Economists in Data Analytics

Steven C. Myers

(updated) May 11, 2016

I reached out to Ken Stanford of SAS Institute back in late 2014 to further development of my econometrics courses and to discuss a presentation of his "Why economists make good data scientists." Our conversation was followed with another and an active correspondence was started including a very informative session with members of the SAS Academic Outreach and Collaborations Manager. I also met with data analytic professionals at or near the CIO level to discuss their perspective of data analytic demands and what is needed on our part to address market needs for data analysts.

Our experience in the economics department is that our graduates, especially from the graduate program, are nearly all employed and doing some form of data analytics. All of our students have experience working in teams, programming, and working with large datasets.[1] They are used to unstructured problem solving, data cleaning and economic and econometric model specification. They have the soft skills as well as the hard analytic skills that employers want. They have an interest in understanding the data generating process and making sense of why a relationship exists. They know that a perfect statistical fit is of little help when that fit goes against and rejects the economic model and hypotheses generated from that modeling. A strong statistical fit that predicts an upward sloping demand curve comes to mind.

Economists are trained in causality and are interested in (if not obsessed with) the data generating process. The data represents actors in the economy who behave in their own interest under constraints and economists are especially well trained to make sense of the underlying behavior leading to the particular collection of data. Why a consumer makes a purchase or not, why an investor decides to buy a property or invest in an instrument, why a worker leaves a job, or why one unemployed refuses an offer of employment are all the purview of the economist. Data

Why economists make good data scientists

Ken Sanford offers 5 reasons why Economists are poised to break into data science roles:

1. We understand objective functions
2. Economists have a very strong linear regression toolkit
3. We own observational data and causality
4. We have experience in articulating the problem and the solution
5. We work with Big Data

Ken Sanford. Why corporate economists are hot again and a great source for analytical talent, Subconscious Musings (blog), August 27, 2014 accessed at http://blogs.sas.com/content/subconsciousmusings/2014/08/27/why-corporate-economists-are-hot-again-and-a-great-source-for-analytical-talent/ on Dec 2, 2014.

---

[1] Large here is typically many thousands to tens of thousands of records. Big data as it is discussed in the popular press may be many times that amount, but the skills are transferable.

represents the observations on purchases, but they come about because of the human behavior economics is able to model.

Economists make good story tellers. Asking "why" and not just accepting facts as stated or statistically revealed is a type of occupational hazard for economists and a good one for employers. One of the former CIOs I spoke to expressed excitement that economists understood causality and the data generating process, as his experience was most analysts do not. Knowing that data has revealed a correlation or trend as one might find in a statistical analysis does not explain why the correlation or trend exists. It also does not necessarily allow itself to be subject to other influences that may matter and that are not examined unless one is asking the "why" question. Additionally, the data is typically never clean as might be experimental data or results of a quality control investigation. In fact, economists have to learn to deal with data that is observational and subject to all sorts of dirtying influences. Harvard Economist, Zvi Griliches, once remarked that if it weren't for dirty data, economists wouldn't have jobs. This "dirty" data is to say that lots of influences are in the data, that one must pay strict attention to the data generation process. Knowing the data well and understanding how it is generated allows an economist to fashion a model solution that cannot just forecast a trend or display a correlation, but to actually predict.

The data analytic world speaks of predictive modeling and predictive accuracy, but this is not what an economist means by prediction. In the data analytic world a predictively accurate analysis would be considered by an economist as the result of a descriptive exercise. In such descriptive analysis, tools of statistical analysis are applied to data in order to mine the data, describe the data or model the patterns in the data. The data analyst so trained would seek a best fit and accurate representation of a next or future data point and provide for statistical significant results. When this analysis is done on time-series data an economist refers to this as a forecast. Many economists do participate and excel in forecasting, but what makes an economist different from a statistician is not the statistical modeling, even though the statistician may be more skilled at that, but that the economist first builds an economic model; a model that explains "why."

A statistical model is not an economic model. Statistical significance is a necessary condition, but not a sufficient condition for economic significance. The economist does use a statistical model, but the econometrician uses it in context with causality and the DGP (data generating process), hereinafter referred to as the econometric model.

So the first step for the economist is to create an economic model of the actors from which the data is generated. So if retail point-of-sale transactions are being analyzed, the economist seeks to model the demand of the consumer for the product and must also take into account the supply of product from the retail establishment. The economic model would apply economic theory of demand and supply and would formulate hypotheses that would be expected to be seen in the data. Essentially the economist is concerned not just with how many items were sold, but why were they bought? The economist is

modeling the casual relationships and this theory may be best thought of as a story. A story about why consumers come into a retail establishment and why they purchase an item or items while simultaneously considering the other influences on the decision such as the income of the consumers, other prices of complementary and substitutive goods, and more. Economists make good story tellers.

What skills then do economists need to be data analysts? In a word they need to be trained in econometrics; you want to hire an econometrician. These econometricians must be economists first, then statisticians, and finally computer programmers. What makes econometricians so valuable in data analysis is that they are highly competent in all three areas: economics, statistics and computer programming.

But you don't want to hire any econometrician, but you want one that is highly trained in applied econometrics. Theoretical econometricians and classically trained econometrician will have strong skills in statistical inference. Statistical inference is the use of statistics to infer results and requires a classical training in how to estimate results (such as correlations and regressions) and how to preform hypothesis testing (such as will sales next period statistically and significantly exceed last period). All economists who have taken econometrics are trained in this classical inference of estimation and hypothesis testing.

Inference	
Classical econometrics	**Applied econometrics**
Estimation techniques	Problem Articulation / Research Hypothesis
Hypothesis testing techniques	Data finding / cleaning / management
	Model Specification

Applied econometrics goes well beyond the classical training in inference briefly discussed above. Peter Kennedy (2011) describes the three pillars of applied econometrics as (1) problem articulation, (2) data cleaning, and (3) model specification. It is these three pillars that a CIO should seek to find among his/her data analysts. Applied econometricians offer all the three.

The first of these is the ability of a researcher to deal with an unstructured problem and make sense out of it. As W. Lee Hansen observed, the highest level of learning is the ability to ask the right question. It is the derivation of the question out of the unstructured problem that is a key asset of an applied econometrician. Put to use will be economic reasoning and a story will emerge as to how and why the economic agents behave. From the economic model will derive the hypotheses that will determine that the economic story is be "accepted" or rejected. It is from this model and understanding that causality is hypothesized and that an explanation of behavior including prediction is forthcoming.

Kennedy refers to the second of these pillars as Data Cleaning. An unsatisfying term as economists regard this as much more than mere cleaning of a dataset, but the wording is instructive in that economists regard all observational data as dirty. And dirty data will need cleaned. But more than that, it is important to understand the data generating process (DGP). Is the data non-random, censored, truncated, self-selected? Are relevant variables available or must proxies or instruments be used. Are

good instruments available? Among the variables prepared, are they exogenous or endogenous, and if the latter can the data support the system wide economic model.

The third pillar of model specification requires the applied econometrician to build a model of available data variables that matches the theoretical model as articulated in the first step of problem articulation. Given the dirty nature of the data and the attempts to clean it, the model specification gives rise to other issues (some of which are a direct result of the DGP and the quality of the variables). Other issues come directly from the nature of the problem as articulated, from the economic model itself. In the latter case if there is known endogeneity of right hand side variables then this effects the model specification and hence the estimation and testing strategy.

Economists as Data Scientists
Evidence from SAS's Econometric Evangelist

Background for why economics must be a part of any data analytic effort!

This is a list of YouTube and blog entries by Ken Sanford, SAS's Econometric Evangelist, on how and why economists make great data analysts.

[1]   Ken Sanford. Econometric reflections from Analytics 2014, Subconscious Musings (blog), Nov 3, 2014 accessed at http://blogs.sas.com/content/subconsciousmusings/2014/11/03/econometric-reflections-from-analytics-2014/ on Dec 2, 2014.

Ken discusses various items, but the first is the most important and is his joint presentation with Jan Chvosta titles, "*Why Econometrics Should Be in Your Analytics Toolkit: Applications of Causal Inference*" (available at http://gozips.uakron.edu/~myers/E627/Sanford_Ken_Chvosta_Jan_A2014.pdf)

Partial quote: "Of note to me was just how many audience members approached us afterward and said that 'causal interpretation' is what they strive for with their predictive modeling. From marketing mix models to CCAR stress testing to price elasticity estimates, I saw many nodding heads when we talked the importance of interpretation in these models. To twist the words of Nobel Laureate Robert Lucas, "once you start thinking about causality, it is hard to think about anything else." It appears to me that there are still many people interested in the meaning of models in this world of 'big data' and 'machine learning.'"

[2]   Ken Sanford. Why corporate economists are hot again and a great source for analytical talent, Subconscious Musings (blog), August 27, 2014 accessed at http://blogs.sas.com/content/subconsciousmusings/2014/08/27/why-corporate-economists-are-hot-again-and-a-great-source-for-analytical-talent/ on Dec 2, 2014.

In this video, Ken offers 5 reasons why Economists are poised to break into data science roles:

1. We understand objective functions
2. Economists have a very strong linear regression toolkit

3. We own observational data and causality
4. We have experience in articulating the problem and the solution
5. We work with Big Data

[3]   Ken Sanford. From Economists to Data Scientist: How our discipline can participate in the growth of analytics. Presentation to AIER. YouTube video (1:22:58), September 9, 2014. https://www.youtube.com/watch?v=KlNkXCkUKAo accessed Dec 2, 2014.

This important 83 minute video is very clear on how economists can become data scientists. Data scientists cover all the areas of data acquisition, data management and transformation, computation and reporting and visualization.

He describes the reasons economists excel and then discusses the barriers to economists entering the data analytic field. They include understanding the technology and the ability to speak with the jargon that business understands. One example, a business many might want data "scored" and an economist won't understand until they learn it is in sample prediction.

[4]   Ken Sanford. Why econometrics is important for business analysis. Youtube Video (2:57) From Analytics 2013 http://youtu.be/ONzG8jJ0i5Y

[5]   Ken Sanford. Economists make Good Data Scientists. YouTube Video (1:29) http://youtu.be/WGvMARRid7w at Analytics 2014