Statistical Science
2005, Vol. 20, No. 4, 401–417
DOI 10.1214/088342305000000331

Fisher and Regression

John Aldrich

Abstract. In 1922 R. A. Fisher introduced the modern regression model, synthesizing the regression theory of Pearson and Yule and the least squares theory of Gauss. The innovation was based on Fisher's realization that the distribution associated with the regression coefficient was unaffected by the distribution of X. Subsequently Fisher interpreted the fixed X assumption in terms of his notion of ancillarity. This paper considers these developments against the background of the development of statistical theory in the early twentieth century.

Key words and phrases: R. A. Fisher, Karl Pearson, M. S. Bartlett, regression, theory of errors, correlation, ancillary statistic, history of statistics.

INTRODUCTION

In the 1920s R. A. Fisher (1890–1962) created modern regression analysis out of two nineteenth century theories: the theory of errors of Gauss and the theory of correlation of Pearson. Although much has been written about these theories, the synthesis has not really been noticed.

Fisher is generally credited with having completed the distribution theory of the theory of errors, the last phase in "the historical development of the Gauss linear model" (cf. Seal, 1967). However, associated with his t and F theory was a reconception of regression. Regression had belonged, not to Gauss' univariate theory of errors, but to the multivariate theory of correlation. Fisher's reconception rested on two innovations: the normal linear regression specification that, conditional on the x's, y is normally distributed with its expectation linear in the x's, and the notion that for inference the x values could be treated as fixed. Historians have passed over the reconception. For Seal (1967, page 16) it was restoration: "with Fisher (1922a) the sampling theory of regression equations returned to the Gaussian model." For Hald (1998, page 616) it was seeing the obvious: the statisticians who worked on the multivariate theory "did not realise that their regression analysis was a version of the linear model and the linear estimation theory." Yet the realization did not come easily, or at all, to the statisticians who lived through the change—Karl Pearson, Yule and Gosset—and how could the theory return to where it had never been?

The account below begins before Fisher and follows the new regression from its emergence in the 1920s to Fisher's last presentation of the theory more than 30 years later. Section 1 describes the situation in *Britain* in 1900; I emphasize Britain because the situation was different elsewhere as Hald (1998) makes clear. In Britain there was an almost moribund univariate theory and a new and thriving multivariate theory. There was work to do in the multivariate theory—on goodness of fit and inference about regression coefficients—but there was no sign that solving these problems would lead to a reverse in which the univariate theory would take over much of business of the multivariate theory. These challenges and Fisher's responses are described in Sections 2–4. Fisher went on to put his idea of regression to a broader audience in his *Statistical Methods for Research Workers* (Fisher, 1925a). The new unified theory relied on the argument that the distribution of the test statistics is the same whether the x's are fixed or random. Continuing research on the multivariate theory of regression—notably by M. S. Bartlett (1910–2002)—supported the argument. Sections 5 and 6 describe the new idea of regression and the continuing work on the old.

In the late 1930s, after Fisher had added a conditional inference chapter to his theory of estimation, a new idea about the legitimacy of fixed x analysis appeared: the x's provide "ancillary information" about regression coefficients. Sections 7–10 follow

John Aldrich is Reader, Division of Economics, School of Social Sciences, University of Southampton, Highfield, Southampton SO17 1BJ, UK (e-mail: john.aldrich@soton.ac.uk).

this development. Bartlett again made an important contribution: he was the first to write about the regression/ancillarity connection—in 1936. In Fisher's last engagement with fixed x regression—discussed in Section 11—regression is pressed into the campaign against Neyman's emphasis on "repeated sampling from the same population." Section 12 has some comments on the entire story and on where we stand now. First, though, there is a sketch of the theories Fisher unified; for further details see Seal (1967), Stigler (1986), Aldrich (1995, 1998, 1999), Hald (1998) and Farebrother (1999).

1. THE NINETEENTH CENTURY BACKGROUND

The theory of errors or the "Gauss linear model" had been devised for combining astronomical observations. Using modern notation, equations from dynamical theory ($\theta = X\beta$) connect the nonstochastic θ and X with β, a vector of unknown quantities. The elements of X are observed, but not those of θ; a vector of measurements y deviates from θ by the vector of unobserved errors, ε. In Gauss' (1809) first treatment the error vector is distributed $N(0, \sigma^2 I)$ and the least squares estimate is the mode of the posterior of β obtained from a uniform prior. For Karl Pearson (1857–1936) and Fisher—the main protagonists of the regression story—this combination of specification and inference procedure *was* the theory of errors, *was* Gauss; see Aldrich (1997, pages 162–164). Neither referred to Gauss' second proof—without normality and treated by the Gauss–Markov theorem—although Pearson used books that presented it. The proof had come to Britain in the nineteenth century, but in its second coming it was called the Markoff theorem in Neyman (1934).

Extensions of the error theory setup in which the elements of X might also be measured with error or the elements of θ might be points in space were developed, but they did not become an essential part of the teaching. The theory of errors was also applied to fitting empirical formulae. Merriman's (1884/1911) textbook (the least squares reference for Pearson and Yule) gives many examples; the x values are either time trends or quantities selected by the scientist—for example, recorded depth and depth squared in an equation for water velocity (Merriman, 1884/1911, page 131)—not values of observational variables. Surprisingly, perhaps, the application of least squares to observational data appears to have begun only in the correlation era.

Multivariateness was essential to Galton's (1877) normal autoregressive process of reversion and his bivariate normal correlation (Galton, 1886). In developing these specifications Galton used distributional results from the theory of errors but not its inference theory. Edgeworth (1893) applied least squares theory to the estimation of the correlation coefficient when he took a weighted average of the ratios y/x, treating the x values as nonrandom. However, as Stigler (1986, page 321, 2001) noted, Edgeworth gave no justification for this procedure and dropped it when Pearson's approach arrived.

Pearson (1896) and Pearson and Filon (1898) applied a large-sample Bayesian argument to the parameters of the multinormal data density, including the regression and partial regression coefficients; the argument probably derived ultimately from Gauss (see Aldrich, 1997, page 170, 1993 and Hald, 1999). The formulae for the regression coefficients and their probable errors were the same as those for least squares values, although the difference in derivation and notation obscured the fact. If Pearson had noticed and investigated the point, the modern Bayesian rationalization of regression analysis—as in Gelman, Carlin, Stern and Rubin (1995, page 235)—might have arrived in the nineteenth century. Instead Pearson emphasized the differences between the theory of correlation and the theory of errors. Pearson (1920, pages 25–27) described how the theories have different domains—the X and y of Gauss' theory are *not* correlated—and multinormality enters in one theory as the (posterior) distribution of β and in the other as the distribution of the observables. Pearson was not dissatisfied with his large-sample results, but he was dissatisfied with the underlying assumption of normality, so he worked on a theory of skew correlation to complement his (Pearson, 1895) theory of skew curves; see Section 2 below.

In the Introduction I described the normal linear regression specification as one of Fisher's innovations. Seal (1967, page 15) described how Pearson (1896) was aware that conditional normality could hold without joint normality, in particular when the values of the conditioning variables are selected. Furthermore, Seal mentioned Pearson's "On the reconstruction of the stature of prehistoric races" (Pearson, 1899), which presents a regression formula for the most probable value of an organ B given the values of other organs $A_1, \ldots, A_n$. However, in the 1896 paper Pearson does not provide any inference theory for those nonnormal cases and in the 1899 applied paper he does not use any inference theory. Fisher made the regression specification central and provided an inference theory for it.

G. U. Yule (1871–1951) also wanted to escape from multinormality. His work needs close examination because it looks like Fisher's regression analysis without the small-sample refinements. Yule's use of least squares in regression, his inventiveness in least squares technique and his interpretations of causal relationships have been lovingly detailed by Stigler (1986, pages 348–353) and Aldrich (1995, 1998). Yet Yule did not make a transforming contribution to the inference theory of regression. His (Yule, 1897, pages 813–817) way of going beyond the multinormal specification was to point out that the regression curve of conditional expectations existed in any multivariate population and to choose linear least squares for estimating this curve for "convenience of analysis." *All* Yule took from the theory of errors was the idea of least squares and the statement of the first-order conditions for a minimum. He took no further inference theory nor did he devise any, which is not surprising, because to apply the method of Pearson and Filon, Yule's (1897, 1907) authority on inference, would require a parametric form for the joint density and he did not have one.

Yule's (1899) major empirical paper shows how his regression methods anticipated the Fisher synthesis and how his theory did not. He used least squares without reservation—seeing the regressors as not normal—but he presented the Pearson and Filon normal theory probable errors, warning "so far as they are valid for these cases of nonnormal correlation" (Yule, 1899, page 277). Yule did not develop his idea of regression and it appears unchanged in his textbook (Yule, 1911). His regression work began to have an influence only around the time of the Fisher synthesis, for example, in Tolley and Ezekiel (1923). His use of least squares may have influenced Fisher, but the latter's theory drew on Pearson (1896, 1916), Slutsky (1913) and textbook Gauss.

Stigler's history ends on a least squares high with a "second great synthesis" (Stigler, 1986, page 360), but Yule's intuitions had limited influence and elsewhere least squares was under pressure. Pearson (1902a) argued that the method of moments was superior in curve fitting and Pearson (1900, 1902b) argued that observational errors in astronomy are *not* normally distributed. In early twentieth century Britain the theory of errors was applied and taught—to Fisher with spectacular consequences!—but there was no sustained theoretical research. Students of biometry and statistics were not taught the theory of errors. The textbooks mentioned least squares for the sake of the mathematician reader but did not expound least squares or presuppose it. The Edgeworthian Bowley (1901, page 284) mentioned least squares twice, noting that his way of rationalizing his estimate of the modulus of the normal distribution came from that literature and remarking (Bowley, 1901, page 177) that least squares might be used to obtain a relationship between the marriage rate and foreign trade. The Pearsonian Elderton (1906, page viii) skipped least squares because "the range of its applicability is so limited that there is a growing tendency to put it aside in curve fitting." Yule (1911, page 233) mentioned the method of least squares, but did not suggest the reader study it. Meanwhile Pearsonian notions were entering courses for astronomers on the combination of observations as in Brunt (1917, Chapters IX and X). There were a few harbingers of Fisher's Gauss revival. Student's (1908a) problem of the "probable error of the mean" belonged to the theory of errors, although it was written for Pearson's journal and in his language; see Aldrich (2003). Student was also applying the theory of errors to agricultural experiments, as was Fisher's tutor, the astronomer F. J. M. Stratton; see E. S. Pearson (1990, page 47). However, the most powerful force—Pearson's laboratory—was being applied in a different direction.

2. THE TWENTIETH CENTURY: REGRESSION AND GOODNESS OF FIT

At the beginning of the twentieth century Yule had moved on to other aspects of covariation—to association and, eventually, to time series analysis. The only continuing regression project was Pearson's, and his nonlinear regression generated the first of the problems that Fisher solved—the testing of goodness of fit.

Pearson (1923) reviewed the years of struggle to develop a "general theory of skew correlation and nonlinear regression" based on a surface relating to univariate skew curves as the multinormal surface relates to the normal curve. In earlier days he (Pearson 1905, Section 4) had generalized Yule and presented the regression curve, admitting that the full theory of skew correlation surfaces "has not yet been worked out owing to difficulties of analysis, and their complete discussion must be postponed." Pearson did not write down a specification for the regression curve, but something very commodious is implied, say

$$Y = \mu(X) + \varepsilon(X),$$

where the regression curve $\mu(X) = E(Y|X)$ may be linear, quadratic, ... or quartic. My symbol $\varepsilon(X)$ marks the possibility that the distribution of the deviation can vary with X; in particular, skewness and

scedasticity can vary. The data on X are grouped and consist of repeated values of x—an array—with associated values of y. If Y_p is associated with x_p and μ_p is its expected value, then the regression curve of y on x expresses the relationship between μ_p and x_p. There are n_p replicates of x_p, where the numbers n_p are random variables. Pearson does not restrict the distribution of Y_p around μ_p and he certainly did not want to assume the normal homoscedastic case. Blyth (1994) presents Pearson's project and his data analysis from the perspective of Bjerve and Doksum's (1993) theory of correlation curves.

Pearson (1905, Section 4) gave a string of propositions that led up to the probable error of the "correlation ratio," a measure of correlation that in the case of linear regression equals the correlation coefficient. Pearson proposed testing for linearity by comparing the two quantities and Blakeman (1905) developed the suggestion. Pearson did not give probable errors for the regression coefficients that are to be estimated by the method of moments.

A goodness of fit test was provided by E. E. Slutsky (1880–1948), an economist correlator in Kiev, remembered today for his work on probability and stochastic processes; see Seneta (1988) for a brief biography. Slutsky (1913) proposed a χ^2 test for the skew correlation setup. The test was a contribution to the Pearson regression project and used Pearson's (1900) χ^2 test, yet its assumption of normality was a departure from Pearson's practice. Pearson's χ^2 ventures involved the multinormal, but as an approximation to the multinomial, not as a data distribution in its own right: normality, even conditional normality, was not acceptable.

To formulate Slutsky's test, denote by $\overline{y}_p$ the mean of the y's associated with μ_p and by e_p the deviation of $\overline{y}_p$ from its expected value μ_p. Appealing to one of Pearson's propositions, Slutsky stated that the standard deviation of e_p is given by $\sigma_p/\sqrt{n_p}$, where σ_p is the standard deviation of y in the pth array. Appealing to another, he states, "Now it is known that there is no correlation between the deviations in the mean of an x-array and in the mean of a second x-array." Slutsky altered Pearson's specification, retaining heteroscedasticity but assuming that each Y_p is normally distributed around the appropriate μ_p. Slutsky concluded that the quantity

$$\chi^2 = \sum \frac{n_p(\overline{y}_p - \mu_p)^2}{\sigma_p^2}$$

is distributed as chi squared with the number of degrees of freedom equal to the number of arrays. In the test statistic, estimates replace unknowns.

The first of Slutsky's examples is the cubic Pearson (1905, Section 9) fitted to the height and age of 2272 girls classified into 20 age groups. The other is a linear relationship fitted to 124 observations classified into 11 groups on the price of rye in pairs of adjacent months (the first fitting of an autoregressive model to time series data?). Whereas the numbers in the groups are small, heteroscedasticity cannot be established, so for this case Slutsky reworked the test assuming homoscedasticity.

Pearson reacted to Slutsky's procedure first (Pearson, 1914, page xxxii) by offering a "word of caution" and then by setting out his own ideas in a 1916 paper. Pearson (1916, page 256) warned that "very fallacious results" can be reached by Slutsky's test. Pearson criticized the presumption of normality, but adopted the assumption nevertheless. He was content with replacing population quantities with sample quantities, but considered Slutsky's replacements unsatisfactory. In the case of a random array size, the $\sigma_p/\sqrt{n_p}$ quantity could be improved on, but Pearson (1916, page 248) had a more general objection: Slutsky's arbitrary practice of estimating σ_p and n_p from data on the pth array but estimating μ_p from all the data—all the quantities should be estimated from all the data. The realized value n_p is replaced by an estimate obtained from fitting a distribution to the x values, and σ_p is estimated from the entire sample by using the heteroscedasticity relationship between σ_p and x. Compromises are necessary when working with Slutsky's price data (Pearson, 1916, pages 250–253), but the full scheme is demonstrated on the abundant height/age data (Pearson, 1916, pages 253–256).

The Pearson archives at University College London have a letter from 1912 in which Slutsky outlined his test but not Pearson's reply. Slutsky told Pearson that "quite analogous" would be a criterion to be applied to the physical sciences for testing whether a given system of measurements can reasonably be supposed to correspond to a certain functional relationship. Slutsky's published paper does not consider this application and restricts itself to observational (statistical) data. However Pearson's (1916) "On the application of 'goodness of fit' tables to test regression curves and theoretical curves used to describe observational or experimental data" considered both and rejected any analogy.

Pearson (1916, page 256) wrote of the physicist who makes a few measurements of a variate A for each of a series of a variate B: "there is no question in the ordinary sense of a frequency surface." For Pearson

there was an essential difference between the physical and the statistical cases: in the former, the numbers in each array are nonstochastic. For the category of "physical, technical and astronomical measurements," Pearson's (1916, pages 256–258) procedure is the same as Slutsky's except that σ_p is estimated by a different method. Pearson (1916, page 247) remarked, "It is singular that the goodness of fit theory can actually be applied with greater accuracy to test physical laws than to test regression lines." Presumably the reasoning behind Slutsky's analogy was the same as Fisher later gave—that the distribution of the test statistic is the same. To judge from the 1913 paper, Slutsky was not as deeply immersed in the theory of errors as Fisher.

3. FISHER ON THE FIT OF REGRESSION FORMULAE

Fisher's first years at Rothamsted were spectacularly productive; see Box (1978, Chapters 3–5) and, for Fisher more generally, Aldrich (2003–2005). In 1922 he was working on χ^2 theory, agricultural meteorology, genetics, the theory of estimation and the analysis of variance—projects which were more interrelated than they may sound; there are sketches of some of them in Fienberg and Hinkley (1980). Regression goodness of fit was a minor division of χ^2 theory. Work in another division brought Fisher recognition from the statisticians because Bowley and Yule were both dissatisfied with Pearson's contingency table theory. The regression work, however, made no immediate mark; it was not mentioned in the new editions of the Bowley and Yule textbooks.

The main business of the paper (Fisher, 1922a) "The goodness of fit of regression formulae, and the distribution of regression coefficients" was to sort out the regression goodness of fit issue. Fisher's core model was the normal linear regression model: conditional on the x's, y is normally distributed with its expectation linear in the x's. Expressed in modern notation the goodness of fit analysis uses

$$y \sim N(X\beta, \sigma^2 I),$$

where the N rows of X (the x's may be powers of some underlying variable) comprise a distinct vectors, the pth of which is replicated n_p times, where n_p is random. Fisher focusses on the homoscedastic version of Slutsky's statistic, namely

$$\chi^2 = \sum \frac{n_p(\overline{y}_p - \mu_p)^2}{\sigma^2}.$$

Fisher treated explicitly only the statistical case of random n's, though it is obvious that the results also hold for the physical case. The key step is establishing the distributions of the components of the χ^2 statistic. Fisher (1922a, page 598) wrote:

> For such samples of n_p, therefore, the mean, $\overline{y}_p$, will vary about the same mean m_p [my μ_p], and since this mean of $\overline{y}_p$ is independent of the number in the array, m_p [my μ_p] will be the mean of all values of $\overline{y}_p$ from random samples, however the number n_p may vary.

Fisher took $\sqrt{n_p}(\overline{y}_p - \mu_p)$ to be normal with mean zero and standard deviation σ, so the numerator of Slutsky's statistic was σ^2 times a χ^2. When μ_p has to be estimated, a degrees of freedom adjustment is necessary; in this regression setup there is the further distributional complication associated with s^2 replacing σ^2 in the test statistic,

$$\chi^2 = \sum \frac{n_p(\overline{y}_p - \widehat{\mu}_p)^2}{s^2}.$$

The estimate s^2 is obtained by combining the within-array estimates of σ^2. The combining rule is based on a marginal maximum likelihood argument; s^2 has a χ^2 distribution with $N - k$ degrees of freedom. Fisher derived the exact distribution of the test statistic and identified it as a Pearson Type VI curve—as distinct from the Type III, which is appropriate when σ^2 is known. When Fisher (1924–1928, page 812, 1925a, pages 214–218) presented the test in the format of analysis of variance, he introduced the numbers of degrees of freedom associated with the deviation of the array mean from the formula and rescaled the statistic to become $F(a - k, N - k)$. *Statistical Methods for Research Workers* (Fisher, 1925a) has tables for $z = 1/2 \ln F$. Hald (1998, Section 27.6) has a more detailed discussion.

Fisher compared his statistic with Slutsky's and with Pearson's statistic for the experimental case. The point he stressed was that neither Slutsky nor Pearson adjusted the degrees of freedom for the estimated parameters; the need for such an adjustment was the theme of his χ^2 work (see, e.g., Fisher, 1922c and the discussions by Lancaster, 1969, Chapter 1, Fienberg, 1980 and Hald, 1998, Section 27.4). It is very clear that for Fisher the observational and experimental cases should *not* be treated differently. For example, he (Fisher, 1925a, page 607) comments on the limitation of the

analysis to the case of groups of y-values that correspond to identical values of x: "little statistical or physical data is strictly of this kind although the former may in favourable cases be confidently grouped, so as to simulate [this] kind of data." When he illustrated the method in *Statistical Methods* (Fisher, 1925a, Example 42 of Section 44 in all editions), it was for a (nonrandomized) experiment on the influence of temperature on the number of eye facets in drosophila.

4. THE DISTRIBUTION OF REGRESSION COEFFICIENTS

Fisher's treatment of the first topic of "The goodness of fit of regression formulae, and the distribution of regression coefficients" (1922a) was an incremental improvement, a resolution of a disagreement in the literature. The second topic also arose from the multivariate theory—from his own paper (Fisher, 1915) on the exact distribution of the correlation coefficient and ultimately from Pearson (1896)—but the treatment came out of nowhere.

The story of how Gosset asked Fisher for the regression counterpart of the 1915 result and how a new role for Student's (1908a) distribution was found is familiar from Box (1978, page 115), Eisenhart (1979, pages 7–8), E. S. Pearson (1990, page 48) and Lehmann (1999, pages 420–421). In April 1922 Gosset wrote (letter 5 of McMullen, 1970):

> I want to know what is the frequency distribution of $r\sigma_x/\sigma_y$ for small samples, in my work I want that more than the r distribution now happily solved. . . .

Apparently Fisher sent his answer by return and then included it in the goodness of fit paper. In a letter from 1954 (see Bennett, 1990, page 214) Fisher refers to applying the principle used in treating the goodness of fit test to the distribution of regression coefficients. In fact he never answered Gosset's question because his (Fisher, 1922a, page 598) "exact solution of the distribution of the regression coefficients" proved to be the distribution of the regression coefficient t-ratio, to use modern (post-1925) terminology.

Fisher's argument belongs to the theory of errors and follows now-familiar lines; see Seal (1967, page 17). In the simple case with which he began the dependent variable y is normally distributed with expectation $a + b(x - \overline{x})$ and standard deviation σ. The "coefficients a and b are calculated by the equations"

$$a = \overline{y}, \quad b = \frac{\sum y(x - \overline{x})}{\sum (x - \overline{x})^2}.$$

The assumption that x is *given* slips out when Fisher (1922a, page 608) notes that "a and b are orthogonal functions, in that given the series of x observed, their sampling variation is independent." Fisher gives only the derivation of the Student distribution associated with the parameter α, but he states the results for all the coefficients in multiple regression. It may be worth remarking on the notational innovations here, for example, the use of Greek and Latin letters for statistics and parameters; for more on this theme see Aldrich (2003) or Miller (1999–2005) (continuing) *Earliest Uses of Symbols*. . . .

Everything in Fisher's paper—apart apparently from the argument—indicates that Fisher was talking about regression as it had been traditionally understood: the language of "regression coefficients," the bundling with the regression goodness of fit test, the reference (Fisher, 1925a, page 612) to "agricultural meteorology," where x's are weather variables (as in Hooker, 1907), and finally the emphasis in the statement (Fisher, 1925a, page 611), "the accuracy of the regression coefficients is only affected by the correlations which appear *in the sample*," which makes no sense unless there is a population of x's.

Gosset was gratified by Fisher's extension of his (Student, 1908a) distribution, but he was not convinced. Through 1922 he kept asking for the marginal distribution for b; in November he told Fisher that the proof of the distribution of b is limited to "given $\overline{x}$ and σ_x." Fisher's replies have not survived, but Fisher (1925c) contains an answer. Fisher (1925c, page 96) begins his derivation of the distribution of the t-ratio, emphasizing that he is "confining attention to samples having the same value of x." The work done, he reflects (Fisher, 1925c, page 99):

> The quantity t involves no hypothetical quantities, being calculable wholly from the observations. It is the point of the method, as of 'Student's' original treatment of the probable error of the mean, to obtain a quantity of known distribution expressible in terms of the observations only. If we had found the distribution of b for samples varying in the values of x observed, we should have been obliged to express the distribution in terms of the unknown standard deviation σ_x in the population sampled; moreover since σ_x is unknown, we should have been obliged to substitute for it an estimate based on $S(x - \overline{x})^2$; the inexactitude

> of the estimate would have vitiated our solution, and required us to make allowance for the sampling variation of $S(x - \bar{x})^2$; finally this process, when allowance had been accurately made would lead us back to the 'Student's' distribution found above. The proof given above has, however the advantage that it is valid whatever may be the distribution of x, provided that y is normally and equally variable in each array, and the regression of y on x is linear in the population sampled.

While it is not clear that the proof is "valid," the answer to Gosset is clear: the marginal distribution of b is no use on its own and the usable form—the t-ratio—is available whether x is normally distributed or not.

5. THE IDEA OF REGRESSION

Statistical Methods for Research Workers (1925a, page 114) presents Fisher's idea of regression:

> The idea of regression is usually introduced in connection with the theory of correlation, but it is in reality a more general, and, in some respects a simpler idea, and the regression coefficients are of interest and scientific importance in many classes of data where the correlation coefficient, if used at all, is an artificial concept of no real utility.

Excluding Fisher's own work, "usually" can be read as "invariably." Regression and correlation were related features of a joint distribution.

Fisher took the situations for which Pearson and Yule had used correlation/regression and Merriman has used least squares, and treated them together. For Yule, least squares was outside statistics: correlation is "an application [of least squares] to the purposes of statistical investigation" (Yule, 1909, page 722). He did not conflate the situations where least squares was traditionally used with those to which correlation/regression was appropriate, nor did he conflate the sampling theories. Fisher did both. As "classes of data," he made no distinction between observational and experimental material: his examples (Fisher, 1925a, pages 114–136) of x and y include age and height of children, height of fathers and sons, fertilizers, and yield, time and yield, position and rainfall. One sampling theory does for all. Curiously Fisher, so prolific in creating new terms, retained the term "regression," extending its range into the theory of errors. Years before Yule (1897, page 814) had wanted to use the colorless term "characteristic line" instead of "regression line" which had unwanted associations with biological "stepping back."

Fisher's (1925a, pages 114–115) only restrictions on the use of the model arise from the "very different relations" the independent and dependent variables bear to the regression line. If errors occur in the former, the regression line will be altered; if they occur in the latter, the regression line will not be altered, provided the errors "balance in the averages"; so the errors in variables case was *not* covered. Second, "the regression function does not depend on the frequency distribution of the independent variable, so that a true regression line may be obtained even when the age groups are arbitrarily selected. . . ." On the other hand, a selection of the dependent variate will "change the regression line altogether."

The book *Statistical Methods. . .* has a chapter on correlation as an aspect of the bivariate normal; evidently correlation coefficients may be of "interest and scientific importance." The results of several of his papers (Fisher, 1915, 1921a, 1925d) are presented and illustrated. The examples that illustrate the significance of a correlation and a partial correlation (Fisher, 1925a, pages 158–161) are from agricultural meteorology and Yule's work on pauperism. Fisher mentions that the partial correlation depends on the "assumption that the variates correlated (but not necessarily those eliminated) are normally distributed." Fisher was attached to the idea of investigating the existence of dependence between variables by testing hypotheses about the correlation coefficient rather than the regression coefficient. The t-test that modern packages offer (of $\beta = 0$) appears in *Statistical Methods. . .* as a test on the correlation coefficient, a coefficient only meaningful in the bivariate normal setting.

Statistical Methods. . . was a very busy book and the reviewers, including Student (1926) and E. S. Pearson (1926), had plenty to discuss without mentioning regression. The book went through 14 editions and came to be recognized as epoch-making; an issue of the *Journal of the American Statistical Association* marked its silver jubilee. Yet it did not make a good platform for a new idea of regression. Fisher (1925a, page 16) had discovered that the same few distributions turn up "again and again," and his book consists of a few tables each prefaced by a chapter surveying its many uses. The idea of regression appears in the chapter on the t-distribution and the regression goodness of fit test appears in the chapter on the z-distribution. The methods are not documented; "references" are

listed, but, apart from the data sources, not referred to. The crucial papers (Fisher, 1924–1928 and 1925c) were not published in time for the first edition and are not listed. Fisher's early readers had to discover for themselves that his regression a's and b's are least squares/maximum likelihood values; only after 1934 was there a historical note (Section 5) mentioning Gauss, least squares and maximum likelihood. Perhaps Fisher was responding to grumbles like Schultz's (1929, page 86): "it is to be regretted that Dr. Fisher did not see fit clearly to separate the propositions which are due to him from the general body of statistical theory." However, Dr. Fisher never presented an integrated account of his new methods and the theory underlying them.

The new regression was crowded out of Fisher's empirical work. In 1922 he had mentioned agricultural meteorology as an application of regression and his first job at Rothamsted was analyzing historical data on yields and weather. The task may have inspired the new regression, but the main product—the orthogonal polynomials of Fisher (1921b and 1925a)—belonged to the old least squares, to fitting empirical formulae; Hald (1998, Section 25.7) placed Fisher's work in a literature that goes back to Chebyshev. In the most ambitious study, "The influence of rainfall on the yield of wheat at Rothamsted," Fisher (1925d, page 96) did not regress yield on weekly rainfall directly, but made an ingenious use of orthogonal polynomials in a discrete approximation to a continuous time formulation in which yield depends on the entire past rainfall record. The regressors are time trends!

"Studies in crop variation. I" (Fisher, 1921b) analyzed historical data, but Study II (Fisher and Mackenzie, 1923) analyzed Fisher's own experiments, and soon observational studies were eclipsed by experiments in the work of Fisher and other statisticians. The new experimentation was not the kind familiar to Merriman or Pearson, because now randomization was involved. Fisher's work on experiments did not affect his regression theory—it was already done and the later conditional inference theory owed nothing to experiments—but there was probably an influence the other way. The randomized experiment setup resembles Pearsonian regression, with the statistician randomizing rather than nature. The analysis of variance in Study II is fixed x analysis.

The new regression went forward without further contributions from Fisher. In econometrics, the field where regression was most used, a practical synthesis of regression and least squares had been proceeding independently with Tolley and Ezekiel (1923) and others applying least squares algorithms to Pearson–Yule regression. Ezekiel's (1930) standard work, *Methods of Correlation Analysis*, written at the end of the decade, taught Fisher's methods, and the first adequate account of the new regression theory appeared in Koopmans' (1937) *Linear Regression Analysis of Economic Time Series*.

6. REGRESSION OLD AND NEW

Pearson did not visibly react to the new regression. He continued to publish on frequency surfaces (Pearson, 1923), although the grand theory projected long before never materialized. More surprisingly, he started working in the vein of Student (1908b) and Fisher (1915). His first contributions (Pearson, 1925, 1926) gave Gosset what he had asked Fisher for—the marginal distribution of b for the bivariate normal. By the early 1930s there was a complete account of the exact distribution of the statistics for the multinormal distribution—statistics introduced in the 1890s; Fisher, Wishart and Bartlett (Wishart's first mathematical postgraduate student) all contributed.

The results were consolidated in Bartlett's (1933a) "On the theory of statistical regression." Part I surveyed the statistics associated with the multivariate normal distribution. Although Bartlett was born into the new regression, he was not satisfied with the treatment in Fisher (1925c): "[Fisher] seems to suggest that... his test holds under somewhat wider conditions than he assumed." Part II considered which of the results survive if all that is normal is the conditional distribution of one of the variables. Crucial to the analysis were factorizations of the joint distribution. Among numerous results, Bartlett (1933a, page 278) showed that the t-test of significance of b is "valid, with no restrictions on x." Bartlett had shown how Fisher's regression theory could be integrated with the Pearson regression, crossing all of the t's. Sampson (1974) presented Bartlett's results for the multinormal distribution in modern notation, although curiously his tale of two regressions does not include Bartlett's interest in integrating them.

Pearson (1934, page li) eventually conceded the goodness of fit point, writing that Fisher's test applies whether the array totals "are kept the same or vary in a random manner." However, he (Pearson, 1931, pages cxxxii–cxl) gave a very negative evaluation of Student's t-work and did not mention the

extension to regression. He finally engaged Fisherian regression—without mentioning Fisher—in a very long comment on Welch (1935) and Kołodziejczyk (1935). They had applied the test theories of Neyman and E. S. Pearson (1928 and 1933), and used Fisher's regression results. Welch was explicitly concerned with fixed x *regression* (his y and x have a joint distribution), while Kołodziejczyk's "linear hypothesis" belongs with Neyman (1934) in descending from Markov's statement of the theory of errors, although with normality restored. Pearson (1935) argued that the generality of the "Welch–Kołodziejczyk frequency surface"—the frequency surface underlying the normal linear regression specification—is illusory because the only important case is the bivariate normal and that is best treated *without* using the Fisher–Student apparatus.

7. ESTIMATION: POPULATION, INFORMATION AND SUFFICIENCY

Fisher's first justification for fixing x was the distribution theory he produced for the tests proposed or inspired by Pearson and Student. His later justifications derived from his own "theory of estimation." Fisher worked on this theory while he worked on fixed x regression. Originally the two did not fit, but eventually he produced a conditional inference theory in which they did. Fisher did not develop a conditional theory to solve the regression problem, but his sense of the rightness of the regression practice may have guided his thinking. However, it cannot be seen in what he wrote.

The theory of estimation is more a theory of the *information* that estimation, in its usual sense, exploits. On the mathematical foundations of theoretical statistics" (Fisher, 1922b, page 311) describes the statistician's task as the "reduction of data," ideally without loss of information. The statistician specifies a "hypothetical infinite population" to which the observed sample is referred and calculates a statistic which "should summarise the whole of the relevant information supplied by the sample." This is the supreme "criterion of sufficiency" (Fisher, 1922b, page 316): when such a statistic is found, "the problem of estimation is completely solved" (Fisher, 1922b, page 315). See Aldrich (1997, pages 171–173) for an account of the paper and its criteria of estimation.

The application of sufficiency to regression was problematic. In the regression paper, Fisher (1922a, page 598) reflected on his handling of the randomness of n (see Section 3 above):

> [We] have not attempted to eliminate known quantities, given by the sample, from the distribution formulae of the statistics studied, but only the unknown quantities—parameters of the population from which the sample is drawn—which have to be estimated somewhat inexactly from the given sample.

A footnote ties the point to the "problem of estimation":

> Statistics whose sampling distribution depends upon other statistics given by the sample cannot, in the strict sense, fulfil the Criterion of Sufficiency. In certain cases evidently no statistic exists, which strictly fulfils this criterion. In these cases statistics obtained by the Method of Maximum Likelihood appear to fulfil the Criterion of Efficiency; the extension of this criterion to finite samples thus takes a new importance.

Fisher's (1925b) "Theory of statistical estimation" extended "efficiency" to finite samples—measured by the information in the statistic's sampling distribution—but it did nothing about the ineligibility due to the use of a conditional distribution. In his note to the 1950 reprint, Fisher described the Theory as "more compact and businesslike" than the foundations; it was, because it shelved many of the problems. Bartlett's notion of "quasi-sufficiency" (see Section 9 below) better addressed the regression difficulty.

Fisher applied efficiency and consistency—the two lesser criteria—to regression in an unpublished critique (Fisher, 1924–1925) of Campbell's (1924) alternative to least squares, a variant of the method of averages (see Farebrother, 1999, pages 236–237). Fisher stated that both methods are consistent and asymptotically normal under general conditions, but that least squares is more efficient. If the errors in y are normally distributed, "it may be shown" that the estimate b has 100% efficiency. In the 1922 theory, "efficiency" is a large sample property delivered by maximum likelihood and "showing" presumably used the fact that least squares is maximum likelihood in fixed or random x situations. Fisher gave two examples that quantify the inefficiency of Campbell's method. In the first, illustrative of experimental work, the x is in a (nonstochastic) arithmetic progression; in the other, illustrative of "observational studies," x is normally distributed. The second analysis is curious because Fisher does *not* condition on x when he calculates the variance of the estimator.

The *infinite* in "hypothetical infinite population" was criticized by William Burnside and Fisher (1925b, page 700) offered a clarification. Only in the 1950s (see Section 11) did he press himself to clarify the *hypothetical*. He (Fisher, 1922b, page 313) had written, "any such set of numbers [observations] are a random sample from the totality of numbers produced by the same matrix of causal conditions." Naturally any hypothesized population had to face a "rigorous and objective test of the adequacy with which the proposed population represents the whole of the available facts," but that was the end of it.

Some students had paused over the hypotheticalness of the regression population. Working and Hotelling (1929, page 82), who made the first extension to Fisher's regression t-results, were fitting time trends by least squares:

> The fiction is conventionally adopted that the sampling might be repeated indefinitely with new and independent values of the random part of y, but with the same fundamental trend.

Koopmans (1937, pages 1–8) discussed the interpretation of the fixed x population. He sent his book to Fisher, but they seem not to have discussed the regression population. Koopmans was to have a strong influence on econometric thinking on the subject, but that is another story; see Aldrich (1993).

8. ANCILLARY INFORMATION

Ancillarity reconciled regression with the theory of estimation. Ancillarity had been trailed in Fisher (1925b, page 724), but it only became prominent in "Two new properties of mathematical likelihood" (Fisher, 1934) and "The logic of inductive inference" (Fisher, 1935); see Hinkley (1980a, b) and Hald (1998, pages 729–733) for discussion. The help an ancillary provides is in reducing the "loss of accuracy" associated with the use of a single estimate; the loss is the difference between the information in the entire sample and in the estimate. The ancillaries that materialized in 1934 were for the location and location/scale families. For the location case, Fisher showed how conditioning on the "configuration" leads to the full recovery of the information lost; the configuration is the set of $n-1$ differences between the median and the other observations.

Fisher's practice was to work through "trivial but representative" (Fisher, 1956, page 158) problems, without proving or even stating precisely any theorem of which they are representative instances. To help fix these notions of loss and recovery I present an argument which underlies much that he wrote, but which he seems never to have written down. The formulation is from Kalbfleisch (1982, page 78).

The information in the sample X is

$$I_X(\theta) = -E\frac{\partial^2 \ln f_X(x;\theta)}{\partial\theta^2}.$$

In the case of interest there is no single sufficient statistic. If T is the maximum likelihood estimator of θ, then $I_T(\theta)$, the information in T, calculated from the sampling distribution of T, will be less than that in the sample, $I_X(\theta)$. Fisher calls the information "lost" in using T rather than X the difference.

Suppose there is a statistic A (for ancillary) such that (T, A) is jointly sufficient for θ and the distribution of A is free from θ. Consider now the information in the conditional distribution of T given the realized value of A,

$$\begin{aligned} I_{T|A=a}(\theta) &= -E\left[\frac{\partial^2 \ln f_{T|A}(t;x;\theta)}{\partial\theta^2}\Big|A=a\right] \\ &= -E\left[\frac{\partial^2 \ln f_{T,A}(t;x;\theta)}{\partial\theta^2}\Big|A=a\right], \end{aligned}$$

since $f_A(a)$, the density of A, is free from θ.

Average these conditional informations across A and use the joint sufficiency of $(T,\ A)$ to obtain the information measure for the sample,

$$EI_{T|A}(\theta) = I_{T,A}(\theta) = I_X(\theta).$$

Fisher's (1934, page 303) gloss is that "the process of taking account of the distribution of our estimates in samples of the particular configuration [A for the location problem] observed has therefore recovered the whole of the information available."

Fisher (1935, page 48) emphasizes two further points: ancillary statistics tell us nothing about the value of the parameter; their function is to tell us what "reliance" to place on the estimate. Regression is not mentioned, but the ideas seem obviously applicable and indeed Bartlett applied them. Fisher also initiated a second life for ancillarity with an example showing that ancillarity is "useful not only in questions of estimation proper" (Fisher, 1935, page 78). This is the test for independence in the 2×2 table, obtained by conditioning on the margins (Fisher, 1935, page 48):

> If it be admitted that these marginal frequencies by themselves supply no information... as to the proportionality of the frequencies in the body of the table we may recognize the information they supply as wholly ancillary; and therefore recognize that we are concerned only with the relative probabilities of occurrence of the different ways in which the table may be filled in, subject to these marginal frequencies.

Information is mentioned here but not loss and recovery. Fisher eventually applied to regression both this recovery-free notion of ancillarity (and information) and the original notion; see Sections 10 and 11 below.

9. QUASI-SUFFICIENCY AND STATISTICAL REGRESSION

For the next few years Bartlett was writing more about conditional inference than Fisher; see Fraser (1992) for a brief review. The first paper, "Statistical information and properties of sufficiency" (Bartlett, 1936), is the most relevant to regression. Bartlett took Fisher's idea away from "the theory of estimation," because he did not share Fisher's "reduction of data" viewpoint, or his enthusiasm for information calculations in small-sample work. Relations between Fisher and Bartlett were variable; sometimes they were in accord but more often not (see Bartlett, 1965, 1982, Olkin, 1989 and Zabell, 1992, pages 377–378).

Bartlett (1936, page 131) restates Fisher's analysis of the location problem: the distribution of each item in the sample S is of the form $f(x-m)$, the chance of a configuration C is independent of the parameter m and there is a T such that

$$p(S|m) = p(S|C,m)p(C) = p(T|C,m)p(C).$$

"Hence all the information on m is given by $T|C$." Such estimates as T are called *quasi-sufficient* statistics by analogy with the factorization condition for sufficiency.

Bartlett thus made explicit the scheme implicit in Fisher (1934)—or rather half explicit; the reader has to define quasi-sufficiency in general. That done, Bartlett (1936, page 135) pointed out:

> The important practical illustration of the use of *quasi*-sufficient statistics occurs in the theory of statistical regression. In the simplest case [σ^2 known] our estimate b_{yx} of the coefficient β_{yx} is accompanied by a specification of the value of $\sum(x-\overline{x})^2$ obtained, the distribution of $b_{yx}|\sum(x-\overline{x})^2$ being normal (for normal y), whatever the distribution of x.

Behind the correspondence between $(b_{yx}, \sum(x-\overline{x})^2)$ and (T,C) is a certain amount of calculation which is not given; it draws on the factorization analysis of the 1933 paper on statistical regression (see Section 6 above). I do not think the sufficiency of b in the fixed x regression model had been noted before.

Bartlett's later papers show the influence of Neyman and Pearson (1933) as well as of Fisher. As Fraser (1992, page 110) noted with regard to the 1937 paper, Bartlett (1937) "uses concepts and theory from both the Fisher and Neyman–Pearson schools in a manner that might now be called unified." Bartlett had an appetite for unification: earlier he (Bartlett, 1933b) had tried to explain Fisher and Jeffreys to one another. [For the relationship between Fisher and Jeffreys, see Howie (2002) and Aldrich (2005).] Bartlett returned to quasi-sufficiency and the regression example after Welch (1939) identified a conflict between conditioning and power. Welch (1939, page 66) had concluded "that certain methods, for which properties analogous to those of sufficiency have been claimed, do not satisfy conditions which I think they should, if these claims are to be upheld." [See Fraser (2004) for a recent discussion of Welch's argument.] The "claims" were Fisher's, the "conditions" related to Neyman and Pearson's power, but Bartlett felt the criticism.

Bartlett (1940, page 392) did not directly defend conditioning, but turned the objection by showing how, if the size of the test is varied with the value of the ancillary, conditional tests can achieve maximum power for a given unconditional (long run) size. He illustrates as follows:

> The orthodox theory is to consider the conditional statistic $b|\sum(x-\overline{x})^2$.... Suppose for the sake of argument that the true variance of... y_x was known to be unity, and the x's are such that $\sum(x-\overline{x})^2 = 1$ on Mondays and 1.44 on Tuesdays. Then for an 0.025 significance level (one tail), the usual practice would be to take 1.96 as the significance level for b (from $b_0 = 0$) on Mondays, and $1.96/1.2 = 1.633$ on Tuesdays. The power of the test in relation to the alternative that $b_1 = 3.92$ is 0.9860.

> But if we were satisfied with adjusting the significance level to be 0.025 merely in the long run for Mondays and Tuesdays together, we may raise the power of the test to its maximum value of 0.9878 by taking the Monday significance level at $b = 1.87$ ($\alpha = 0.0307$) and the Tuesday level at $b = 1.723$ ($\alpha = 0.0194$).

Bartlett returned to regression and conditioning in an obituary of E. S. Pearson. He (Bartlett, 1981, page 3) mentioned a "formidable logical criticism" of the concept of power: in regression the conditional power for a test about β_{yx} depends on $\sum(x - \overline{x})^2$ and "we usually consider it irrelevant to ask whether we can obtain a better procedure based on 'absolute power' by considering the sampling variation of $\sum(x - \overline{x})^2$."

In the 1930s neither Fisher nor Bartlett articulated Birnbaum's (1962, page 271) "conditionality principle," although Bartlett wrote as though he accepted it. Fisher was more equivocal: he recognized a world beyond "questions of estimation proper," but information extraction came first.

10. REGRESSION AND ANCILLARY INFORMATION

Fisher may well have regarded the application of ancillarity to regression as obvious; he first mentions it in a 1939 letter to Jeffreys (see Bennett, 1990, page 173):

> I regard regression work... as a good example of ancillary information, in that the precision of the regression does not really depend on the number in the sample, but only on the sum of squares of the independent variate, or, in general, on the dispersion.... In fact the whole work is completely independent of how they may be distributed in the population sampled....

Fisher made the same point to Darmois in August 1940. In an earlier letter, Fisher (see Bennett, 1990, page 70) had criticized Bartlett's use of the phrase "conditional sufficiency." Bartlett had not actually used the term, although years later Cox and Hinkley (1974, page 32) did. Fisher never referred to Bartlett's treatment of regression, but he must have been aware of it.

The Fisher–Darmois correspondence (see Bennett, 1990, pages 65–79) is particularly rich and many of the ideas sketched there went into "Conclusions fiduciaires" (Fisher, 1948) and then into *Statistical Methods and Scientific Inference*. One of the innovations of "Conclusions fiduciaires" was a new definition of ancillarity. Instead of implicitly defining a ancillary by its role in recovering lost information, Fisher (1948, page 193) defined it as a variation-free statistic:

> Tout ensemble de statistiques dont la distribution simultanée est indépendante des paramètres, est appelé un ensemble "ancillaire" des statistiques.

In effect this was the Bartlett (1936) definition and presumably it was what Fisher (1935) had in mind when he wrote that marginal frequencies "supply no information."

In "Conclusions fiduciaires," the technique of the theory of estimation was applied to regression. The paper's second example (Fisher, 1948, page 197) considered the bivariate normal regression model with known variance σ^2. In 1922 Fisher had written that the least squares estimator b is normally distributed with variance σ^2/A, where A is the sum of squared deviations of x. He now supposed that x is normally distributed with known variance α, so that A is α times a χ^2 with $N - 1$ degrees of freedom. With this specification the marginal distribution of b is a noncentral t with $N - 1$ degrees of freedom with parameters which are functions of the known quantities α and σ^2 and the unknown β. Fisher (1921a, page 21) stated that there is less information in this unconditional distribution than in the normals of which it is a mixture. By ignoring the value of A and using the value of α and the sample size N, the information has been reduced in the ratio $N/N + 2$.

Fisher obtained this value by applying results from his first example, which was itself based on a weighting argument that went back to 1925. However, an argument can be based on the "implicit theorem" of Section 8 above, identifying β with θ and (b, A) with (T, A). The information in a sample of size N conditional on the value of A is A/σ^2. Taking the expectation of these conditional informations yields $\alpha(N - 1)/\sigma^2$ as the information in the entire sample. If now we compute the information in b from its marginal distribution, we obtain a smaller value: the information in the sample is reduced in the ratio $N/N + 2$.

These information calculations—unlike most of the paper—did not find their way into *Statistical Methods and Scientific Inference*. Indeed regression does not appear in the estimation chapter, but in the chapter on "Misapprehensions about tests of significance."

11. A MULTIPLICITY OF POPULATIONS

Fisher's campaign against the Neyman–Pearson theory of testing and the notion of repeated sampling from a fixed population was a reply to criticisms of his treatment of the 2×2 table and of the Behrens–Fisher problem. Thus on the latter he (Fisher, 1946) wrote—rather surprisingly—against Bartlett:

> I am quite aware that Bartlett, following Neyman, feels bound to identify the population of samples envisaged in tests of significance with those generated by repeated sampling of a fixed hypothetical population. . . .

In the polemics of the 1950s Fisher argued that his treatment must be correct because it follows the regression pattern which everyone knows is correct. The controversy was over testing and the theory of estimation aspect recedes from attention.

Fisher's article, "Statistical methods and scientific induction," and his book, *Statistical Methods and Scientific Inference*, stress the *hypotheticalness* of the statistician's population. The root difficulty with the formula, "repeated sampling from the same population," is that there is "a multiplicity of populations to each of which we can regard our sample as belonging" (Fisher, 1955, page 71). In an "acceptance sampling" (quality control) situation the population has an "objective reality," but in the natural sciences the population is a "product of the statistician's imagination" and "the first to come to mind may be quite misleading" (Fisher, 1956, pages 77 and 78). Fisher was criticizing Neyman, but his own formulation in the "foundations" (see Section 7 above) was as vulnerable.

Setting up the regression model, Fisher (1955, page 71) stated "the qualitative data may also tell us how x is distributed with or without specific parameters; this information is irrelevant." He (Fisher, 1956, page 72) continued:

> The normal distribution of b about β with variance σ^2/A does not correspond with any realistic process of sampling for acceptance but to a population of samples in all relevant respects like that observed, neither more precise nor less precise, and which therefore we think it appropriate to select in specifying the precision of the estimate b. In relation to the value of β the value A is known as an *ancillary statistic*.

However, there is no appeal to information calculations.

In the book, Fisher does not use the word "ancillary" with regression, perhaps to make the attack on repeated sampling less dependent on the theory of estimation. He (Fisher, 1956, page 82) presented regression, adding the fiducial distribution of β and this pointed introduction:

> A case which illustrates well how misleading the advice is to base the calculations on repeated sampling from the same population, if such advice were taken literally, is that of data suitable for the estimation of a coefficient of linear regression.

The regression material appears in the book's chapter on misapprehensions about significance tests and the material is organized around the t-distribution, that is, Fisher's first regression theory, where it is shown that the distribution is not affected by the distribution of x. However, the "advice" for testing and fiducial inference resulting from failure to condition is not going to be "misleading"; it is going to be the same.

For Bartlett (1940) and for Fisher (1948), conditioning had to be related to power or to information. In 1956, Fisher (1956, page 84) made a more direct appeal:

> To judge of the precision of a given value of b, by reference to a mixture of samples having different values of A, and therefore different precisions for the values of b they supply, is erroneous because these other samples throw no light on the precision of that value which we have observed.

This is an eloquent amplification of the 1922 proposition: "the accuracy of the regression coefficients is only affected by the correlations which appear *in the sample*."

Fisher wrote about fixed x regression over a period of more than 30 years. He produced three justifications: from t distribution theory, from the theory of estimation and from the application of the conditionality principle to the choice from the multiplicity of populations. He saw these justifications not as alternatives but as reinforcing each other.

12. RETROSPECTS

The story of the Gauss–Pearson–Fisher triangle and the reconception of regression presents some paradoxes. There is Seal's (1967, page 16) tribute to

Pearson, who "must... be given the credit for extending the Gauss linear model to a much broader class of problems than those of errors of measurement." Such an acknowledgment to the model's greatest critic became possible only because Fisher moved Pearson's regression, with its applications, into the orbit of Gauss' model. Fisher's contribution even got short measure from Fisher! In 1956 he (Fisher, 1956, page 84) did not even mention it:

> [I]n repeated sampling from the bivariate distribution of x and y, the value of A would vary from sample to sample. The distribution of $(b - \beta)$ would no longer be normal, and before we knew what is was, the distribution of A, which in turn depends on that of x would have to be investigated. Indeed, at an early stage Karl Pearson did attempt the problem of the precision of a regression coefficient in this way, assuming x to be normally distributed. The right way had, however been demonstrated many years before by Gauss, and his method only lacked for completeness the use of 'Student"s distribution, appropriate for samples of rather small numbers of observations.

In the Foreword Fisher (1956, page 3) remarked that Pearson cared little for the past, instancing the "Gaussian tradition of least square techniques." Yet when Fisher was pulling regression into that tradition, Gauss was not to be seen; only Fisher's first (preregression) paper (Fisher, 1912) has working references to its literature.

In 1956 Fisher thought fixed x regression beyond dispute. It was clearly a presence, although to investigate its standing would be a project in itself. Section 6 gave some views from the Rothamsted/University College "inside" and I will add some examples from outside to illustrate further possibilities. Hotelling was a born-again Fisherian of the 1920s, an early contributor to regression t-theory and ought to have been an insider. Yet when he (Hotelling, 1940, pages 276–277) weighed the merits of the fixed x and joint multinormality assumptions in the regressor/predictor selection problem, he did not consider Fisher's distribution argument:

> The advantages of exactness and of freedom from the somewhat special trivariate normal assumption are obtained at the expense of sacrificing the precise applicability of the results to other sets of values of the predictors.

The attitude recalls Yule on normal theory probable errors (Section 1 above): they are not perfect, but they are all—or the best—there is. Fixed x regression had not yet established itself as doing what comes naturally. In the 1950s there was much soul-searching about the treatment of relationships between variables (see, e.g., Berkson, 1950 and Kendall, 1951), and in these discussions the models and techniques favored by Fisher had no special ascendancy.

Cramér (1946) noticed the distribution theory argument or at least part of it. His Chapter 29 considers regression inference for the multinormal distribution case and Chapter 37 discusses regression with nonrandom x's. Cramér (1946, page 550) recorded the "formal identity" of the t-results in the two cases: he noticed Part I of Bartlett (1933a), but did not mention the results in Part II.

The conditioning arguments were less visible and less noticed. The underlying theory of estimation was not accepted, understood or even widely known. Hotelling (1948, page 867) complained after reading Kendall's (1946) *Advanced Theory*, "it is still not clear what the statistician is supposed to do with ancillary statistics." Before 1955–1956 the regression applications were footnotes, not headlines; thus the earliest reference in Barndorff-Nielsen's (1978, page 36) historical note on regression in relation to ancillarity is to Fisher (1956). Modern high theory views on conditioning in regression (see Barndorff-Nielsen and Cox, 1994, page 39 and Gelman et al., 1995, page 235) are linked to Fisher and to Bartlett. Thus Cox (1958, page 360) restated Fisher's point about the multiplicity of populations and his weighing machine example is clearly a parable for regression: his references include Bartlett (1940); see Reid (1994) for the background to this paper. Savage (1962, page 19) referred to Cox when he gave his Bayesian view of ancillarity and regression. Of course, the general topic of ancillarity and conditional inference has received plenty of attention in recent years (see, e.g., the review by Reid, 1995) and Brown (1990) has reopened the question of conditioning in regression.

Although a subject was built around t- and F-tests, the fixed x assumption was not a central issue in Anglo-American statistics. However, from the 1930s into the 1970s econometricians made a profession out of *not* fixing x—with errors in variables and simultaneity; see Morgan (1990) for an account. Naturally they discussed whether the fixed x practice could pass (see Aldrich, 1993), but their discussions did not influence

the literature treated here. In their discussions the possible *causal* nature of the relationship between y and x was also an issue. From Yule onwards regression was used for investigating causal relationships but in the statistics tradition the causal interest was not intrinsic to the statistical analysis but something apart.

ACKNOWLEDGMENTS

This paper is a revised version of a Southampton Department of Economics discussion paper, "The origins of fixed X regression" (2000). I am grateful for advice and suggestions to the editors and referees who have looked at it.

REFERENCES

Fisher's published papers appear in J. H. Bennett, ed. (1971–1974). *Collected Papers of R. A. Fisher*, 5 vols. Adelaide Univ. Press. Bennett (1990) and nearly all of the papers referred to here are available from the University of Adelaide R. A. Fisher Digital Archive at http://www.library.adelaide.edu.au/ual/special/fisher.html.

ALDRICH, J. (1993). Cowles exogeneity and CORE exogeneity. Discussion Paper 9308, Dept. Economics, Southampton Univ.

ALDRICH, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statist. Sci.* **10** 364–376.

ALDRICH, J. (1997). R. A. Fisher and the making of maximum likelihood 1912–1922. *Statist. Sci.* **12** 162–176.

ALDRICH, J. (1998). Doing least squares: Perspectives from Gauss and Yule. *Internat. Statist. Rev.* **66** 61–81.

ALDRICH, J. (1999). Determinacy in the linear model: Gauss to Bose and Koopmans. *Internat. Statist. Rev.* **67** 211–219.

ALDRICH, J. (2003–2005). A guide to R. A. Fisher. Available at http://www.economics.soton.ac.uk/staff/aldrich/fisherguide/rafreader.htm.

ALDRICH, J. (2003). The language of the English biometric school. *Internat. Statist. Rev.* **71** 109–129.

ALDRICH, J. (2005). The statistical education of Harold Jeffreys. *Internat. Statist. Rev.* **73** 289–308.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

BARTLETT, M. S. (1933a). On the theory of statistical regression. *Proc. Royal Soc. Edinburgh* **53** 260–283.

BARTLETT, M. S. (1933b). Probability and chance in the theory of statistics. *Proc. Roy. Soc. London Ser. A* **141** 518–534.

BARTLETT, M. S. (1936). Statistical information and properties of sufficiency. *Proc. Roy. Soc. London Ser. A* **154** 124–137.

BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. London Ser. A* **160** 268–282.

BARTLETT, M. S. (1940). A note on the interpretation of quasi-sufficiency. *Biometrika* **31** 391–392.

BARTLETT, M. S. (1965). R. A. Fisher and the last fifty years of statistical methodology. *J. Amer. Statist. Assoc.* **60** 395–409.

BARTLETT, M. S. (1981). Egon Sharpe Pearson, 1895–1980. *Biometrika* **68** 1–7.

BARTLETT, M. S. (1982). Chance and change. In *The Making of Statisticians* (J. Gani, ed.) 42–60. Springer, New York.

BENNETT, J. H., ED. (1990). *Statistical Inference and Analysis*: *Selected Correspondence of R. A. Fisher*. Oxford Univ. Press.

BERKSON, J. (1950). Are there two regressions? *J. Amer. Statist. Assoc.* **45** 164–180.

BIRNBAUM, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.* **57** 269–326.

BJERVE, S. and DOKSUM, K. A. (1993). Correlation curves: Measures of association as functions of covariate values. *Ann. Statist.* **21** 890–902.

BLAKEMAN, J. (1905). On tests for linearity of regression in frequency distributions. *Biometrika* **4** 332–350.

BLYTH, S. (1994). Karl Pearson and the correlation curve. *Internat. Statist. Rev.* **62** 393–403.

BOWLEY, A. L. (1901). *Elements of Statistics*. King, London.

BOX, J. F. (1978). *R. A. Fisher*: *The Life of a Scientist*. Wiley, New York.

BROWN, L. D. (1990). An ancillarity paradox which appears in multiple linear regression (with discussion). *Ann. Statist.* **18** 471–538.

BRUNT, D. (1917). *The Combination of Observations*. Cambridge Univ. Press.

CAMPBELL, N. (1924). The adjustment of observations. *Philosophical Magazine* (*6*) **47** 816–826.

COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29** 357–372.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press, Princeton, NJ.

EDGEWORTH, F. Y. (1893). Exercises in the calculation of errors. *Philosophical Magazine* (*5*) **36** 98–111.

EISENHART, C. (1979). On the transition from 'Student's z' to 'Student's t.' *Amer. Statist.* **33** 6–10.

ELDERTON, W. P. (1906). *Frequency Curves and Correlation*. Layton, London.

EZEKIEL, M. (1930). *Methods of Correlation Analysis*. Wiley, London.

FAREBROTHER, R. W. (1999). *Fitting Linear Relationships*: *A History of the Calculus of Observations*. Springer, New York.

FIENBERG, S. E. (1980). Fisher's contribution to the analysis of categorical data. *R. A. Fisher*: *An Appreciation. Lecture Notes in Statist.* **1** 75–84. Springer, New York.

FIENBERG, S. E. and HINKLEY, D. V., EDS. (1980). *R. A. Fisher*: *An Appreciation. Lecture Notes in Statist.* **1**. Springer, New York.

FISHER, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics* **41** 155–160.

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. (1921a). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron* **1** 3–32.

FISHER, R. A. (1921b). Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *J. Agricultural Science* **11** 107–135.

FISHER, R. A. (1922a). The goodness of fit of regression formulae, and the distribution of regression coefficients. *J. Roy. Statist. Soc.* **85** 597–612.

FISHER, R. A. (1922b). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368.

FISHER, R. A. (1922c). On the interpretation of χ^2 from contingency tables, and the calculation of P. *J. Roy. Statist. Soc.* **85** 87–94.

FISHER, R. A. (1924–1925). Note on Dr. Campbell's alternative to the method of least squares. Unpublished manuscript, Barr Smith Library, Univ. Adelaide.

FISHER, R. A. (1924–1928). On a distribution yielding the error functions of several well known statistics. In *Proc. International Mathematical Congress* **2** 805–813. Univ. Toronto Press, Toronto.

FISHER, R. A. (1925a). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

FISHER, R. A. (1925b). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.* **22** 700–725.

FISHER, R. A. (1925c). Applications of 'Student's' distribution. *Metron* **5** 90–104.

FISHER, R. A. (1925d). The influence of rainfall on the yield of wheat at Rothamsted. *Philos. Trans. Roy. Soc. London Ser. B* **213** 89–142.

FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc. London Ser. A* **144** 285–307.

FISHER, R. A. (1935). The logic of inductive inference (with discussion). *J. Roy. Statist. Soc.* **98** 39–82.

FISHER, R. A. (1946). Testing the difference between two means of observations of unequal precision. *Nature* **158** 713.

FISHER, R. A. (1948). Conclusions fiduciaires. *Ann. Inst. H. Poincaré* **10** 191–213.

FISHER, R. A. (1955). Statistical methods and scientific induction. *J. Roy. Statist. Soc. Ser. B* **17** 69–78.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh.

FISHER, R. A. and MACKENZIE, W. A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *J. Agricultural Science* **13** 311–320.

FRASER, D. A. S. (1992). Introduction to reprint of "Properties of sufficiency and statistical tests" [Bartlett (1937)]. In *Breakthroughs in Statistics* (S. Kotz and N. L. Johnson, eds.) **1** 109–112. Springer, New York.

FRASER, D. A. S. (2004). Ancillaries and conditional inference (with discussion). *Statist. Sci*. **19** 333–369.

GALTON, F. (1877). Typical laws of heredity. *Nature* **15** 492–495, 512–514, 532–533.

GALTON, F. (1886). Family likeness in stature. *Proc. Roy. Soc. London* **40** 42–73.

GAUSS, C. F. (1809/1963). *Theoria Motus Corporum Coelestium* (C. H. Davis, transl.). Dover, New York, reprinted 1963.

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.

HALD, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley, New York.

HALD, A. (1999). On the history of maximum likelihood in relation to inverse probability and least squares. *Statist. Sci*. **14** 214–222.

HINKLEY, D. V. (1980a). Theory of statistical estimation: The 1925 paper. *R. A. Fisher*: *An Appreciation. Lecture Notes in Statist.* **1** 85–94. Springer, New York.

HINKLEY, D. V. (1980b). Fisher's development of conditional inference. *R. A. Fisher*: *An Appreciation. Lecture Notes in Statist.* **1** 101–108. Springer, New York.

HOOKER, R. H. (1907). Correlation of the weather and crops. *J. Roy. Statist. Soc.* **70** 1–51.

HOTELLING, H. (1940). The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. *Ann. Math. Statist.* **11** 271–283.

HOTELLING, H. (1948). Review of *The Advanced Theory of Statistics* **2**, by M. G. Kendall. *Bull. Amer. Math. Soc.* **54** 863–868.

HOWIE, D. (2002). *Interpreting Probability*: *Controversies and Developments in the Early Twentieth Century*. Cambridge Univ. Press.

KALBFLEISCH, J. (1982). Ancillary statistics. *Encyclopedia of Statistical Sciences* **1** 77–81. Wiley, New York.

KENDALL, M. G. (1946). *The Advanced Theory of Statistics* **2**. Griffin, London.

KENDALL, M. G. (1951). Regression, structure and functional relationship. I. *Biometrika* **38** 11–25.

KOŁODZIEJCZYK, S. (1935). On an important class of statistical hypotheses. *Biometrika* **27** 161–190.

KOOPMANS, T. C. (1937). *Linear Regression Analysis of Economic Time Series*. Bohn, Haarlem, Netherlands.

LANCASTER, H. O. (1969). *The Chi-Squared Distribution*. Wiley, New York.

LEHMANN, E. L. (1999). 'Student' and small-sample theory. *Statist. Sci.* **14** 418–426.

MCMULLEN, L. (1970). *Letters from W. S. Gosset to R. A. Fisher 1915–1936*: *Summaries by R. A. Fisher with a Foreword by L. McMullen*, 2nd ed. Printed by Arthur Guinness for private circulation and placed in a few libraries.

MERRIMAN, M. (1884/1911). *A Textbook on the Method of Least Squares*. Wiley, New York. References are to the eighth edition, 1911.

MILLER, J., ED. (1999–2005). Earliest uses of symbols in probability and statistics. Available at http://members.aol.com/jeff570/stat.html.

MORGAN, M. S. (1990). *The History of Econometric Ideas*. Cambridge Univ. Press.

NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection (with discussion). *J. Roy. Statist. Soc.* **97** 558–625.

NEYMAN, J. and PEARSON, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. I, II. *Biometrika* **20A** 175–240, 263–294.

NEYMAN, J. and PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. Roy. Soc. London Ser. A* **231** 289–337.

OLKIN, I. (1989). A conversation with Maurice Bartlett. *Statist. Sci.* **4** 151–163.

PEARSON, E. S. (1926). Review of *Statistical Methods for Research Workers*, by R. A. Fisher. *Science Progress* **20** 733–734.

PEARSON, E. S. (1990). '*Student*', *A Statistical Biography of William Sealy Gosset* (R. L. Plackett, ed.; G. A. Barnard, assist.). Oxford Univ. Press.

PEARSON, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philos. Trans. Roy. Soc. London Ser. A* **186** 343–414.

PEARSON, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos. Trans. Roy. Soc. London Ser. A* **187** 253–318.

PEARSON, K. (1899). Mathematical contributions to the theory of evolution. V. On the reconstruction of the stature of prehistoric races. *Philos. Trans. Roy. Soc. London Ser. A* **192** 169–244.

PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* (5) **50** 157–175.

PEARSON, K. (1902a). On the systematic fitting of curves to observations and measurements. I, II. *Biometrika* **1** 265–303, **2** 1–23.

PEARSON, K. (1902b). On the mathematical theory of errors of judgment, with special reference to the personal equation. *Philos. Trans. Roy. Soc. London Ser. A* **198** 235–299.

PEARSON, K. (1905). On the general theory of skew correlation and non-linear regression. *Drapers' Company Research Memoirs, Biometric Series II*. Cambridge Univ. Press.

PEARSON, K., ED. (1914). *Biometrika Tables for Statisticians and Biometricians*. Cambridge Univ. Press.

PEARSON, K. (1916). On the application of 'goodness of fit' tables to test regression curves and theoretical curves used to describe observational or experimental data. *Biometrika* **11** 239–261.

PEARSON, K. (1920). Notes on the history of correlation. *Biometrika* **13** 25–45.

PEARSON, K. (1923). Notes on skew frequency surfaces. *Biometrika* **15** 222–230.

PEARSON, K. (1925). Further contributions to the theory of small samples. *Biometrika* **17** 176–200.

PEARSON, K. (1926). Researches on the mode of distribution of the constants of samples taken at random from a bivariate normal population. *Proc. Roy. Soc. London Ser. A* **112** 1–14.

PEARSON, K., ED. (1931). *Tables for Statisticians and Biometricians, Part II*. Cambridge Univ. Press.

PEARSON, K., ED. (1934). *Tables of the Incomplete Beta-Function*. Cambridge Univ. Press.

PEARSON, K. (1935). Thoughts suggested by the papers of Messrs. Welch and Kołodziejczyk. *Biometrika* **27** 227–259.

PEARSON, K. and FILON, L. N. G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philos. Trans. Roy. Soc. London Ser. A* **191** 229–311.

REID, N. (1994). A conversation with Sir David Cox. *Statist. Sci.* **9** 439–455.

REID, N. (1995). The roles of conditioning in inference (with discussion). *Statist. Sci.* **10** 138–157, 173–196.

SAMPSON, A. R. (1974). A tale of two regressions. *J. Amer. Statist. Assoc.* **69** 682–689.

SAVAGE, L. J. (1962). Subjective probability and statistical practice. In *The Foundations of Statistical Inference*: *A Discussion* (L. J. Savage et al., eds.) 9–35. Methuen, London.

SCHULTZ, H. (1929). Applications of the theory of error to the interpretation of trends: Discussion. *J. Amer. Statist. Assoc. Suppl.* **24** 86–89.

SEAL, H. (1967). The historical development of the Gauss linear model. *Biometrika* **54** 1–24.

SENETA, E. (1988). Slutsky (Slutskii), Evgenii Evgenievich. *Encyclopedia of Statistical Sciences* **8** 512–515. Wiley, New York.

SLUTSKY, E. E. (1913). On the criterion of goodness of fit of the regression lines and on the best method of fitting them to the data. *J. Roy. Statist. Soc.* **77** 78–84.

STIGLER, S. M. (1986). *The History of Statistics*. *The Measurement of Uncertainty before 1900*. Belknap, Cambridge, MA.

STIGLER, S. M. (2001). Ancillary history. In *State of the Art in Probability and Statistics*: *Festschrift for Willem R. van Zwet* (M. deGunst, C. Klaassen and A. van der Vaart, eds.) 555–567. IMS, Beachwood, OH.

STUDENT (1908a). The probable error of a mean. *Biometrika* **6** 1–25.

STUDENT (1908b). Probable error of a correlation coefficient. *Biometrika* **6** 302–310.

STUDENT (1926). Review of *Statistical Methods for Research Workers*, by R. A. Fisher. *Eugenics Review* **18** 148–150.

TOLLEY, H. R. and EZEKIEL, M. J. B. (1923). A method of handling multiple correlation problems. *J. Amer. Statist. Assoc.* **18** 993–1003.

WELCH, B. L. (1935). Some problems in the analysis of regression among k samples of two variables. *Biometrika* **27** 145–160.

WELCH, B. L. (1939). On confidence limits and sufficiency, with particular reference to parameters of location. *Ann. Math. Statist.* **10** 58–69.

WORKING, H. and HOTELLING, H. (1929). Applications of the theory of error to the interpretation of trends. *J. Amer. Statist. Assoc. Suppl.* **24** 73–85.

YULE, G. U. (1897). On the theory of correlation. *J. Roy. Statist. Soc.* **60** 812–854.

YULE, G. U. (1899). An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades (part I). *J. Roy. Statist. Soc.* **62** 249–295.

YULE, G. U. (1907). On the theory of correlation for any number of variables, treated by a new system of notation. *Proc. Roy. Soc. London Ser. A* **79** 182–193.

YULE, G. U. (1909). The applications of the method of correlation to social and economic statistics. *J. Roy. Statist. Soc.* **72** 721–730.

YULE, G. U. (1911). *An Introduction to the Theory of Statistics*. Griffin, London.

ZABELL, S. (1992). R. A. Fisher and the fiducial argument. *Statist. Sci.* **7** 369–387.