International Statistical Review (1999), 67, 2, 211–219, Printed in Mexico

Determinacy in the Linear Model: Gauss to Bose and Koopmans

John Aldrich

University of Southampton, Southampton SO17 1BJ, UK. E-mail: jca1@soton.ac.uk

Summary

Gauss showed that least squares fails to produce a unique solution only when the problem is indeterminate. This note considers his argument and the notion of indeterminacy underlying it. It also relates the argument to twentieth-century discussions of estimability and identifiability.

Key words: Gauss; Least squares; Collinearity; Rank; Estimability; Identifiability; Fidelity.

Introduction

This note discusses a neglected aspect of Gauss's least squares theorising—his demonstration that "our principle" only fails when the problem is indeterminate. It examines Gauss's demonstration and the underlying notion of determinacy and indicates some links with twentieth century work.

Gauss applied least squares to measurement problems in astronomy and geodesy. The "Gauss linear model", as Seal (1967) called it, was developed (1809 and 1823) for use in astronomy. There is an observable vector **y** with

$$\mathbf{y} = \mu + \varepsilon;\ \mu = \mathbf{X}\beta,\ E\varepsilon = \mathbf{0},\ E\varepsilon\varepsilon' = \sigma^2\mathbf{I},$$

where **X** is known. The model for geodesy (1828) has

$$\mathbf{y} = \zeta + \varepsilon;\ \mathbf{A}\zeta = \mathbf{d},\ E\varepsilon = \mathbf{0},\ E\varepsilon\varepsilon' = \sigma^2\mathbf{I},$$

where **A** and **d** are known. In the first case the object is to estimate β, in the second to estimate ζ. I will concentrate on the astronomy model which has the non-trivial determinacy analysis.

Gauss's two expositions of the astronomy model have the same three part structure: a foreword treating the possibility of making inferences from observations to unknown quantities, a justification of the use of least squares and an afterword detailing the circumstances in which a unique solution of the normal equations cannot be obtained. Sections 1–3 below follow this plan.

Gauss's justification of least squares was Bayesian in 1809 and Gauss–Markov in 1823, to use familiar, if not quite appropriate, labels. Yet the fore- and afterwords hardly changed. My exposition combines the fuller commentary of 1809 with the more transparent mathematical argument of 1823. Gauss was a virtuoso in handling systems of equations, in work for which we would use matrices. I have put his argument into matrices and used modern regression terminology. So the formalism recalls Aitken (1934 and 1945) and the terminology recalls Fisher (1922 and 1925), who drew together the least squares and correlation/regression traditions. Seal (1967) surveys the "Gauss Linear Model" from Gauss to Fisher. Stigler (1986) reviews the least squares and correlation traditions. Further background is given by Dale (1991) and Hald (1998).

Sections 5 and 6 relate Gauss's "determinability" to two creations of the 1940s, R.C. Bose's "estimability" and T.C. Koopmans's "identifiability". Bose's work belonged to the least squares renaissance set off by Aitken and Neyman. Koopmans emphasised other models but his notion of identifiability is rather closer to Gauss's determinability. Section 4 digresses to consider an example of how indeterminacy was treated in nineteenth century empirical work.

Gauss expounded least squares in the *Theoria Motus* (1809) and the *Theoria Combinationis* (1823; 1828); note that I refer to Gauss's works by date of publication—the (1809) was written in 1806, the first part of the (1823) was read in 1821, the second part read in 1823 and the supplement (1828) read in 1826. The *Theoria Motus* was translated into English in the nineteenth century by Rear-Admiral Davis and there is a welcome new edition of the *Theoria Combinationis* with translation and notes by G.W. Stewart. I have used these translations and my page references refer to them; I have also checked the translations against those into French by Bertrand which Gauss approved—see Gauss (1855).

1 Problems Indeterminate and Determinate

Gauss (1809, pp. 253–4; cf. 1823, p. 37) begins by discussing the possibility of finding the k-vector β of unknown quantities given n observations $\mathbf{y}$ on the functions $\mathbf{f}(\beta)$ when there is error ε:

$$\mathbf{y} = \mathbf{f}(\beta) + \varepsilon.$$

"Generally speaking, determination of the unknown quantities will constitute a problem, indeterminate, determinate, or more than determinate according as $n < k$, $n = k$, or $n > k$." Note that here, as in future quotations, I have altered the author's notation so it matches that used in this note.

The terminology of "more than determinate" is connected to the yardstick of error-free observation. This is clear from the footnote where Gauss (1809, p. 254) explains how things can go wrong even in the $n > k$ case:

> If ... $n + 1 - k$ of the functions $\mathbf{f}(\beta)$ could be regarded as functions of all the others, the problem would become over-determinate relative to these functions, but indeterminate relative to β. One would not then be able to deduce the value of the latter, even if the values of the functions $\mathbf{f}(\beta)$ were known absolutely exactly; we shall however exclude this special case from our investigations.

I will refer to this passage as the 'footnote' although Gauss (1823, p.37) incorporated it into the text and applied the observation to the $n = k$ case as well. Gauss never actually said that k independent functions are needed but he clearly knew it. He doubtlessly expected the reader to return to this prologue after working through the arguments that follow, which we shall do. From now on we will replace the nonlinear functions by the linear functions $\mathbf{X}\beta$ where $\mathbf{X}$ is an $n \times k$ matrix.

2 To Inference

Gauss published two different probabilistic arguments for least squares. The arguments are too well-known to need detailing but it is worth noticing the passages linking the foreword and estimation sections. The bridge passage of 1809 (p. 254) starts from a gloss on the $n > k$ case:

> evidently, an exact representation of all the observations would only be possible when they were all absolutely free from error. And since this cannot, in the nature of things, happen, every system of values of the unknown quantities $\beta_1, \ldots, \beta_k$, must be regarded as possible, which gives the values of the functions $\varepsilon_1, \ldots, \varepsilon_n$, within the limits of the possible errors of observation; this, however, is not to imply that each of these systems would possess an equal degree of probability.

Gauss goes on to discuss reasonable properties for ε and to show how to calculate the probability that ε takes a specific value. Then by means of Bayes' theorem he established a connection between the "degree of probability" of the error ε and the posterior distribution of β for the case of a uniform prior. He also presented a justification for the normality of ε—this has been re-examined recently by Waterhouse (1990). Gauss obtained the least squares value as the maximum of the posterior density for β when the distribution of $\mathbf{y}$ is

$$\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}). \tag{2.1}$$

and the prior for β is uniform. The error variance, σ^2, is an unknown constant and is not estimated.

In 1823 Gauss changed the model and gave a new justification for least squares. He dispensed with normality and the new model is

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,\ E\varepsilon = \mathbf{0},\ E\varepsilon\varepsilon' = \sigma^2\mathbf{I}. \tag{2.2}$$

There is a new bridge between the foreword and the estimation section:

> if k is less than n, then there are infinitely many ways of reducing $\beta_1, \ldots, \beta_k$ to functions of $f_1, \ldots, f_n$ and thereby calculating their values. All these calculations would agree exactly if the observations were absolutely precise; otherwise, different expressions will give different values, and the estimates from these various combinations will have different precisions (1821, p. 37).

Gauss (1823, pp. 39–43) then showed that the least squares estimator has highest precision (smallest variance) amongst estimators of the form of $\tilde{\beta}$, where $\mathbf{G}$ is a matrix of constants,

$$\tilde{\beta} - \beta = \mathbf{G}\varepsilon.$$

Plackett (1949) found implicit in this condition the requirements that $\tilde{\beta}$ be linear in $\mathbf{y}$ and unbiased. Sprott (1979, p. 193), referring to the bridge passage, identified the force of the condition as

$$\tilde{\beta} = \beta \text{ when } \varepsilon = \mathbf{0}.$$

This interpretation as a "fidelity" requirement, a term used by twentieth century actuarial authors—see Section 5 below—seems the more natural one.

3 Normal Equations and Indeterminacy

After separating for estimation, the 1809 and 1823 accounts come together when the k normal equations are in place

$$\mathbf{d} = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\beta = \mathbf{0}. \tag{3.1}$$

Gauss (1809, p. 261; cf. 1823, pp. 50–51) asked whether "solution is always possible or whether it may give indeterminate values or be impossible". He attacked the last two cases as follows:

> As a consequence of the theory of elimination it follows that the second or third case will occur if after omitting one of the equations
>
> $$\mathbf{d} = \mathbf{0}$$
>
> one can deduce from the remaining equations an equation identical or contradictory to that which has been omitted, or, what comes to the same thing, if one can determine a linear function $\lambda'\mathbf{d}$ which is either identically zero or does not contain any of the unknowns. Thus suppose one has
>
> $$\lambda'\mathbf{d} = r.$$

Following now the later version, Gauss (1823, p. 51) supposes that

$$r = \lambda'\mathbf{d} = \lambda'(\mathbf{X}'\mathbf{X}\beta - \mathbf{X}'\mathbf{y}) = (\beta'\mathbf{X}'\mathbf{X} - \mathbf{y}'\mathbf{X})\lambda$$

for all values of β and some scalar r which does not depend on β. Equating coefficients (of each element of β) produces

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} \\ \cdots\cdots \\ \mathbf{y}'\mathbf{X} \end{pmatrix} \lambda = \begin{pmatrix} 0 \\ \cdots \\ -r \end{pmatrix}. \tag{3.2}$$

Now introduce θ defined by

$$\mathbf{X}\lambda = \theta. \tag{3.3}$$

Combining (3.2) and (3.3) gives

$$\begin{pmatrix} \mathbf{X}' \\ \cdots \\ \mathbf{y}' \end{pmatrix} \theta = \begin{pmatrix} 0 \\ \cdots \\ -r \end{pmatrix}. \tag{3.4}$$

Multiplying the upper block of (3.4) by λ' and using (3.3) gives

$$\theta'\theta = \lambda'\mathbf{X}'\mathbf{X}\lambda = 0.$$

Thus all the elements of θ must be zero and so must be $r(= -\mathbf{y}'\theta)$. Gauss does not mention this point but clearly the third case—of *no* solution to the normal equations—can be ruled out. Now since

$$\theta = \mathbf{X}\lambda = \mathbf{0} \tag{3.5}$$

it follows that

$$\mathbf{X}\beta = \mathbf{X}(\beta + \kappa\lambda) \tag{3.6}$$

where κ is any scalar. The value is unchanged even if "the quantities β_i receive any increments or decrements proportional to the numbers λ_i" (1809, p. 262). It would be impossible to determine β even if the value of $\mathbf{X}\beta$ were known exactly:

> but we have already mentioned before, that cases of this kind, in which evidently the determination of the unknown quantities would not be possible, even if the true values of the functions ... should be given, do not belong to this subject.

Implicit in the afterword and the foreword is a notion of "determinability"—to rhyme with "estimability" and "identifiability". β is determinable if no two values of β generate the same value of $\mathbf{y}$ in the error-free case, i.e the same values of $\mu = \mathbf{X}\beta$. Equation (3.6) shows determinability failing. A natural interpretation of "more than determinate" is that β is determinable and would remain so if some elements of μ were discarded: compare the modern econometric term "over-identified". In the footnote Gauss mentions the determinability of elements of μ; μ can always be obtained from itself so it is always determinable. When $n + 1 - k$ of its elements depend on the remaining $k - 1$ then β is not determinable but μ is more than determinate because elements could be discarded without loss of determinability.

Gauss's discussion brings three elements into relation with each other: the determinacy of the problem, the independence of the functions and the unique solvability of the normal equations. His main effort goes into showing that least squares fails to produce a unique solution only when the problem is indeterminate. His successors often considered only two of the elements. They also usually focussed on (3.5) and the properties of the columns while Gauss emphasised the rows, the independence of k of the functions. The focus on columns was natural to twentieth century students interested in the inter-relation of regressors—e.g. Aitken's regressors (see Section 5) may be orthogonal and Stone's (see Section 4) are certainly highly correlated.

In two respects Gauss's determinacy analysis stands apart from his other least squares work: it does not use the triangularisation technique Gauss used for theory in 1809 and for computing later (see Stewart (1995) and Aldrich (1998)); it has no probabilistic content and so may pre-date the development of a probabilistic basis for least squares. Thus it is curious that the analysis makes a neater connection with the later estimation theory than with the earlier—asking whether faithful estimation is possible is a natural preliminary to a search for the best of all faithful estimators. However the neater connection seems to have been only a by-product of a change directed for other reasons. Thus Gauss (1823, pp. 31–33) mentioned that the new argument did not require normality and it worked for all sample sizes, unlike Laplace's large sample argument for least squares. In correspondence (see Plackett 1972) Gauss raised other objections to maximising posterior densities.

4 Indeterminacy and Practical Least Squares

Gauss's mathematical treatment of indeterminacy is definitive or contains the elements of a definitive treatment but it does not indicate how the functions can fail to be independent or whether this is a genuine possibility. Nineteenth-century numerical analysis and empirical work in astronomy are outside my competence and I glance at these points through the discussion in Chauvenet's nineteenth century astronomy textbook, which has a fuller treatment of these issues than others I have consulted. It is interesting to see that preoccupations with collinearity go back a long way. Incidentally Goldstine (1977, pp. 258–260) finds in a part of the *Theoria Motus* not concerned with least squares what he believes is the first systematic treatment of rounding errors as they affect the accuracy of calculations.

Chauvenet's (pp. 477–478) 'foreword' is like Gauss's but lacks the 'footnote'. Chauvenet derives the normal equations on the 1809 pattern but he does not mention Bayes' theorem. His (p. 513) account of indeterminacy follows Gauss and clears the method of any inadequacy:

> The normal equations will give determinate values of β provided they are really independent. If, however, any two of them become identical by the multiplication of either of them by a constant, the number of independent equations is, in fact, one less than that of the unknown quantities, and the problem becomes indeterminate. This difficulty does not arise from the method by which the normal equations are formed, but from the nature of the equations of condition. In any such case, additional observations are necessary, for which the coefficients have such varied values as to lead to independent equations.

In this last sentence the "coefficients" in the "equations of condition" refer to the x's in the equations $\mu = \mathbf{X}\beta$. As in Gauss the focus is on the row condition.

Chauvenet treats near dependence of the equations of condition and practical indeterminacy of the normal equations as real possibilities and discusses them when he treats computation. He (p. 548) recognises that the uncertainties associated with the β's are linked and that a much more precise determination of $\beta_1, \ldots, \beta_{k-1}$ may be possible if an extraneous estimate of β_k is used. However, he does not relate this observation to Gauss's notion of independence, either row-wise or column-wise. Indeed his discussion of the sources of uncertainty about β_k seems correct only for the case of a single regressor $\mathbf{x}_k$.

Chauvenet's treatment contrasts interestingly with Stone's (1954) post-Aitken treatment. Both authors recommend the use of an extraneous estimator and their proposals are essentially the same, although Chauvent puts everything in terms of Gaussian elimination. Stone (pp. 302–305) was concerned with a near linear dependence in the columns of $\mathbf{X}$, with "multicollinearity" in the form of a high degree of correlation between income and price; see Aldrich (1994) for some history of the curious term "multicollinearity". Stone recommends the use of an extraneous estimate of the income coefficient. In the matrix language of this note (and of Stone, who did much to popularise it in econometrics), let $\mathbf{X}_1$ be the matrix of the $k-1$ well-behaved regressors and let $\mathbf{x}_k$ be the vector whose coefficient is the "uncertain quantity" β_k . If we could solve the normal equations we would have

$$\mathbf{X}_1'\mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_1'\mathbf{x}_k b_k = \mathbf{X}_1'\mathbf{y}.$$

with $\mathbf{b}_1$ expressed in terms of b_k,

$$\mathbf{b}_1 = \mathbf{b}_1^1 - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{x}_k b_k \tag{4.1}$$

where $\mathbf{b}_1^1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$. The quantity b_k is unavailable but suppose a quantity $\hat{\beta}_k$ is found by a "subsequent independent determination" then this can replace b_k in (4.1) and an estimate of β_1 can be obtained. Chauvenet (p. 549) and Stone (p. 305) express the variance of the estimator of β_1 in terms of the variance of $\mathbf{b}_1^1$ and the variance of $\hat{\beta}_k$. Stone's treatment is much more careful; he also recognises that the method is not optimal, referring to Durbin (1953).

5 The Renaissance of Least Squares Theory

The merits of Gauss's two arguments for least squares were debated through the nineteenth century; Merriman's annotated bibliography records the contest up to 1877. Merriman (1877a, p. 142) himself judged the second argument "entirely untenable" and did not put it into his widely-used textbook (1884). Textbook theory seems to have settled on the Bayes with normality version. This version fed the development of test theory in the linear model and perhaps also the development of maximum likelihood—developments associated with Fisher. However, explicit consideration of determinability or of the solvability of the normal equations was more associated with the revival of the second version.

Modern Gauss–Markov theory—the material in, say, Chapter 4 of Rao (1973)—is a continuous development from Neyman (1934) and Aitken (1935). These papers make it seem that the theory originated in the twentieth century, for Aitken's earliest reference is to Sheppard (1912) and Neyman's to Markov (1908). Neyman thought "the Markoff method and theorem" were Markov's work and were new to the west. Aitken probably knew of Gauss's "Markoff theorem" for it is given—with a different proof—in E.T. Whittaker's *Calculus of Observations* (1924 pp. 224–226) and Whittaker was Aitken's past supervisor and present colleague. Whittaker (p. 300) also remarked that Sheppard's graduation problem could be solved by Gauss's methods. Lidstone (1933, p. 155) noticed the remark but could not develop it and devised a new proof for orthogonal polynomial regressors. Aitken (1935, p. 42) aimed to treat Lidstone's problem "from a more general standpoint". Plackett (1949) finally brought out the connections with Gauss.

Aitken and Neyman use Gauss's second model (2.2), although Aitken generalised it to the case of a known arbitrary error variance matrix. Gauss (1809, 1823) had treated the case of observations with unequal weight, so the novelty was in the provision for correlated observations. Aitken, following Sheppard and Lidstone, was concerned with approximating a time series by a time polynomial, i.e. with graduation. This graduation objective resembled the adjustment of observations objective associated with Gauss's geodesy model for the focus was not on Gauss's "unknown quantities" β but on the "functions", $\mu_1, \ldots, \mu_n$. Lidstone (1933, p. 154) called the requirement that $\hat{\mu}_i = \mu_i$ when $\varepsilon = \mathbf{0}$ a "fidelity-condition"; in this spirit Aitken's result could be phrased: $\mathbf{x}_i'\mathbf{b}$ (where $\mathbf{x}_i'$ is the i-th row of $\mathbf{X}$ and $\mathbf{b}$ the least squares estimator of β) is the minimum variance member of the class of faithful linear graduators $\hat{\mu}_i$, where

$$\hat{\mu}_i = \mathbf{a}'\mathbf{y} \text{ such that } \hat{\mu}_i = \mu_i \text{ when } \varepsilon = \mathbf{0}.$$

In a later contribution Aitken (1945) focussed on β and called the condition $\tilde{\beta} = \beta$ when $\varepsilon = \mathbf{0}$ a "consistency" condition, a usage complying with the interpretation of least squares as a way of finding a (usually) approximate solution to a set of (usually) inconsistent equations.

In a famous contribution to sample survey theory Neyman (1934) sketched the extension from Markov's unbiased estimation of an element of β to that of a linear combination of the elements of β. David & Neyman (1938) later gave the details and a proof. Aitken's paper had already appeared and David and Neyman presented their determinant analysis as an elementary introduction to his matrix treatment. David & Neyman (1938) consider estimating a linear combination of the β's, say $\tau = \mathbf{t}'\beta$, and show that $\mathbf{t}'\mathbf{b}$ has minimum variance in the class of linear unbiased estimators of τ.

The papers by Aitken and David & Neyman return to Gaussian rigour for, unlike the textbooks for astronomers, they give conditions ensuring the determinacy of the least squares problem. Aitken was a matrix specialist and Neyman was—by the standards of the 1930s—a very mathematical statistician. Their writings incidentally also record the changes in the "theory of elimination" since Gauss's time, notably the rise of determinants and matrices. Some of these changes are discussed by Farebrother (1996 and 1997) and by Aldrich (1998). Aitken's paper was a major force in the propagation of matrix methods in least squares.

Both papers lead with a condition on the $\mathbf{X}$ matrix guaranteeing a unique solution of the normal equations for β. The i-th row of Aitken's $\mathbf{X}$ is given by

$$\mathbf{x}_i' = [1\ x_1(t_i)\ x_2(t_i)\dots x_{k-1}(t_i)]$$

where each element is a function of t. In earlier work the regressors had been orthogonal polynomials but here (p. 42) the functions have only to be "linearly independent over the n values of t"—i.e. the column equation (3.5) holds only for λ the zero vector. Aitken does not mention the condition once he has introduced it but it is obviously needed for the formation of $(\mathbf{X}'\mathbf{X})^{-1}$. Aitken was only concerned with the passage in Gauss from (3.1) to (3.5).

Neyman (1934, p. 595) requires there be "at least" k independent equations; of course, there are at most k. Among the "ifs" in David & Neyman's (p. 107) list for the "Markoff Theorem (extended)" is the condition:

> if out of the n equations
> $$E\mathbf{y} = \mathbf{X}\beta$$
> it is possible to select at least one system of k equations which is soluble with respect to the β's, i.e. if at least one of the determinants of the k-th order of the matrix $\mathbf{X}$ is different from zero.

The first formulation could be described as requirement that β is determinable from the expected value of $\mathbf{y}$ and matrix insiders would recognise the determinantal formulation as the standard way of saying that the rank of $\mathbf{X}$ is k; MacDuffee (1933, p. 10) attributes the definition to Frobenius.

6 Estimability & Identifiability

Aitken and Neyman work with two out of the three elements of Gauss's argument for they do not explicitly consider the determinability and estimability of their particular linear combinations of β. Bose (1944)—see also Rao's (1945) elaboration of this work which appeared only as an abstract—followed up the work of David and Neyman but returned unknowingly (?) to Gauss by considering the existence question. Bose introduced the notion of an estimable function: a function of the β's is estimable if there exists an unbiased estimator of it. Bose recognised that the assumption that $\mathbf{X}$ has full rank is not necessary for a solution of the problem. The work could be taken as a bold variation on Gauss's footnote which recognised that the requirements for estimating functions of β are not the same as those for estimating β but of course Gauss saw no point in estimating these other quantities.

It is easy to show that the conditions given by Bose and Rao for estimability of a linear combination of β's are identical to those for Gaussian determinability—the problem of estimating $\tau = \mathbf{t}'\beta$ is determinate when $\hat{\tau} = \mathbf{T}y$ is such that $\hat{\tau} = \tau$ when $\varepsilon = \mathbf{0}$.

Gauss considered whether distinct values of β could generate the same vector μ, which in the case of no error is identical to the observation vector. This idea can be extended in various ways to the with-error case: can distinct values of β generate the same vector of expectations, μ again? Can distinct values of β generate the same distribution of $\mathbf{y}$? This last question addresses whether β is identifiable and it seems closer to the spirit of Gauss than the question of estimability. Koopmans introduced identifiability in the context of the simultaneous equations model of econometrics and then, with Reiersøl (1950), discussed it in general terms. More recently Leamer (1978, section 5.9) has discussed the linear model in terms of identification while Chipman (1964) related Koopmans's results on identification in the simultaneous equations model to the estimability of the linear model.

Acknowledgements

I am grateful to Philip Cooper for information about twentieth century actuarial work and to the referee for helpful suggestions.

References

Aitken, A.C. (1935). On Least Squares and Linear Combinations of Observations. *Proceedings of the Royal Society of Edinburgh*, **55**, 42–48.

Aitken, A.C. (1945). Studies in Practical Mathematics. IV. On Linear Approximation by Least Squares. *Proceedings of the Royal Society of Edinburgh*, **62**, 138–146.

Aldrich, J. (1994). Haavelmo's Identification Theory. *Econometric Theory*, **10**, 198–219.

Aldrich, J. (1998). Doing Least Squares: Perspectives from Gauss and Yule. *International Statistical Review*, **66**, 61–81.

Bose, R.C. (1944). The Fundamental Theorem of Linear Estimation. *Proceedings of the 31st Indian Scientific Congress*, (abstract) 2–3.

Chauvenet, W. (1891). *A Manual of Spherical and Practical Astronomy, volume 2*, fifth edition, reprinted by Dover 1960.

Chipman, J.S. (1964). On Least Squares with Insufficient Observations. *Journal of the American Statistical Association*, **59**, 1078–1111.

Dale, A.I. (1991). *A History of Inverse Probability from Thomas Bayes to Karl Pearson*. New York: Springer-Verlag.

David, F.N. & Neyman, J. (1938). Extension of the Markoff Theorem on Least Squares. *Statistical Research Memoirs*, University College London, **2**, 105–116.

Durbin, J. (1953). A Note on Regression when there is Extraneous Information about One of the Coefficients. *Journal of the American Statistical Association*, **48**, 799–808.

Farebrother, R.W. (1996). Some Early Statistical Contributions to the Theory and Practice of Linear Algebra. *Linear Algebra and its Applications*, **237**, 205–224.

Farebrother, R.W. (1997). A.C. Aitken and the Consolidation of Matrix Theory. *Linear Algebra and its Applications*, **264**, 3–12.

Fisher, R.A. (1922). 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. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.

Gauss, C.F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Hamburg: Perthes & Besser. English translation by C.H. Davis, reprinted 1963 Dover, New York.

Gauss, C.F. (1823/8). Theoria Combinationis Observationum Erroribus Minimus Obnoxiae, in two parts and a supplement. *Commentatines Societatis Regiae Scientarum Gottingensis Recentiores*, **5**, 33–62 & 63–90, **6**, 57–93. Reprinted with facing English translation and notes by G.W. Stewart, (1995). Philadelphia: SIAM.

Gauss, C.F. (1855). *Méthode des Moindres Carrés*, a translation into French of Gauss's memoirs on least squares by J.Bertrand. Paris: Mallet-Bachelier.

Goldstine, H. (1977). *A History of Numerical Analysis from the 16th through the 19th Century*. New York: Springer-Verlag.

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

Koopmans, T.C. & Reiersøl, O. (1950). The Identification of Structural Characteristics. *Annals of Mathematical Statistics*, **21**, 165–181.

Leamer, E.E. (1978). *Specification Searches*. New York: Wiley.

Lidstone, G.J. (1933). Notes on Orthogonal Polynomials and their Application to Least-Square Methods etc. *Journal of the Institute of Actuaries*, **64**, 128–164.

MacDuffee, C.C. (1933). *The Theory of Matrices*. Berlin: Springer.

Markov, A.A. (1912). *Wahrscheinlichkeitsrechnung*, second edition. Leipzig: Teubner.

Merriman, M. (1877). A List of Writings Relating to the Method of Least Squares, with Historical and Critical Notes. *Transactions of the Connecticut Academy of Arts and Sciences*, **4**, reprinted in volume I of Stigler (1980).
Merriman, M. (1877a). On the History of the Method of Least Squares. *The Analyst*, **4**, reprinted in volume I of Stigler (1980).
Merriman, M. (1884). *A Textbook on the Method of Least Squares*. New York: Wiley.
Neyman, J. (1934). On the Two Different Aspects of the Representative Method. *Journal of the Royal Statistical Society*, **97**, 558–606.
Plackett, R.L. (1949). A Historical Note on the Method of Least Squares. *Biometrika*, **36**, 458–460.
Plackett, R.L. (1972). The Discovery of the Method of Least Squares. *Biometrika*, **59**, 239–251.
Rao, C.R. (1945). Generalisation of Markoff's Theorem and Tests of Linear Hypotheses. *Sankhyā*, **7**, 9–16.
Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition. New York: Wiley.
Seal, H.L. (1967). The Historical Development of the Gauss Linear Model. *Biometrika*, **54**, 1–24.
Sheppard, W.F. (1912). Reduction of Errors by Means of Negligible Differences. *Proceedings of the Fifth International Congress of Mathematicians* (Cambridge), **2**, 348–384.
Sprott, D.A. (1978). Gauss's Contributions to Statistics. *Historia Mathematica*, **5**, 183–203.
Stewart, G.W. (1995). Gauss, Statistics, and Gaussian Elimination. *Journal of Computational and Graphical Statistics*, **4**, 1–11.
Stigler, S.M. (1980). *American Contributions to Mathematical Statistics in the Nineteenth Century*, in two volumes. New York: Arno.
Stigler, S.M. (1986). *The History of Statistics*. Cambridge: Harvard University Press.
Stone, R. (1954). *The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom 1920–1938*, volume 1. Cambridge: Cambridge University Press.
Waterhouse, W.C. (1990). Gauss's First Argument for Least Squares. *Archive for the History of the Exact Sciences*, **41**, 41–52.
Whittaker, E. & Robinson, G. (1924). *The Calculus of Observations*. London: Blackie.

Résumé

Gauss montre que la méthode du moindres carrés ne produit pas une solution unique quand le problème est indéterminé. Cette note considere l'argument du Gauss et indique certains relations avec les notions modernes de l'estimabilité et l'identifiabilité.

[*Received October 1997, accepted October 1998*]