PART I

IDENTIFICATION AND EFFICIENT ESTIMATION

CHAPTER 1

Incredible Structural Inference

Thomas J. Rothenberg

1. INTRODUCTION

In the course of their everyday work, economists routinely employ statistical techniques to analyze data. Typically, these techniques are based on probability models for the observations and justified by an appeal to the theory of statistical inference. An important example is the estimation of structural equations relating economic variables. Such equations are interpreted as representing causal mechanisms and are widely used for forecasting and policy analysis. This econometric approach is arguably the dominant research methodology today among applied economists both in and out of academia.

The econometric approach is not without its critics. Scholars from other disciplines often seem puzzled by the emphasis that economists place on regression analysis. Statisticians express surprise that their techniques should be applicable to so many situations. Recently, a number of leading econometricians have added to the critique. In his paper "Let's Take the Con Out of Econometrics," Ed Leamer (1983) chides economists for ignoring the fragility of their estimates. The title of this paper comes from Christopher Sims's (1980) paper "Macroeconomics and Reality," which argues that the economic and statistical assumptions underlying most macromodels are not believable. They are, he asserts, literally "incredible."

Although my purpose is similar to that of Leamer and Sims, my approach will be rather different. In any area of application there will always be differences of opinion on what constitutes a reasonable set of assumptions on which to base the statistical analysis. Particularly in macroeconomics, where one is trying to summarize in a manageable aggregate model the behavior of millions of decision makers with regard to thousands of products, the disagreements are bound to be enormous. Therefore, instead of discussing typical economic examples

Presented at the International Symposium on Foundations of Statistical Inference, December 1985, Tel Aviv, Israel. This paper evolved from a series of lectures given in June 1985 at the University of Canterbury, Christchurch, New Zealand. I am grateful to Yoel Haitovsky and Richard Manning for providing me with the opportunity to discuss these ideas in such marvelous settings.

where assumptions are always controversial, I shall go to the other extreme and discuss two very simple, almost trivial, examples of statistical inference where the assumptions are quite conventional yet the inferences could naturally be called incredible. Although the examples have nothing to do with economics, I hope to persuade the reader that the key problems with econometric inference are illuminated by their analysis.

2. EXAMPLE ONE: A MEASUREMENT PROBLEM

In order to learn the dimensions of a rectangular table, I ask my research assistant to measure its length and width a number of times. The measuring device is imperfect, so the measurements do not yield the exact length and width. I believe, however, that the measurement errors behave like unpredictable random noise, with any particular error having equal probability of being positive or negative. Therefore, I decide to treat the measurement errors as independent, identically distributed random variables, each with median zero. In addition, I assume that the common error distribution is symmetric and possesses finite fourth moment. For example, the normal probability curve (truncated to insure the measurements are positive) might serve as an approximate model for the error distribution.

These assumptions would not usually be called incredible. They might not be valid for every measurement situation, but they could be reasonable for many such situations. (One might worry about my ruling out thick-tailed distributions that could capture the effects of gross measurement errors. I do that to simplify my story; the analysis could be conducted using medians rather than means, but only with harder distribution theory.) Now I shall make one further assumption. My research assistant mistakenly thinks I care only about the *area* of the table and hence multiplies the length and width measurements. Instead of receiving n length measurements $L_1, L_2, \ldots, L_n$ and n width measurements $W_1, W_2, \ldots, W_n$, I get only n area measurements $A_1 = L_1W_1, A_2 = L_2W_2, \ldots, A_n = L_nW_n$. Worse yet, my research assistant throws away the original data so they are lost forever.

Can I get reasonable estimates of the true length and width of the table using only these area measurements? Can I salvage anything from this badly reported experiment? If there were no measurement error, the answer is clearly no; I will learn the true area of the table, but there are an infinity of length and width pairs that are consistent with any given area. Length and width are simply not identifiable in this experiment. In the presence of measurement error, the answer is quite different. Both length and width are identifiable and can be well estimated from a moderately large sample. In this case credible assumptions seem to lead us to incredible inference!

To demonstrate that inference about length and width is possible, some notation will prove useful. Suppose α is the true length of the table and β is the true width. Let u_i be the error in the ith length measurement, let v_i be the error

in the ith width measurement, and let σ^2 be the common error variance. Then we can write

$$A_i = \alpha\beta + \alpha v_i + \beta u_i + u_i v_i. \tag{1.1}$$

Given the assumption that u_i and v_i are independent random variables distributed symmetrically about zero and possessing third moments, we find:

$$E[A_i] = \alpha\beta, \quad \text{Var}[A_i] = \sigma^2(\alpha^2 + \beta^2 + \sigma^2)$$
$$E(A_i - \alpha\beta)^3 = 6\alpha\beta\sigma^4.$$

By convention, $\alpha \geq \beta > 0$. Simple algebra demonstrates that the three population moments uniquely determine the three parameters α, β, and σ^2. Furthermore, under our assumptions, the sample moments converge in probability to the population moments as n tends to infinity. Denoting the sample mean of the area measurements by M_1, the sample variance by M_2, and the sample third central moment by M_3, a natural method of moments estimator of σ^4 is $M_3/6M_1$. Assuming this is positive and denoting its square root by S, we can estimate $(\alpha + \beta)^2$ by the equation

$$(\alpha + \beta)^2 = \frac{M_2}{S} - S + 2M_1. \tag{1.2}$$

If $\sigma^2 > 0$, the probability that both estimates are positive goes to 1 as n tends to infinity. Define A to be the square root of expression (1.2) if real, and zero otherwise. Then A is a consistent estimate of $\alpha + \beta$. A natural estimate of $(\alpha - \beta)^2$ is

$$(\alpha - \beta)^2 = \frac{M_2}{S} - S - 2M_1. \tag{1.3}$$

If this expression is positive, its square root is a consistent estimate of $\alpha - \beta$. However, if the table is almost square, a negative value for (1.3) is quite likely. Define B to be the square root of expression (1.3) if real, and zero otherwise. Then $(A + B)/2$ and $(A - B)/2$ should be reasonable estimates of α and β.

These method of moments estimates will converge in probability to the true values as long as there is some measurement error. Central limit theory can be employed to develop large sample approximations of their sampling distributions. These approximate distributions are typically normal, although things get slightly more complicated when the table is square (because then the length and width estimates are confounded). To avoid this technical problem in the asymptotic distribution theory, I shall continue the discussion using $\alpha + \beta$ as the parameter of interest and A as my estimate. The essential feature of my example – that the parameter is estimable in the presence of measurement error but not otherwise – is unchanged.

If $\sigma > 0$ and the errors possess finite sixth moments, then the standardized estimator $\sqrt{n}(A - \alpha - \beta)$ converges in distribution to a zero-mean normal

Table 1.1. *Asymptotic relative standard errors for estimates of* $\alpha + \beta$

		Relative standard error[a]		
σ/α	β/α	σ unknown	σ known	Original data
0.01	1.0	14.44	0.35	0.01
0.05	1.0	2.91	0.36	0.04
0.10	1.0	1.50	0.37	0.07
0.20	1.0	0.84	0.41	0.14
0.30	1.0	0.69	0.47	0.21
0.40	1.0	0.68	0.54	0.28
0.50	1.0	0.73	0.62	0.35
0.60	1.0	0.82	0.71	0.42
1.00	1.0	1.53	1.15	0.71
2.00	1.0	7.87	2.67	1.41
0.01	0.5	15.86	0.39	0.01
0.05	0.5	3.23	0.40	0.05
0.10	0.5	1.70	0.42	0.09
0.20	0.5	1.05	0.48	0.19
0.30	0.5	0.96	0.67	0.28
0.40	0.5	1.03	0.68	0.38
0.50	0.5	1.20	0.81	0.47
0.60	0.5	1.44	0.94	0.57
1.00	0.5	3.39	1.60	0.94
2.00	0.5	22.91	4.12	1.89
0.01	0.2	39.13	0.51	0.01
0.05	0.2	8.04	0.52	0.05
0.10	0.2	4.35	0.54	0.12
0.20	0.2	2.86	0.62	0.24
0.30	0.2	2.74	0.74	0.35
0.40	0.2	3.05	0.89	0.47
0.50	0.2	3.65	1.05	0.59
0.60	0.2	4.53	1.25	0.71
1.00	0.2	11.67	2.19	1.18
2.00	0.2	84.67	5.98	2.34

[a] Standard deviation of the limiting distribution of $\sqrt{n}(A - \alpha - \beta)/(\alpha + \beta)$ for alternative estimates A. Approximate relative standard errors for any given sample size n are obtained by dividing by $\sqrt{n}$.

random variable as the sample size n tends to infinity. The asymptotic variance is a complicated function of α, β, σ^2, and the higher moments of the error distribution. Table 1.1 gives the asymptotic relative standard error for the estimate of $\alpha + \beta$ [i.e., the standard deviation of the limiting distribution of $\sqrt{n}(A - \alpha - \beta)/(\alpha + \beta)$] for the special case where the fourth and sixth moments are equal to those of a normal random variable. Also given in the table are the asymptotic relative standard errors for the estimate using the true variance

σ^2 in place of the estimate S in (1.2) and for the estimate using the sample means of the original length and width data. (Note that the tabulated values must be divided by $\sqrt{n}$ to get approximate standard errors for sample size n.)

The table suggests the following conclusions. Depending on the values of α, β, and σ, the efficiency loss from having only the area data ranges from very large to quite modest. If one knows σ^2, the best results are obtained by measuring the table as carefully as possible (but not perfectly!). If one does not know σ^2, the best results are obtained by measuring the table rather badly; for a table that is nearly square, σ/α should be approximately 0.5. In this latter case, there is a simple moral to the story: if one cannot have a smart research assistant, at least have a sloppy one. Truly, an incredible result! Needless to say, I do not seriously propose estimating the length and width of a table from area measurements. My point is quite different. No sensible person would ever use the estimation method derived here. Yet many sensible people would use the sample means of the original observations – if they were available. The assumptions made are not incredible. But they are also not credible enough to justify the inference procedure described. I shall return to this point in a moment, but let me first develop another example.

3. EXAMPLE TWO: A REGRESSION PROBLEM

In order to estimate the gravitational constant I ask another of my research assistants to drop a coin from various heights and to report how long it takes before the coin hits the ground. On the basis of my study of physics, I believe that the true time ought to be proportional to the square root of the distance the coin travels and that the constant of proportionality is related in a simple way to the gravitational constant. This particular research assistant is very good at measuring lengths, but not so good at stopping the stopwatch at the right moment. I therefore propose the regression model

$$y_i = \alpha + \beta x_i + u_i \qquad (i = 1, \ldots, n), \tag{1.4}$$

where y_i is interpreted as the measured time on the ith trial, x_i is the (correctly measured) square root of distance, and u_i is the error in measuring the time. (Of course, here I know α is zero, but I shall not use that fact). I am tempted to estimate β by the least-squares slope coefficient $b = \sum(x_i - \bar{x})y_i / \sum(x_i - \bar{x})^2$ and to form a confidence interval using the statistic

$$T = (b - \beta)\left[\sum \frac{(x_i - \bar{x})^2}{s^2}\right]^{1/2}, \tag{1.5}$$

where s^2 is the sum of squared residuals divided by $n - 2$.

If the errors are independent, identically distributed random variables with mean zero and finite variance, the least-squares estimates are unbiased and have small variance as long as the sample is reasonably large and there is sufficient

variation in x_i. Furthermore, if the errors are normal, these estimates are best unbiased and the statistic T is distributed exactly as Student's t with $n-2$ degrees of freedom.

It would be nice to assume that the measurement errors behave like zero-mean random noise. But what if my research assistant is not so regular in making errors? Maybe he sometimes forgets to stop the stopwatch when he goes out for coffee. Maybe he forgets to reset the watch at zero when he starts a new trial. Given my previous experience with research assistants, anything is possible! I would not like to assume any more than that his errors are a sequence of unobserved numbers. It would be more attractive if the analysis could be conducted on the basis of assumptions on observables, like the regressors, rather than on these mysterious unobserved errors. In fact, as R. A. Fisher (1939) showed many years ago, this can easily be done. Least-squares regression can be justified with almost no assumptions on the errors if we are willing to make some assumptions about the process generating the regressors. The following is a special case of a general result on linear models with multivariate normal regressors:[1]

Theorem 1.1. *In the regression model (1.4), suppose the x_i are i.i.d. normal random variables with variance σ^2 and are distributed independently of the errors. Then the least-squares slope estimate b is distributed symmetrically about β and the statistic T is distributed exactly as Student's t with $n-2$ degrees of freedom,* no matter how the errors are generated. *If the errors have second moments, the mean and variance of b are given by*

$$E(b) = \beta, \quad Var(b) = E\sum \frac{(u_i - \bar{u})^2}{\sigma^2(n-3)(n-1)}.$$

When the Eu_i^2 are uniformly bounded, the variance is $O(n^{-1})$ as n tends to infinity and b is a consistent estimate of β.

Thus, I have a simple solution to my problem of coping with a research assistant whose errors cannot be easily modeled. Before the experiment begins, I randomly draw n numbers from a normal distribution with large mean and unit variance. (I truncate to avoid negative outcomes, but with a large mean the results will look almost normal.) I then instruct my research assistant to use the square of these numbers as the heights (in meters) in the coin-dropping experiment. If the sample is large enough, I can rely on Theorem 1.1 to convince myself that I will get good estimates of the gravitational constant no matter how badly my assistant botches the time measurements. Statistical theory triumphs over a flawed experiment. Once again, credible assumptions lead to incredible inference.

[1] Although this result is not new, I have not found a good reference. A multivariate version is derived in the mimeographed paper by C. Cavanagh and T. Rothenberg (1984). See also Box and Watson (1962). For asymptotic results, the normality assumption can be dropped; symmetry of the x distribution is all that really matters.

Of course, I do not really believe that I can get good estimates of the gravitational constant without making any assumptions about the errors. Indeed, the point of the example is to emphasize that the key assumption in the linear model is that the errors are independent of the regressors. The other assumptions about the errors are easily dispensed with. Moreover, if I am really unhappy about modeling the error process, I should be just as unhappy about trying to model the relation between the errors and the regressors. Independence between regressors and errors is a powerful assumption and cannot be taken lightly.

4. IMPLICATIONS FOR ECONOMETRICS

What has any of this to do with econometrics? Measuring the length of a table and the time it takes a coin to drop seem totally unrelated to the activities that occupy economists. Nevertheless, these examples are, I believe, relevant. Actual econometric models are much more complicated than the ones I have presented and concern more important phenomena. But, deep down, they possess the same key features that drive the examples.

Economists are usually interested in parameters that have a structural or causal interpretation. If, other things equal, the price of coffee doubles, by how much would price-taking consumers decrease their purchases? Such numbers are treated by economists in much the same way as the length of the table and the gravitational constant in my examples. They are parameters of interest that could in principle be determined by very carefully conducted experiments. Unfortunately, these experiments are much too difficult, so we have to rely on different ones. Usually, the actual data we have available were generated by someone else using methods very far from the ones we would have used in our ideal experiment. Instead of actually changing the price of coffee, we simply observe the historical variation that has taken place over time. Just as in the artificial examples given earlier, structural inference in economics involves the analysis of data from flawed experiments.

The error terms in econometric equations represent misspecifications of functional form, omitted variables, and pure measurement error. It is not hard to make assumptions about these errors that are moderately plausible. Unfortunately, their converses are often also moderately plausible. Most econometric models are reasonable, but they are not compelling. There always exist alternative models that are just as reasonable. Yet, as in the examples, the results are often very sensitive to the assumptions. If we are suspicious of estimating the length of a table from area measurements and if we are suspicious of estimating the gravitational constant from an experiment where the measurement errors may take on arbitrary values, then we should be even more suspicious of structural econometric inference in models where the number of unknown parameters and the number of unverified assumptions are much larger. The estimation of separate supply and demand curves from equilibrium market data is considerably more difficult than the estimation of the length of a table.

Using cross-sectional variation in the crime rate to determine the effect of longer jail sentences on the level of crime is certainly just as difficult as using flawed experimental data to determine the gravitational constant. Relevant research is not easier than trivial research.

Not all econometrics involves structural inference. Sometimes we collect economic data just to describe the current state of affairs or to indicate trends. Sometimes we run regressions simply to summarize the pattern of correlations in a data set or to take advantage of stable relations for use in forecasting. Many of the most successful applications of statistics in economics have nothing to do with the estimation of structural relations. Nevertheless, the temptation to interpret empirical regularities as representing causal mechanisms is overwhelming. For better or worse, econometrics is generally viewed as a method for learning about the underlying structure of the economy.

Structural inference in econometrics, like the structural inference in my simple examples, is indeed incredible. Surprisingly strong conclusions about causal mechanisms can be drawn from seemingly weak assumptions. Unfortunately, the conclusions are often not very robust to changes in these assumptions. In those cases, it is difficult to put much credence in the results. More emphasis by applied econometricians on presenting alternative estimates based on alternative models might help make econometrics less incredible.

References

Box, G., and G. Watson (1962), "Robustness to Non-normality of Regression Tests," *Biometrika*, 49, 93–106.

Cavanagh, C., and T. Rothenberg (1984), Linear Regression with Non-normal Errors, mimeo.

Fisher, R. A. (1939), "The Sampling Distributions of Some Statistics Obtained from Nonlinear Equations," *Annals of Eugenics*, 9, 238–49.

Leamer, E. (1983), "Let's Take the Con Out of Econometrics," *American Economic Review*, 73, 31–43.

Sims, C. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

CHAPTER 2

Structural Equation Models in Human Behavior Genetics

Arthur S. Goldberger

1. INTRODUCTION

That IQ is a highly heritable trait has been widely reported. Rather less well known are recent reports in major scientific journals such as those announcing that the heritability of controllable life events is 53 percent among women and 14 percent among men (Saudino et al. 1997), while the heritabilities of inhibition of aggression, openness to experience, and right-wing authoritarianism are respectively 12, 40, and 50 percent (Pedersen et al. 1989; Bergeman et al. 1993; McCourt et al. 1999). It seems that milk and soda intake are in part heritable, but not the intake of fruit juice or diet soda (de Castro 1993).

These reported heritabilities are parameter estimates obtained in structural modeling of measures taken on pairs of siblings – prototypically, identical (monozygotic) twins and fraternal (dizygotic) twins, some reared together and others reared apart. The models are of the linear random effects type, in which an observed trait – a phenotype – is expressed in terms of latent factors – genetic and environmental – whose prespecified cross-twin correlations differ by zygosity and rearing status. Estimation is by maximum likelihood applied to the phenotypic variances and covariances. Heritability, the key parameter of interest, refers to the proportion of the variance of the phenotype that is attributable to the variance of the genetic factors.

Regarding these studies, various issues arise. Those that I will touch on here include identification, nonnegativity constraints, alternative estimators, pretest estimation, conditioning of the design matrix, multivariate analyses, and the objectives of structural modeling. Some of these issues were featured in Thomas Rothenberg's dissertation (1972), a remarkable book that led me to appreciate the generality of the minimum chi-square principle in estimation, and the contrast between equality and inequality constraints in efficient estimation.

In the present chapter, I will focus on the SATSA project – the Swedish Adoption/Twin Study of Aging – which, from the early 1980s on, has assembled a sample of adult twin pairs: approximately 200 MZT (identical twins reared together), 200 DZT (fraternal twins reared together), 100 MZA (identical twins reared apart), and 150 DZA (fraternal twins reared apart). The fraternal twins