing in the NASDAQ exchange.



The regulators are eager to have more statisticians work on these problems, and vast quantities of data can be made available.

What I have in mind is building a new subfield of statistics that can assist financial regulators in monitoring markets to ensure that they are fair and orderly, to flag suspicious behavior and to detect new sources of systemic risk. Someday, I hope that enrollment in financial regulatory statistics rivals that in financial engineering. The students may not be as well remunerated, but their value to society will be higher.

## **Conclusion**

Let me conclude by emphasizing that we are meeting at a time when there is great opportunity to use statistics for the good of society. Research on the challenges that I mentioned – Big Data, sample weights, seasonal adjustment, evidence-based policy, social mobility, and statistical tools for financial market surveillance – can serve our increasingly data dependent country.

But we are also meeting at a time when our statistical infrastructure is under pressure. A budget bill passed by the House of Representatives in May would eliminate the American Community Survey and the Economic Censuses. These data sets are essential public goods that help households and businesses make better decisions. Private label data that are becoming available are not a substitute for core government surveys. In fact, the ACS and Economic Censuses are even more valuable to businesses and households because they provide the scaffolding to interpret and benchmark the big private data sets that are becoming available.

Another threat that we face is that special interests have been trying to reduce funding for the financial regulators to undermine the Dodd-Frank Wall Street Reform Act. If successful, this would make it even harder to police financial manipulation and reign in systemic risk, and it would reduce the resources needed to develop and implement statistical tools to ensure that markets operate fairly and efficiently.

As a researcher, I spent much of my time and energy on methodological debates. As a policy advisor, I have learned to be much less doctrinaire about economic and statistical methods. Some tools are useful for some questions, and other tools useful for other questions. And sometimes it is useful to represent the range of estimates across a variety of methods. What matters more than I previously appreciated is the general problem to which we apply and develop tools. So, in conclusion, my last challenge to you is to pursue work in areas that, as President Obama requested, “will do the most good for the most people in this country.”