lostr.tex, 7/8/1996.

COMMENTARY ON TWO ARTICLES BY C. A. LOS *

E. T. Jaynes
Wayman Crow Professor of Physics
Washington University
St. Louis MO 63130, U. S. A.

Abstract: Recently, C. A. Los has recommended that all previous methods for estimating linear relations from data with unknown errors be scrapped, and proposes to find an 'exact objective mathematical solution' that depends only on the data. We discuss this with the conclusions that (1) his criticisms of previous methods, while overly severe, make some important points worth noting; (2) however, his solution ignores relevant information and does not exist except in very special cases; (3) the Los solution, when it exists, may be characterized as the one which would be correct if we had an infinite amount of data which led to the same data correlation matrix; (4) methods for dealing with the problem which are optimal in all cases have been known for 20 years.

INTRODUCTION

The problem of estimating linear relations from a scatter plot of data with unknown errors has been discussed, in one form or another, by physicists, astronomers, mathematicians, statisticians, economists, biologists, and psychologists, since at least the time of Gauss (1809). What is for all practical purposes the correct general solution has been available for 20 years [1], but conceptual misunderstandings have prevented some from recognizing it as the solution, and attempts to solve it by other means still persist.

In two recent articles, C. A. Los [2], [3] has proposed a radical new approach, which starts by rejecting all previous methods (least squares, maximum likelihood, principal components, Bayes' theorem, etc.) by which these problems have been solved in the past, on the grounds that they are 'subjective' and 'prejudiced'. Repeatedly, his own proposal is called 'exact'. In our view, the work

* To be published in Volume 3 of special issues, "On System-theoretic Methods in Economic Modelling", S. Mittnik, Editor, in *Computers and Mathematics with Applications*, and subsequently as a monograph by Pergamon Press.

does contain some valid criticisms which could lead to better practice in the future, so it is worth while to examine it in some detail.

The most general problem considered is that where we have T observations of an n component vector, thus a $(T \times n)$ matrix X of data values. The object is to see whether there is evidence for q linear relations between the variables, of the form $X\beta = 0$, where β is an $(n \times q)$ matrix of rank q, and $0 \le q \le n$.

In conventional approaches one does this by assuming such a relation (that is, adopting a model with prescribed q), not as a 'subjective prejudice', but as a *tentative working hypothesis*, then finding the resulting 'best' estimates of q and β, as the values for which one achieves the best fit to the data. As Los stresses, this is usually done in a way which implies an assumption that some elements of X are known exactly (we agree that this is often done without justification; but see it as a misuse, rather than a defect, of the underlying principle). It is essential to get also an indication of the accuracy of the estimates, since this is the basis for judging whether the model may lead to useful predictions.

In the scheme of Los none of this is permitted; the concepts of a 'model' and a 'sampling distribution' are rejected and it is required that we determine q and β directly from the data. Since his determinations are called 'exact' there is no mention of their accuracy. To fix ideas, we note first the specific example which he considers, then proceed to the general problem.

THE EXAMPLE

Los reports that the data refer to the year 1985 and T = 32 Bank holding companies; x_{t1} is the net interest margin of the t'th company, x_{t2} is "the consumer loans in percent of total loans to the U.S. addresses"; x_{t3} is "the net purchases funds in percent of total assets." The raw data, with other explanatory remarks, are presented in [3], Appendix A. Then he gives the (3×3) data covariance matrix

$$\Sigma = X'X = \begin{pmatrix} 0.7022 & 6.9040 & -10.6826 \\ 6.9040 & 99.0556 & -114.7687 \\ -10.6826 & -114.7687 & 259.2516 \end{pmatrix} \tag{1}$$

His matrix X is not the actual data, but those data with mean values over t subtracted off; i.e. denoting the raw data values by D_{ti}, we have $X_{ti} = D_{ti} - T^{-1}\sum_t D_{ti}$. Los takes it for granted without discussion, that Σ "contains the information needed for all computations." This implies that he will come to the same conclusions whether that covariance matrix was generated by three data points or three million.

Suppose we were forced to operate under the Los restrictions; then what could we do with this problem? Since all properties of a real symmetric matrix can be found, and visualized intuitively, from its principal axes and eigenvalues, it must be true that everything we can infer is contained in the diagonalization of (1). We find that the diagonalization $\Sigma = U\Lambda U'$ is achieved by the components:

$$\Lambda = \begin{pmatrix} .163573 & 0 & 0 \\ 0 & 39.23379 & 0 \\ 0 & 0 & 319.6120 \end{pmatrix}, \qquad U = \begin{pmatrix} 0.998756 & .030199 & -.039681 \\ -.045142 & .885617 & -.462218 \\ .021184 & .463434 & .885878 \end{pmatrix} \tag{2}$$

U is a real orthogonal matrix, so its columns are the normalized eigenvectors of Σ. The data concentration ellipsoid (locus of vectors x satisfying $x'\Sigma^{-1}x = 1$) has semi-axes (0.4, 6.3, 17.9), oriented parallel to those eigenvectors.

Los demands that we choose our linear relations solely from (2) without making any use of what we may know about the measurement errors in the three directions, or other relevant evidence.

But even under these handicaps, we can still ask for small residuals; we observe that the component of data parallel to the first eigenvector (the one with the smallest eigenvalue)

$$u_t \equiv 0.999\, x_{t1} - 0.045\, x_{t2} + 0.021\, x_{t3} \tag{3}$$

has by construction the mean value zero, and shows a smaller mean–square variation across t than does any other component; so in this state of self–imposed ignorance about the nature of the problem, we can do little else than to conclude that the linear relation most strongly indicated by the data alone is $u_t = 0$; *i.e.*, we estimate β as the first column of U. Of course, if we knew that the measurement errors in x_1 were only about a tenth as great as those in x_2, common sense would lead us to a different conclusion; but Los denies us the use of such information.

If we want to find two relations, it seems that we can do little else than take β as the first two columns of U. In that case (*i.e.* if one believes that these data give good evidence for the presence of two linear relations), the explained part of the data would be orthogonal to the first two columns, thus parallel to the third column of (2); indeed, this component accounts for nearly 90% of the total data variance. Let us call this the Simple Solution; we do not see how it is possible to do any better than this under the constraints imposed by Los, so it will be interesting to compare the Simple Solution with the Los solution.

CRITICISM

But these results are quite arbitrary; with noisy data one cannot judge linear relations merely from Σ because it depends on our units. For example, suppose we expressed interest rates in hundredths of a percent instead of percents. Then all values of x_{t1} would be increased by a factor of 100, and the eigenvalues and eigenvectors would be totally different; we now find in place of (2):

$$\Lambda = \begin{pmatrix} 29.589 & 0 & 0 \\ 0 & 95.547 & 0 \\ 0 & 0 & 7255.17 \end{pmatrix}, \qquad U = \begin{pmatrix} -.076403 & -.163329 & -.983609 \\ .987469 & .124221 & -.097330 \\ .138082 & -.978720 & .151790 \end{pmatrix} \tag{4}$$

and if we are not permitted to use the knowledge that the measurement errors in x_1 are now 100 times greater than before, the linear relation 'most strongly indicated' would correspond to the first column of the new U, almost orthogonal to the first estimate (3). Likewise, the new Simple Solution is nearly orthogonal to the old one.

But if our variables have different physical natures, there is no unique "correct" system of units. For example, in a different problem x_1 might be acres planted, x_2 tons shipped, and x_3 dollars received, for T different farms. But the same data could be represented equally well in terms of hectares, bushels, and drachma. Then the data concentration ellipsoid would be quite different, but any rational method of inference ought to lead us to the same substantive conclusions.

Conventional methods of inference achieve this necessary invariance by using sampling distributions; any change in the units of measurement is reflected by a change in the sampling variance that automatically compensates for it. Common sense might tell us that it is not the absolute size of the data covariance, but the data covariance *in comparison with the sampling variances* that is relevant to inference; then we reach the same substantive conclusions whatever units of measurement we choose. By rejecting all use of sampling distributions, Los denies himself this way of correcting the arbitrariness; but there are other ways.

Los also rejects the practice of using correlation functions instead of covariance functions, on the grounds that this is a nonlinear transformation that 'introduces distortions of the true covariation'. Our position is just the opposite; since correlation functions, being dimensionless, are independent of the system of units, their use is another way of *correcting* the arbitrary distortions present in covariance functions. So he also denies himself this way of achieving invariance.

These considerations make it clear that we started with an ill–posed (mathematically underdetermined) problem. Given only the data, or the data covariance matrix, many different linear relations are equally compatible with them, and there is nothing to guide our choice. Even the criterion of smallness of the residuals leads to arbitrary results, as we have just seen. The difficulty here is that, if our variables have different physical dimensions, then the idea of 'small' residuals is basically meaningless; if the residuals r_1, r_2 have different physical meanings, then in writing $r_1^2 + r_2^2$ we are trying to add apples and oranges. Put differently, the data space is not a metric space, but an affine space; there is no meaning to such geometrical notions as 'distance' and 'perpendicular'.

There can be no defensible solution until the data are supplemented by additional information about the nature of the problem. What we know about the meaning of the variables and their measurement errors is crucially important, as the comparison of (2) and (4) shows. Likewise, prior information may tell us whether there is any rational reason for expecting a relation between the variables, and previous data may give us prior estimates of their parameters, which can improve the accuracy of the new estimates.

Los rejects all such information, and so he will be obliged, inevitably, to invoke arbitrary conditions in order to get any solution. But if all the conditions usually used, which express our information about relevant properties of the real problem, are to be dismissed as 'prejudice', then what epithet shall we apply to conditions which do *not* express any such information?

THE GENERAL PROBLEM

In the Los formulation, the goal is to separate the reduced data matrix X into an 'exact' or 'explained' component $\hat{X}$ of interest (for example, vectors orthogonal to the above u_t), and another component $\overline{X}$ variously termed 'inexact', 'unexplained', 'noise', 'error', or 'residual'. Only linear relations are permitted, by imposing a condition

$$A\,\hat{X}' = 0 \tag{5}$$

where A is a $(q \times n)$ matrix of rank q. Note that if we partition off the first q columns by writing $A = (I| - \beta')$ and $\hat{X} = (y|X_2)$, then (5) takes the more familiar form $X_2\,\beta = y$.

From (5) he then obtains

$$A\,\hat{X}'\hat{X} = A\,\hat{\Sigma} = 0 \tag{6}$$

where $\hat{\Sigma}$ is the reduced covariance matrix. The nature of the problem facing us can be seen at once by dealing directly with the data vectors $\hat{X}$ without introducing $\hat{\Sigma}$ at all; but Los chooses to take this detour, which obscures that information.

Next he partitions the columns of A in the aforementioned way $A = (I|B)$ where I is a $(q \times q)$ matrix and B a $q \times (n-q)$ one, and partitions $\hat{\Sigma}$ similarly:

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} \tag{7}$$

where $\hat{\Sigma}_{12} = \hat{\Sigma}'_{21}$ is a $(q \times n)$ matrix, etc., and $\hat{\Sigma}_{22}$ is to have full rank $(n-q)$. Then writing out (6) we find that B is determined to be $B = -I\,\hat{\Sigma}_{12}\,\hat{\Sigma}_{22}^{-1}$. If we choose I to be the $(q \times q)$ unit matrix, as Los does, then (6) reduces to the condition

$$\hat{\Sigma}_{11} - \hat{\Sigma}_{12}\,\hat{\Sigma}_{22}^{-1}\,\hat{\Sigma}_{21} = 0\,. \tag{8}$$

When first introduced, this appears to be the principal relation that Los proposes to use to determine his solution. Thus he describes (8) as imposing a set of "exact, simultaneous, nonlinear equality constraints on the variances and covariances of the reduced covariance matrix $\hat{\Sigma}$, i.e. on the explained part of the data covariance matrix Σ." Then he states that (8) poses (italics his): "a formidable intellectual problem. *A general solution has not yet been found*". We are happy to give that general solution below.

In the case $(n = 3,\ q = 2)$, [3] devotes half a page to studying (8), which is thought to impose complicated conditions on the solution to his problem, noting some implied sign conditions but not finding any specific solution. However, it is evident from inspection that in this case (or in any case where $q = n - 1$), the missing general solution is simply a projection operator:

$$\hat{\Sigma}_{ij} = a_i\, a_j\ , \qquad\qquad (a_1 \cdots a_n) \quad \text{arbitrary} \tag{9}$$

showing that (8) imposes no condition at all on the solution! By the choice $q = n - 1$ we have already required that the 'exact' component of the data $\hat{X}$ must be parallel to some fixed vector a, but (8) allows its direction to be arbitrary. Factoring (9) we have $\hat{X}_{tj} = a_j$, just what we knew before taking this detour.

For that general solution, we extend this result to any n and q. Since all matrix elements are real, the symmetric matrix $\hat{\Sigma}$ has $n(n+1)/2$ independent components and (8) imposes $q(q+1)/2$ conditions on them. Therefore by the 'general solution' of (8) we mean any algorithm that enables us to write a specific solution $\hat{\Sigma}$ as a function of

$$\frac{n(n+1)}{2} - \frac{q(q+1)}{2} \tag{10}$$

arbitrary real constants. To do this, choose $\hat{\Sigma}_{22}$ as an arbitrary symmetric positive definite matrix; this requires $(n-q)(n-q+1)/2$ quantities. Then define $\hat{\Sigma}_{12}$ arbitrarily; this requires $q(n-q)$ more arbitrary quantities. Indeed,

$$\frac{(n-q)(n-q+1)}{2} + q(n-q) \quad = \quad \frac{n(n+1)}{2} - \frac{q(q+1)}{2} \tag{11}$$

so all degrees of freedom are now specified. As the final step, then, we need only *define*

$$\hat{\Sigma}_{11} \equiv \hat{\Sigma}_{12}\, \hat{\Sigma}_{22}^{-1}\, \hat{\Sigma}_{21} \tag{12}$$

and we have produced that 'general solution' by construction.

But this shows that (8) is always empty; it states only what we already knew, that the n − dimensional space has been partitioned into a q − dimensional 'forbidden' manifold spanned by the q eigenvectors with eigenvalue zero; and the $(n-q)$ − dimensional 'allowed' manifold M orthogonal to it, but it says nothing about the directions of those subspaces. Indeed, those directions ought to depend in some way on the data; but (8) makes no reference to the data.

The moral of this is that in order to carry out a well motivated calculation one must be aware of what *information* is and is not contained in our equations. Information is never created by mere mathematical manipulations, but it may be destroyed by irreversible ones. Since (8) was derived from (5), it cannot tell us anything that was not already apparent from inspection of (5). This detour was an attempt to extract information (restricting conditions) from an equation which does not contain it; the basic ambiguity of the problem has not been reduced below that indicated by (10), so the search resumes for other conditions to impose.

THE PROPOSED SOLUTION

To see the situation we are in more clearly, note that if there were no 'noise', the problem would indeed be one of exact mathematical deduction not requiring probability theory at all; if the data obey any linear relations, any method of solution would find them. In other words, our uncertainty as to the existence of linear relations arises solely from the fact that the measurement errors are unknown. In order to get that 'exact' solution demanded by Los, it is therefore a mathematical necessity that we impose $n(n+1)/2 - q(q+1)/2$ exact conditions *on those measurement errors.* However this is done, it requires us to assume a great deal of information which we do not possess; how could one ever justify any such assumption?

The difference $\overline{\Sigma} = \Sigma - \hat{\Sigma}$ is called the "residual covariance matrix", although it is not necessarily the covariance matrix of the residuals. In any event, the condition he now imposes is not that the residuals be small, but that $\overline{\Sigma}$ shall be a diagonal matrix! Far from being able to explain a reason for this, we are bewildered by it.

To see why, suppose that $\overline{\Sigma}$ is actually the covariance matrix of the residuals. For 'purely random' errors we expect the off diagonal elements of $\overline{\Sigma}$ to have an average magnitude of the order of $1/\sqrt{T} = 1/\sqrt{32} \simeq 1/6$ of the diagonal elements. To require them all to be zero is to make a very strong arbitrary assumption about the noise, which is almost certainly false, and which must have serious consequences.

Those consequences are not hard to find; this assumption imposes $n(n-1)/2$ new conditions on the solution, so the number of degrees of freedom is reduced to

$$\frac{n(n+1)}{2} - \frac{q(q+1)}{2} - \frac{n(n-1)}{2} = n - \frac{q(q+1)}{2} \tag{13}$$

Unless $q(q+1)/2 \le n$ (which Los calls "Wilson's inequality") the originally underdetermined problem becomes overdetermined, and there is no solution at all. Indeed, when the strict inequality holds the problem is still underdetermined; this new condition determines a unique (although possibly multiple) solution *only* when $q(q+1) = 2n$.

To see what this implies, it is quite possible to have $n = 5$ variables that satisfy $q = 2$ linear relations, so that the noiseless data vectors lie in a manifold of dimension $5 - 2 = 3$. One can easily produce computer simulated data that conform to this. In such a case, if the noise level is small enough, the conventional methods that Los rejects (least squares, maximum likelihood, Bayesian) can easily find the correct linear relations; but the proposed solution of Los does not exist.

The Wilson inequality means that Los has no unique solution for $n =$ 4, 5, 7, 8, 9, 11, and so on. Only in a few special cases [$(n, q) =$ (3,2), (6,3), (10,4), *etc.*] does such a solution exist; it is not surprising that his example is (3,2). It seems to us that at this point he might have perceived that the Wilson inequality, far from causing difficulty for the conventional solutions as he alleges, is what makes *his* proposed solution indefensible and unusable in general.

In any event, his final solution for the current problem ([3], Table 6) is that $\hat{\Sigma}$ is the projection operator (9) with the vector $a' = (-.80162, -8.61237, 13.32603)$. Normalized to unit length, this vector is $\hat{x}' = (-.05046, -.54210, .83880)$. Comparing with the Simple Solution noted after (3) we see that they are nearly the same; the Los solution is rotated from the third column of U by $\arccos(\sum \hat{x}_i U_{i3}) = \arccos 0.99565 = 5.35$ degrees. Pragmatically, we have not advanced very far from our first crude guess.

However, we must concede a surprising point: Los does manage to achieve invariance under a change of units, in spite of the fact that he takes no note of the need for it. Had we used the covariance matrix corresponding to (4) instead of (1), the Los solution would have stretched the component a_1 by a factor of 100, and thus would still yield the same substantive conclusions. So, it appears that the Los solution has one desirable property which the Simple Solution lacks.

In understanding the reason for this, we get a much clearer characterization of the Los solution; in effect, supplying the missing rationale for it. If our model was indeed correct (that is, it accounted for every systematic effect present in the data) and we had the correct values of its parameters expressed in $\hat{\Sigma}$, then the residuals would be 'pure random noise'. As noted, we would not expect the off-diagonal elements of $\overline{\Sigma}$ to be zero, but they would be as likely to be positive as negative. So if we were to accumulate more and more data with this correct model, all elements of $\hat{\Sigma}$ and the diagonal elements of $\overline{\Sigma}$ would grow like T, while the off-diagonal elements of $\overline{\Sigma}$ would grow like $\sqrt{T}$. They would indeed tend to become negligible compared to the other elements of these matrices.

Therefore we can explain the Los solution in much simpler terms: when it exists, it is the solution that would indeed be 'exactly correct' *if we had an arbitrarily large amount of data, leading to the same covariance matrix.* Or, of course, the same matrix with all elements magnified by any constant factor. When we have a finite amount of data, it requires correction to take account of our uncertainty as to the noise. We consider below the methods that would be appropriate in all cases, which contain the Los solution as a limiting form for those special values of (n, q).

Of course, if we had perfect noiseless data, then as noted the data concentration ellipsoid would flatten into a disk or line that identifies the correct allowed manifold M exactly without any need for us to specify it in a model; the problem would reduce to one of pure mathematical deduction of the kind which Los demanded, and the Los solution, the Simple Solution – or any other method of mathematical solution – would give the same results and achieve automatically this invariance under a change of units. Since the Los solution, in effect, assumes this case, it too achieves that invariance.

This still leaves us with one question: "Why does the Los solution work only for certain special values of n and q?" Answer: however large T, the contribution of noise to the diagonal elements of Σ never becomes negligible; and so without a sampling distribution one has no way to separate them into 'signal' and 'noise' components. Therefore the Los solution must seek to determine all elements of $\hat{\Sigma}$ from the off-diagonal elements of Σ. Only for these special values of n and q do the off–diagonal elements contain that information. In all other cases one must resort to a different algorithm, which makes use of a sampling distribution.

WHAT IS THE CORRECT METHOD?

After all these criticisms, let us now try to make some positive, constructive remarks about such problems. If the available information were sufficient to determine those 'exact, objective mathematical solutions' that Los seeks, they would surely have been found by Gauss 180 years ago. The reason why these problems still cause trouble is that they are mathematically ill–posed. When we have a finite amount of noisy data with unknown errors, finding a linear relation is not a problem of mathematical deduction at all; it is problem of *inference.*

Scientific inference is concerned necessarily, not with empty assertions of 'objectivity' but with information processing; how to extract the best conclusions possible from the incomplete information available to us. Surely, the fundamental basis of scientific inference – almost a principle of morality – is that we should

(1) take into account *all the relevant information we have*, of whatever type,

(2) carefully *avoid* assuming information that we do not have.

but as we have seen, Los commits egregious violations of both of these rules.

In contrast, Bayesian and Maximum–Entropy methods are algorithms for information processing, uniquely determined by these rules and a few other elementary desiderata of rationality and consistency. Both their theoretical basis and their pragmatic success in applications are now established by overwhelming masses of evidence, of which Los takes no note. So let us summarize the historical development that culminated in this.

We have seen that use of a sampling distribution, which describes our information about the measurement errors in the variables, is essential for any rational inference. In many simple cases the method of maximum likelihood will then yield all the information we need.

Least squares is a special case of maximum likelihood, valid when the sampling distribution is gaussian. When derived as it should be from probability theory, it is always *weighted* least squares, the weighting coefficients being determined by the sampling distribution. Then the weighted sum of squares has the needed invariance under change of units, and it interpolates between the extreme correctly criticized by Los, in which one supposes some variables known exactly, and his opposite extreme, in which weights are not mentioned at all (and are therefore, in effect, equal by default).

However, sampling distributions may not contain all the information that is relevant to the problem, and they do not provide all the technical apparatus needed for calculations. Full power to deal with all aspects of the real problem requires four further advances beyond maximum likelihood:

(1) One may have highly cogent prior information about the likely linear relations, and prior estimates of their parameters. For example, one may know in advance that the sum of two variables must be nearly constant, simply because the agent's commission is relatively small (or perhaps because wheat is not wilfully created or lost in the transactions). It would be irrational to fail to take such knowledge into account in judging functional relations.

(2) One may have a sequence of data sets – perhaps from successive years – and the previous data yield good prior estimates of the parameters. We need some algorithm to update our estimated relations when new data become available, in a way that takes full account of all the past data. A useful example of the process is the Kalman filter, which we would like to refine and generalize.

(3) The real problem may have 'nuisance parameters' of no interest to us; indeed, regression problems may have more nuisance parameters than data points. Use of maximum likelihood or least squares would require us to fit all the parameters, interesting or not. Computationally, one would have to find a global maximum in a space of high dimensionality; and we may then find that the maximum occupies not a point, but a region. If nuisance parameters can be eliminated at the start, one may achieve orders of magnitude reduction in computation by estimating only the parameters of interest.

(4) Merely estimating the parameters of a given model is only the first step in a real research program; one needs also a way to judge the relative merits of different models, in the light of the data so that cumulative improvements can take place over long times.

Now all of these advances over maximum likelihood are provided automatically and elegantly, by Bayesian methods. In particular, the Bayesian 'errors in variables' (EVM) models, which Los dismisses as "*not* fruitful to discuss" are nevertheless fruitful enough to provide the information handling ability and computational power that all other methods lack. The Los solution, when it exists, is the limit of the Bayesian solution as $T \to \infty$, $\Sigma/T \to const.$

It is not our present topic to go into details of these Bayesian solutions; we hope to present them in a later article. However, several examples of such solutions were given long ago by Zellner [1]. If one will investigate the properties of those solutions, they will be found to do every reasonable thing that one might hope for in these problems. On very similar problems, including economic time series, the analytical theory and many fully worked–out numerical examples to illustrate its use in all four of the above extensions of maximum likelihood, are given by Bretthorst [4].

REFERENCES

1. A. Zellner (1971), *An Introduction to Bayesian Inference in Econometrics*, J. Wiley & Sons, Inc., New York, (1971). Reprinted by R. E. Krieger Pub. Co., Malabar, Florida (1987); Chapter 5.

2. C. A. Los (1989a), "The Prejudices of Least Squares, Principal Components and Common Factors Schemes", Computer Math. Applic. **17**, 1269-1284.
3. C. A. Los (1989b), "Identificaton of a Linear System from Inexact Data: a Three-variable Example" Computer Math. Applic. **17**, 1285-1304 (1989).
4. G. L. Bretthorst (1988), *Bayesian Spectrum Analysis and Parameter Estimation*, Springer Lecture Notes in Statistics, Vol. 48.