University of Illinois
Fall 2016

Department of Economics
Roger Koenker

Economics 536
Lecture 8

The Regression Fallacy

1. Galton's Regression to Mediocrity

Arguably, the most important statistical graphic ever produced is Galton's (1885) figure illustrating "regression to the mean", reproduced badly below as Figure 1. In it Galton plots childrens' height versus parents' height for a sample of 928 children. He begins by dividing the plane into one inch squares and entering the frequency counts for each square. The resulting "histogram" appeared too rough so he smoothed the plot by averaging the counts within each group of four adjacent squares and plotting the averaged count at the intersection of the cell boundaries. Not content to invent "regression" in one plot, he managed to invent bivariate kernel density estimation, too! After smoothing, the counts appeared more regular and he enlisted the help of the Cambridge mathematician, J.H. Dickson, to draw elliptical contours corresponding to level curves of the underlying population density.

Now suppose we wished to predict children's height based on parental height, say the average height of the parents which we will call, following Galton, the height of the midparent. what would we do? One approach, given the graphical apparatus at hand would be to find the "most likely" value of the child's height given the parents' height, that is, for any given value of the mid-parent height we could ask, what value of the child's height puts us on the highest possible contour of the joint density. This obviously yields a locus of tangencies of the ellipses with horizontal lines in the figure. These conditional modes, given the joint normality implicit in the elliptical contours, are also the conditional medians and means. The slope of the line describing this locus of tangencies is roughly 2/3 so a child with midparent 3 inches taller than average can expected to be (will most probably be) only 2 inches taller than average.

Galton termed this regression towards mediocrity, and paraphrasing Abraham Lincoln we might strengthen this to regression of the mediocre, to the mediocre, and for the mediocre. Children are more mediocre than their parents: they tend to be "on average" closer to the mean height, the mean weight, the mean intelligence of the population than were their parents. In the case of height we have seen that children of parents who are one inch taller than the general population tend on average to be only about two-thirds of an inch taller than the population. But before you despair, I should hasten to point out, as Galton did, that parents are also more mediocre than their children – if we run the usual conventions of temporal causality backward and ask: how do the heights of parents compare to the heights of their children we find (looking at the figure) that children who are unusually tall have parents that are closer than they to the mean height of the population, and children who are unusually short also have parents closer than they are to the mean.

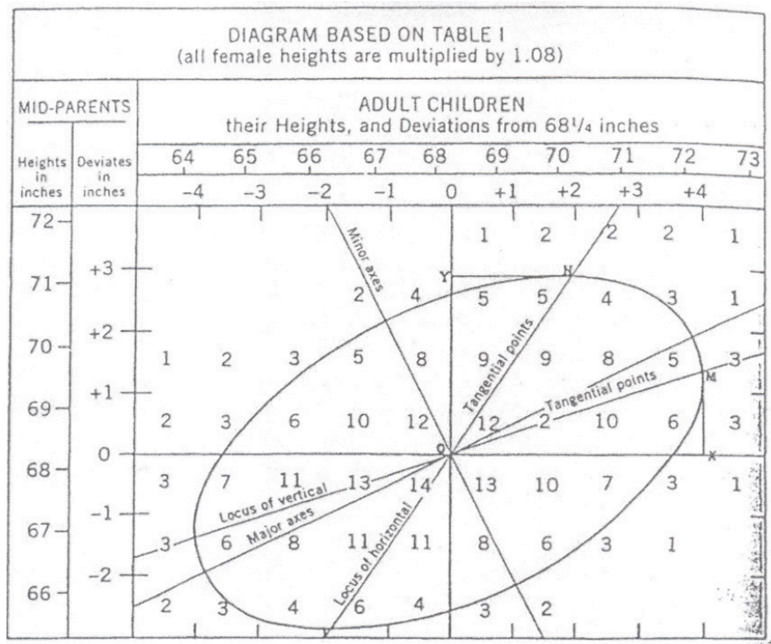

Figure 1. Galton's (1889) Regression to the Mean Plot

Stigler (1997) provides a fascinating guide to Galton's own thinking about this idea, and to the illusive nature of its reception in subsequent statistical research.

It is a remarkable feature of the conditional densities of jointly Gaussian random variables that the conditioning induces what we may call a "pure location shift". In Galton's original example the height of the midparent alters only the location of the center of the conditional density of the child's height; dispersion and shape of the conditional density is invariant to the height of the midparent. This is, of course, the essential feature of the classical linear regression model – the entire effect of the covariates on the response is captured by the location shift

$$E(Y|X = x) = x^\top \beta$$

while the remaining randomness of Y given X may be modeled as an additive error *independent* of X. Just to confirm that this empirical regularity hasn't been repealed over the intervening 100 years I illustrate in Figure 2 a similar plot for a sample of Finnish boys. The results are quite similar to those obtained by Galton.

2. Secrist's Regression to Mediocrity

What does this have to do with econometrics? To answer this question we need to step forward in time about 50 years and consider a book published in 1933 by Horace Secrist. Secrist was a professor of economics at Northwestern University trained at Chicago and an expert in what we would now refer to as Industrial Organization. The book was titled *The Triumph of Mediocrity in Business* and had occupied 10 years of this research. In it Secrist showed in excrutiating detail if you grouped firms into performance categories in some initial year, and then

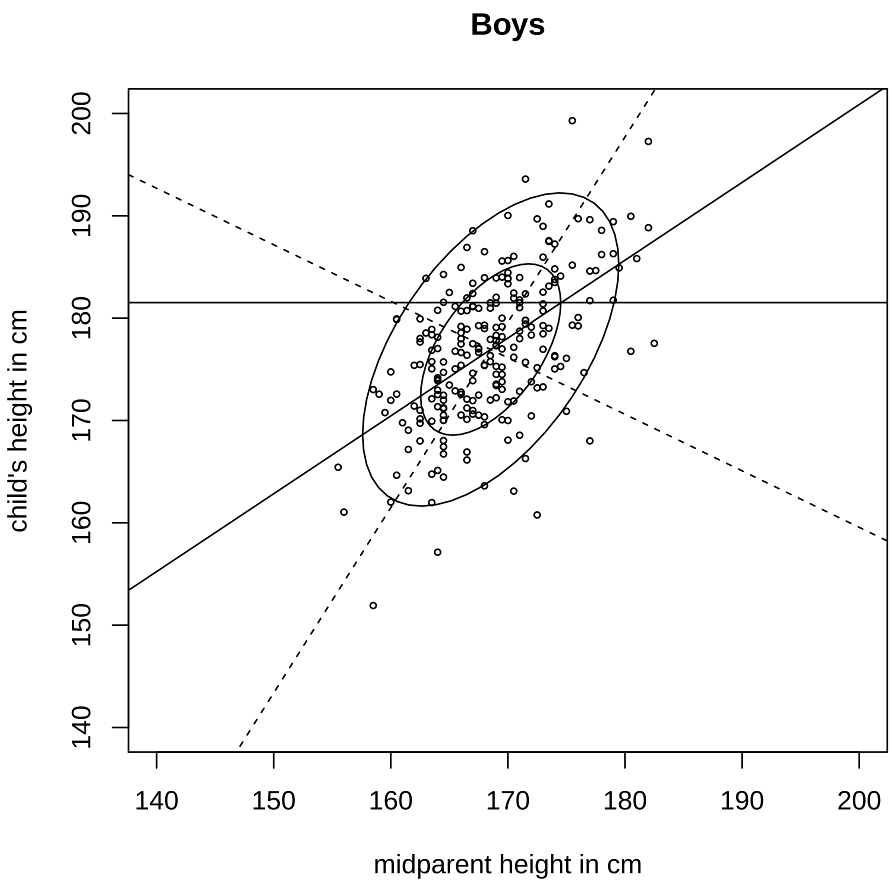

Figure 2. A Modern Galton Inheritance-of-Height Plot: The plot depicts the heights of 236 Finnish boys at age 17 versus the mean height of their parents. The two ellipses represent 50 and 90 percent confidence regions estimated for the pairs of points based upon the conventional bivariate Gaussian model for the data. The dotted lines depict the major and minor axes of the ellipses. The solid line represents the least squares fit; that is, it is the line that minimizes the sum of squared vertical distances from the points to the line. The slope of the line is $\beta = .76$, somewhat larger than the slope of $\frac{2}{3}$ obtained by Galton.

followed them in subsequent years, that the initially most successful tended to do worse over time, while the least successful tended to improve. Secrist's conclusion was that American business was "converging to mediocrity." As Stigler notes, the book was very favorably reviewed by the JPE the AER and several other journals, for JASA however, the reviewer was Harold Hotelling, who was not nearly so easily impressed.

FIGURE 3. Harold Hotelling was born in Fulda Minnesota in 1895, grew up in Seattle, and was educated at the University of Washington, and Princeton University. He taught at Stanford from 1924-31, Columbia from 1931-46, and U. of North Carolina until his retirement. He made important contributions in economics to the theory of utility and demand, public economics, and game theory. His contributions to statistics range extremely widely from the theory of rank tests, and simultaneous confidence regions to multivariate analysis and regression. The festscrift published in 1960 for Hotelling by his students and "long-time associates" in economics includes 6 Nobel Prize winners: Arrow, Friedman, Frisch, Klein, Samuelson, and Vickrey.

Hotelling wrote that "the seeming convergence is a statistical fallacy, resulting from the method of grouping. [Secrist's] diagrams really prove nothing more than that the ratios in question have a tendency to wander about." To illustrate Hotelling's point we may consider the following very simple panel data model. Suppose the profits of firm i at time t evolve according to,

$$y_{it} = \alpha + \rho y_{i,t-1} + u_{it} \tag{1}$$

where for convenience suppose u_{it} is iid $\mathcal{N}(0, \sigma^2)$. Provided that $|\rho| < 1$, we can easily deduce that the stationary distribution of the firms profits is at any time, $\mathcal{N}(\alpha/(1-\rho), \sigma^2/(1-\rho^2))$. Thus, whatever the initial distribution of firms profits might be, if we let the population of firms evolve it will eventually exhibit this distribution. And if we proceed as Horace Secrist did to construct a group of firms that did particularly well in 1920, and another group that did particularly badly, and follow their performance, we will find that the mean and variance of profits

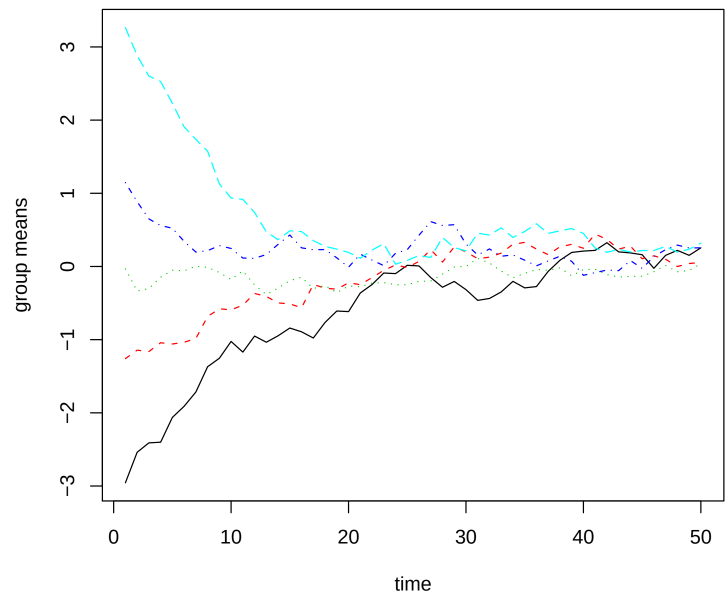

FIGURE 4. Electronic Elephants on Parade: This figure illustrates a simple AR(1) version of the Hotelling-Secrist regression fallacy. 500 AR(1) series of length $T = 100$ were generated with $\rho = .9$. At time $t = 50$ the series were grouped into quintiles and the group means of these quintiles were plotted following Secrist's approach with department stores' profitability. As in Secrist, the group means "converge to mediocrity", which in this case may be interpreted simply as the common stationary mean of the processes.

for all of these groups converges to this same distribution. We illustrate this in Figure 4 where we have generated 500 AR(1) time series of length 100 according to (1). In period 50 the "firms" were grouped into 5 performance categories and then the means of the five groups were followed for the next 50 periods. The means converge even though the over all variability of performance doesn't change at all over this period. Hotelling suggested that if Secrist had grouped firms according to their performance at the end of the period, and repeated the exercise of following them – now backwards in time – they would appear to diverge. This is illustrated for our model (1) in Figure 5.

Secrist was not very happy with Hotelling's review and managed to convince the editors of JASA to publish a letter responding to Hotelling's criticism. This was a tactical blunder since it gave Hotelling an excuse and an opportunity to highlight his attack. He did so in the following devastating paragraph:

> Consider a statistical variate x whose variance does not change from year to year, but for which there is a correlation r between

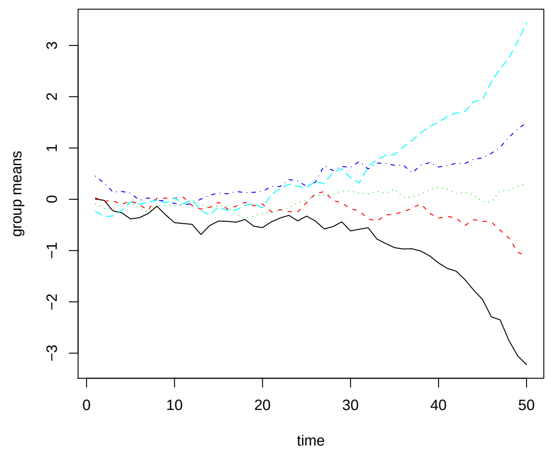

FIGURE 5. Electronic Elephants on Parade (Part 2): This figure illustrates a simple AR(1) version of the Hotelling-Secrist regression fallacy. 500 AR(1) series of length $T = 100$ were generated with $\rho = .9$. This time, at time $t = 100$ the series were grouped into quintiles and the group means of these quintiles were plotted following Hotelling's suggestion for evaluating Secrist's approach with department stores' profitability. As suggested by Hotelling, the group means now "diverge from mediocrity", which in this case may be interpreted simply as the common stationary mean of the grouped processes gradually approaching their known conditional means at the end of the period.

> successive values for the same individual. Let the individuals be grouped so that in a certain year all those in a group have values of x with a narrow range. Then among the mean values in these groups, the variance (calculated with the group frequencies as weights) will in the next year be less than that in the first year in a ratio of which the mean value for linear regression and fine grouping is r^2, but in any case is less than unity. This theorem is proved by simple mathematics. It is illustrated by genetic, astronomical, physical, sociological and other phenomena. To "prove" such a mathematical result by a costly and prolonged numerical study of many kinds of business profit and expense ratios is analogous to proving the multiplication table by arranging elephants in rows and columns, and then doing the same for numerous other kinds of animals. The performance, though perhaps entertaining, and having a certain pedagogical value, is not an important contribution either to zoology or to mathematics.

In a Figures 6 and 7 we illustrate a variant of model (1) in which we have introduced an individual specific fixed effect, α_i for each firm. As the figures illustrate

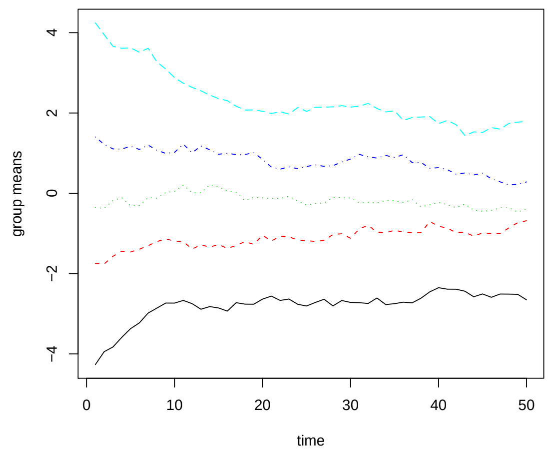

Figure 6. Electronic Elephants on Parade (Part 4): This figure illustrates a simple AR(1) version of the Hotelling-Secrist regression fallacy. 500 AR(1) series of length $T = 100$ were generated with $\rho = .9$. In contrast to the previous example, now each cross-sectional unit has a fixed effect which is drawn from a $\mathcal{N}(0, 2)$ distribution. As in Secrist, the group means still "converge to mediocrity". But note that now the convergence is less pronounced than in the previous case since the groups tend to different group means because of the fixed effects.

this tends to attenuate the effect but one still sees the pronounced tendency of the group means to "converge" or "diverge" depending on whether the grouping was done at the beginning or end of the sample period. Hotelling suggested that a reasonable test of the convergence hypothesis would be to examine the cross-sectional variance to see whether it was becoming smaller over time as the convergence hypothesis would suggest.

There are lots of other examples of research that seems deeply confused about the issues I have tried to highlight above. One example that is a little too close to home is Bass (1958) who basically replicates the Secrist nonsense on drugstores from 1948-52, in conclusion he comments that Robert Ferber suggested to him that the Secrist conclusion could be attributed to "fortuitous circumstances" an interpretation completely consistent with the Hotelling review. It is embarrassing to admit that Bass was an UIUC Phd, but I am happy to report that Bob Ferber was a distinguished member of the faculty in the business school here for many years.

Fast-forwarding another 60 years from Secrist it is fitting to conclude by noting that the estimable Milton Friedman in 1992 felt inspired to revive the Hotelling-Secrist confrontation in a short note in the *Journal of Economic Literature*. In this note he recalls the Hotelling review and comments that it lay behind his formulation of the permanent income hypothesis. Friedman was a student of Hotelling's at

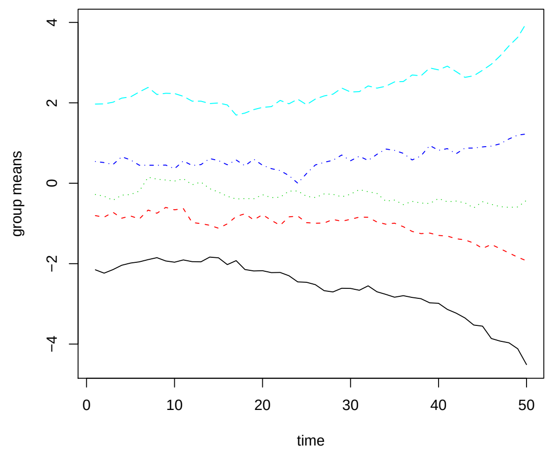

FIGURE 7. Electronic Elephants on Parade (Part 5): This figure illustrates a simple AR(1) version of the Hotelling-Secrist regression fallacy. 500 AR(1) series of length $T = 100$ were generated with $\rho = .9$. This time, there are fixed effects and at time $t = 100$ the series were grouped into quintiles and the group means of these quintiles were plotted following Hotelling's suggestion for evaluating Secrist's approach with department stores' profitability.

Columbia in the early 1930's. Friedman notes that recent work on empirical growth indulge in flights of rhetoric that come perilously close to those of Secrist. He mentions explicitly a review by Jeffrey Williamson of a book by William Baumol and others, but one could now add Robert Barro work with Sala-i-Martin to this list of dubious interpreters of regression. Another example discussed in the blog piece "Mauboussin on Strategy" `http://www.lmcm.com/868299.pdf` is the book on firm growth and decline by Clayton Christensen, called "The Innovator's dilemma." This sounds quite a lot like a rehash of Secrist. An interesting even more recent example that was pointed out to me by Taehyun Kim is the paper of Lemmon, Roberts and Zender (2008) which won a Brattle Group Prize as a "Distinguished Paper" in the *J. of Finance*. Regression is a powerful tool of science, but it can also be a powerful tool for nonsense; it is important to know the difference.

3. Epilogue

A somewhat related "fiasco of econometrics" was recently brought to my attention by Dan Karney, a UIUC econ Phd student. Kane and Staiger (2002) describe an evaluation procedure for schools: those making large improvements in average test scores in a given year are rewarded with additional resources. The procedure, however, tends to reward small schools inordinately often; to see why consider the following graphic. Changes in test scores are plotted against school size and there is a clear pattern showing that dispersion of the mean change decreases with size.

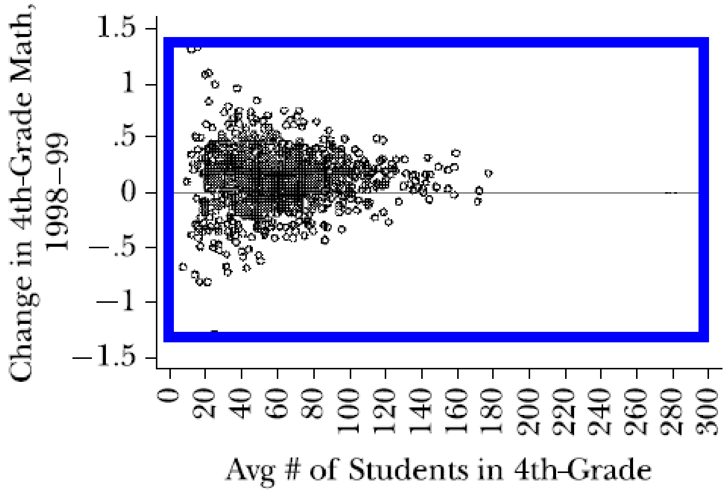

Figure 8. Mean 4th Grade Math Score Changes by School Size: Source Kane and Staiger (2002)

Why? Because means inherently have dispersion that decreases like $1/n$ and therefore standard deviations that decrease like $1/\sqrt{n}$. Here variability is a good thing since the small schools are much more likely to have years in which a few good students pull their performance into the "best schools" category. The paper by Kane and Staiger offers some useful cautionary advice about relying too heavily on short term performance measures as incentive mechanisms for schools, a policy issue that is very much alive in the current political climate.

References

Bass, F.M. (1958) Mediocrity versus Chance in Drug-Store Operating Ratios, J. of Marketing, 22, 307-312.

Christensen, Clayton M., (1997) The Innovators Dilemma: When New Technologies Cause Great Companies to Fail, Boston: MA: Harvard Business School Press.

Stigler, S. (1996) The History of Statistics in 1933, Statistical Science, 11, 244-252.

Friedman, M. (1992) Do Old Fallacies Ever Die? J. of Econonomic Literature, 30, 2129-2132.

Hotelling, H. (1933) Review of Secrist, JASA, 28, 463-4.

Kane T.J. and D. O. Staiger, (2002) The Promise and Pitfalls of Using Imprecise School Accountability Measures, *Journal of Economic Perspectives* 16, 91-114.

Hotelling, H. (1934) Response to Secrist, JASA, 29, 198-9.

Secrist, H. (1933) *The Triumph of Mediocrity*, Northwestern U. Press.