

# Econometrics: An Historical Guide for the Uninitiated



**Stephen Pollock**, University of Leicester, UK

Working Paper No. 14/05

February 2014

# ECONOMETRICS: AN HISTORICAL GUIDE FOR THE UNINITIATED

by D.S.G. Pollock

University of Leicester

This essay was written to accompany a lecture to beginning students of the course of Economic Analytics, which is taught in the Institute of Econometrics of the University of Lodz in Poland. It provides, within a few pages, a broad historical account the development of econometrics. It begins by describing the origin of regression analysis and it concludes with an account of cointegration analysis. The purpose of the essay is to provide a context in which the students can locate various aspects of econometric analysis. A distinction must be made between the means by which new ideas were propagated and the manner and the circumstances in which they have originated. This account is concerned primarily with the propagation of the ideas.

## **Introduction: The Business of Statistical Inference**

The business of statistical inference is predicated upon the metaphysical notion that, underlying the apparent randomness and disorder of events that we observe in our universe, there is a set of regular and invariant structures. In attempting to identify its underlying structure, we may imagine that a statistical phenomenon is composed of a systematic or determinate component and a component that is essentially random or stochastic. The fundamental intellectual breakthrough that has accompanied the development of the modern science of statistical inference is the recognition that the random component has its own tenuous regularities that may be regarded as part of the underlying structure of the phenomenon.

In the sphere of social realities, statistical science has uncovered many regularities in the behaviour of large aggregates of apparently self-willed individuals. Examples spring readily to mind. Consider the expenditure on food and clothing of a group of individual households that might be observed over a given period. These expenditures vary widely, yet, when family income and other measurable factors are taken into account, evident regularities emerge.

Denoting the expenditure of the  $i$ th family by  $y_i$  and its income by  $x_i$ , we might postulate a statistical relationship of the form

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

where  $\alpha$  and  $\beta$  are parameters, which are described, respectively, as the intercept and the slope, and  $\varepsilon_i$  is a random variable. In fitting the data to the model, we would find that the systematic component  $\mu_i = \alpha + \beta x_i$  would, in many cases, amount to a large proportion of  $y_i$ . The residual part of  $y_i$  would be attributed to the random variable  $\varepsilon_i$ . The precise details of the decomposition of each  $y_i$  would depend upon the values attributed to the parameters  $\alpha$  and  $\beta$ . These values can be assigned only in view of the more or less

specific assumptions that are made about the regularities inherent in the random variable  $\varepsilon_i$ .

It might be assumed, for example, that the  $\varepsilon_i$  are distributed independently of each other with a common expected value of  $E(\varepsilon_i) = 0$  and a common variance of  $V(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$ . Then, as Chebyshev's inequality shows, there is an upper bound on the probability of large deviations of  $\varepsilon_i$  from zero; and it is appropriate to attribute to  $\alpha$  and  $\beta$  the values that minimise the quantity

$$S = \sum_i (y_i - \alpha - \beta x_i)^2,$$

which is the estimated sum of squares of the deviations. Alternatively, it might be more realistic to assume that the dispersion of the random component  $\varepsilon_i$  is related to the size of family income  $x_i$ . Then, we might specify that  $V(\varepsilon_i) = \sigma^2 x_i$ ; and the values of  $\alpha$  and  $\beta$  would be found by minimizing

$$S = \sum_i x_i^{-1} (y_i - \alpha - \beta x_i)^2.$$

The crucial assumptions concerning  $\varepsilon_i$  describe the stochastic structure of the model.

Doubtless, many would contend that the randomness in the variation of household expenditures is more apparent than real. For they would argue that the appearance of randomness is due to our failure to take into account a host of other factors contributing to this behaviour. They might suggest that, if every factor were taken into account, a perfect description of the behaviour could be derived.

Fundamental though this objection might be, we can afford to ignore it; for it makes little difference to the practice of statistical inference whether the indeterminacy of the behaviour is the result of pure randomness or the result of our inability to comprehend more than a few of an infinite number of peculiar factors and circumstances affecting each household.

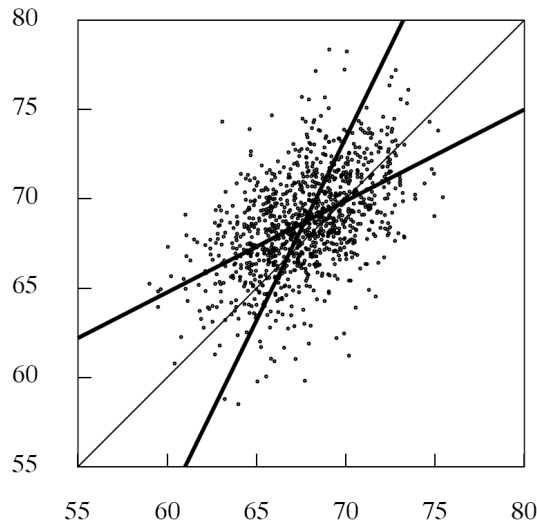
## **Francis Galton and the Origins of the Regression Equation**

The equation that we have proposed for determining the household expenditure is described as a regression equation. This odd terminology is due a Victorian man of science by name of Francis Galton (1811–1911). Galton, who was a first cousin of Charles Darwin (1809–1882), was also the man who invented the word Eugenics to describe the science of improving the human population by a programme of controlled breeding, designed to promote desirable inherited characteristics.

As is well known, the ideas of Eugenics were widely popular in the early decades of the 20th century; and they sustained many of the fatal aberrations that accompanied the era of the Nazis. Galton first propounded his theory in 1865 in an article titled *Hereditary Talent and Character*. He elaborated it further in his 1869 book *Hereditary Genius*, where he studied the way in which intellectual traits, moral traits and traits of personality tend to run in families.

Galton noted that, as one generation succeeds another, the outstanding qualities that have characterised the earlier generations will tend to be found in an attenuated form in later generations. He described this phenomenon as a tendency to mediocrity. Although

the etymology of his words is impeccable, they have a negative connotation. Nowadays, we prefer to talk of a central tendency.



**Figure 1.** Pearson’s data comprising 1078 measurements of on the heights of father (the abscissae) and of their sons (the ordinates), together with the two regression lines. The correlation coefficient is 0.5013.

Galton was fully aware that the process works in both directions. He declared that

The law of regression is even handed; it levies the same succession tax on the transmission of badness as well as goodness. It discourages the extravagant expectations of gifted patents that their children will inherit their powers, it no less discountenances extravagant fears that they will inherit all their weaknesses and diseases.

This implies that, as one generation succeeds another, the heights within families that have been characterised by lack of stature will progress towards the mean values. Thus, a regression analysis could be described, with equal justification, as a progression analysis. Also, if the distributions of heights is to remain the same from one generation to the next, then the two central tendencies of progression and regression must be balanced by equal and opposite dispersive tendencies.

Galton conducted a detailed statistical investigation of such tendencies; and his methodology reached its maturity in 1886 in an article titled *Regression Towards Mediocrity in Hereditary Stature*. The data that were collected by Galton were later supplemented by Karl Pearson (1857–1936), who was another prominent early statistician and eugenicist and who was Galton’s disciple and his biographer. Pearson collected the heights of 1,078 fathers and their full-grown sons, in England, at the end of the 19th century. (See Pearson 1896 and Pearson and Lee 1896, 1903.)

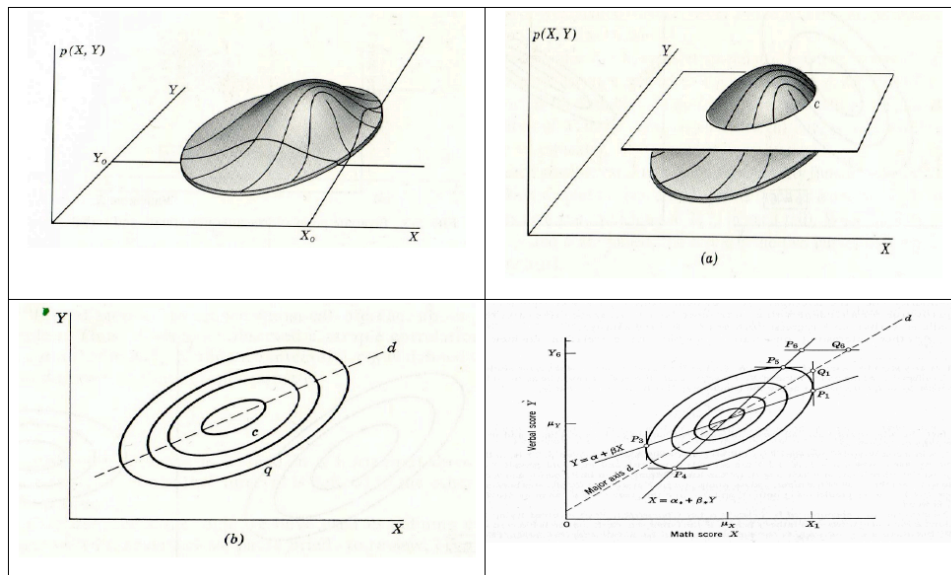
In Figure 1, which displays Pearson’s data, each dot represents a father-son pair. The heights of the fathers are measured on the horizontal axis and those of their sons are on the vertical axis. Running through the scatter of points at an angle of 45 degrees to the horizontal is what seems, at first glance, to be an axis of symmetry—if the diagram were flipped around this axis, so that the heights of the sons are on the horizontal and those of

their the fathers are on the vertical, then we might not be able to discern much difference. That would be the expected outcome, if the distribution of the heights of adult males were the same in each generation.

We might also be tempted to use the 45 degree line as the means of predicting the heights of the sons, knowing the heights of their fathers. However, as we have already asserted, there is a tendency for tall fathers to produce sons whose tallness is less accentuated. In fact, the best line for predicting the height of a son is the thick black line that rises at an angle of 35 degrees from the horizontal.

A ready explanation of this feature is provided by taking account of the role of the mother in determining the height of the son. If we attribute an average height to the mother, in place of her actual height, which has not been revealed by the data, then we can imagine that the tall father and the average mother are likely to produce a male child whose height, which is the product of the heights of his two parents, is somewhere between the observed height of the father and the average height of men.

Another point to note is that the average height of men in late Victorian England was increasing over time. This can be seen clearly in Figure 1. If the distributions of the heights of the fathers and of their sons were identical, then the 45 degree line would indeed be an axis of symmetry, and the regression of the heights of the sons on the heights of their fathers and vice versa would be indistinguishable. The two regression lines would intersect at a point on 45 degree line.



**Figure 2.** The Bivariate Normal Gaussian Distribution, which has elliptical contours.

In fact, the intersection is at a significant distance from the line; which is evidence of an increasing average height from one generation to the next. One may attribute this to improvements in nutrition and health. Disregarding such refinements, the relationship of the heights can be summarized fairly well in reference to the properties of a bivariate normal or Gaussian distribution, of which an example is depicted in Figure 2. This may seem remarkable in view of the disparity between the theory and the realities of human stature. In theory, the Gaussian distribution extends over the entire plane. In practice,

human heights are bounded both from below, by the height of Tom Thumb, and above, by the height Fingal, a giant of Celtic mythology.

### **R.A. Fisher and the Experimental Regression Model**

The regression model that strongly influenced the early development of econometrics has features that set it apart from the bivariate statistical model of Galton, which might be classed as a descriptive regression. This was the regression model of agronomy wherein the variable  $x$  that is plotted on the horizontal axis is subject to experimental manipulation. The model was used extensively by R.A. Fisher (1890–1962) in his work at the Rothamsted Experimental Station.

The Rothamsted Station was founded in 1843, which makes it one of the oldest agricultural research institutions in the world. From its inception, the research entailed of a series of long-term field experiments in which the object was to measure the effect on crop yields of inorganic and organic fertilisers. By the end of the 19th century, these so-called Classical Field Experiments had generated a vast amount of data that demanded a systematic analysis.

Fisher's first task on joining in 1919 was to grapple with these data. Over the years, he generated many of the statistical methods that are associated with the linear regression model. These include the theory of the design of experiments and the theory of hypothesis testing. Fisher also propounded the method of maximum-likelihood estimation. His first book, *Statistical Methods for Research Workers*, which was published 1925, became a standard reference work for scientists in many disciplines, and the traces of it were readily discernible in the early texts of econometrics.

R.A. Fisher also made significant contributions to the science of genetics to the extent that, together with Sewall Wright and J.B.S. Haldane, he can be regarded as one of the founders of what has been described as the modern evolutionary synthesis.

It must also be recorded that Fisher was an ardent eugenicist. In 1911, he was involved in forming the Cambridge University Eugenics Society with John Maynard Keynes, R.C. Punnett and Horace Darwin (Charles Darwin's son). He went so far as to propose that the state should afford him a bursary to sustain his prolific procreation, on the grounds that he was contributing to the improvement of the race—he had two sons and seven daughters.

The basic experimental regression model of Fisher embodies a linear relationship between a dependent variable and one or more independent variables. In Fisher's context, the dependent variable might be the yield of a crop by weight and the independent variables might be the quantity of fertiliser applied to a crop, the salinity of the soil, the amount of water available to the crop, and so forth.

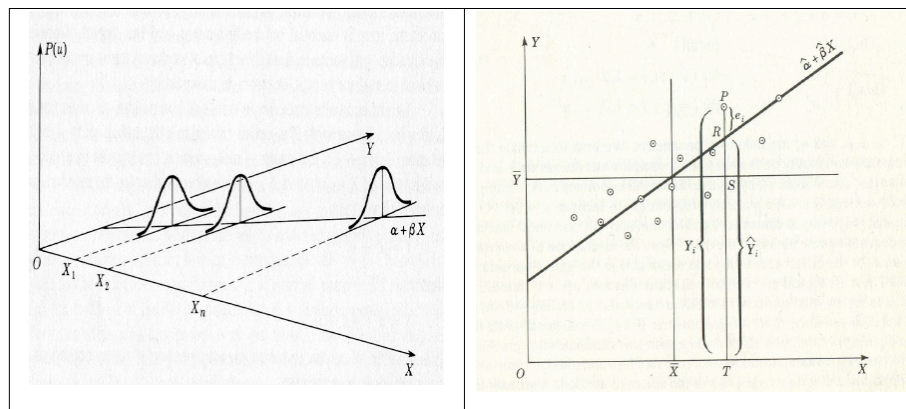
It is supposed that there is a firm underlying relationship between the dependent and the independent variables. The relationship is also subject to additive disturbances, which summarise the effect on the dependent variable of other factors that have not been included explicitly in the analysis. The assumption is made that these additional and unobserved factors are unaffected by the values of the independent variables that are included in the systematic part of the model. Thus, in aggregate, they constitute a random variable that, according to the classical assumptions, is independently and identically distributed at each instant of observation.



The parameters of the regression relationship are invariably estimated by the method of least squares. In the case of the model with a single explanatory variable, this entails interpolating a straight line through the scatter of the data points in such a way as to minimise the sum of squares of the deviations of the points from the line, when these distances are measured parallel to the vertical  $y$ -axis.

### Explanatory Variables: Limits to the Range of Variation

The descriptive regression model of Galton can be reconciled with the experimental regression model if we are prepared to overlook the fact that, in the descriptive model, the  $x$  variable has its own Gaussian distribution. By taking a specific value of  $x$  and asking what the consequent value of  $y$  is likely to be, we are asking a question about the conditional distribution of  $y$  given  $x$ .



**Figure 3.** The conditional distributions of  $y$  given the values  $x_1$ ,  $x_2$  and  $x_3$  of the explanatory variable and the least-squares estimation of the regression line from a sample of observations.

This distribution has exactly the appearance that it would have in the experimental model. In effect, when the  $x$  variables are divested of their probability measures, the resulting conditional distributions look the same for all values of  $x$ , with the exception that they have differing mean values, which are given by  $\mu_x = \alpha + \beta x$ .

The contours of the bivariate distribution of Galton's regression model give way to the contours of a ridge-like structure erected over the regression line, which is depicted in Figure 3. However, notwithstanding these appearances, the conditioning on  $x$  does not give us the freedom to manipulate the values of the fathers' heights in an experimental manner. The idea of stretching or compacting their bodies in order to observe the effect on their offspring hardly bears a moment's thought. This would kill the fathers and it would preclude the possibility of their having any offspring.

Nevertheless, we are able to manipulate height within a thought experiment. In this sense, we may talk of the conjectural variation of  $x$  as opposed to its actual manipulation. We can easily imagine fathers of various heights. However, it is clear that such conjectural variation ought to be bounded, at least, by the height of the micro dwarf Tom Thumb and by that of the Irish giant Fingal.

By the same token, the variation of  $x$  within the experimental model is limited. Most of the explanatory variables that we have thought of in connection with the biometric

version of the model are bounded away from zero. There are also upper bounds on their quantities. Too much fertiliser would scorch the plants, too much salt would kill them and too much water would destroy the structure of the soils. Indeed, it would be wise to declare the regression model to be valid only within the ranges of variation that are represented by the sample from which it has been estimated.

In econometrics, it is uncommon to impose such limits on the ranges of the variations of the explanatory variables. In a non-experimental situation, the actual variations in the sample are liable to be limited. Often, the purpose of the model is to form conjectures regarding the effects of wider variations. Econometric regression models are frequently used to investigate what, in philosophical jargon, are called counterfactual conditional propositions.

These are statements indicating what would be the case if something were true that is not actually true, or if something were to be made to happen that might not happen otherwise. In other words, econometric models are liable to be used for speculative purposes rather than for descriptive purposes. For that reason, they should often be treated with caution.

The difficulties with econometric modelling are exacerbated when there are numerous explanatory variables on the right hand sides of the regression equations. In the process of formulating economic policies, one is liable to speculate about the effects of unprecedented combinations of the values of such variables. The variables are embedded in a complex interdependent system, and they might not be amenable to independent variation, either in an experimental manner or in a speculative counterfactual sense.

To cater to such complex interdependencies has been one of the objectives of classical econometric theory. One of the essential difficulties came to light in the early years of the 20th century in connection with the statistical estimation of demand curves.

### **Statistical Demand Curves and the Simultaneous-Equations Model**

The estimation of demand curves was a major preoccupation of American statisticians and economists throughout the 1920's and 1930's. Many of their studies focussed on the demand for agricultural produce. A typical example is provided by a study of the demand for potatoes that was published in 1935 by Holbrook Working (1895–1985) in a paper titled *The Statistical Determination of Demand Curves*. The paper expresses some of the difficulties and the unease that accompanied such exercises.

The principal difficulty arises from the fact that the prices and quantities of economic goods are jointly determined within their markets by the interaction of demand and supply. Therefore, it may be inappropriate to view the matter from one side only. Alfred Marshall (1842–1924) used a metaphor to convey the fact that economic values are determined simultaneously by supply and demand. According to Marshall (1920, p.348)

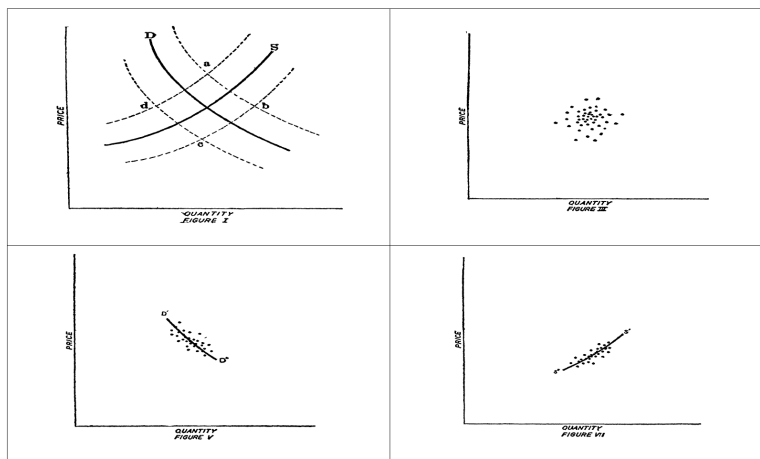
We might as reasonably dispute whether it is the upper or the under blade of a pair of scissors that cuts a piece of paper, as whether value is governed by utility or cost of production. It is true that, when one blade is held still and the cutting is effected by moving the other, we may say with careless brevity that the cutting is done by the second; but the statement is not strictly accurate, and is to be excused only so long as it claims to be merely a popular and not a strictly scientific account of what happens.



The implication of this truism for the statistical determination of demand curves was fully revealed by Elmer Joseph Working, (1900–1968), the younger brother of Holbrook Working, who published an article in 1927 titled *What do Statistical Demand Curves Show?*

Historical research has shown that others had reached conclusions similar to those of Elmer Working at earlier dates. (For a detailed account of the early history of demand analysis, see Morgan 1990.) Also, it is now known that the Danish economist Jens Warming reached the same conclusions as Working in his commentary on a book published in 1906 by the Danish economist Edvard P. Mackeprang, which contained estimates of demand curves that showed perverse slopes. (See Kørsgaard 1984 and Kørsgaard, Andersen and Topp 1998.)

Elmer Working produced a series of diagrams that can still be found in many textbooks of econometrics. These diagrams are reproduced in Figure 4.



**Figure 4.** Diagrams of supply and demand from the paper of E.J. Working *What do Statistical Demand Curves Show?*

He began by noting that, if both the demand and the supply curves are liable to shift over time to the same degree, then their points of intersection will form a seemingly unstructured scatter. If the shifting of the supply curve is greater than the shifting of the demand curve, then a scatter of points will be generated that will reveal the shape of the demand curve. Conversely, if the demand curve is shifting over a stationary supply curve, then the scatter of points will serve reveal the latter.

It might be inferred that, when both curves are shifting to an equal degree, it will be impossible to disentangle them from the data. Nevertheless, Working declared that

By intelligently applying proper refinements, and making corrections to eliminate separately those factors which cause demand curves to shift and those factors which cause supply curves to shift, it may be possible even to obtain both a demand curve and a supply curve for the same product and from the same original data.

What he was implying is that, if the shifts in the supply and demand curves are due

to separate variables that have no interaction, then it will be possible to identify both curves or schedules. In econometric jargon, we may describe the prices and quantities that are determined in the market as the jointly dependent variables in the supply and demand equations.

The two schedules will be capable of separate identification if each equation contains exogenous variables that are not in the other equation. The term exogenous denotes the fact that the variables are generated outside the system of demand and supply. They will affect the endogenous variables that are within the system, but they will not be affected in turn by the latter variables.

As an example, which may serve to fix these ideas more firmly, consider the following system:

$$y_1 = \gamma_1 y_2 + \varepsilon_1 : \text{The Demand Equation,}$$

$$y_2 = \gamma_2 y_1 + x\beta + \varepsilon_2 : \text{The Supply Equation,}$$

where

- $y_1$  represents the quantity of popcorn consumed and produced,
- $y_2$  represents the price of popcorn and
- $x$  represents the cost of maize.

These variables, which are deviations from mean values, have expected values of zero. The effect of taking deviations is to simplify the algebra; for the intercept terms are thereby eliminated from the equations. It is assumed that the disturbances  $\varepsilon_1$  and  $\varepsilon_2$  are independent of the variable  $x$ , which is an exogenous variable that it is generated in a context that lies outside the model.

In this model, the consumers of popcorn, whose behaviour is represented by the demand equation, respond to the price of popcorn, whereas the producers, whose behaviour is represented by the supply equation, set the price in view of the demand for their product and in view of their costs of production. The market is in a state of equilibrium, where the quantity produced is equal to the quantity consumed.

Although the cost of maize is not the only cost of production, it can be assumed, for present purposes, that it is the only one that varies significantly. The other costs, which are fixed, will have an effect that is subsumed in an intercept term, which has been eliminated. The factors, other than price, that determine the demand for popcorn are likewise assumed to be constant and are subsumed in another intercept.

The effect of the variability of the cost of maize will be to shift the supply curve over the more-or-less static demand curve, thereby allowing its parameters to be determined from the observed variables.

The recognition of the possibility of achieving the identification of the parameters of models with jointly dependent variables motivated a search for efficient methods of simultaneous-equation estimation. The problem was addressed and solved in theory by a formidable group of young statisticians working under the auspices of the Cowles Commissions in the late 1940's and the early 1950's

The Cowles Commission for Research in Economics was founded in Colorado Springs in 1932 by Alfred Cowles, a businessman and economist. In 1939, it moved to the University of Chicago.

In 1943, Jacob Marschak, who had left Oxford University to come to the United States in 1939, became the director of the Cowles Commission, which was a post he held until 1948, when he was succeeded by Tjalling Koopmans. Under Marschak's directorship, the Commission began an intensive study of the problems of estimation and identification associated with simultaneous equation systems. An inspiration for this line of research had been the work of Ragnar Frish (1895--1973), who was one of the founders of the Econometric Society and editor of the journal *Econometrica* for over twenty years.

In 1948, Tjalling C. Koopmans became the director of the Commission; and it was in the period of his tenure that the definitive work on simultaneous-equation estimation was published in three Monographs. These were Monograph 10, *Statistical Inference in Dynamic Economic Models*, edited by Koopmans (1950), Monograph 11, *Economic Fluctuations in the United States, 1921–1941*, by Klein (1950) and Monograph 14, *Studies in Econometric Method*, edited by Hood and Koopmans (1953), which gathered some important papers, some of which had been published previously elsewhere.

The methods for estimating simultaneous-equation systems were the indirect least-squares method, the method of limited-information maximum likelihood (Anderson and Rubin 1949, Anderson 1950 and Koopmans and Hood 1953) and the method of full-information maximum-likelihood (Koopmans, Rubin, and Leipnik 1950).

Rising opposition to the Cowles Commission by the department of economics at University of Chicago during the 1950's encouraged the Cowles group to move to Yale University in 1955, where the Commission was reincarnated as the Cowles Foundation. Electronic copies of the monographs of the Commission reside on the website of the Foundation at the address <http://cowles.econ.yale.edu/>, where other historical documents and histories can be found. The history of the Cowles Commission has also been discussed by Qin (1993).

It is fair to say that the maximum-likelihood methods defeated the practitioners at the time of their inception. In the first place, the methods demanded computing facilities that were not widely available. Also, their mathematical derivations were demanding and they were not widely understood. However, in the late 1950's and the early 1960's these methods were effectively reinvented under the guise of two-stage least squares (Basmann 1957 and Theil 1953, 1958) and three-stage least squares (Zellner and Theil 1962) and in a manner that rendered them widely intelligible.

### **The Advent and the Demise of the Large-Scale Econometric Model**

The idea of the joint or simultaneous determination of certain economic variables met with a degree of resistance from several parties, most notably from Wold (1949, 1954), who argued that multi-equation econometric models should be structured in a recursive manner, such that the inputs to each equation would be predetermined by preceding equations.

In a paper titled *Causal Ordering and Identification*, which was published in the Cowles Commission Monograph No.14, Herbert Simon (1953) provided the following

simple example: *poor growing weather* → *small crop yield* → *increase in price of wheat*.  
 The associated system of equations may be written as follows:

$$a_{11}x_1 = \varepsilon_1,$$

$$a_{21}x_1 + a_{22}x_2 = \varepsilon_2,$$

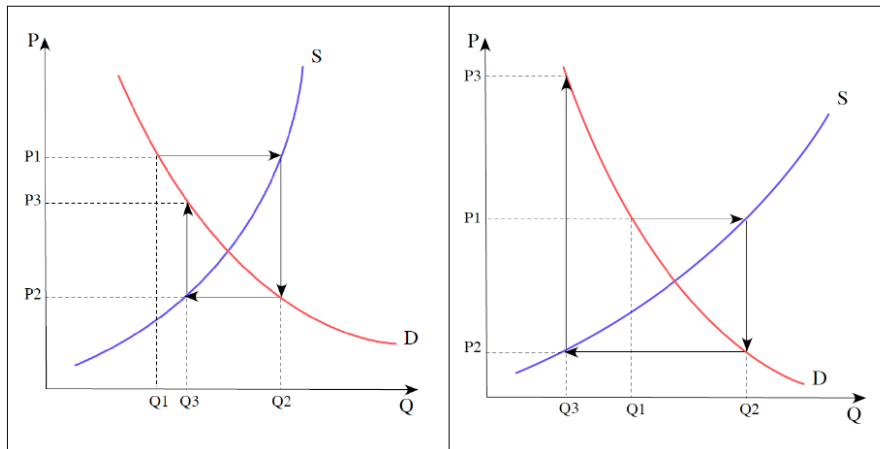
$$a_{32}x_2 + a_{33}x_3 = \varepsilon_3.$$

Here,  $x_1$  is an index of the weather,  $x_2$  is the size of the crop and  $x_3$  is the price of wheat. The three equations determine these three variables in sequence, which can be portrayed by writing  $x_1 \rightarrow x_2 \rightarrow x_3$ . It can be seen that this is also a temporal sequence, albeit that Simon was keen to assert that a causal ordering is not necessarily predicated on a temporal ordering.

Without doubt, the existence of a temporal ordering amongst the equations can greatly alleviate the problems of identification and estimation, as well as lending a dynamic quality to the models. To see this, we need only consider a simple elaboration of Simon's model in which the farmers growing this year's crop are reacting to last year's prices.

The so-called cobweb model depicts farmers who overreact, to the extent that a high price, last year, for their crops will lead, this year, to an oversupply and to a glut on the market. The consequent fall in price will lead the farmers, next year, to reduce the size of their crop, which will occasion a rise in prices.

These circumstances will generate a repetitive cycle of increasing and declining crop yields in successive years, accompanied by falling and rising prices. An issue arising is whether, over time, the cycles will be of an increasing or of a decreasing amplitude. In theory, the matter depends on the geometry of the diagram of supply and demand, as Figure 5 serves to illustrate.



**Figure 5.** The cobweb model, which explains how the prices  $P$  and quantities  $Q$  in agricultural markets may show repeated annual fluctuations by the interaction of the supply schedule  $S$  and the demand schedule  $D$ . The geometry of the diagrams determine whether the amplitude of the fluctuations will be increasing or diminishing

The work of the Cowles Commission on simultaneous-equation modeling was accompanied by an increasing interest in large-scale macroeconomic models, which were proposed as indispensable aids to governmental decision making.

The Dutch economist Jan Tinbergen developed the first comprehensive national model, which he first built for the Netherlands (Tinbergen 1937) and later applied to the United States (Tinbergen 1939) and to the United Kingdom after World War II (Tinbergen 1951). The first global macroeconomic model, the Wharton Econometric Forecasting Associates' LINK project, was initiated by Lawrence Klein. The model was cited in 1980, when Klein was awarded the Nobel Prize in Economics, as Tinbergen had been, together with the Norwegian economist Ragnar Frishe, in 1969.

In the hands of the statisticians of the national central statistical offices, such models grew to include hundreds of equations, each of which proposed to depict some structural aspect of the economy or the behaviour of some well-defined group of economic agents. There were equations for estimating a variety of consumer demands, equations for describing industrial capital investments, equations to describe house building activities in detail, and so forth.

In reality, much of the modelling was on an ad-hoc basis, and it ignored the methodology of the Cowles Commission. (An exception to this was the small model that Klein proposed in the Monograph No. 11 of the Cowles Commission.) The beautiful and grandiose model of linear simultaneous-equation systems was insufficiently flexible for the purposes of the macroeconomic models, which embodied intractable non-linearities in many of their sections.

Such models survive to this day within various governmental agencies, in highly refined forms. However, the prestige of large-scale macroeconomic models was severely damaged in the 1970's, when it was revealed that their forecasting performance was often far inferior to that of the simple unconditional time-series models of the autoregressive-moving average (ARMA) variety.

Part of the problem here was that the equations of the macroeconomic models had paid insufficient attention even to the simple laws of linear dynamic systems, which is precisely what the ARMA models are. Such models were popularised in the 1970's with the publication of the book of Box and Jenkins (1970).

## **Dynamic Economic Models**

In an experimental situation, where one might be investigating the effects of an input variable  $x$  on a mechanism or on an organism, one can set the value of  $x$  and then wait until the system has achieved an equilibrium before recording the corresponding value of the output variable  $y$ .

In economics, we are often interested in the dynamic response of  $y$  to changes in  $x$ ; and, given that  $x$  is continuously changing, the system might never reach an equilibrium. Moreover, it is in the nature of economic relationships that the adjustments of  $y$  to changes in  $x$  are distributed widely through time.

In the early days of econometrics, attempts were made to model the dynamic responses primarily by including lagged values of  $x$  on the RHS of the regression equation; and the so-called distributed-lag model was commonly adopted, which takes the form of

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_k x_{t-k} + \varepsilon_t.$$

(For an example of this approach, see Alt 1942.) Here, the sequence of coefficients  $\{\beta_0, \beta_1, \dots, \beta_k\}$  constitutes the impulse-response function of the mapping from  $\{x_t\}$  to  $\{y_t\}$ . That is to say, if we imagine that, on the input side, the signal sequence is of the form

$$\{x_t\} = \{\dots, 0, 1, 0, \dots, 0, 0, \dots\},$$

which comprises a single unit impulse preceded and followed by zero values, then the output of the transfer function would be

$$\{r_t\} = \{\dots, 0, \beta_0, \beta_1, \dots, \beta_k, 0, \dots\}.$$

A problem with the distributed-lag formulation is that it is profligate in its use of parameters; and given that, in a dynamic econometric context, the sequence  $\{x_t\}$  is likely to show strong serial correlation, we may expect to encounter problems of multicollinearity—which is to say that the estimates of the parameters will be ill-determined with large standard errors.

It is difficult to specify *a priori* what form a lag response will take in any particular econometric context. Nevertheless, there is a common presumption amongst economists that the coefficients will all be of the same sign, and that, if this sign is positive, their values will rise rapidly to a peak before declining gently to zero.

One such lag scheme, which accords with these expectations and which is sparing in its use of parameters, is the geometric lag scheme. Its impulse response is a geometrically declining sequence

$$\{r_t\} = \{\dots, 0, \beta, \beta\varphi, \beta\varphi^2, \dots, \beta\varphi^k, \dots\},$$

which continues indefinitely, albeit that it may decline quite rapidly.

Assuming that the sample runs from time  $t = 0$ , the effect of the pre-sample data values can be summarised with the help of a parameter  $\theta = \beta\{x_{-1} + \varphi x_{-2} + \varphi^2 x_{-3} + \dots\}$ , which will enable the model to be written as

$$y_t = \beta\{x_t + \varphi x_{t-1} + \dots + \varphi^{t-1} x_0\} + \varphi^t \theta + \varepsilon_t.$$

This is the model of Koyk (1954). Although the equation is not amenable to ordinary least-squares regression, it is reasonably straightforward to estimate its parameters.

A simpler way of creating a geometric lag scheme is to include a lagged value of the dependent variable on the RHS of the equation to give

$$y_t = \varphi y_{t-1} + \beta x_t + \varepsilon_t.$$

This is described, alternatively, as a model with a single lagged dependent variable or as a model with first-order feedback. By repeated substitution, we obtain



$$\begin{aligned}
y_t &= \varphi y_{t-1} + \beta x_t + \varepsilon_t \\
&= \varphi y_{t-2} + \beta \{x_t + \varphi x_{t-1}\} + \varepsilon_t + \varphi \varepsilon_{t-1} \\
&\vdots \\
&= \varphi^n y_{t-n} + \beta \{x_t + \varphi x_{t-1} + \cdots + \varphi^{n-1} x_{t-n+1}\} \\
&\quad + \varepsilon_t + \varphi \varepsilon_{t-1} + \cdots + \varphi^{n-1} \varepsilon_{t-n+1}.
\end{aligned}$$

If  $|\varphi| < 1$ , then  $\lim(n \rightarrow \infty) \varphi^n = 0$ ; and it follows that, as the number of repeated substitutions increases indefinitely, the equation will tend to the limiting form of

$$y_t = \beta \sum_{i=0}^{\infty} \varphi^i x_{t-i} + \sum_{i=0}^{\infty} \varphi^i \varepsilon_{t-i}.$$

Both the systematic and the stochastic part of this equation comprise the geometric lag scheme, whereas, in the previous model, the disturbance term  $\varepsilon_t$  was assumed to be an element of a sequence of independently and identically distributed random variables.

Nerlove (1956) was an influential advocate of models with lagged dependent variables. In their original form, before the expansion that gives rise to the impulse response function, they are amenable to estimation by ordinary least-squares regression.

Whereas it can be justified by the fact that it is the consequence of some simple behavioural models—namely of the partial adjustment mechanism and the adaptive expectations mechanism—the geometric lag scheme has a limited ability to model the variety of dynamic behaviour that one is liable to discover amongst economic variables. For that reason, one might be advised, at least, to adopt a second-order feedback scheme of the sort that is embodied in the following equation:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t.$$

The second-order model is capable of generating a variety of impulse responses, including a damped sinusoidal response.

The failure of its equations to accommodate an adequate range of dynamic behaviour may have been one of the main reasons for the demise of the largescale macro-econometric models. The simple autoregressive moving-average models, which often show superior predictive powers, lack any explicit explanatory variables. Such models are aimed at capturing the inherent dynamics of the data sequences. They take the generic form of

$$\begin{aligned}
y_t &= \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} \\
&\quad + \mu_0 \varepsilon_t + \mu_1 \varepsilon_{t-1} + \cdots + \mu_q \varepsilon_{t-q}.
\end{aligned}$$

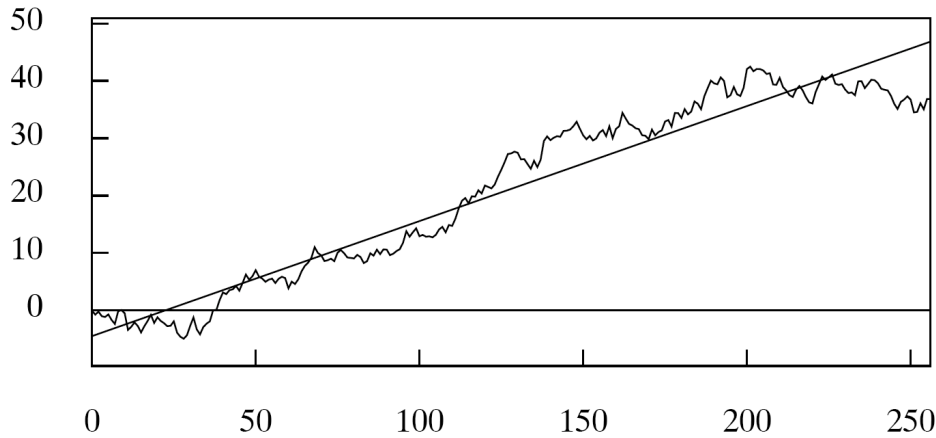
This equation comprises an autoregressive component with  $p$ th-order feedback and a moving-average component formed as a weighted sum of  $q$  successive independently and identically distributed random variables.

Additional flexibility arises from the possibility of replacing the variable  $y_t$  by its first difference  $d_t = y_t - y_{t-1}$  or by its second difference  $d_t - d_{t-1} = y_t - 2y_{t-1} + y_{t-2}$  to create the so-called autoregressive integrated moving average (ARIMA) models.

The integrated models that account for the differences of a random variable have provided effective means of mimicking the trajectories of macroeconomic variables, such as income and consumption. The simplest of such models, which defines a random-walk process in one dimension, has the equation

$$d_t = y_t - y_{t-1} = \varepsilon_t,$$

wherein  $\varepsilon_t$  is an element from a sequence of independently and identically distributed random variables.



**Figure 6.** A random walk generated by the equation  $y_t = y_{t-1} + \alpha + \varepsilon_t$  together with an interpolated regression line. The variance of the white-noise disturbance is  $V(\varepsilon_t) = 1$  and the drift parameter is  $\alpha = 0.2$ .

By applying the process of back substitution that has generated the geometric distributed lag from the model with first-order feedback, it can be shown that

$$y_t = \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1 + y_0.$$

Here,  $y_0$  represents the starting value at time  $t = 0$  of a process that wanders haphazardly in random steps. A random walk in two dimensions, which resembles a drunkard's walk, is described as Brownian motion.

Although the process defined above has a zero expected value, it has no central tendency and its variance or dispersion increases with each step. A process that has a tendency to give rise to increasing values can be obtained by including of a constant term  $\alpha > 0$ , to create the equation

$$d_t = y_t - y_{t-1} = \alpha + \varepsilon_t.$$

The outcome is described as a random walk with drift. This is illustrated in Figure 6.

The process that begins at time  $t = 0$  has an expected value of  $E(y_t) = \alpha t$  for its level. The variance of  $y_t$  continues to increase with each step. Therefore, unless the steps are very small in comparison to the value of  $\alpha$ , the expected value of  $y_t$  will be a poor predictor of its actual value.

## Problems with Trended Data

Many econometric data sequences are nonstationary, with trends that persist for long periods. However, the usual linear regression procedures presuppose that the relevant moments of the data, i.e. their variances and covariances etc., will converge asymptotically to fixed limits as the sample size increases. This cannot happen if the data are trended, in which case, the standard techniques of statistical inference will not be applicable.

In order to apply the regression procedures successfully, it is necessary to find some means of reducing the data to stationarity. A common approach is to subject the data to as many differencing operations as may be required to achieve stationarity. Often, only a single differencing is required.

A problem with differencing is that it tends to remove, or at least to attenuate severely, some of the essential information regarding the behaviour of economic agents. There may be processes of equilibration by which the relative proportions of econometric variables are maintained over long periods of time. The evidence of these will be lost in the process of differencing the data.

When the original undifferenced data sequences share a common trend, the coefficient of determination in a fitted regression, which measures the goodness of fit, is liable to be high; but it is often discovered that the regression model loses much of its explanatory power when the differences of the data are used instead.

In such circumstances, one might use the so-called cointegrating error-correction model. The model depicts a mechanism whereby two or more trended economic variables maintain an enduring long-term proportionality with each other. Moreover, the data sequences comprised by the model are stationary, either individually or in an appropriate combination; and this enables us to apply the standard procedures of statistical inference that are available to models comprising data from stationary processes.

Consider taking  $y_{t-1}$  from both sides of the equation of  $y_t = \phi y_{t-1} + \beta x_t + \varepsilon_t$ , which represents the first-order dynamic model. This gives

$$\begin{aligned}d_t &= y_t - y_{t-1} = (\phi - 1)y_{t-1} + \beta x_t + \varepsilon_t \\ &= (1 - \phi) \{ [\beta x_t / (1 - \phi)] - y_{t-1} \} + \varepsilon_t \\ &= \lambda \{ \gamma x_t - y_{t-1} \} + \varepsilon_t,\end{aligned}$$

where  $\lambda = 1 - \phi$  and where  $\gamma = \beta / (1 - \phi)$  is the gain of the transfer function, which is described by economists as the long-term multiplier in the relationship between  $x$  and  $y$ . This is the so-called error-correction form of the equation; and it indicates that the change in  $y_t$  is a function of the extent to which the proportions of the values  $x_t$  and  $y_{t-1}$  differs from those that would prevail in the steady state.

The error-correction form provides the basis for estimating the parameters of the model when the signal sequence  $\{x_t\}$  is trended or nonstationary. A pair of nonstationary series that maintain a long-run proportionality are said to be cointegrated. It is easy to obtain an accurate estimate of  $\gamma$ , which is the coefficient of proportionality, simply by running a regression of  $y_{t-1}$  on  $x_t$ .

Once a value for  $\gamma$  is available, the remaining parameter  $\lambda$  may be estimated by regressing  $d_t = y_t - y_{t-1}$  upon the composite variable  $\{\gamma x_t - y_{t-1}\}$ . This amounts to the two-step procedure of Engle and Granger (1987). However, if the error-correction model is an unrestricted reparametrisation of an original model in levels, then the parameters of the latter can be estimated by ordinary least-squares regression. The estimates of the parameters of the error-correction form of the model can be inferred directly from the least-squares estimates of the parameters of the original model in levels.

The developments that culminated in cointegration analysis occurred on many fronts. However, it is commonly agreed that a significant impetus was given by Granger and Newbold (1978), who revealed the dangers of reaching spurious conclusions when regression models are fitted to the levels of trended economic variables. It was revealed that the success of econometric analyses that had delivered models that seemed to fit such data exceedingly well was illusory. Another article by Granger (1981) paved the way.

The analysis of Engle and Granger (1987), for which they were jointly awarded a Nobel prize, was extended by others, including Johansen (1988), who showed how to estimate multi-equations systems that comprise several cointegrating relationships amongst their variables.

In the context of a multi-equation linear system, a cointegrating relationship amounts to a linear combination of the observed trended variables that creates an untrended resultant. Such relationships exist only to the extent that the variables within the economy obey stringent laws of proportionality. That is to say, multi-equation models of cointegration create straightjackets that often cannot easily be fitted to the economic variables. Econometricians continue to seek more flexible ways of modelling the complex interdependencies of economics variables.

## References

Alt, F.L., (1942), Distributed Lags, *Econometrica*, 10, 113–128.

Anderson, T.W., (1950), Estimation of the Parameters of a Single Equation by the Limited-Information Maximum-Likelihood Method, Chapter 9 in *Statistical Inference in Dynamic Economic Models*, T. C. Koopmans (editor), Cowles Foundation for Research in Economics, Monograph No. 10, New York, John Wiley and Sons.

Anderson, T.W., and H. Rubin, (1949), Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, *Annals of Mathematical Statistics*, 20, 46–63.

Basmann, R.L., (1957), A Generalised Classical Method of Linear Estimation of Coefficients in a Structural Equation, *Econometrica*, 25, 77–84.

Box, G.E.P., and G.M. Jenkins, (1970), *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day.

Engle, R.F., and C.W.J. Granger, (1987), Co-integration and Error-Correction, Representation, Estimation and Testing, *Econometrica*, 55, 251–276.

Fisher, R.A., (1925), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh. An internet version of the book is available at the address <http://psychclassics.yorku.ca/Fisher/Methods/>

Galton, F., (1865), Hereditary Character and Talent, *Macmillan's Magazine*, 12, 157–166.

- Galton, F., (1869), *Hereditary Genius*, London, Macmillan.
- Galton, F., (1886), Regression Towards Mediocrity in Hereditary Stature, *Journal of the Anthropological Institute*, 15, 246–63.
- Granger, C.W.J. and P. Newbold, (1974), Spurious Regressions in Econometrics, *Journal of Econometrics*, 2, 111–120.
- Granger, C.W.J., (1981), Time Series Data and Econometric Model Specification, *Journal of Econometrics*, 16, 121--30.
- Hood, W.C., and T.C. Koopmans (editors), (1953), *Studies in Econometric Method*, Cowles Commission for Research in Economics, Monograph No. 14, New York, John Wiley and Sons.
- Johansen, S., (1988), Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, 12, 231–254.
- Klein, L.R., (1950), *Economic Fluctuations in the United States, 1921--1941*, Monograph No. 11 of the Cowles Commission for Research in Economics, New York, John Wiley and Sons.
- Koopmans, T.C., (editor) (1950), *Statistical Inference in Dynamic Economic Models*, Cowles Commission for Research in Economics, Monograph No. 10, New York, John Wiley and Sons.
- Koopmans, T.C., and W.M. Hood (1953), The Estimation of Simultaneous Linear Economic Relationships, Chapter 6 in *Studies in Econometric Method*, W.C. Hood and T.C. Koopmans (editors), Cowles Commission for Research in Economics, Monograph No. 14, New York, John Wiley and Sons.
- Koopmans, T.C., H. Rubin, and R.B. Leipnik (1950), Measuring the Equation Systems of Dynamic Economics, Chapter 2 in *Statistical Inference in Dynamic Economic Models*, T. C. Koopmans (editor), Cowles Commission for Research in Economics, Monograph No. 10, New York, John Wiley and Sons.
- Kørgaard, N., (1984), The Earliest History of Econometrics: Some Neglected Danish Contributions, *History of Political Economy*, 16, 437--444.
- Kørgaard, N., P. Andersen, and N. Topp (1998), The Danish Economist Jens Warming---An Odd Genius, in Vol. 1 of *European Economists of early 20th Century*, edited by Samuels and Cheltenham, Edward Elgar Publishing, UK.
- Koyck, L.M., (1954), *Distributed Lags and Investment Analysis*, Amsterdam, North-Holland Publishing Co.
- Marshall. A., (1920), *Principles of Economics*, London, Macmillan.
- Morgan, Mary S., (1990), *The History of Econometric Ideas*, Cambridge University Press, Cambridge,
- Nerlove, M., (1956), Estimates of the Elasticities of Supply of Selected Agricultural Commodities, *Journal of Farm Economics*, 38, 496–509.
- Pearson, K., (1896), Mathematical Contributions to the Mathematical Theory of Evolution: III. Regression, Heredity, and Panmixia, *Philosophical Transactions of the Royal Society of London*, 187, 253–318.
- Pearson, K., and A. Lee, (1896), Mathematical Contributions to the Theory of Evolution, On Telegony in Man, &c., *Proceedings of the Royal Society of London*, 60, 273–283.

- Pearson, K., and A. Lee, (1903), On the Laws of Inheritance in Man: I. Inheritance of Physical Characters, *Biometrika*, 2, 357–462.
- Qin, Duo, (1993), *The Formation of Econometrics*, Oxford, Clarendon Press.
- Raj, B., and J. Koerts (eds.), (1992), *Henri Theil's Contribution to Economics and Econometrics: Volume I, Econometric Theory*, Dordrecht, Kluwer.
- Simon, H., (1953), Causal Ordering and Identifiability, Chapter 3 in *Studies in Econometric Method*, W.C. Hood and T.C. Koopmans (editors), Cowles Commission for Research in Economics, Monograph No. 14, New York, John Wiley and Sons.
- Theil, H., (1953), Estimation and Simultaneous Correlation in Complete Equation Systems, mimeographed memorandum of the Central Planning Bureau, The Hague. Reprinted as Chapter 6 in Raj and Koerts (eds.) 1992, Volume I. Theil, H., (1958), *Econometric Forecasts and Policy*, Amsterdam, North-Holland Publishing Co.
- Tinbergen, J., (1937), *An Econometric Approach to Business Cycle Problems*, Paris, Hermann et Cie.
- Tinbergen, J., (1939), *Statistical Testing of Business Cycle Theories. Vol. I: A Method and its Application to Investment Activity. Vol II: Business Cycles in the United States of America, 1919–1932*, Geneva, League of Nations.
- Tinbergen, J., (1951), *Business Cycles in the United Kingdom 1870–1914*, Amsterdam, North-Holland Publishing Co.
- Wold, H.O.A., (1949), Statistical Estimation of Economic Relationships, *Econometrica*, 17, (Supplement), 1--22.
- Wold, H.O.A., (1949), Causality and Econometrics," *Econometrica*, 22, 162--177
- Working, E.J., (1927), What Do Statistical "Demand Curves" Show? *The Quarterly Journal of Economics*, Vol. 41, No. 2. (Feb., 1927), pp. 212–235.
- Working, H., (1925), The Statistical Determination of Demand Curves, *The Quarterly Journal of Economics*, 39, 503–543.
- Zellner, A., and H. Theil (1962), Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations, *Econometrica*, 30, 54–78.