S

Statistics and Economics

Aris Spanos

Abstract

Some statisticians and economists might find it surprising to learn that statistics and economics share common roots going back to 'Political Arithmetic' in the mid-17th century. The primary objective of this article is to revisit the common roots and trace the parallel development of both disciplines up to and including the 20th century, and to attempt to signpost certain methodological lessons that were missed along the way to the detriment of both disciplines. The emphasis is primarily on methodological developments, with less attention paid to institutional developments.

Keywords

ARIMA models; Bayes, T.; Bernoulli, J.; Bowley, A. L.; Central limit theorems; Cointegration; Convergence in distribution; Cournot, A. A.; Cowles Commission; Davenant, C.; Econometric Society; Edgeworth, F. Y.; Error-correction models; Farr, W.; Fisher, I.; Fisher, R. A.; Frequentist approach to inference; Galton, F.; Gauss, C. F.; Gauss–Markov theorem; Generalized method of moments; Graphical techniques; Graunt, J.; Haavelmo, T.; Heckman, J. J.; Hume, D.; Identification; Index numbers; Induction; Inverse probability; Jevons, W. S.; King, G.; Koopmans, T. C.; Laplace, P.-S.; Law of large numbers; Least squares; Legendre, A.-M.; Life tables; Marginal revolution; Mathematics and economics; Mills, F. C.; Mortality; Neyman, J.; Nonparametric methods; Pearson, K.; Petty, W.; Playfair, W.; Political arithmetic; Political economy; Probability; Quetelet, A.; Reliability of inference; Royal Statistical Society; Semiparametric methods; Simultaneous equations models; Specification; Spurious regressions; Statistical adequacy; Statistical description; Statistical inference; Statistical models; Statistical Society of London; Statistics and economics; Stochastic processes; Structural models; Unit roots; Walras, L.; Yule, G. U.

JEL Classifications

B0

The close interrelationship between economics and statistics, going back to their common roots in 'Political Arithmetic', played a crucial role in availing the development of both disciplines

This chapter was originally published in *The New Palgrave Dictionary of Economics*, 2nd edition, 2008. Edited by Steven N. Durlauf and Lawrence E. Blume

© The Author(s) 2008
Palgrave Macmillan (ed.), *The New Palgrave Dictionary of Economics*,
DOI 10.1057/978-1-349-95121-5_1935-1

during their practical knowledge (pre-academic) period. Political economy was first separated from political arithmetic and became an academic discipline – the first social science – at the end of the 18th century, partly as a result of political arithmetic losing credibility. Statistics emerged as a 'cleansed' version of political arithmetic, focusing on the collection and tabulation of data, and continued to develop within different disciplines including political economy, astronomy, geodesy, demography, medicine and biology; however, it did not become a separate academic discipline until the early 1900s.

During the 19th century the development of statistics was institutionally nurtured and actively supported by the more empirically oriented political economists such as Thomas Malthus who helped to create section F of the Royal Society, called 'Economic Science and Statistics', and subsequently to found the Statistical Society of London. The teaching of statistics was introduced into the university curriculum in the 1890s, primarily in economics departments (see Walker 1929).

The close relationship between economics and statistics was strained in the first half of the 20th century, as the descriptive statistics tradition, associated with Karl Pearson, was being transformed into modern (frequentist) statistical inference in the hands of Fisher (1922, 1925, 1935a, 1956) and Neyman and Pearson (1933), and Neyman (1935, 1950, 1952). During the second half of the 20th century this relationship eventually settled into a form of uneasy coexistence. At the dawn of the 21st century there is a need to bring the two disciplines closer together by implementing certain methodological lessons overlooked during the development of modern statistics.

The 17th Century: Political Arithmetic, the Promising Beginnings

If one defines statistics broadly as 'the subject matter of collecting, displaying and analysing data', the roots of the subject are traditionally traced back to John Graunt's (1620–74) *Natural and Political Observations upon the Bills of Mortality*, published in 1662 (see Hald 1990; Stigler 1986), the first systematic study of demographic data on birth and death records in English cities. Graunt detected surprising regularities stretching back over several decades in a number of numerical aggregates, such as the male/female ratio, fertility rates, death rates by age and location, infant mortality rates, incidence of new diseases and epidemics, and so on. On the basis of these apparent regularities, Graunt proceeded to draw certain tentative inferences and discuss their implications for important public policy issues. Hald summarized the impact of this path-breaking book as follows:

> Graunt's book had immense influence. Bills of mortality similar to the London bills were introduced in other cities, for example, Paris in 1667. Graunt's methods of statistical analysis were adopted by Petty, King and Davenant in England; Vauban in France; by Struyck in the Netherlands; and somewhat later by Sussmilch in Germany. Ultimately, these endeavours led to the establishment of governmental statistical offices. Graunt's investigation on the stability of the sex ratio was continued by Arthuthnott and Nicolas Bernoulli. (Hald 1990, p. 103)

Graunt's book had close affinities in both content and objectives to several works by his close friend William Petty (1623–87) on 'Political Arithmetick' published during the 1670s and 1680s; Graunt and Petty are considered joint founders of the 'political arithmetic' tradition (Redman 1997). The fact that Graunt had no academic credentials and published only the single book led to some speculation in the 1690s, which has persisted to this day, that Petty was the real author of *The Bills of Mortality*. The current prevailing view (see Greenwood 1948; Kreager 1988) is that Petty's potential influence on Graunt's book is marginal at best. Stone aptly summarizes this view as follows:

> Graunt was the author of the book associated with his name. More than likely, he discussed it with his friend; Petty may have encouraged him to write it, contributed certain passages, helped obtaining the Bills for the county parish. . .at Romsey, the church in which Petty's baptism is recorded and in which he is buried; he may even have suggested the means of interpolating the numbers of survivors between childhood and old age. But all this does not amount

> to joint let alone sole authorship. (Stone 1997, p. 224)

Hull (1899), one of Petty's earliest biographers and publisher of his works, made a strong case against Petty being the author of the 'Bills of Mortality' by comparing his methodological approach to that of Graunt:

> Graunt exhibits a patience in investigation, a care in checking his results in every possible way, a reserve in making inferences, and a caution about mistaking calculation for enumeration, which do not characterize Petty's work to a like degree.
>
> The spirit of their work is often different when no question of calculation enters. Petty sometimes appears to be seeking figures that will support a conclusion which he has already reached; Graunt uses his numerical data as a basis for conclusions, declining to go beyond them. He is thus a more careful statistician than Petty, but he is not an economist at all. (Hull 1899, pp. xlix and lxxv)

Both Graunt and Petty used limited data to draw conclusions and make predictions about the broader populations, exposing themselves to severe criticisms as to the appropriateness and reliability of such inferences. For instance, using data on christenings and burials in a single county parish in London, they would conjure up estimates of the population of London (which included more than 130 parishes), and then on the basis of those estimates, and certain contestable assumptions concerning mortality and fertility rates, proceed to project estimates of the population of the whole of England. The essential difference between their approaches is that Graunt put enough emphasis on discussing the possible *sources of error* in the collection and compilation of his data, as well as in his assumptions, enabling the reader to assess the reliability (at least qualitatively) of his inferences. Petty, in contrast, was more prone to err on the side of political expediency by drawing inferences that would appeal to the political powers of his time (see Stone 1997).

Graunt and Petty considered statistical analysis a way to draw *inductive inferences* from observational data, analogous to performing experiments in the physical sciences (see Hull 1899, p. lxv). Political arithmetic stressed the importance of a new method of quantitative measurement – 'the art of reasoning by figures upon things relating to the government' – and was instrumental in the development of both statistics and economics (see Redman 1997, p. 143). The timing of this emphasis on quantitative measurement and the collecting of data was not coincidental. The *empiricist* turn pioneered by Francis Bacon (1561–1626) had a crucial impact on intellectual circles such as the London Philosophical Society and the British Association, with which Graunt and Petty were associated – these circles included Robert Boyle, John Wallis, John Wilkins, Samuel Hartlib, Christopher Wren and Isaac Newton. As summarized by Letwin:

> The scientific method erected by Bacon rested on two main pillars: natural history, that is, the collection of all possible facts about nature, and induction, a careful logical movement from those facts of nature to the laws of nature. (Letwin 1965, p. 131)

Graunt and Petty were also influenced by philosopher John Locke (1632–1704), through personal contact. Locke was the founder the British empiricist tradition, which continued with George Berkeley (1685–1753) and David Hume (1711–76). Indeed, all three philosophers wrote extensively on political economy as it relates to empirical economic phenomena, and Locke is credited with the first use of the most important example of analytical thinking in economics, the demand-supply reasoning in determining price (see Routh 1975).

Graunt's and Petty's successors in the political arithmetic tradition, Gregory King (1648–1712) and Charles Davenant (1656–1714) continued to emphasize the importance of collecting data as the only objective way to frame and assess sound economic policies. Their efforts extended the pioneering results of Grant and Petty and provided an improved basis for some of the original predictions (such as the population of England), but they did not provide any new methodological insights into the analysis of the statistical regularities originally enunciated by Graunt. The enhanced data collection led to discussions of how certain economic variables should be measured over time, and a new literature on index numbers was pioneered by William Fleetwood (1656–1723). The roots of national income accounting, which eventually led to the current standardized macro-data time series, can be traced back to the efforts of these early pioneers in political arithmetic (see Stone 1997).

According to Hald:

> His [Graunt's] life table was given a probabilistic interpretation by the brothers Huygens; improved life tables were constructed by de Witt in the Netherlands and by Halley in England and used for the computation of life annuities. The life table became a basic tool in medical statistics, demography, and actuarial science. (Hald 1990, p. 1034)

The improved life tables, with proper probabilistic underpinnings, were to break away from the main political arithmetic and become part of a statistical/probabilistic tradition that would develop independently in Europe in the next two centuries, giving rise to a new literature on life tables and insurance mathematics (see Hald 1990).

A methodological digression. This was a crucial methodological development for data analysis because it was the first attempt to provide probabilistic underpinnings to Graunt's statistical regularities. Unfortunately, the introduction of probability in the life tables was of limited scope and had no impact on the broader development of political arithmetic, which was growing during the 18th century without any concerns for any probabilistic underpinnings. Without such underpinnings, however, one cannot distinguish between real regularities and artifacts.

The 18th Century: The Demise of Political Arithmetic

At the dawn of the 18th century political arithmetic promised a way to provide an objective basis for more reliable framing and assessment of economic and social policies. As described by Petty, the method of political arithmetic replaces the use of 'comparative superlative words, and intellectual arguments' with 'number, weight, or measure; to use only arguments of sense; and to consider only such causes as have visible foundations in nature, leaving those that depend on the mutable minds, opinions, appetites, and passions of particular men, to the consideration of others' (Hull 1899, p. 244).

English political institutions, including the House of Commons, the House of Lords and the monarchy, took full advantage of the newly established methods of political arithmetic and encouraged, as well as financed, the collection of new data as needed to consider specific questions of policy (see Hoppit 1996). Putting these methods to the (almost exclusive) service of policy framing by politicians carried with it a crucial danger for major abuse. An inherent problem for social scientists in general has always been to distinguish between inferences relying on sound scientific considerations and those motivated by political or social preferences and leanings.

The combination of (*a*) the absence of sound probabilistic foundations that would enable one to distinguish between real regularities and artefacts, and (*b*) the inbuilt motivation to abuse data in an attempt to make a case for one's favourite policies, led inevitably to extravagant and unwarranted speculations, predictions and claims. These indulgences eventually resulted in the methods of political arithmetic losing credibility. The extent of the damage was such that Greenwood, in reviewing 'Medical Statistics from Graunt to Farr', argued:

> One may fairly say on the evidence here summarized that the eighteenth-century political arithmeticians of England made no advance whatever upon the position reached by Graunt, Petty and King. They were second-rate imitators of men of genius. (Greenwood 1948, p. 49)

An important component of the evidence provided by Greenwood was the 'population controversy', which often involved idle speculation in predicting the population of England. This speculation began with Graunt with a lot of cautionary notes attached, but it continued into the 18th century with much less concern about the possible errors that could vitiate such inferences. The discussions were from two opposing schools of thought: the *pessimists*, who claimed that the population was decreasing, and the *optimists*, who argued the opposite; their conflicting arguments were based on the same bills of mortality popularized by Graunt. Neither side had reliable evidence for its predictions because the data provided no sound basis for reliable inference. All predictions involved highly conjectural assumptions of fertility and mortality rates, the average number of people living in each house, and so on. The acrimonious arguments between the two sides revealed the purely speculative foundations of all such claims and contributed significantly to the eventual demise of political economy (see Glass 1973, for a detailed review).

The above quotation from Greenwood might be considered today as an exaggeration, but it describes accurately the prevailing perception at the end of the 18th century. An unfortunate consequence of disparaging the methods of political arithmetic was the widely held interpretation that it provided decisive evidence for the ineffectiveness of Bacon's *inductive method*. Indeed, one can argue that this cause was instrumental in the timing of the emergence of *political economy* at the end of the 18th century, as the first social science to break away from political arithmetic. Adam Smith (1723–90) declared: 'I have no great faith in political arithmetick' (1776, p. 534). James Steuart (1712–80) was even more critical:

> Instead of appealing to political arithmetic as a check on the conclusions of political economy, it would often be more reasonable to have recourse to political economy as a check on the extravagances of political arithmetic. (quoted by Redman 1997)

During the late 18th century, political economy defined itself by contrasting its methods with those of political arithmetic, arguing that it did not rely only on tables and figures in conjunction with idle speculation, but was concerned with the theoretical issues, causes and explanations underlying the process that generated such data. Political economists contrasted their primarily deductive methods to the discredited inductive methods utilized by political arithmeticians. As argued by Hilts:

> Of importance to the history of statistics in England was the fact that the political economists were fully conscious of their deductive proclivities and saw political economy as methodologically distinct from the inductive science of statistics. (Hilts 1978, p. 23)

At this point it should be emphasized that the terms induction and deduction had different connotations during the 18th century, and care should be taken when interpreting some of the claims of that period (see Redman 1997). Despite the criticisms by leading political economists of the inductive method, broadly understood as using the data as a basis of inference, the tradition of collecting, compiling and charting data as well as drawing inferences concerning broad tendencies on such a basis, continued to grow throughout the 18th and 19th centuries, and was influential in the development of political economy. Some political economists such as Thomas Malthus (1766–1834) and John McCulloch (1789–1864) continued to rely on the British empiricist tradition of using data as a basis of inference, but were at great pains to separate themselves from the 18th century's discredited political arithmetic tradition. Indeed, the leading political economists of that period, including Adam Smith and David Ricardo (1772–1823), used historical data extensively in support of their theories, conclusions and policy recommendations developed by deductive arguments (see Backhouse 2002a).

At the close of the 18th century, the only bright methodological advance in the withering tradition of political arithmetic was provided by William Playfair's (1759–1823) *The Commercial and Political Atlas*, published in 1786. This book elevated the analysis of tabulated data to a more sophisticated level by introducing the power of graphical techniques in displaying and analysing data. Playfair introduced several innovating techniques such as hachure, shading, colour coding, and grids with major and minor divisions of both axes to render the statistical regularities in the data even more transparent. In a certain sense, the graphical techniques introduced by Playfair made certain empirical regularities more transparent and rendered certain conclusions easier to draw. The graphs in this book represent economic time series, measuring primarily English trade (imports/exports) with other countries during the 18th century. Indeed, Playfair's writings were mainly on political economy; his first book, *Regulation of the Interest of Money*, was published in 1785 (see Harrison 2004).

In what follows the developments in probability theory will be discussed only when they pertain to the probabilistic underpinnings of statistical analysis; for a more detailed and balanced discussion see Hald (1990, 1998, 2007). The probabilistic underpinnings literature on probability developed independently from political arithmetic in England, and there was no interaction between the two until the mid-19th century.

Viewed from today's vantage point, the primary problem with Grant's inferences based on data pertaining to a single parish in London, was how 'representative' the data were for the population of London as a whole, which included more than 130 other parishes. This problem was formalized much later in terms of whether the data can be realistically viewed as a 'random sample' from the population of London. Defining what a random sample is, however, requires probability theory, which was not adequately understood until the late 19th century (see Peirce 1878).

Jacob Bernoulli. The first important result relating to the probabilistic underpinnings of statistical regularities was Jacob Bernoulli's (1654–1705) Law of Large Numbers (LLN), published posthumously in 1713 by his nephew Nicolas Bernoulli (1687–1759). Bernoulli's theorem showed that *under certain circumstances*, the relative frequency of the occurrence of a certain event A, say $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{m}{n}$ (m occurrences of $\{X_i = 1\}$ and $n{-}m$ occurrences of $\{X_i = 0\}$ in n trials) provides an estimate of the *p*robability $\mathbb{P}(A)$ p whose accuracy increases as n goes to infinity. In modern terminology $\bar{X}$ constitutes a consistent estimator of p. Bernoulli went on to use this result in an attempt to provide an interval estimator of the form: p is in $\bar{X} \pm \varepsilon$ for some $\varepsilon > 0$, but his estimator was rather crude (see Hald 1990).

A methodological digression. The *circumstances* assumed by Bernoulli were specified in terms of the trials being *independent and identically distributed* (IID). It turned out that the same probabilistic assumption defines the notion of a *random sample* mentioned in relation to the probabilistic underpinnings concerning Graunt's statistical regularities, though the two literatures were developing independently. The role of these probabilistic underpinnings was not made explicit, however, until the early 1920s (see section "The Fisher–Neyman–Pearson Approach"). Indeed, the role of the IID assumptions is often misunderstood to this day. For instance, Hilts argues:

> Mathematically the theorem stated [LLN], in very simplified language, that an event which occurs with a certain probability, appears with a frequency approaching that probability as the number of observations is increased. (Hilts 1973, p. 209)

Strictly speaking, the LLN says nothing of a sort, because, unless the trials are IID, the result does *not* follow. This insight was clearly articulated by Uspensky:

> It should, however, be borne in mind that little, if any, value can be attached to the practical applications of Bernoulli's theorem, unless the conditions presupposed in this theorem are at least approximately fulfilled: independence of trials and constant probability of an event for every trial. (Uspensky 1937, p. 104)

Laplace. The first successful attempt to integrate data analysis with the probabilistic underpinnings should be credited to Pierre-Simon Laplace (1749–1827), a famous French mathematician and astronomer, and Thomas Bayes (1702–61), a British mathematician and Presbyterian minister. In papers published in 1764 and 1765 (see Hald 2007) respectively, they proposed the first inverse probability (posterior-based) interval for p for the form p of the form 'p is in $(\bar{x} \pm [\varepsilon|\bar{x}])$' for some $\varepsilon > 0$, by assuming a prior distribution $p \sim U(0,1)$ that is p is a uniformly distributed random variable (see Hacking 1975). This gave rise to the *inverse probability* approach (known today as the Bayesian approach) to statistical inference, which was to dominate statistical induction until the 1920s, before the Fisherian revolution. In 1812 Laplace (see Hald 2007) also provided the first frequentist interval estimator of p of the form p is in $(\bar{X} \pm \varepsilon)$ for some $\varepsilon > 0$. The difference between this result and a similar result by Bernoulli is that Laplace used a more accurate approximation based on *convergence in distribution* as the basis of his result; the first *central limit theorem* supplying an asymptotic approximation of the binomial by the Normal distribution (see Hald 1990).

The 19th Century: Political Economy and Statistics

The demise of political arithmetic by the early 19th century was instrumental in contributing to the creation of two separate fields: political economy and statistics. Political economy was created to provide more reasoned explanations for the causes and contributing factors giving rise to economic phenomena. Statistics was demarcated by the narrowing down of the scope of political arithmetic in an attempt to cleanse it from the unwarranted speculation that undermined its credibility during the 18th century.

The Statistical Society of London

Given their common roots, the first institution created to foster the development of the field of statistics, the Statistical Society of London, was created in 1834 with the active participation of several political economists, including Thomas Malthus and Richard Jones (1790–1855), who, together with John Drinkwater (1801–51), Henry Hallam (1777–1859) and Charles Babbage (1791–1871), were to found the Society after some prompting from Quetelet, who visited England in 1833. Other political economists who played very active roles in the early stages of the Society included Thomas Tooke (1774–1858), John R. McCulloch (1789–1864) and Nassau Senior (1790–1864). The first council included notable personalities such as Earl FitzWilliam (1748–1833), William Whewell (1794–1866), G.R. Porter (1792–1852) and Samuel Jones-Loyd (1796–1883).

In an attempt to protect themselves from the disrepute on speculation based on data brought about by political arithmeticians, the new society was founded upon the explicit promise to put the emphasis, not on inference, but upon the collection and tabulation of data of relevance to the state. The founding document stated:

> The Statistical Society of London has been established for the purposes of procuring, arranging, and publishing Facts calculated to illustrate the condition and prospects of the Society. (Journal of the Statistical Society of London 1834, p. 1)

The seal on the cover of the *Journal of the Statistical Society of London* (*JSSL*) was a wheatsheaf around which was written '*aliis exterendum*' ('to be threshed by others'). That is, the aim of the society is to painstakingly gather the facts and let others draw whatever conclusions might be warranted:

> The Statistical Society will consider it to be the first and most essential rule of its conduct to exclude carefully all Opinions from its transactions and publications – to confine its attention rigorously to facts – and, as far as it may be found possible, to facts which can be stated numerically and arranged in tables. (JSSL 1834, pp. 1–2)

Of particular interest is the way the statement of the aims of the society separated statistics from political economy:

> The Science of Statistics differs from Political Economy because although it has the same end in view, it does not discuss causes, nor reason upon probable effects; it seeks only to collect, arrange, and compare, that class of facts which alone can form the basis of correct conclusions with respect to social and political government. (JSSL 1834, p. 2)

The overwhelming majority of the published papers in the *JSSL* were in the political arithmetic tradition of Graunt, relating primarily to economic, medical and demographic data, with two major improvements: ameliorated methods for the collection and tabulation of data giving rise to more accurate and reliable data, and more careful reasoning being used to yield less questionable inferences. This is particularly true for data relating to life tables and mortality rates associated with epidemics. The best examples of such an output are given by William Farr (1807–83), who is considered to be the founder of medical statistics because his analysis of such data contributed to medical advances and crucial changes in policies concerning public health (see Greenwood 1948). For a more extensive discussion of the methodological and institutional developments associated with data collection and tabulation in England and France see Schweber (2006) and Desrosières (1998).

By the 1850s it became apparent that the early founding declaration of the society to publish papers that stay away from 'Opinions' – drawing conclusions on the basis of data – was unrealistic, unattainable and unjustifiable in the minds of the members of the society. Despite this initial promise, slowly but surely *JSSL* publications began to go beyond the mere reporting and tabulation of data relating to economic, political, demographic, medical, moral and intellectual issues, including poverty figures and education statistics. The motto '*aliis exterendum*' was removed in 1857 from their seal to reflect the new vision of the society (see RSS 1934).

The Probabilistic Underpinnings in the 19th Century

During the early 19th century, a completely separate tradition in statistical analysis of data was being developed in Europe (mainly in France and Germany) in the fields of *astronomy* and *geodesy*. This literature was developing completely independently of political arithmetic, but by the 1840s the two traditions had merged in the hands of Adolphe Quetelet (1796–1874): see Porter (1986).

Legendre and Gauss. In the early 19th century the analysis of astronomical and geodesic data by Adrien-Marie Legendre (1752–1833), Carl Friedrich Gauss (1777–1855) and Laplace introduced curve-fitting as a method to summarize the information in data (see Farebrother 1999). In modern notation the simplest form of curve-fitting can be expressed in the form of a linear model $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, where $\mathbf{y}: = (y_1, y_2, \ldots, y_n)$ and $\mathbf{X}: = (\mathbf{x}_1, \mathbf{x}_2, \ldots, x_n)$ denote a vector and a matrix of observations, respectively, $\beta := (\beta_1, \beta_2, \ldots, \beta_m)$ a vector of unknown parameters and $\varepsilon := (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ a vector of errors. Legendre (1805) is credited with inventing least squares as a mathematical approximation method, by proposing the minimization of $l(\beta) = (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}}(\mathbf{y} - X\beta)$ as a way to estimate β. Gauss (1809) should be credited with providing the probabilistic underpinnings for this estimation problem by transforming the mathematical approximation error into a generic statistical error:

$$\varepsilon_k(x_k, m) = \varepsilon_k \sim \mathrm{NIID}\left(0, \sigma^2\right), k = 1, 2, \ldots, n, \ldots, \tag{1}$$

where NIID(0, σ^2) stands for 'Normal, Independent and Identically Distributed with mean 0 and variance σ^2'. Laplace provided the first justification of the Normality assumption based on the central limit theorem in 1812 (see Hald 2007). What makes Gauss's contribution all-important from today's vantage point is that the probabilistic assumptions in (1) provide the framework that enables one to assess the reliability of inference. Ironically, Gauss's embedding of the mathematical approximation problem into a statistical model is rarely appreciated as the major contribution that it is (see Spanos 2008). Instead, what Gauss is widely credited with is the celebrated Gauss-Markov theorem (see section "Demarcating the Boundaries of Modern Statistics").

Quetelet. The 'law of error' was elevated to a most important method in analysing social phenomena by Adolphe Quetelet (1796–1874), a Belgian astronomer and polymath, in the 1840s. His statistical analysis of data differed in that his methods were integrated with the probabilistic underpinnings that were lacking in the analysis of political arithmeticians; his probabilistic perspective was primarily influenced by the work of Joseph Fourier (1768–1830), a French mathematician and physicist. Quetelet's most important contribution was to explicate Graunt's regularities in terms of the notion of probabilistic (chance) regularity which combined the unpredictability at the individual level with the abiding regularity at the aggregate level. By fitting the Normal curve over the histogram of a great variety of social data, his objective was to eliminate 'accidental' influences and determine the average physical and intellectual features of a human population, including normal and abnormal behaviour. His *modus operandi* was the notion of the 'average man' (see Desrosières 1998). The 'average man' began as a simple way of summarizing the systematic characteristic of a population, but in some of Quetelet's later work, 'average man' is presented as an ideal type, and any deviations from this ideal were interpreted as *errors* of nature.

A methodological digression. In addition to the substantive issues raised by his approach to 'social physics' (see Cournot 1843), the methodological underpinnings of Quetelet's statistical analysis were rather weak. When the Normal curve is fitted over a histogram of data $\mathbf{x}:=(x_1, x_2,\ldots,x_n)$ in an attempt to summarize the statistical regularities, one implicitly assumes that data $\mathbf{x}$ constitutes a realization of an IID process $\{X_k, k=1,2,\ldots,n,\ldots\}$ (see Spanos 1999); these are highly questionable assumptions for most of the data used in Quetelet (1942). The concern to evaluate the precision (reliability) of an inference, introduced earlier by Laplace and Gauss, was absent from Quetelet's work. Hence, his analysis of statistical regularities did not give rise to any more reliable inferences than those of the political arithmetic tradition a century earlier; the necessity to assess the validity of the premises (NIID assumptions) for inductive inference was not clearly understood at the time.

The 'Mathematical' Turn

In the last quarter of the 19th century there was a concerted effort to render both statistics and economics more rigorous by introducing the language of mathematics into both disciplines. In statistics this effort was spearheaded by Edgeworth, Galton and Pearson and in economics by Edgeworth, Jevons, Walras and Irving Fisher. The mathematical turn of this period was motivated by the strong desire to emulate the physical sciences and introduce quantification into these fields, which involved both calculus and probability theory (see Backhouse 2002a, b).

Galton. Quetelet's use of the Normal curve to analyse social data had a powerful influence on Francis Galton (1822–1911), who provided a different interpretation to the 'law of error'. Galton (1869) interpreted the variation around the mean, not as errors from the ideal type, but as the very essence of nature's variability. Using this variability he introduced the notions of regression and correlation in the 1890s as a way to determine relationships between different data series $\{(x_k, y_k), k=1,2,\ldots,n\}$. Regression and correlation opened the door for providing statistical explanations which revolutionized statistical modelling in the biological and social sciences (see Porter 1986). Retrospectively, Galton was the founder of the biometrics tradition, which had a great influence on the development of statistics in the 20th century in the hands of Karl Pearson (1857–1936) and Udny Yule (1871–1951).

Pearson significantly extended the summarization of data in the form of smoothing histograms, by introducing a whole family of new frequency curves – known today as the Pearson family – to supplement the Normal curve, and applied these techniques extensively to biological data, with notable success. He also provided clear probabilistic underpinnings for Galton's regression and correlation methods. *Yule* (1897) established a crucial link between the Legendre–Gauss least-squares and the linear regression model, by showing that least-squares can be used to estimate the parameters of linear regression, bringing together two seemingly unrelated literatures (see Spanos 1999). This was an important breakthrough that, unfortunately, also introduced a confusion between two different perspectives on empirical modelling: curve-fitting as a mathematical approximation method, and the probabilistic perspective where regression is viewed as a purely probabilistic concept defined in terms of the first moment of a conditional distribution (see Stigler 1986; Spanos 2008). Yule published a highly influential textbook in statistics in 1911 in which he successfully blended the biometric with the 'economic statistics' tradition.

Edgeworth. Of particular interest for the fundamental interaction between statistics and economics is the case of Francis Edgeworth (1845–1926), primarily an economist. His mathematical self-training enabled him to provide a bridge between the theory of errors tradition going back to Gauss and Laplace, the biometric tradition of Galton and Pearson, and the more traditional economic statistics of the 19th century focusing on economic time series data and index numbers. His direct influence on statistics, however, was rather limited because the style and mathematical level of his writings were too demanding for the statisticians of the late 19th century. Bowley (1928), ‘at the request of the Council prepared a summary of his mathematical work which may have served to make his achievement known to a wider circle’ (see RSS 1934, p. 238). Edgeworth contributed crucially to the mathematization of economics and the theory of index numbers (see Backhouse 2002b).

William Stanley Jevons (1835–82), English economist and logician. In his book *The Theory of Political Economy* (1871), he used calculus to formulate the marginal utility theory of value, and the notion of partial equilibrium, which provided the foundation for the marginalist revolution in economics (see Backhouse 2002a).

Léon Walras (1834–1910) was a French mathematical economist, one of the protagonists in the marginalist revolution and the innovator of general equilibrium theory. His perspective on the use of mathematics in economics was greatly influenced by *Augustin Cournot* (1801–77), a French philosopher, mathematician and economist. Cournot is credited with the notion of functional relationships among economic variables, which led him to the supply and demand curves (see Backhouse 2002a).

In the United States the process of mathematization began somewhat later with *Irving Fisher* (1867–1947), who followed in the footsteps of Walras, Jevons and Edgeworth in introducing mathematics into economics and making significant contributions to the theory of index numbers (see Backhouse 2002b).

These early pioneers in the mathematization of economics shared a vision of using statistics to provide pertinent empirical foundations to economics (see Moore 1908). Fisher described this goal as a life-long ambition:

> I have valued statistics as an instrument to help fulfill one of the great ambitions of my life, namely, to do what I could toward making economics into a genuine science. (Fisher 1947, p. 74)

The same vision was clearly articulated much earlier by Jevons:

> The deductive science of Economics must be verified and rendered useful by the purely empirical science of statistics. (Jevons 1871, p. 12)

Indeed, Neville Keynes attributed to *statistics* a much greater role in the quantification of economics than hitherto:

> The functions of statistics in economic enquiries are:. . . descriptive,. . . to suggest empirical laws, which may or may not be capable of subsequent deductive explanation,. . . to supplement deductive reasoning by checking its results, and submitting them to the test of experience,. . . the elucidation and interpretation of particular concrete phenomena,. . . enabling the deductive economist to test and, where necessary, modify his premisses,. . . measure the force exerted by disturbing agencies. (Keynes 1890, pp. 342–346)

At the dawn of 20th century, pioneers such as Moore (1908, 1911), who aspired to help in securing empirical foundations for economics, had several advantages – for example, the institutionalization of the collection and compilation of economic data via the establishment of government statistical offices, the systematic development of index numbers, and so on. The mathematization of economics provided them with economic models amenable to empirical enquiry (see Backhouse 2002a, b). In addition, at the end of the 19th century there were several developments in statistical methods, including least-square curve-fitting, regression, correlation, periodogram analysis and trend modelling that seemed tailor-made for analysing economic data (see Mills 1924; Stigler 1954; Heckman 1992; Hendry and Morgan 1995).

The 20th Century: A Strained Relationship

To enliven the discussion of the tension created in the 1920s between economic statistics and statistical inference, the account below refers to the confrontation between the two protagonists who represented the different perspectives, Bowley and Fisher.

Economic Statistics as Against Statistical Inference

The early 20th century statistics scene was dominated by Karl Pearson (1857–1936) and his research in biology at the Galton Laboratory established in 1904. Pearson's research at this laboratory consolidated the biometrics tradition, whose primary outlet was the in-house journal *Biometrika*. Pearson established the department of 'Applied Statistics' at University College in 1911, which, at the time, was the only place one could study for a degree in statistics (see Walker 1958).

Arthur Bowley (1869–1957) was a typical successful 'economic statistician' of the early 20th century who authored one of the earliest textbooks in statistics, *Elements of Statistics* (1901), while a part-time lecturer at the London School of Economics. Bowley understood statistics as comprising two different but interrelated components, the *arithmetic* and the *mathematical*. The former was concerned with statistical techniques as they relate to measurement, compilation, interpolation, tabulation and plotting of data, as well as the construction of index numbers; this constitutes *Part I – General Elementary Methods*, and comprises the first 258 pages of Bowley (1902). The mathematical dimension (*Part II – The Application of the Theory of Probability to Statistics*, the last 74 pages of Bowley 1902) was concerned with the use of probability theory in minimizing and evaluating the errors associated with particular inferences. The last 12 pages of Bowley (1902) are devoted to a discussion of 'regression and correlation' as expounded by Pearson (1896) and Yule (1897).

Bowley (1906) illustrated what he meant by 'errors' using the 'probable error' for the arithmetic average $\bar{x}_n := \frac{1}{n}\sum_{k=1}^{n} x_k$ of the data $(x_1, x_2, \ldots, x_n)$ as:

$$\bar{x} \pm SD(\bar{x}) \tag{2}$$

with $SD(\bar{x})$ denoting the standard deviation of $\bar{x}$. Taking the Normal distribution as an example he argued that the claim in (2) can be interpreted as saying that 'the chance that a given observation should be within this distance of the true average is 2:1' (1906, p. 549). This interpretation can be best understood as based on a Bayesian credible interval evaluation, instead of a frequentist confidence interval developed in the 1930s.

From this perspective, Bowley interpreted the work of Pearson and Edgeworth as concerned with providing different ways to evaluate these 'probable errors' (for example, $SD(\bar{x})$) using either a fitted frequency curve or an asymptotic approximation, respectively (see Bowley 1906, p. 550). It is interesting to note that in the 5th edition of Bowley's statistics book, published in 1926, *Part II* increased threefold to 210 pages, but contains no reference to Fisher's work, which, at the time, was well on its way to transform Karl Pearson's descriptive statistics into modern statistical inference.

Ronald Fisher (1893–1962) pioneered a recasting of statistics (1915, 1921, 1922), moving away from the Edgeworth–Pearson reliance on large sample approximations based on inverse probability (Bayesian) methods, and focusing on finite sample frequentist inference relying on sampling distributions. This recasting was initially inspired by Gossett's (1908) derivation of the student's t distribution for a given sample size n. Fisher made this recasting explicit in his 1921 paper by severely criticizing the inverse probability (Bayesian) approach and articulating a more complete picture of his approach to statistical inference in his 1922 classic paper.

In the early 1920s Bowley was a professor of statistics (second in fame only to Karl Pearson) at the London School of Economics (LSE), known primarily for his contributions in the area of survey sampling, and Fisher was a young statistician at Rothamstead Agricultural Station trying to make sense of a 200-year accumulation of experimental data. Bowley was aware of Fisher's early work: we know that in 1924 Bowley requested and promptly received Fisher's offprints for the LSE library (see Box 1978, p. 171). Moreover, by some accident of faith, the two were neighbours at Harpenden, interacting socially as bridge companions (see Box 1978, p. 85). Indeed, Bowley encouraged Fisher to publish his correction of Pearson's (1900) evaluation of degrees of freedom associated with his goodness-of-fit test (see Fisher 1922a).

The next academic encounter between the professor and the young aspiring statistician was in 1929 when Fisher applied for an academic position in Social Biology at the LSE, but was turned down in favour of Lancelot Hogben (see Box 1978, p. 202). Fisher's first academic position was at University College as Professor of Eugenics, in 1933. The tension between their different perspectives on statistics became public in their first showdown at Fisher's presentation to the Royal Statistical Society in 18 December 1934 entitled 'The Logic of Inductive Inference', where he attempted to explain his published work on recasting the problem of statistical induction since his 1922 paper. Bowley was appointed to move the traditional vote of thanks and open the discussion, and after some begrudging thanks for Fisher's 'contributions to statistics in general' – by then Fisher's 1925 book had made him famous – he went on to disparage his new approach to statistical inference based on the likelihood function by describing it as abstruse, arbitrary and misleading. His comments were predominantly sarcastic and discourteous, and went as far as to accuse Fisher of giving insufficient credit to Edgeworth (see Fisher 1935, pp. 55–57). The litany of churlish comments and currish remarks continued with the rest of the old guard: Isserlis, Irwin and the philosopher Wolf (1935, pp. 57–64), who was brought in by Bowley to undermine Fisher's philosophical discussion on induction. Jeffreys complained about Fisher's criticisms of the Bayesian approach (1935, pp. 70–72). To Fisher's support came Egon Pearson, Neyman and, to a lesser extent, Bartlett. Pearson (1935, pp. 64–65) argued that:

> When these ideas [on statistical induction] were fully understood… it would be realized that statistical science owed a very great deal to the stimulus Professor Fisher had provided in many directions. (Pearson 1935, pp. 64–65)

Neyman was equally supportive, praising Fisher's path-breaking contributions, and explaining Bowley's reaction to Fisher's critical review of the traditional view of statistics as understandable attachment to old ideas (1935, p. 73).

Fisher, in his reply to Bowley and the old guard, was equally contemptuous:

> The acerbity, to use no stronger term, with which the customary vote of thanks has been moved and seconded… does not, I confess, surprise me. From the fact that thirteen years have elapsed between the publication, by the Royal Society, of my first rough outline of the developments, which are the subject of to-day's discussion, and the occurrence of that discussion itself, it is a fair inference that some at least of the Society's authorities on matters theoretical viewed these developments with disfavour, and admitted with reluctance.... However true it may be that Professor Bowley is left very much where he was, the quotations show at least that Dr. Neyman and myself have not been left in his company.... For the rest, I find that Professor Bowley is offended with me for 'introducing misleading ideas'. He does not, however, find it necessary to demonstrate that any such idea is, in fact, misleading. It must be inferred that my real crime, in the eyes of his academic eminence, must be that of 'introducing ideas'. (Fisher 1935, pp. 76–82)

Fisher's reference to 'his academic eminence', although containing a dose of sarcasm, it was not totally out of place. Bowley became a member of the Council of the Royal Statistical Society as early as 1898, served as its Vice-President in 1907–8 and again in 1912–14, and President in 1938–40. He was awarded the society's highest honour, the Guy Medal in gold, in 1935; he received the Guy in silver as early as 1895. In contrast, Fisher had no academic position until 1933, and even that came with the humiliating stipulation that he would *not* teach statistics from his new position as Professor of Eugenics at University College (see Box 1978, p. 258).

Fisher made it clear that he associated the 'old guard' in statistics with Bowley-type economic statistics:

> Statistical methods are essential to social studies, and it is principally by the aid of such methods that these studies may be raised to the rank of sciences. This particular dependence of social studies upon statistical methods has led to the unfortunate misapprehension that statistics is to be regarded as a branch of economics, whereas in truth methods adequate to the treatment of economic data, in so far as these exist, have mostly been developed in the study of biology and the other sciences. (Fisher 1925, p. 2)

The unbridgeable gap between Bowley and the 'old guard' on one side, and Fisher, Neyman and Pearson on the other, was apparent 6 months earlier when Bowley was assigned the same role for Neyman's first presentation. Despite the fact that Neyman began his presentation by praising

Bowley for his earlier contributions to survey sampling methods, he grouped him with Fisher and accused him of the same abstruseness:

> I am not certain whether to ask for an explanation or to cast a doubt. It is suggested in the paper that the work is difficult to follow and I may be one of those who have been misled by it. I can only say I have read it at the time it appeared and since, and I have read Dr Neyman's elucidation of it yesterday with great care. I am referring to Dr Neyman's confidence limits. I am not at all sure that the 'confidence' is not a 'confidence trick'. (Neyman 1934, pp. 608–609)

His 'confidence trick' remark is not very surprising in view of Bowley's own interpretation of (2) in inverse probabilistic (Bayesian) terms. Predictably, Egon Pearson and Fisher came to Neyman's rescue from the rebukes of old guard.

Retrospectively, Bowley's charge of abstruseness, levelled at both Fisher and Neyman, might best be explained in terms of David Hume's (1711–76) 'tongue in cheek' comment two centuries earlier:

> The greater part of *mankind* may be divided into two classes; that of *shallow* thinkers, who fall short of the truth; and that of *abstruse* thinkers, who go beyond it. The latter class are by far the most rare; and I may add, by far the most useful and valuable. They suggest hints, at least, and start difficulties, which they want, perhaps, skill to pursue; but which may produce fine discoveries, when handled by men who have a more just way of thinking.... All people of *shallow* thought are apt to decry even those of *solid* understanding, as *abstruse* thinkers, and methaphysicians, and refiners; and never will allow any thing to be just which is beyond their own weak conceptions. (Hume 1987, pp. 253–254)

In summary, the pioneering work of Fisher, Egon Pearson and Neyman, was largely ignored by the Royal Statistical Society (RSS) establishment until the early 1930s. By 1933 it was difficult to ignore their contributions, published primarily in other journals, and the 'establishment' of the RSS decided to display its tolerance to their work by creating 'the Industrial and Agricultural Research Section', under the auspices of which both papers by Neyman and Fisher were presented in 1934 and 1935 respectively. In their centennial volume published in 1934, the RSS acknowledged the development of 'mathematical statistics', referring to Galton, Edgeworth, Karl Pearson, Yule and Bowley as the main pioneers, and listed the most important contributions in this sub-field which appeared in its *Journal* during the period 1909–33, but the three important papers by Fisher (1922a, b, 1924) are conspicuously absent from that list. The list itself is dominated by contributions in vital, commercial, financial and labour statistics (see RSS 1934, pp. 208–223). There is only one reference to Egon Pearson, for his 1933 paper 'Control and Standardization of Quality of Manufactured Products' – the very paper used as self-justification by the RSS in creating the new section. It is interesting to note that by the late 1920s the revolutionary nature of Fisher's new approach to statistics was clearly recognized by many. Tippet (1931) was one of the earliest textbook attempts to blend the earlier results on regression and correlation within Fisher's new approach. In the United States, Hotelling (1930) articulated a most elucidating perspective on Fisher's approach.

The Fisher–Neyman–Pearson Approach

The main methods of the Fisher–Neyman–Pearson (F–N–P) approach to statistical inference, point estimation, hypothesis testing and interval estimation, were in place by the late 1930s. The first complete textbook discussion of this approach, properly integrated with its probabilistic underpinnings, was given by Cramer (1946). The methodological discussions concerning the form of inductive reasoning underlying the new frequentist approach, however, were to linger on until the 1960s and beyond; see the exchange between Fisher (1955), Pearson (1955) and Neyman (1956).

One of the most crucial insights of the F–N–P approach to statistical inference, which set it apart from previous approaches to statistics, was the explicit specification of the premises of statistical induction in terms of the notion of a *statistical model*:

> The postulate of randomness thus resolves itself into the question, 'Of what population is this a random sample?' which must frequently be asked by every practical statistician. (Fisher 1922, p. 313.

He defined the initial choice of the statistical model in the context of which the data will be interpreted as a 'representative sample' as the problem of *specification*, emphasizing the fact that: 'the adequacy of our choice may be tested *posteriori*' (1922, p. 314). Indeed, the first three tests discussed in Fisher (1925, pp. 78–94) are misspecification (M-S) tests for the Normality, Independence and Identically Distributed assumptions. Fisher (1922, 1925, 1935), and later Neyman (1938/1952, 1950), emphasized the importance of both model specification and validation vis-à-vis the data:

> Guessing and then verifying the 'chance mechanism', the repeated operations of which produces the observed frequencies. (Neyman 1977, p. 99)

Pearson (1931a, b) was among the first to discuss the implications of non-Normality as well as develop M-S tests for it; see Lehmann (1999) for the early concern about the consequences of misspecification in the 1920s.

The F–N–P discernments concerning statistical model *specification, M-S testing*, and *respecification* can be summarized in the form of what might be called the F–N–P *perspective* (articulated in Spanos 2006a) which can be summarized as follows:

1. Every statistical (inductive) inference is based on certain *premises*, in the form of (*a*) a *statistical model M* parameterizing the probabilistic structure of an observable stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$, and (*b*) a set of data $\mathbf{Z}: = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$, viewed as a 'typical realization' of this process.
2. A statistical model is specified in terms of a complete and internally consistent set of probabilistic assumptions concerning the underlying stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$. For example, the Normal/linear regression model is specified in terms of assumptions [1]–[5] (Table 1) concerning the observable process $\{(y_t \,|\mathbf{X}_t = \mathbf{x}_t); t \in \mathbb{N}\}$, and not the errors.
3. *Statistical adequacy.* Securing the validity of assumptions [1]–[5] vis-à-vis the data in question is necessary for establishing 'statistical regularities' and ensuring the reliability of inference (see Spanos 2006a, b, c).

The importance of the F–N–P perspective stems from the fact that the *statistical model* enables one:

(i) to assess the validity (statistical adequacy) of the premises for inductive inference – by testing the assumptions using misspecification tests; and
(ii) to provide relevant error probabilities for appraising the reliability of the associated inference (see Spanos 2006a).

It is well known that the *reliability* of any inference procedure depends crucially on the validity of the pre-specified *statistical model* vis-à-vis the data in question. The optimality of these procedures is defined by their capacity to give rise to valid inferences (*trustworthiness*), which is calibrated in terms of the associated error probabilities – how often these procedures lead to erroneous inferences (see Mayo 1996). In the case of confidence interval estimation the calibration is usually gauged in terms of minimizing *the coverage error probability*: the probability that the interval does *not* contain the true value of the unknown parameter(s). In the case of hypothesis testing the calibration is ascertained in terms of minimizing *the type II error probability* – the probability of accepting the null hypothesis when false, for a given *type I error probability* (see Cox and Hinkley 1974). It is also known, but often insufficiently appreciated, that when any of the model assumptions are invalid, the *reliability of inference* is called into question (see Pearson 1931a; Bartlett 1935, for early discussions). Departures from the model assumptions will give rise to a discrepancy between the *nominal* error probabilities (valid premises), and the *actual* error probabilities (misspecified premises), giving rise to unreliable inferences (see Spanos and McGuirk 2001; Spanos 2005).

Although the nature of the F–N–P statistical induction became clear by the late 1930s, the form of the underlying *inductive reasoning* was

Statistics and Economics, Table 1 The normal/linear regression model

Statistical GM:	$y_t = \beta_0 + \boldsymbol{\beta}_1^T \mathbf{x}_t + u_t,\ t \in \mathbb{N}$,
[1] Normality:	$(yt\ \vert \mathbf{X}\mathrm{t} = \mathbf{x}_t) \sim \mathrm{N}(.,.)$,
[2] Linearity:	$E(y_t \vert \mathbf{X}_t = \mathbf{x}_t) = \beta_0 + \boldsymbol{\beta}_1^T \mathbf{x}_t$, linear in $\mathbf{x}_{t,}$
[3] Homoskedasticity:	$Var(y_t \vert \mathbf{X}_t = \mathbf{x}_t) = \sigma^2$, free of $\mathrm{x}_{t,}$
[4] Independence:	$(y_t \vert \mathbf{X}_t = \mathbf{x}_t),\ t \in \mathbb{N}$ an independent process
[5] t-invariance:	$\boldsymbol{\theta} : (\beta_0, \boldsymbol{\beta}_1, \sigma^2)$ do not change with t:

clouded by a disagreement between the two protagonists (see Mayo 2005). Fisher argued for 'inductive inference' spearheaded by his significance testing (see Fisher 1955, 1956), and Neyman argued for 'inductive behaviour' based on Neyman–Pearson testing (see Neyman 1956; Lehmann 1993; Cox 2006). Neither account, however, gave satisfactory answers to the question 'when do data $\mathbf{Z}$ provide *evidence* for (or against) a hypothesis or a claim H?' The *pre-data* error–probabilistic account of inference seemed inadequate for a *post-data* evaluation of the inference reached to provide a clear evidential interpretation of the results (see Hacking 1965).

The F–N–P paradigm, in addition to (*a*) the pre-data as against post-data error probabilities, still grapples with some additional philosophical/methodological issues including (*b*) the fallacies of acceptance and rejection (for example statistical as against substantive significance), (*c*) double use of data, (*d*) statistical model selection (specification) as against model validation, (*e*) structural as against statistical models. These and other methodological issues have been extensively debated in other social sciences such as psychology and sociology (see Morrison and Henkel 1970; Lieberman 1971; Godambe and Sprott 1971), but largely ignored in economics until recently.

Mayo (1996) argued convincingly that some of these chronic methodological issues and problems can be addressed by supplementing the Neyman–Pearson approach to testing (see Pearson 1966) with a post-data assessment of inference based on *severe testing reasoning*. This extended frequentist approach to inference, called *the error-statistical approach*, has been used by Mayo (1991) to address (*c*), by Mayo and Spanos (2006) to address the fallacies of acceptance and rejection, and by Spanos (2006b, 2007) to deal with the issues (*d*) and (*e*), respectively.

Economic Statistics in the Early 20th Century

In the 1930s applied economists were more keyed to Bowley's traditional view of economic statistics than to F–N–P statistical inference perspective. Indeed, Bowley was elected president (the first from Britain) of the Econometric Society for 1938–9. The more economics-oriented 'statistics textbooks' written in the 1920s and 1930s, including Bowley (1920/1926/1937), Mills (1924/1938), Ezekiel (1930), Davis and Nelson (1935) and Secrist (1930), largely ignored the new statistical inference paradigm. Their perspective was primarily one of 'descriptive statistics', supplemented with the Pearson–Yule curve-fitting perspective on correlation and regression, and certain additional focus on the analysis of time series data, including index numbers (see Persons 1925).

Economic statistics, as exemplified in Mills (1924), provided the framework for the work at the National Bureau of Economic Research (NBER), of which Mills was a staff member. The empirical work on business cycles by Burns and Mitchell (1946) represents an excellent use of descriptive statistics in conjunction with graphical methods, as understood at the time. Their detailed, carefully crafted and painstaking statistical analysis of business cycles, however, suffers from the same crucial weakness as all descriptive statistics: the premises for inductive inference (the underlying statistical model) is not explicitly specified, and as a result one cannot assess the reliability of inferences based on such statistics. For instance, without clearly specified probabilistic premises one can easily misidentify temporal dependence type cycles with regular business cycles (see Spanos 1999).

The conventional wisdom at the time is summarized by Mills (1924) in the form of a distinction between 'statistical description vs. statistical induction'. In statistical description measures such as the sample mean $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ the sample variance $s_x^2 = \frac{1}{n-1}\sum_{k=1}^{n} (x_k - \bar{x}_n)^2$, the correlation

$$r = \frac{\sum_{k=1}^{n} (x_k - \bar{x}_n)(y_k - \bar{y}_n)}{\sqrt{\left[\sum_{k=1}^{n} (x_k - \bar{x}_n)^2\right]\left[\sum_{k=1}^{n} (y_k - \bar{y}_n)^2\right]}},$$

and so on, 'provide just a summary for the data in hand' and 'may be used to perfect confidence, as accurate descriptions of the given characteristics' (1924, p. 549). However, when the results are to be extended *beyond* the data in hand – statistical induction – their validity depends on certain inherent a priori assumptions such as (*a*) the 'uniformity' for the *population* and (*b*) the 'representativeness' of the *sample* (1924, pp. 550–552).

A methodological digression. Unfortunately, Mills's misleading argument concerning descriptive statistics lingers on even today. The reality is that there are *appropriate* and *inappropriate* summaries of the data, which depend on the inherent probabilistic structure of the data. For instance, if data $\{(x_k, y_k), k = 1..., n\}$ are *trending*, like most economic time series, the summary statistics $\left(\bar{x}, s_x^2, \bar{y}, s_y^2, r\right)$ represent artefacts – highly misleading descriptions of the features of the data in hand. When viewed in the context of a probabilistic framework, $\left(\bar{x}, s_x^2, \bar{y}, s_y^2, r\right)$ are unreliable estimators of $E(X_k)$ $E(Y_k)$, $Var(X_k)$, $Var(Y_k)$, $Corr(X_k, Y_k)$; they provide reliable and precise estimates only when certain probabilistic assumptions concerning the underlying the vector process $\{(X_k ; Y_k)\ k \in \mathbb{N}\}$ such as independent and identically distributed (IID), are valid for the data in hand. Any departures from these premises require one to qualify the reliability and precision of these estimates. In an important sense one of Fisher's lasting contribution to statistics was to (*a*) make the IID assumptions explicit as part of the problem of specification, by formalizing Mills's a priori 'uniformity' and 'representativeness' assumptions, and (*b*) render them empirically testable. It is important to note that ignoring statistical adequacy is a very different criticism of Burns and Mitchell than that of Koopmans (1947); see below.

The paper by Yule (1926), entitled 'Why Do We Sometimes get Nonsense Correlations between time series'?, provided a widely discussed wakeup call in economics, because it raised serious doubts about the appropriateness of the linear regression model when the data $\{(x_k, y_k), k = 1,...,n\}$ constitute time series, by pointing out the risk of getting *spurious* results. As commented in Spanos (1989b), the source of the spurious (nonsense) correlation problem is *statistical inadequacy* (see section "Unit Roots and Cointegration" below). Yule's (1927) autoregressive (AR(p)) and Slutsky's (1927) moving average (MA(q)) models can be viewed as attempts to specify statistical models to capture the temporal dependence in time series data.

Stochastic processes. The AR(p) and MA (q) models were given proper probabilistic underpinnings by Wold (1938) using the newly developed theory of stochastic processes by Kolmogorov and Khitchin in the early 1930s (see Doob 1953). This was a crucial and timely development in probability theory which extended significantly the intended scope of the F–N–P approach beyond the original IID frame-up, by introducing several dependence and heterogeneity concepts, such as Markov dependence, stationarity and ergodicity (see Spanos 1999, ch. 8).

The Econometric Society and the Cowles Commission

The vision statement of the Econometric Society founded in 1930 read:

> Its main object shall be to promote studies that aim at a unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems. (Frisch 1933, p. 106.

The impression among quantitatively oriented economists in the early 1930s was that the F–N–P sampling theory methods were inextricably bound up with agricultural experimentation. It was generally believed that these methods are relevant only for analysing 'random samples' of experimental data, as Frisch argued:

> In problems of the kind encountered when the data are the result of *experiments* which the investigator can control, the sampling theory may render very valuable services. Witness the eminent works of R.A. Fisher and Wishart on problems of agricultural experimentation. (Frisch 1934, p. 6)

In place of the statisticians' linear regression Frisch proposed his errors-in-variables scheme, which treated all observable variables symmetrically by decomposing them into a latent systematic (deterministic) component and a white-noise error with economic theory providing relationships among the systematic components. Fisher's reaction to Frisch's scheme was that economists were perpetuating a major *confusion* between 'statistical' regression coefficients and 'coefficients in abstract economic laws' (see Bennett 1990, p. 305).

Tinbergen's (1939) empirical modelling efforts were in the spirit of the Pearson–Yule curve-fitting tradition, which paid little attention to the validity of the premises of inference. In reviewing this work Keynes and Tinbergen (1939) destructively criticized the use of regression in econometrics and raised numerous substantive and statistical problems, but not the reliability of inference problem (see Spanos 2006a).

The first attempt to bring together Frisch's errors-in-variables scheme with Fisher's linear regression model was made by Koopmans (1939), which had no success. Koopmans' primary influence on econometrics was as a leading figure in the Cowles Commission in Chicago in the 1940s (see Heckman 1992).

The first successful attempt to bring the F–N–P methods into econometrics modelling was made by Haavelmo (1944), who argued convincingly against the prevailing view that sampling methods are only applicable to random samples of experimental data (see Spanos 1989a). Contrary to this view, the F–N–P perspective provides the proper framework for modelling time series data which exhibit both dependence and heterogeneity:

> For no tool developed in the theory of statistics has any meaning... without being referred to some stochastic scheme. (Haavelmo 1944, p. iii)
>
> ... economists might get more useful and reliable information (and also fewer spurious results) out of their data by adopting more clearly formulated probabilistic models. (1944, p. 114)

The part of Haavelmo's monograph that had the biggest impact on the development of econometrics was, however, the technical 'solution' to the simultaneity problem that was formalized and extended by the Cowles Commission in the form of the simultaneous equations model (SEM): see Koopmans 1950.

Despite the introduction of frequentist methods of inference by the Cowles Commission, the theory-driven specification of the *structural model*:

$$\Gamma^T \mathbf{y}_k = \Delta^T \mathbf{x}_k + \varepsilon_k, \quad \varepsilon_k \sim N(0, \Omega), \quad k \in \mathbb{N} \quad (3)$$

(Using the traditional notation, see Spanos 1986), leaves any inferences concerning the structural parameters (Γ, Δ, Ω) highly susceptible to the unreliability of inference problem.

Methodological digression. The unreliability of inference arises primarily because it is often insufficiently appreciated that the statistical reliability of such inference depends crucially on the *statistical adequacy* of the (implicit) *reduced form model*:

$$\mathbf{y}_k = \mathrm{B}^T \mathbf{x}_k + \mathbf{u}_k, \mathbf{u}_k \sim N(0; \Sigma), \quad k \in \mathbb{N}. \quad (4)$$

That is, unless (4), viewed as multivariate linear regression model (assumptions [1]–[5] in Table 1 in vector form), is statistically adequate ([1]–[5] are valid for the data in question), any inference based on (3) is likely to be unreliable. Note that *identification* refers to being able to define the structural parameters (Γ, Δ, Ω) uniquely in terms of the statistical parameters $(\mathrm{B}; \Sigma)$. In practice (4) is not even estimated explicitly, let alone have its assumptions [1]–[5] tested thoroughly before drawing any inferences concerning (Γ, Δ, Ω) (see Spanos 1986, 1990). A more expedient way one that highlights the reliability issue, is to view (3) as a structural model which is embedded into the statistical model (4), giving rise to a special type of substantive information restrictions. Hence, the theory-dominated perspective of the Cowles Commission, despite the importance of the technical innovations introduced in dealing with simultaneity, has (inadvertently) undermined the problem of statistical adequacy in empirical modelling (see Spanos 2006a). As argued by Heckman:

> The Haavelmo–Cowles way of doing business – to postulate a class of models *in advance of looking at the data* and to consider identification problems within the prescribed class – denies one commonly used process of inductive inference that leads to empirical discovery.... The Haavelmo program as interpreted by the Cowles Commission scholars refocused econometrics away from the act of empirical discovery and toward a sterile program of hypothesis testing and rigid imposition of a priori theory onto the data. (Heckman 1992, pp. 883–884)

Koopmans (1947), in reviewing Burns and Mitchell (1946), criticized their focusing on the purely empirical nature of their results without any guidance from economic theory. He pronounced their empirical findings as representing the 'Kepler stage' of data analysis, in contrast to the 'Newton stage', where the original empirical regularities were given a structural (theoretical) interpretation using the law of universal gravitation (LUG). What Koopmans (1947) neglected to point out is that it was not the theory that guided Kepler to the regularities, but the statistical regularities exhibited by the data. Indeed, Kepler established these regularities 60 years before Newton was inspired by them to come up with his LUG. The Cowles Commission approach, which Koopmans misleadingly associates with the Newton stage, was equally (if not more) vulnerable to the reliability of inference problem. There is no reason to believe that the reduced form (4) implied by the structural form (3), which was specified in complete ignorance of the probabilistic structure of the data, will constitute a statistically adequate model. The specification of statistical models relying exclusively on substantive information is not conducive to reliable/precise inferences. The crucial difference between Kepler's empirical results and those in Burns and Mitchell (1946) and Klein (1950) – based largely on Koopmans's preferred approach – is that Kepler's constitute real statistical regularities in the sense that his estimated model of elliptical motion, viewed retrospectively in the context of the linear regression model (Table 1), is statistically adequate; assumptions [1]–[5] are valid for his original data (see Spanos 2008).

Textbook Econometrics. The Gauss–Markov Perspective

The textbook approach to econometrics was largely shaped in the early 1960s by two very successful textbooks by Johnston (1963) and Goldberger (1964) by viewing the SEM as an extension/modification of the classical linear model. These textbooks demarcated the intended scope of econometrics to be the 'quantification oftheoretical relationships', and reverted back to the 'curve-fitting' perspective of the Legendre–Gauss 19th century tradition, instead of adopting the F–N–P perspective (see Spanos 1995, 2007).

The cornerstone of textbook econometrics is the so-called *Gauss–Markov theorem*, which is based on the linear model:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, E(\varepsilon) = 0, E\left(\varepsilon\varepsilon^T\right) = \sigma^2\mathbf{I}n, \qquad (5)$$

where $\mathbf{I}_n$ is the identity matrix. In the context of (5), Gauss in 1823 (see Hald 2007) proved that the least squares estimator $\widehat{\beta}_{LS} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ has minimum variance within the class of *linear* and *unbiased* estimators of β. For the sake of historical accuracy it is important to point out that Markov had nothing to do with this theorem (see Neyman 1952, p. 228). This theorem, and the perspective it exemplifies, provide the central axis around which textbook econometrics revolves (see Greene 2003).

A methodological digression. Spanos (1986) challenged the traditional interpretation that the Gauss–Markov theorem provides a formal justification for least squares via the optimality of the estimators it gives rise to, arguing that the results of this theorem provide a poor basis for *reliable* and *precise* inference. This is primarily because the Gauss–Markov theorem yields the mean and variance of $\widehat{\beta}_{LS}$ but *not* its sampling distribution, that is $\widehat{\beta}_{LS} \sim D^{?}\left(\beta, \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)$. Hence, even the simplest forms of inference, like testing

$H_0 = \beta = \mathbf{0}$ would require one to use either inequalities like Chebyshev's to approximate the relevant error probabilities (Spanos 1999, pp. 550–552), or invoke asymptotic approximations; neither method would, in general, give rise to reliable and precise inferences (Spanos 2006a, pp. 46–47).

The Gauss–Markov 'curve-fitting' perspective promotes 'saving the theory' by attributing the stochastic structure to the error term and favouring broad premises (weak assumptions) in an attempt to protect the inference from the perils of misspecification. This move, however, relegates the essentialness of ensuring the reliability and precision of inference. Weak assumptions, such as the Gauss–Markov assumptions in (5), do not guarantee reliable inferences, but they usually give rise to much less precise inferences than specific premises comprising assumptions such as [1]–[5] (Table 1): Spanos 2006a. As perceptively noted by Heckman:

> In many influential circles, ambiguity disguised as simplicity or 'robustness' is a virtue. The less said about what is implicitly assumed about a statistical model generating data, the less many economists seem to think is being assumed. The new credo is to let sleeping dogs lie. (Heckman 1992, p. 882.

In addition, the 'error-fixing' strategies of the textbook approach, designed to deal with departures from the linearity, homoskedasticity, no-autocorrelation assumptions, do not usually address the reliability of inference problem (Spanos and McGuirk 2001).

Some of the important technical developments in both econometrics and statistics since the 1980s, such as the *generalized method of moments* (see Hansen 1982), as well as certain *nonparametric* (see Pagan and Ullah 1999) and *semiparametric* methods (see Horowitz 1998), are motivated by this Gauss–Markov perspective. These methods, although very useful for a number of different aspects of empirical modelling, do not provide the answer to statistical misspecification, and often compromise the reliability/precision of substantive inferences (see Spanos 1999, pp. 553–555).

Demarcating the Boundaries of Modern Statistics

As argued above, the F–N–P perspective has been largely ignored in empirical modelling in economics, despite the wholesale adoption of Fisher's estimation and the Neyman–Pearson testing methods. One of the primary obstacles has been the problem of blending the substantive subject matter and statistical information and their roles in empirical modelling. Many aspects of empirical modelling, in both the physical and social sciences, implicate both sources of information in a variety of functions, and others involve one or the other, more or less separately. For instance, the development of *structural* (theoretical) *models* is primarily based on *substantive information*; that activity, by its very nature, cannot be separated from the disciplines in question, but where does this leaves statistics? It renders the problem of demarcating its boundaries as a separate discipline extremely difficult (see Lehmann 1990; Cox 1990).

A methodological digression. Spanos (2006c) argued that the lessons learned in blending the substantive and statistical information in econometric modelling can help delineate the boundaries of statistics as a separate discipline. Certain aspects of empirical modelling, which focus on *statistical information* and are concerned with the nature and use of statistical models, can form a body of knowledge that is shared by all applied fields. *Statistical model specification*, *the use of graphical techniques* (going back to Playfair), *misspecification* (*M-S*) *testing and respecification*, together with the relevant inference procedures, constitute aspects of statistical modelling that can be developed *generically* without requiring any information concerning 'what substantive variables the data $\mathbf{Z}$ quantify or represent'. All these aspects of empirical modelling belong to the *realm of statistics* and can be developed generically without any reference to substantive subject matter information. This, in a sense, will broaden the scope of modern statistics because the current literature and textbooks pay little attention to some of these aspects of modelling (see Cox and Hinkley 1974).

The statistical and substantive information can be amalgamated, without compromising their integrity, by embedding structural models into adequate statistical models, which would provide the premises for statistical inference. That is, the substantive restrictions need to be thoroughly tested and accepted in the context of the statistical model in order for the resulting empirical model to enjoy both statistical and substantive meaning (see Spanos 2006b, 2007).

The Box–Jenkins Turn in Statistics

An important development in statistics that had a lasting effect on econometrics and created a tension with textbook econometrics, was the publication of Box and Jenkins (1970). Building on the work of Wold (1938), they proposed a new statistical perspective on time series modelling which placed it within the F–N–P modelling framework where the premises of inference is specified by a *statistical model*. In addition to transforming descriptive time series analysis into statistical inference proper, the Box–Jenkins approach introduced several noteworthy innovations into empirical modelling that influenced empirical modelling in economics.

(i) Modelling begins with a family of *statistical models* in the form of the ARIMA(p, d, q):

$$y_t^* = \alpha_0 + \sum_{k=1}^{P} \alpha_k y_{t-k}^* + \sum_{l=1}^{q} \beta_l \varepsilon_{t-l} + \varepsilon_t, \ \ \varepsilon_t \sim \text{NIID}\left(0, \sigma^2\right), \ \ t \in \mathbb{N} \tag{6}$$

where $y_t^* := \Delta^d y_t$, that was thought to capture adequately the temporal dependence and heterogeneity (including seasonality) in time series data.

(ii) Statistical modelling was viewed as an *iterative process* that involves several stages, *identification*, *estimation*, *diagnostic checking*, and *prediction*.

(iii) *Diagnostic checks*, based on the residuals from the fitted model, offered a way to detect model inadequacies with a view to improve the original model.

(iv) *Exploratory data analysis* (EDA) was legitimized as providing an effective way to select (identify) a model within the ARIMA (p, d, q) family.

(v) The deliberate choice of a more *general specification* in order to put the model 'in jeopardy' (see Box and Jenkins 1970, p. 286) is exploited in assessing the adequacy of a selected model.

The Box–Jenkins approach constituted a major departure from the rigid textbook approach, where the model is assumed to be specified by economic theory in advance of any data. Indeed, the predictive success of the ARIMA(p, d, q) models in the 1970s exposed the statistical inadequacy of traditional econometric models, sending the message that econometric models could ignore the temporal dependence and heterogeneity of times series data at their peril (see Granger and Newbold 1986).

The weaknesses of traditional econometric modelling techniques brought out by the Box–Jenkins modelling motivated several criticisms from within econometrics, including those by Hendry (1977) and Sims (1980), that led to the autoregressive distributed lag (ADL(p, q)) and the vector autoregressive (VAR(p)) family of models, respectively. The LSE tradition (see Hendry 1993), embraced and extended the Box–Jenkins innovations (i)–(v), rendering the general-to-specific approach the backbone of its empirical modelling methodology (see Hendry 1995).

Unit Roots and Cointegration

The Box–Jenkins ARIMA(p, d, q) modelling approach raised the question 'how does one decide on the value of $d \geq 0$ in $\Delta^d y_t$, that is appropriate to induce stationarity?' It turned out that the value of d is related to the number of unit roots in the $AR(m)$ representation, $y_t = \gamma_0 + \sum_{k=1}^{m} \gamma_k y_{t-k} + u_t$, of the underlying stochastic process $y_t, t \in \mathbb{N}$. Efforts to answer this question led to the unit root 'revolution', initiated by Dickey and Fuller (1979) in the statistics literature. This had an immediate impact on the econometrics literature, which generalized and extended the initial results in a number of different directions (see Phillips and Durlauf 1986; Phillips 1987). This literature eventually led to further important developments, which brought out a special relationship (cointegration) among unit root processes and error-correction models (see Engle and Granger 1987; Johansen 1991; Hendry 1995).

A methodological digression. The (non-standard) sampling distribution results associated with unit roots were used by Phillips (1986) to shed light on the chronic problem of spurious regression raised by Yule (1926). This problem was revisited by Granger and Newbold (1974) using simulations of time series data $\{(x_t, y_t),$ $t = 1,\ldots,n\}$ generated by two *uncorrelated* Normal unit root processes:

$$y_t = y_{t-1} + \varepsilon_{1t},\ \ x_t = x_{t-1} + \varepsilon_{2t}, E(\varepsilon_{1t}) = 0,\ \ E(\varepsilon_{2t}) = 0,\ E\left(\varepsilon_{1t}^2\right) = \sigma_{11},\ \ E\left(\varepsilon_{2t}^2\right) = \sigma_{22},\ E(\varepsilon_{1t}\varepsilon_{2t}) = 0.$$

Their results demonstrated that when these data were used to estimate the linear regression model, $y_t = \beta_0 + \beta_1 x_t + u_t$, the inferences based on the estimated model were completely unreliable. In particular, they noted a huge discrepancy between the *nominal* ($\alpha = .05$) and *actual* ($\widehat{\alpha} = .76$) error probabilities when testing the hypothesis $\beta_1 = 0$.

In a very influential paper, Phillips (1986) explained this by deriving analytically the (non-standard) sampling distributions of the least-squares estimators $(\widehat{\beta}_0, \widehat{\beta}_1)$ under the above unit root scheme, showing how different they were from the assumed distributions. What was not sufficiently appreciated was that the discrepancy between the nominal and actual error probabilities is a classic symptom of unreliable inferences emanating from a statistically misspecified model, that is *misspecification*, due to ignoring the temporal dependence/heterogeneity in the data, is the real source of spurious regression. One would encounter similar unreliabilities when the data exhibit deterministic trends or/and Markov dependence, or/and non-Normalities (see Spanos and McGuirk 2001; Spanos 2005). Deriving the sampling distributions under all scenarios of possible misspecifications is impractical (there is an infinity of such scenarios), and does *not* address the unreliability of inference issue. What is needed is to respecify the original model to account for the disregarded information that gave rise to the detected departures. For instance, for the above Granger and Newbold data, if one were to estimate the dynamic linear regression model:

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 y_{t-1} + \varepsilon_t,\ \ t \in \mathbb{N},$$

the above noted unreliabilities would disappear (see Spanos 2001).

Recent Developments in Microeconometrics

Arguably, some of the most important developments in econometrics since 1980 have taken place in an area broadly described as microeconometrics (see Manski and MacFadden 1981; Heckman and Singer 1984; Cameron and Trivedi 2005) for a recent textbook survey. This area includes *discrete and limited dependent* and *duration models* for *cross-section data*, as well as *panel data* models. The roots of these statistical models go back to the statistical literature on the probit/logit and analysis of variance models (see Agresti 2002, ch. 16), but they have been generalized, extended and adapted for economic data.

A welcome facet of microeconometrics is the specification of statistical models that often takes into consideration the probabilistic structure of the data (see Heckman 2001). Unfortunately, this move does not often go far enough in securing statistical adequacy. This becomes apparent when one asks, ‘what are the probabilistic assumptions providing a complete specification for the probit/logit, duration and the fixed and random effect models?’ Without such complete specifications, one would not even know what potential errors to probe for to secure statistical adequacy.

While these developments in microeconometrics are of great importance, their potential value has been offset by the insufficient attention paid to the task of ensuring reliability and precision of inference. Their statistical results are still largely dominated by the Gauss–Markov perspective, in the sense that:

(i) the probabilistic structure of the models in question is specified, almost exclusively, in terms of unobservable *error terms*,
(ii) the error probabilistic assumptions are often vague and incomplete, and invariably involve non-testable orthogonality conditions,
(iii) the statistical analysis focuses primarily on constructing consistent and asymptotically Normal estimators, and
(iv) respecification is often confined to 'error-fixing'.

In view of (i)–(iv), even questions of ensuring statistical adequacy cannot be posed unequivocally for these statistical models.

Spanos (2006a, d) proposed complete specifications for these statistical models in terms of probabilistic assumptions relating to the observable stochastic processes involved, but there is a long way to go to develop adequate misspecification testing and the respecification results needed to ensure the reliability and precision of inference when applying these statistical models to actual data.

Conclusion

The demise of political arithmetic by the end of the 18th century, due to the unreliability of the inferences its methods gave rise to, contains important lessons for both economics and statistics. Petty's attitude of 'seeking figures that will support a conclusion already reached by other means' lingers on in applied econometrics more than three centuries later. The problem then was that, in addition to the quality and the accuracy of data, the probabilistic underpinnings of establishing statistical regularities were completely lacking. Fisher's recasting of statistical induction has changed that, and it is now known that the explicit specification of the *statistical model* enables one to (*a*) assess the validity of the premises for inductive inference, and (*b*) provide relevant error probabilities for assessing the reliability of ensuing inferences. It has taken several decades to understand how one can assess the model assumptions vis-à-vis the observed data using *misspecification tests* (see Spanos 1999), but one hopes it will take less time before modellers understand the necessity to implement such tests with the required care and thoroughness to ensure the *reliability* of the resulting statistical inferences (see Spanos 2006a).

The Box–Jenkins modelling approach exposed the inattention to statistical adequacy in traditional econometric modelling and strengthened the call for adopting the F–N–P perspective. This will bring modern statistical inference closer to econometrics to the benefit of both disciplines. Careful implementation of this perspective will certainly improve the reliability of empirical evidence in economics and other applied disciplines. Moreover, the *ab initio* separation of the statistical and substantive information can help demarcate and extend the intended scope of statistics. The error-statistical extension/modification of frequentist statistics (Mayo 1996) can address some of the inveterate problems concerning inductive reasoning and broaden the intended scope of statistical inference in these disciplines by enabling one to consider questions of substantive adequacy, shedding light on causality issues, omitted variables and confounding effects (see Spanos 2006b).

See Also

▶ Bowley, Arthur Lyon (1869–1957)
▶ Davenant, Charles (1656–1714)
▶ Edgeworth, Francis Ysidro (1845–1926)
▶ Fisher, Ronald Aylmer (1890–1962)
▶ King, Gregory (1648–1712)
▶ Petty, William (1623–1687)

Bibliography

Agresti, A. 2002. *Categorical data analysis*, 2nd ed. New York: Wiley.
Backhouse, R.E. 2002a. *The ordinary business of life: A history of economics from the ancient to the twenty-first century*. Princeton: Princeton University Press.
Backhouse, R.E. 2002b. *The Penguin history of economics*. London: Penguin Books.

Bartlett, M.S. 1935. Some aspects of the time-correlation problem in regard to tests of significance. *Journal of the Royal Statistical Society* 98: 536–543.

Bennett, J.H. 1990. *Statistical inference and analysis: Selected correspondence of R.A. Fisher*. Oxford: Clarendon Press.

Bernoulli, J. 1713. *Ars Conjectandi*. Basilea: Thurnisius. Trans. E.D. Sylla. Baltimore: Johns Hopkins University Press, 2006.

Bowley, A.L. 1902/1920/1926/1937. *Elements of statistics*, 2nd, 4th, 5th, and 6th ed. London: Staples Press.

Bowley, A.L. 1906. Address to the economic science and statistics. *Journal of the Royal Statistical Society* 69: 540–558.

Bowley, A.L. 1928. *F.Y. Edgeworth's contributions to mathematical statistics*. Clifton: Augustus M. Kelley.

Box, J.F. 1978. *R.A. Fisher: The life of a scientist*. New York: Wiley.

Box, G.E.P., and G.M. Jenkins. 1970. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Burns, A.F., and W.C. Mitchell. 1946. *Measuring business cycles*. New York: NBER.

Cameron, A.C., and P.K. Trivedi. 2005. *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press.

Cournot, A. 1843. *Exposition de la théorie des chances et des probabilités*. Paris: Hachette.

Cox, D.R. 1990. Role of models in statistical analysis. *Statistical Science* 5: 169–174.

Cox, D.R. 2006. *Principles of statistical inference*. Cambridge: Cambridge University Press.

Cox, D.R., and D.V. Hinkley. 1974. *Theoretical statistics*. London: Chapman & Hall.

Cramer, H. 1946. *Mathematical methods of statistics*. Princeton: Princeton University Press.

Davis, H.T., and W.F.C. Nelson. 1935. *Elements of statistics*. Indiana: Principia Press.

Desrosières, A. 1998. *The politics of large numbers: A history of statistical reasoning*. Cambridge, MA: Harvard University Press.

Dickey, D.A., and W.A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427–431.

Doob, J.L. 1953. *Stochastic processes*. New York: Wiley.

Engle, R.F., and C.W.J. Granger. 1987. Co-integration and error-correction: Representation, estimation and testing. *Econometrica* 55: 251–276.

Ezekiel, M. 1930. *Methods of correlation analysis*, 2nd ed. New York: Wiley.

Farebrother, R.W. 1999. *Fitting linear relationships: A history of the calculus of observations 1750–1900*. New York: Springer.

Fisher, R.A. 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10: 507–521.

Fisher, R.A. 1921. On the 'probable error' of a coefficient deduced from a small sample. *Metron* 1: 2–32.

Fisher, R.A. 1922a. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A* 222: 309–368.

Fisher, R.A. 1922b. On the interpretation of χ^2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society* 85: 87–94.

Fisher, R.A. 1922c. The goodness of fit of regression formulae and the distribution of regression coefficients. *Journal of the Royal Statistical Society* 85: 597–612.

Fisher, R.A. 1924. The conditions under which χ^2 measures the discrepancy between observation and hypothesis. *Journal of the Royal Statistical Society* 87: 442–450.

Fisher, R.A. 1925. *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.

Fisher, R.A. 1935a. The logic of inductive inference. *Journal of the Royal Statistical Society* 98(39–54): 55–82.

Fisher, R.A. 1935b. *The design of experiments*. Edinburgh: Oliver & Boyd.

Fisher, I. 1947. Response of Irving Fisher. *Journal of the American Statistical Association* 42: 4–5.

Fisher, R.A. 1955. Statistical methods and scientific induction. *Journal of the Royal Statistical Society. Series B* 17: 69–78.

Fisher, R.A. 1956. *Statistical methods and scientific inference*. Edinburgh: Oliver & Boyd.

Frisch, R. 1933. Editorial. *Econometrica* 1: 1–4.

Frisch, R. 1934. *Statistical confluence analysis by means of complete regression schemes*. Oslo: Universitetets Okonomiske Institutt.

Galton, F. 1869. *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan.

Gauss, C.F. 1809. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburg: F. Perthes & I.H. Besser.

Glass, D.V. 1973. *Numbering the people*. Farnborough: Saxton House.

Godambe, V.P., and D.A. Sprott. 1971. *Foundations of statistical inference: A symposium*. Toronto: Holt, Rinehart and Winston.

Goldberger, A.S. 1964. *Econometric theory*. New York: Wiley.

Gossett, W.S. 1908. The probable error of the mean. *Biometrika* 6: 1–25.

Granger, C.W.J., and P. Newbold. 1974. Spurious regressions in econometrics. *Journal of Econometrics* 2: 111–120.

Granger, C.W.J., and P. Newbold. 1986. *Forecasting economic time series*, 2nd ed. London: Academic Press.

Graunt, J. 1662. *Natural and political observations upon the bills of mortality*. London: John Martyn. Repr. in Hull (1899).

Greene, W.H. 2003. *Econometric analysis*, 5th ed. Englewood Cliffs: Prentice-Hall.

Greenwood, M. 1948. *Medical statistics from Graunt to Farr*. Cambridge: Cambridge University Press.

Haavelmo, T. 1944. The probability approach to econometrics. *Econometrica* 12(Suppl): 1–115.

Hacking, I. 1965. *Logic of statistical inference*. Cambridge: Cambridge University Press.

Hacking, I. 1975. *The emergence of probability*. Cambridge: Cambridge University Press.

Hald, A. 1990. *History of probability and statistics and their applications before 1750*. New York: Wiley.

Hald, A. 1998. *A history of mathematical statistics from 1750 to 1930*. New York: Wiley.

Hald, A. 2007. *A history of parametric statistical inference from Bernoulli to Fisher*. New York: Springer.

Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 97: 93–115.

Harrison, B. 2004. *Oxford dictionary of national biography*. Oxford: Oxford University Press.

Heckman, J.J. 1992. Haavelmo and the birth of modern econometrics: A review of the history of econometric ideas by Mary Morgan. *Journal of Economic Literature* 30: 876–886.

Heckman, J.J. 2001. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy* 109: 673–748.

Heckman, J.J., and B. Singer. 1984. Econometric duration analysis. *Journal of Econometrics* 24: 63–132.

Hendry, D.F. 1977. On the time series approach to econometric model building. In *New methods in business cycle research*, ed. C.A. Sims. Minnesota: Federal Reserve Bank of Minneapolis.

Hendry, D.F. 1993. *Econometrics: Alchemy or science?* Oxford: Blackwell.

Hendry, D.F. 1995. *Dynamic econometrics*. Oxford: Oxford University Press.

Hendry, D.F., and M.S. Morgan. 1995. *The foundations of econometric analysis*. Cambridge: Cambridge University Press.

Hilts, V.L. 1973. Statistics and social science. In *Foundations of scientific method: The nineteenth century*, ed. R.N. Giere and R.S. Westfall. Bloomington: Indiana University Press.

Hilts, V.L. 1978. Aliss exterendum, or, the origins of the Statistical Society of London. *Isis* 69: 21–43.

Hoppit, J. 1996. Political arithmetic in eighteenth-century England. *Economic History Review* 49: 516–540.

Horowitz, J.L. 1998. *Semiparametric methods in econometrics*. New York: Springer.

Hotelling, H. 1930. British statistics and statisticians today. *Journal of the American Statistical Association* 25: 186–190.

Hull, H.C. 1899. *The economic writings of Sir William Petty*. Cambridge: Cambridge University Press.

Hume, D. 1987. In *Essays, moral, political and literary*, ed. E.F. Miller. Indianapolis: Liberty Fund.

Jevons, W.S. 1871. *The theory of political economy*. New York: Augustus M. Kelley, 1911.

Johansen, S. 1991. Estimation and hypothesis testing of cointegrating vector of Gaussian vector autoregressive models. *Econometrica* 59: 1551–1580.

Johnston, J. 1963. *Econometric methods*. New York: McGraw-Hill.

JSSL (Journal of the Statistical Society of London). 1834. Prospects of the objects and plan of operation of the statistical society of London. Reprinted in *Journal of the Statistical Society of London* 1869: 385–387.

Keynes, J.N. 1890/1917. *The scope and method of political economy*. Fairfield: Augustus M. Kelley.

Keynes, J.M., and J. Tinbergen. 1939–40. Professor's Tinbergen's method. *Economic Journal* 49: 558–568; A Reply, by J. Tinbergen, and Comment by Keynes, 50: 141–156.

Klein, L.R. 1950. *Economic fluctuations in the United States 1921–1941*, Cowles commission for research in economics, monograph no. 11. New York: Wiley.

Kolmogorov, A.N. 1933. *Foundations of the theory of Probability*, 2nd ed. New York: Chelsea.

Koopmans, T.C. 1939. *Linear regression analysis of economic time series*, Netherlands Economic Institute, publication no. 20. Haarlem: F. Bohn.

Koopmans, T.C. 1947. Measurement without theory. *Review of Economics and Statistics* 29: 161–172.

Koopmans, T.C. 1950. *Statistical inference in dynamic economic models*, Cowles commission monograph, no. 10. New York: Wiley.

Kreager, P. 1988. New light on Graunt. *Population Studies* 42: 129–140.

Legendre, A.M. 1805. *Nouvelles Méthodes pour la détermination des orbites des comètes*. Paris: Courcier.

Lehmann, E.L. 1990. Model specification: The views of fisher and Neyman, and later developments. *Statistical Science* 5: 160–168.

Lehmann, E.L. 1993. The Fisher and Neyman–Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association* 88: 1242–1249.

Lehmann, E.L. 1999. 'Student' and small-sample theory. *Statistical Science* 14: 418–426.

Letwin, W. 1965. *The origins of scientific economics*. New York: Anchor Books.

Lieberman, B. 1971. *Contemporary problems in statistics: A book of readings for the behavioral sciences*. Oxford: Oxford University Press.

Manski, C.F., and D. MacFadden. 1981. *Structural analysis of discrete data with econometric applications*. Cambridge, MA: MIT Press.

Mayo, D.G. 1991. Novel evidence and severe tests. *Philosophy of Science* 58: 523–552.

Mayo, D.G. 1996. *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.

Mayo, D.G. 2005. Philosophy of statistics. In *Philosophy of science: An encyclopedia*, ed. S. Sarkar and J. Pfeifer. London: Routledge.

Mayo, D.G., and A. Spanos. 2004. Methodology in practice: Statistical misspecification testing. *Philosophy of Science* 71: 1007–1025.

Mayo, D.G., and A. Spanos. 2006. Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *British Journal for the Philosophy of Science* 57: 323–357.

Mills, F.C. 1924. *Statistical methods*. New York: Henry Holt, 1938.
Moore, H.L. 1908. The statistical complement of pure economics. *Quarterly Journal of Economics* 23: 1–33.
Moore, H.L. 1911. *The law of wages*. New York: Macmillan.
Morgan, M.S. 1990. *The history of econometric ideas*. Cambridge: Cambridge University Press.
Morrison, D.E., and R.E. Henkel. 1970. *The significance test controversy: A reader*. Chicago: Aldine.
Neyman, J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97: 558–625.
Neyman, J. 1935. On the problem of confidence intervals. *Annals of Mathematical Statistics* 6: 111–116.
Neyman, J. 1950. *First course in probability and statistics*. New York: Henry Holt.
Neyman, J. 1952. *Lectures and conferences on mathematical statistics and probability*, 2nd ed. Washington, DC: U.S. Department of Agriculture.
Neyman, J. 1956. Note on an article by Sir Ronald Fisher. *Journal of the Royal Statistical Society. Series B* 18: 288–294.
Neyman, J. 1977. Frequentist probability and frequentist statistics. *Synthese* 36: 97–131.
Neyman, J., and E.S. Pearson. 1933. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A* 231: 289–337.
Pagan, A.R., and A. Ullah. 1999. *Nonparametric econometrics*. Cambridge: Cambridge University Press.
Pearson, K. 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A* 187: 253–318.
Pearson, K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables in such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* 50: 157–175.
Pearson, E.S. 1931a. The analysis of variance in cases of non-normal variation. *Biometrika* 23: 114–133.
Pearson, E.S. 1931b. Note on tests for normality. *Biometrika* 22: 423–424.
Pearson, E.S. 1955. Statistical concepts in the relation to reality. *Journal of the Royal Statistical Society. Series B* 17: 204–207.
Pearson, E.S. 1966. The Neyman–Pearson story: 1926–34. In *Research papers in statistics: Festschrift for J. Neyman*, ed. F.N. David. New York: Wiley.
Peirce, C.S. 1878. The probability of induction. *Popular Science Monthly* 12: 705–718.
Persons, W.M. 1925. Statistics and economic theory. *Review of Economics and Statistics* 7: 179–197.
Phillips, P.C.B. 1986. Understanding spurious regression in econometrics. *Journal of Econometrics* 33: 311–340.
Phillips, P.C.B. 1987. Time series regressions with a unit root. *Econometrica* 55: 227–301.
Phillips, P.C.B., and S.N. Durlauf. 1986. Multiple time series regression with integrated processes. *Review of Economic Studies* 53: 473–495.
Playfair, W. 1786. *The commercial and political atlas*. London: T. Burton, 1801.
Porter, T.M. 1986. *The rise of statistical thinking 1820–1900*. Princeton: Princeton University Press.
Quetelet, A. 1942. *A treatise on man and the development of his faculties*. Edinburgh: Chambers.
Redman, D.A. 1997. *The rise of political economy as a science*. Cambridge, MA: MIT Press.
Routh, G. 1975. *The origins of economic ideas*. London: Macmillan.
RSS (Royal Statistical Society). 1934. *Annals of the royal statistical society 1834–1934*. London: The Royal Statistical Society.
Schweber, L. 2006. *Disciplining statistics: Demographic and vital statistics in France and England, 1830–1885*. Durham: Duke University Press.
Secrist, H. 1930. *An introduction to statistical methods*, 2nd ed. New York: Macmillan.
Sims, C.A. 1980. Macroeconomics and reality. *Econometrica* 48: 1–48.
Slutsky, E. 1927. The summation of random causes as the source of cyclic processes. In Russian; English trans. in *Econometrica* 5: 1937.
Smith, A. 1776. In *An inquiry into the nature and causes of the wealth of nations*, ed. R.H. Campell, A.S. Skinner, and W.B. Todd. Oxford: Clarendon Press.
Spanos, A. 1986. *Statistical foundations of econometric modelling*. Cambridge: Cambridge University Press.
Spanos, A. 1989a. On re-reading Haavelmo: A retrospective view of econometric modeling. *Econometric Theory* 5: 405–429.
Spanos, A. 1989b. Early empirical findings on the consumption function, stylized facts or fiction: A retrospective view. *Oxford Economic Papers* 41: 150–169.
Spanos, A. 1990. The simultaneous equations model revisited: Statistical adequacy and identification. *Journal of Econometrics* 44: 87–108.
Spanos, A. 1995. On theory testing in econometrics: Modeling with nonexperimental data. *Journal of Econometrics* 67: 189–226.
Spanos, A. 1999. *Probability theory and statistical inference: Econometric modeling with observational data*. Cambridge: Cambridge University Press.
Spanos, A. 2001. Time series and dynamic models. In *A companion to theoretical econometrics*, ed. B. Baltagi. Oxford: Blackwell.
Spanos, A. 2005. Misspecification, robustness and the reliability of inference: The simple t-test in the presence of Markov dependence. Working paper, Virginia Polytechnic Institute and State University.
Spanos, A. 2006a. Econometrics in retrospect and prospect. In *New Palgrave handbook of econometrics*, vol. 1, ed. T.C. Mills and K. Patterson. London: Macmillan.

Spanos, A. 2006b. Revisiting the omitted variables argument: Substantive vs. statistical adequacy. *Journal of Economic Methodology* 13: 179–218.

Spanos, A. 2006c. Where do statistical models come from? Revisiting the problem of specification. In *Optimality: The second Erich L. Lehmann symposium*, Lecture notes-monograph series, vol. 49, ed. J. Rojo. Beachwood: Institute of Mathematical Statistics.

Spanos, A. 2006d. Revisiting the statistical foundations of panel data models. Working paper, Virginia Polytechnic Institute and State University.

Spanos, A. 2008. Curve-fitting, the reliability of inductive inference and the error-statistical approach. *Philosophy of Science* 74: 1046.

Spanos, A., and A. McGuirk. 2001. The model specification problem from a probabilistic reduction perspective. *Journal of the American Agricultural Association* 83: 1168–1176.

Stigler, G.J. 1954. The early history of the empirical studies of consumer behavior. *Journal of Political Economy* 62: 95–113.

Stigler, S.M. 1986. *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.

Stone, J.R.N. 1954. *The measurement of consumers' expenditure and behaviour in the United Kingdom, 1920–1938*. Cambridge: Cambridge University Press.

Stone, R. 1997. *Some british empiricists in the social sciences 1650–1900*. Cambridge: Cambridge University Press.

Tinbergen, J. 1939. *Statistical testing of business cycle research*, vol. 2. Geneva: League of Nations.

Tippet, L.H.C. 1931. *The methods of statistics*. London: Williams & Norgate.

Uspensky, J.V. 1937. *Introduction to mathematical probability.* New York: McGraw-Hill.

Walker, H.M. 1929. *Studies in the history of statistical method*. Baltimore: Williams & Wilkins.

Walker, H.M. 1958. The contributions of Karl Pearson. *Journal of the American Statistical Association* 53: 11–22.

Wold, H.O. 1938. *A study in the analysis of stationary time series*. Uppsala: Almquist & Wicksell.

Yule, G.U. 1897. On the theory of correlation. *Journal of the Royal Statistical Society* 60: 812–854.

Yule, G.U. 1911. *An introduction to the theory of statistics*. London: Griffin & Co.

Yule, G.U. 1926. Why do we sometimes get nonsense correlations between time series? – A study in sampling and the nature of time series. *Journal of the Royal Statistical Society* 89: 1–64.

Yule, G.U. 1927. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A* 226: 267–298.