Courant Institute of Mathematical Sciences

Magneto-Fluid Dynamics Division

ADVANCED STATISTICS FOR TOKAMAK TRANSPORT COLINEARITY AND TOKAMAK TO TOKAMAK VARIATION

Kurt S. Riedel

U.S. Department of Energy

New York University
UNCLASSIFIED

New York University
Courant Institute of Mathematical Sciences
Magneto-Fluid Dynamics Division

ADVANCED STATISTICS FOR TOKAMAK TRANSPORT COLINEARITY
AND TOKAMAK TO TOKAMAK VARIATION

Kurt S. Riedel

U.S. Department of Energy
Grant No. DE-FG02-86ER53223

March 1989

UNCLASSIFIED
This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Printed in U.S.A.

Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161
This is a compendium of three separate articles on the statistical analysis of tokamak transport. The first article is an expository introduction to advanced statistics and scaling laws. The second analyzes two important problems of tokamak data -- colinearity and tokamak to tokamak variation in detail. The third article generalizes the Swamy random coefficient model to the case of degenerate matrices.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Advanced Statistical Techniques for Tokamak Data Analysis</td>
<td>1</td>
</tr>
<tr>
<td>2. Tokamak to Tokamak Variation and Colinearity in Scaling Laws</td>
<td>15</td>
</tr>
<tr>
<td>3. The Swamy Random Coefficient Model with Singular Covariance Matrices</td>
<td>24</td>
</tr>
</tbody>
</table>
ADVANCED STATISTICAL TECHNIQUES FOR TOKAMAK DATA

ANALYSIS

K.S. Riedel

Courant Institute of Mathematical Sciences

New York University

New York, New York 10012

Abstract

The determination of empirical energy confinement scaling laws, using linear regression, involves a number of statistical hypotheses. The assumptions of least squares analysis and possible breakdowns in these assumptions are discussed. The multiplicity of scaling laws, which almost equally well describe the tokamak data, is associated with the lack of data variation along certain data directions. A number of advanced techniques and diagnostics to treat the irregularities associated with fusion data are given.

I. Standard Least Squares Fit to Confinement Data

Empirical scaling laws of energy confinement in tokamaks have been valuable in providing a concise summary of existing experimental results and a standard to measure the performance of different devices. A more controversial use of scaling laws is the prediction of energy confinement for future, large scale devices such as C.I.T. and I.T.E.R.

Typically, the scaling law is determined by performing a least squares regression to a large set of plasma data. In this comment, we discuss the assumptions which are implicit in scaling laws. We then consider advanced
statistical techniques to provide a more accurate analysis of existing data. Finally, we suggest which new data will be of the most value in improving the reliability of predictions of energy confinement.

The most common empirical scaling laws are geometric scalings of the form: $\tau_E = c_0 a^{\beta_1} R^{\beta_2} \kappa^{\beta_3} M^{\beta_4} I^{\beta_5} P^{\beta_6} B_t^{\beta_7} n^{\beta_8}$. Geometric scaling laws are convenient since they are linear in the logarithms of the macroscopic plasma variables. We let $y_i = \ln \tau_{E_i}$ and $\bar{x}_i$ be a vector containing the logarithms of certain physical quantities such as the major and minor plasma radii, $R$ and $a$, the ellipticity, $\kappa$, the effective ion mass, $M$, the input power, $P$, the plasma current, $I_p$, the magnetic field, $B_t$, and the line average density, $\bar{n}$.

We now review the standard least squares analysis as it is commonly applied to scaling law studies. A linear relationship between the logarithm of the energy confinement time and the logarithms of the plasma parameters, $\bar{x}_i$, is assumed:

$$y_i = \bar{\beta} \bar{x}_i + \epsilon_i$$

where $\epsilon_i$ is a Gaussianly distributed independent random variable. The k dimensional free parameter vector $\bar{\beta}$ is calculated by minimizing the sum of squares:

$$\min_{\beta'} \sum_i |y_i - \bar{\beta}' \bar{x}_i|^2 = \min_{\beta'} (Y - X \bar{\beta}')^t (Y - X \bar{\beta}')$$

yielding the least squares estimate:

$$\hat{\beta} = (X^t X)^{-1} (X^t Y)$$

where $X$ and $Y$ are $n \times k$ and $n \times 1$ dimensional matrices whose rows consist of $\bar{x}_i$ and $y_i$. 

2
We note that the least squares estimate is the best possible estimate within the class of unbiased linear estimators.

An unbiased estimator for the variation of the measured logarithms of the energy confinement time is

$$
\hat{\sigma}^2 = \frac{1}{n-k} \sum |y_i - \hat{\beta} \bar{x}_i|^2
$$

and the error in the estimate of the free parameter $\hat{\beta}$ is

$$
cov(\hat{\beta} - \bar{\beta}) = \hat{\sigma}^2 (X^t X)^{-1}
$$

which decreases as $1/n$. Since $\bar{\beta}$ is a k dimensional vector the covariance matrix has dimension $k \times k$. However if the data covariance matrix, $X^t X$ is nearly singular, then the corresponding directions of $\bar{\beta}$ are poorly determined. These directions are related to the lack of independent variation of the "independent" variables. This problem occurs often in fusion databases.

In predicting the energy confinement in a proposed new observation, $\bar{u}$, such as the value in ITER, we must make the distinction between the true value and the observed value. The difference between the predicted value using the estimated parameters $\bar{\beta}$ and the observed value contains not only the uncertainty from the covariance of $\hat{\beta}$ but also the individual uncertainty from $\epsilon$. We define the "leverage", $h_i$ of $\bar{x}_i$ as $h_i = \bar{x}_i (X^t X) \bar{x}_i$, which scales as $1/n$. The predicted values for $\bar{u}$ are

$$
\begin{align*}
\text{True} : & \quad \hat{y}(\bar{u}) = \bar{\beta} \bar{u} \pm \hat{\sigma} \sqrt{\bar{u} (X^t X)^{-1} \bar{u}} \\
\text{Observed} : & \quad \hat{y}(\bar{u}) = \bar{\beta} \bar{u} \pm \hat{\sigma} \sqrt{1 + \bar{u} (X^t X)^{-1} \bar{u}}
\end{align*}
$$
Thus the average error in our knowledge of the true value of \( \hat{y}(\vec{u}) \) is proportional to the square root of its leverage.

II. Implicit Assumptions

We now discuss the assumptions involved in the standard analysis [1].

1. The linear model is correct and only a finite number of free parameters must be determined.

Clearly treating a complex physical system such as a tokamak as a linear function of four to eight variables is only an approximation. The log linear scaling may be thought of as the first term in a Taylor series expansion of an exact nonlinear relationship for \( \tau_E \) in terms of a small number of the most important plasma variables. Higher order corrections may be trivially included by adding additional free parameters corresponding to higher order polynomials in \( \vec{x} \).

Perhaps the most important use of dimensionless variables is in defining the domain of validity of the Taylor series expansion. Luckily in many applications we are interested in the scaling of energy confinement with size and the dimensionless physical variables are varied only slightly. Caution is needed as a small (30\% ) variation in a dimensionless variable, such as the poloidal beta or edge \( q \), may change the physics more than a fivefold increase in device size.

Also contained in the assumption that the model is correct is the statement that other physical variables have no significant impact on energy confinement. Usually the regression analysis ignores all variables related to the edge plasma.

2. Independence of Errors.
We assume the departures of the individual discharges from the log-linear scaling are independent. In particular, we assume that all tokamaks have exactly the same scaling and that all discrepancies in the log-linear relationship are due to discharge to discharge variation and not tokamak to tokamak variation. In particular we assume that one hundred new D-IIID datapoints are equally likely to scale as the last hundred D-IIID points and as any hundred TFTR points!

An alternative hypothesis is that on each tokamak the scalings describe perfectly the energy confinement. The differences in the scaling coefficients in different tokamaks is associated with many small factors. Estimates for the true scaling coefficients and the variance associated with these coefficients are constructed by averaging over all tokamaks:

$$\bar{\beta} = \frac{1}{N} \sum \bar{\beta}, \quad \text{cov} (\bar{\beta} - \bar{\beta}) = \frac{1}{N - 1} \sum (\bar{\beta} - \bar{\beta})(\bar{\beta} - \bar{\beta})^t$$

When both discharge to discharge and tokamak to tokamak differences exist, a generalised least squares problem must be solved\(^2\).

3. The variance of the measurement error, \(\epsilon_i\), is a constant, independent of the discharge number.

The correlation and nonconstancy of the error, \(\epsilon_i\), may be taken into account by using a general linear model of multivariate statistics. In general linear models, an error structure, \(\Sigma = \text{cov} (\epsilon_i, \epsilon_j)\) is postulated or estimated. The problem of estimating the free parameters \(\bar{\beta}\) is reduced to the standard case of uncorrelated, constant variation \(\epsilon_i\) by multiplying \(X\) and \(Y\) by \(\Sigma^{-1/2}\).
If $\Sigma$ is diagonal this corresponds to minimizing

$$\sum_i \left( \frac{y_i - \beta \tilde{x}_i}{\sigma_i} \right)^2$$

Typically the error structure is estimated by fitting the residual error matrix to a model error matrix with only a small number of free parameters and then $\tilde{\beta}$ is then estimated iteratively using the approximate error matrix.

4. *Expected value of errors is zero.*

Statistics cannot eliminate systematic errors.

5. *The regressor variables are measured without error.*

The uncertainty in the isotope ratio in mixed species discharges is probably larger than the uncertainty in $\tau_E$. All other global variables are usually known to good accuracy. The standard assumption that mixed species discharges have an isotope ratio of 1.5 is more a systematic error than a random error. Local transport analysis is marked by uncertainties in deposited power and local profile gradients.

6. *The errors, $\epsilon$, have a Gaussian distribution.*

This final assumption is only used in hypothesis testing and confidence intervals for finite size samples. (By the central limit theorem, all such tests will be asymptotically valid as $n$ tends to infinity.)

III. Advanced Regression Techniques

1. *Reduction of Variance/Colinearity*

Colinearity occurs when the matrix $X'X$ is almost singular. The ordinary least squares estimator, $\beta = (X'X)^{-1}X'Y$ can be ill conditioned since a small error in $X'Y$ will produce a large error in $\beta$. However a change in the value of $\tilde{\beta}$ along the direction of the approximate null vector will only slightly
increase the residual error. *This near degeneracy in the scaling coefficients, \( \tilde{\beta} \), explains why so many scaling laws can explain the experimental data.* We now briefly describe two methods to stabilize the ill conditioned procedure of inverting \( X^t X \).

1a. Principle Components Analysis\(^3\)

The eigenvalues and eigenvectors of \( (X^t X)^{-1} \) are calculated. The smallest eigenvalues correspond to the directions where the data was varied the least and therefore the uncertainty is the highest. The principle components estimator of \( \tilde{\beta} \) is calculated by projecting \( (X^t X) \) onto a subspace of the eigenvectors which varied sufficiently and then inverting \( X^t X \) in that subspace.

1b. Ridge Regression\(^4\)

A small multiple, \( \theta \), of the identity is added to \( X^t X \) and

\[
\beta(\theta) \equiv (X^t X + \theta I)^{-1} X^t Y
\]

For small values of \( \theta \), only the near null eigenvalues of \( X^t X \) are affected. \( \theta \) introduces a biased error into the analysis but reduces the large variation about the expected value of \( \beta \). For sufficiently small \( \theta \), the stabilizing effect of removing the degeneracy always results in a smaller mean squared error. Instead of setting some components of \( \tilde{\beta} \) to zero, ridge regression reduces all components of \( \tilde{\beta} \) proportionally to the uncertainty in that component.

The ridge trace is the plot of \( \beta(\theta) \) versus \( \theta \). In many cases, the value of \( \theta \) which stabilizes the regression coefficients is apparent. Ridge regression has the advantage over principle components that it simultaneously and continuously reduces the importance of the poorly conditioned eigenvectors.
Design of experiment is the branch of statistics which seeks to minimize the effects of colinearity by acquiring data in the directions which correspond to the smallest eigenvalues in $X'X$. By balancing the matrix $X'X$, ill conditioned data directions are minimized.

3. **Stepwise Regressions**

When there are a large number of potential regression variables, a stepwise regression procedure adds (subtracts) the additional regression variable which decreases (increases) the residual sum of squares the most (least). A stepwise regression enables the physicist to try many possibilities in an automated procedure before concentrating on the most successful variables.

3. **Robust Statistics**

Typically, the errors, $\epsilon_i$, have a much larger tail of “unlikely” events than a Gaussian distribution. These statistical outliers can have a large influence on the determination of the fit coefficients since points are weighted according to their least squares error.

Robust statistics seeks to minimize the influence of outliers by minimizing a function that grows more slowly than quadratically. One example is the absolute value estimator where $\tilde{\beta}$ is defined by

$$\min_{\tilde{\beta}} \sum_i \rho(y_i - \tilde{y}_i(\tilde{\beta})) = \min_{\tilde{\beta}} \sum_i |y_i - \tilde{\beta}x_i|$$

4. **Resampling Techniques**

When the distribution of errors is not Gaussian, resampling techniques create an empirical distribution of errors. The basic assumption is that the empirical distribution of errors is a multinominal distribution of the observed errors. The empirical distribution is then created by a Monte Carlo sampling
of the datapoints. Typically, the empirical distribution of points has a much larger tail than a Gaussian. The primary application of resampling techniques, such as the bootstrap and jackknife, is to provide error estimates for robust, non-least squares estimators.

Cross validation is a related technique, where part of the data is withheld from the initial regression and the initial regression is used to predict the new values of the energy confinement. By comparing the predicted values with the new values, an assessment of the reliability of the regression may be obtained.

III. Regression Diagnostics

To understand better which datapoints have a large role in determining the regression and how they influence the fit coefficients a number of regression diagnostics have been developed. These diagnostics can be put into two broad categories. First, diagnostics to assess whether the residual errors, \( \epsilon_i \), are uncorrelated and uniform. Second, diagnostics which assess the extent that one or more datapoints determine the least square fit and the resulting fit coefficients. All of the following diagnostics are available in the statistical regression package SAS [1].

1. Residual Plots

To assess the goodness of fit, statisticians have found it useful to plot the residual error, \( \epsilon_i \), versus a) the predicted value \( \hat{y}_i = \beta \bar{x}_i \), b) the individual independent variables \( \bar{x}_i \).

The residual plots enable the researcher a) to locate outlying datapoints where the fit is bad, b) find clustering of the residual errors into discrete groups, indicative of a correlated error structure, c) identify a heteroscedastic
error structure, i.e. the magnitude of $\epsilon_i$ depends explicitly on the dependent variable $y$ or the independent variables $\bar{x}$, i.e. $\sigma^2(y_i, \bar{x}, \bar{\alpha})$ where $\bar{\alpha}$ is the free parameter vector which describes the spatial variation of $\sigma^2$, d) identify possible higher order (quadratic) terms which should be included, and e) identify terms which have no significant impact on the goodness of fit, except possibly to fit one or two outlying points.

2. Partial Residual Plots

To isolate the dependencies of $y$ on the $j$-th independent variable $x_j$, $y$ and $x_j$ are both regressed against the other $(K - 1)$ independent variables. The residual error of the $y$ regression is then plotted against the residual error of the $x_j$ regression. After the dependencies from the other regression variables have been projected out, a clearer picture of the number of terms necessary for fitting the dependent variable.

4. Indicator Variables

In regression plots, each datapoint can be represented by a different symbol and color. The symbols and colors are chosen to correspond to important categories of data, which may exhibit different types of behavior. These "indicator" variables allow one to check for the clustering of the residual error into discrete groups.

5. Influence Diagnostic [7]

When one datapoint, $(\bar{x}_i, y_i)$, is deleted the following estimates for the free parameters occur

$$\hat{\beta}(i) = \hat{\beta} - \frac{\epsilon_i}{1 - h_i} \frac{(X^t X)^{-1} \bar{x}_i}$$
where $\epsilon_i = y_i - \beta \bar{x}_i$,

$$\hat{\sigma}^2(i) \equiv \frac{1}{(n-1) - k} \sum_{j \neq i} (y_j - \beta(i) \bar{x}_j)^2 = \frac{n - k}{(n-1) - k} \hat{\sigma}^2 - \frac{\epsilon_i^2}{(n-1-k)(1-h_i)}$$

and

$$\hat{y}_i - \hat{y}_i(i) \equiv (\beta - \beta(i)) \bar{x}_i = \frac{h_i \epsilon_i}{1-h_i}$$

To determine if these quantities are unusually large, we normalize them with their expected deviation, defined using the estimator $\hat{\sigma}(i)^2$, not $\hat{\sigma}^2$. Corresponding influence diagnostics are available for the case when a number of points are deleted.

In analyzing the L mode energy confinement data, these influence diagnostics have shown that the ASDEX data has a strong influence in determining the aspect ratio scaling.

**IV. Specialized Tools in Statistics** [3]

We conclude by briefly describing a number of special topics in statistics and their applications in fusion physics.

1. **Discriminant Analysis**

The boundary between two groups of points is determined by a least squares fit. Potential fusion applications include L-H transitions, density limits, and disruption boundaries.

2. **Clustering Analysis**

The data set is grouped into discrete subsets which have distinctly different scalings. Cluster analysis determines new natural subdivisions of the data into discrete classes. A potential fusion application is the grouping of extremely high power discharges into one or more beam-plasma categories.
V. Improved Datasets for Transport Studies

There are four major structural problems with existing tokamak databases. First, the number of free parameters which are determined by multiple tokamak comparisons is approximately equal to the number of tokamaks.

A typical multitokamak database consists of about eight different devices. From this data, seven or eight scaling coefficients plus the overall constant are determined. Unfortunately, between two and four of the plasma variables plus the overall constant are not varied on the individual tokamaks. The problem of determining the scalings of these intertokamak variables is made even more difficult by the large variation tokamak to tokamak variation in scaling laws. Reducing the number of intertokamak variables by varying plasma crosssectional ellipticity and effective ion mass in each tokamak will greatly improve the reliability of the transport analysis.

Second, important physical variables, which could have a significant impact on energy confinement, have been neglected. These include the variables which define the edge plasma such as wall material, relative neutral penetration depth and limiter/divertor configuration. Clearly, detailed single machine studies of edge effects on transport would help to define which new physical variables should be included.

Third, the design matrix for tokamak databases is very colinear. Conducting experiments having significant variation in the data directions which have not been varied can greatly reduce the illconditioning of the linear regression. For tokamak scaling laws, the largest uncertainty is in the geometric direction of simultaneously increasing aspect ratio and the ellipticity of the plasma crosssection.
Finally, the variation of the scaling coefficients from device to device may have systematic trends or a highly non-Gaussian random nature. The most important systematic trend is a decrease in the value of current scaling coefficient, $\beta_5$, with device size. If $\beta_5$ does decrease with device size the impact on C.I.T. and I.T.E.R. performance is large. The non-Gaussian nature of device performance is indicated by some tokamak such as DIII and JET having good performance with nearly identical scalings while other tokamaks have scalings which differ greatly from this typical scaling. Physics studies could elucidate the reason for this difference in performance. In the meantime, we are left with the choice of either including only the typical tokamaks or all tokamaks in our statistical analysis.

2. Riedel K.S. submitted to Nuclear Fusion


5. Riedel K.S. and Kaye S. in progress


Acknowledgment

The author wishes to thank S. Kaye for discussions on his results on scaling laws. The author’s understanding of statistics has benefitted significantly from his previous collaborations with O. Kardaun, K. Lackner and P. McCarthy and from discussions with C. Bolton, W. Miner and H. Weitzner. The manuscript has benefited from a critical reading by H Weitzner. This work was performed under U.S. Dept. of Energy grant # DE-FG02-86ER53223.
TOKAMAK TO TOKAMAK VARIATION AND COLINEARITY IN SCALING LAWS

K.S. Riedel
Courant Institute of Mathematical Sciences
New York University
New York, New York 10012

Abstract

The large multiplicity of scaling laws which almost fit the data is associated with the near degeneracy of the data covariance matrix. The traditional analysis assumes that only discharge to discharge variations are important. Using Swamy's random coefficient model, tokamak to tokamak variations are included in the statistical analysis. Scaling laws for energy confinement are among the oldest and most widely used tools in studying tokamak transport. In this letter, we present two statistical techniques which substantially enhance the reliability of the standard analysis.¹

Traditionally, one assumes that the energy confinement time, $\tau_E$ scales as $\tau_E = c_0 a^{\beta_1} R^{\beta_2} \kappa^{\beta_3} M^{\beta_4} J^{\beta_5} P^{\beta_6} B_t^{\beta_7} \pi^{\beta_8}$. By defining $y_i$ to be $\ln \tau_E$, and $\bar{x}_i$ to contain the logarithms of the global plasma variables, the analysis is reduced to the solution of an overdetermined system of equations:

$$y_i = \bar{x}_i \bar{\beta} + \epsilon_i$$

where $\bar{\beta}$ is a $m$ dimensional vector of free parameters and $\epsilon_i$ is the error associated with the $i$-th datapoint. The matrix whose rows consist of $\bar{x}_i^t$ is called the design matrix and is denoted by $X$. 

15
To proceed further, assumptions on the covariance structure of the errors, \( \epsilon_i \), must be made. We assume the error matrix, \( \Sigma \), is given, where \( \text{cov}(\epsilon_i \epsilon_j) = \Sigma_{ij} \).

The parameter vector, \( \beta \), can now be determined using generalized least squares (GLS). Namely, \( \beta \) is determined by minimizing

\[
\text{min}(\tilde{y} - X\tilde{\beta})^t\Sigma^{-1}(\tilde{y} - X\beta).
\]  

(2)

The GLS estimator is

\[
\tilde{\beta} = (X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}\tilde{y}
\]  

(3)

and is the best unbiased estimator within the class of linear estimators of \( \tilde{\beta} \).

Typically, \(^1\) the error matrix is assumed to be a multiple of the identity matrix, which corresponds to assuming the departures of the observed values, \( y_i \), from the linear scaling are statistically independent. Thus device to device differences are neglected in favor of discharge to discharge differences.

II. Colinearity and the Multiplicity of Scaling Laws

We now distinguish two types of independent variables. Within device variables are plasma parameters which can be varied in a single tokamak, such as toroidal magnetic field \( B_t \), plasma current \( I_p \), line averaged density, \( \bar{n} \), and total injected power \( P \). Interdevice variables are not varied on a given tokamak. The major and minor radii \( R \) and \( a \) are examples. Unfortunately, the plasma cross section ellipticity \( \epsilon \) and the isotope mixture \( M \) are frequently interdevice variables as well. We denote the interdevice variables as \( \tilde{x}_o \) and the within device variables as \( \tilde{x}_1 \).
Multitokamak datasets are often poorly defined in the sense that the number of different tokamaks, $N_T$, is only slightly larger than the number of intertokamak variables, $m_o$. When this occurs, the data covariance matrix, $X' \Sigma^{-1} X$ is usually close to singular.

When the data covariance matrix is almost degenerate, the GLS estimator of $\beta$ is ill-conditioned. This occurs because the inversion of $X' \Sigma^{-1} X$ multiplies components of $X' \Sigma^{-1} y$ by $\lambda_j^{-1}$ where $\lambda_j$ are the eigenvalues of $X' \Sigma^{-1} X$. Thus small errors in the observed values of $\bar{y}$ are multiplied by $\lambda_j^{-1}$.

Statisticians refer to the problem of the near degeneracy of the data matrix as “colinearity”. We briefly describe the two most commonly used statistical techniques to reduce colinearity problems.

Principle component analysis decomposes the data covariance matrix $X' \Sigma^{-1} X$ into $E' \Lambda E$. The columns of $E$ are the eigenvector of the covariance matrix and $\Lambda$ is a $m \times m$ diagonal matrix whose entries consist of the corresponding eigenvalues $\lambda_j$ of $\bar{e}_j$. The eigenvectors $\bar{e}_j$ are also called the principle components. The eigenvalues $\lambda_j$ are equal to the variance of their respective eigenvectors.

Small eigenvalues correspond to data directions which have only been slightly varied in the dataset. The projection of the parameter vector $\bar{\beta}$ onto these directions is poorly determined. If an eigenvalue $\lambda_m$ is identically zero, then an arbitrary multiple of the eigenvector $e_m$ may be added to the parameter vector $\bar{\beta}$ without affecting the goodness of fit.

When one or more eigenvalues $\lambda_j$ are almost zero, then an arbitrary combination of these eigenvectors may be added to $\bar{\beta}$ while only slightly degrading
the goodness of fit. This explains why a large number of distinctly different scaling laws can describe tokamak data with only slightly different goodnesses of fit.

The $k$ principle component estimator is defined by setting the $(m - k)$ most ill defined projections of $\tilde{\beta}$ to zero. Mathematically

$$\tilde{\beta}_{PC}^k = (E'\Lambda_k^{-1}E)X'E \Sigma^{-1} \tilde{y}$$

where $\Lambda_k$ has the $(m - k)$ smallest eigenvalues, $\lambda_j$ set to zero.

Ridge regression is a related method where a small multiple of the identity is added to the data covariance matrix. Thus the ridge estimator is

$$\tilde{\beta}(\theta) \equiv [X'\Sigma^{-1}X + \theta I]^{-1}X'\Sigma^{-1} \tilde{y}$$

The ridge trace is the plot of the individual components, $\tilde{\beta}_i(\theta)$ vs. $\theta$ to determine graphically when $\theta$ is large enough to stabilize the fit coefficients.

Both ridge regression and principle components stabilize the ill conditioned least squares estimator by shrinking the ill conditioned components. Mathematically, there always exists a small enough $\theta$ such that the ridge estimator has a smaller variance than the least squares estimator. In other words, the small amount of bias introduced by $\theta$ is more than compensated by the improvement in numerical stability.

In summary, the near degeneracy of $X' \Sigma^{-1} X$ means that many parameter vectors, $\tilde{\beta}$ (scaling laws) can describe the data. Ridge regression sets the poorly defined components of $\tilde{\beta}$ to a much smaller value than the least squares estimator. Finally, we note that the independent variables $\bar{x}_i$ should be centered about the weighted mean value, $\bar{x}$ for both principle components and ridge regression.
III. Tokamak to Tokamak Variation

As discussed earlier, almost all scaling studies assume that the errors (departures from the fitted values) are independent from discharge to discharge. This neglects the component of error which remains constant on any one device.

If this tokamak to tokamak variation is attributable to one or more important factors such as wall material or limiter/divertor configuration, statistics is of little help in analyzing confinement. If, however, the tokamak to tokamak differences are due to many small factors, then it is possible to analyze the data by averaging over the small factors.

We now present and apply a slight generalization of Swamy's random coefficient model. In this model, we assume that the scaling coefficients for each tokamak, $\bar{\beta}_i$, have an \textit{a priori} distribution where $E[\bar{\beta}_i] = \bar{\beta}$ and

$$E[(\bar{\beta}_i - \bar{\beta})(\bar{\beta}_j - \bar{\beta})^t] = \Delta \delta_{ij}$$

where $\Delta$ is an $m \times m$ covariance matrix.

Given the covariance, $\Delta$, the mean parameter vector $\bar{\beta}$ may be estimated using generalized least squares. The covariance matrix $\Delta$ may then be estimated iteratively from the residuals, $\bar{\beta}_i - \bar{\beta}$.

The inter-tokamak variables, $\bar{x}_o$, are usually too poorly determined to rule out fixed values of the parameters. Thus we adopt a hybrid model where $\bar{\beta}_o$ is fixed and $\bar{\beta}_1$ is distributed about $\bar{\beta}_1$ with covariance matrix $\Delta_1$. We do assume the overall scaling constant is random and included in $\bar{x}_1$.

Following [6], we divide our design matrix into separate blocks, $x_j$, for each tokamak. The block $X_j$ is then a $k_j \times (m_o + m_1)$ matrix where $k_j$ is
the number of datapoints from the j-th tokamak. $X_j$ is then divided into
$(X_{o,j}, X_{1,j})$.

The error matrix for the j-th block reduces to

$$\Sigma_{ii} = \sigma_{ii}^2 I_i + X_{1,i} \Delta_1 X_{1,i}^t$$

The GLS estimator for $\tilde{\beta}$ is

$$\tilde{\beta} = (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} \tilde{y} = \left( \sum_{j=1}^{N} X_j^t \Sigma_{jj}^{-1} X_j \right)^{-1} \sum_{i=1}^{N} X_i^t \Sigma_{ii}^{-1} \tilde{y}_i$$

Using results from [6], this simplifies to

$$\tilde{\beta} = \left( \sum_{j=1}^{N} \left( \frac{F_j^t}{I_m} \right) \left( \sigma_{jj}^2 E_j + \tilde{\Delta}_j \right)^{-1} (F_j, I_i) \right)^{-1} \sum_{i=1}^{N} \left( \frac{F_i^t}{I_m} \right) \left( \sigma_{ii}^2 E_i + \tilde{\Delta}_i \right)^{-1} b_i$$

where $E_j = (X_{1,j}^t X_{1,j})^{-1}$, $F_j = k_j E_j \tilde{x}_j \tilde{x}_j^t$, and $\tilde{\Delta} = \sigma_{jj}^{-1} P_{x,j} \Delta_1 P_{x,j}$. We have used the standard statistical notation that "\~" denotes the Moore-Penrose generalized inverse and $P_{x,j}$ is the projection onto the row space of $X_j$.

The covariance matrix, $\Delta$, may then be iteratively estimated from

$$\sum_{i=1}^{N} Q \Delta_1 Q_i = - \frac{N}{N-1} S_b - \sum_{i=1}^{N} \hat{\sigma}_{ii}(X_{1,i}^t X_{1,i})^{-1}$$

$$S_b = \sum_{i=1}^{N} (b_i^{-} - \hat{b}_i^{-})(\hat{b}_i^{-} - \hat{b}_i^{-})'$$

To take into account the effects of different $\tilde{x}_o$ for different tokamaks $\hat{b}_j^{-}$ is defined as

$$\hat{b}_i^{-} = Q_i \begin{pmatrix} \tilde{\beta}_1 + \tilde{\beta}_o \cdot \tilde{x}_o \ (1) \\ \vdots \\ 0 \end{pmatrix}$$
We note that the error in our estimate of $\tilde{\beta}$ tends to zero as the square root of the number of tokamaks not as the square root of the number of datapoints.

The predicted performance of any new tokamak, with parameters $\tilde{u} = (\tilde{u}_o, \tilde{u}_1)$ is

$$\hat{y}(\tilde{u}) = \tilde{\beta} \cdot \tilde{u} \pm \sqrt{\tilde{u}^2 \left( \sum_{j=1}^{N} \left( \frac{F_j}{I_{m_1}} \right) (\sigma_{jj} E_j + \hat{A}_j)^- (F_j, I_{m_1}) \right)^{-1}} \tilde{u} + \tilde{u}_1 \Delta_1 \tilde{u}_1$$

The first term in the square root represents the uncertainty in the estimate of $\tilde{\beta}$. The second term accounts for the intrinsic variability of tokamaks when only the parameters in $\tilde{u}$ are specified.

We close this section by noting that the ordinary least squares estimator for $\tilde{\beta}$ does tend to the correct value of $\tilde{\beta}$ as the number of tokamaks increases. However the GLS estimator using the random coefficient model is a more efficient estimate of $\tilde{\beta}$. Finally the error estimates for $\tilde{\beta}$ and $\hat{y}(\tilde{u})$ even scale incorrectly when ordinary least squares is used.

IV. Discussion

In no way do we want to pretend that a complex dynamical system such as a tokamak is completely describable as a linear system with six or eight degrees of freedoms. We prefer to view scaling laws as a Taylor series expansion about a particular point in parameter space, including only the most important plasma variables.

Scaling laws describe subregions of parameter space where a single loss mechanism or a special combination of losses is dominant. Thus scaling laws
should not be trusted in extrapolating to new regions in parameter space. Luckily, scaling laws are most often used for the case when the physical, dimensionless variables are only varied slightly and the major extrapolation is in device size. Furthermore in most multimachine databases, the component of the database which varies the most is strongly associated with device size.

Acknowledgment

The author wishes to thank S. Kaye for several detailed discussions of his previous analyses of tokamak data. Dr. Kaye presented the author with a table of scaling laws for individual tokamaks. This table help to stimulate the author to realize tokamak to tokamak correlations need to be included. Professor K. Lackner originally suggested the use of principle components in their joint work on temperature profile shape scalings.

This work was supported under U.S. Department of Energy Grant No. DE-FG02-86ER53223.

REFERENCES


THE SWAMY RANDOM COEFFICIENT MODEL
WITH SINGULAR COVARIANCE MATRICES

K.S. Riedel
Courant Institute of Mathematical Sciences
New York University
New York, N.Y. 10012

Abstract

The Swamy random coefficient model is generalized to the case
where the covariance matrices $X_j'X_j$ for each individual and the dis-
ruption matrix $\Delta$ for the free parameters $\beta$ are singular.

In many physical or economic systems, the free parameters $\beta_j$ vary from
individual/individual device to individual (device). If the differences in the
parameter vectors $\beta_j$ are due to many small differences in the individuals,
instead of one or two significant latent variables, a statistical treatment is
still possible. In these cases, the random coefficient model,\cite{1,2} introduced
by Swamy, is applied. Swamy assumes that the parameter vectors $\beta_j$ are
randomly distributed about a mean vector $\beta$. More precisely,

$$\beta_i = \beta + \mu_i$$  \hspace{1cm} (1)

where $E[\mu_i] = 0$, $E[\mu_i\mu_i^t] = \Delta$, and $E[\mu_i, \mu_j] = 0$, $i \neq j$.

In many situations, not all independent variables can be varied in the in-
dividual experiments. One such problem is the extrapolation of performance
parameters in controlled nuclear fusion experiments.\cite{3} A database consisting
of a large number of datapoints from eight tokamak experiments has been
collected. The energy confinement time is assumed to be a geometric function of certain bulk variables such as plasma current, magnetic field and device size. However, several of these independent variables, namely the device size and device shape, cannot be varied in the individual experiments.

In this note, we generalized Swamy's random coefficient model to the case where a number of coefficients can be determined only through interindividual comparisons. We assume that the $i$-th individual can be modeled as

$$\tilde{y}_i = X_i \tilde{\beta}_i + \tilde{\epsilon}_i$$

where $\tilde{y}_i$ and $\tilde{\epsilon}_i$ are $(K_i \times 1)$ vectors and $\tilde{\epsilon}_i$ is distributed $E[\tilde{\epsilon}_i, \tilde{\epsilon}_i^t] = \sigma_\epsilon \delta_{ij} I_{K_i}$.

In contrast to the standard random coefficient model, we assume the design matrix has a block structure:

$$X_i = (X_{o_i}, X_{1_i})$$

where $X_{o_i}$ is a $k_i \times m_o$ matrix and $X_{1_i}$ is a $k_i \times m_1$ matrix.

We further assume that none of the independent variables in $X_{o_i}$ vary over the $i$-th individual. Thus we can write $X_{o_i}$ as

$$X_{o_i} = \tilde{j}_i \bigotimes \tilde{x}_{o_i}^t$$

where $\tilde{j}_i$ is the $K_i$ vector consisting entirely of ones.

For many cases, including ours, the number of individuals is only slightly larger than $m_o$, the number of parameters which can only be determined by interindividual comparisons. Thus allowing these parameters, $\beta_{o_i}$ to be random is ill conditioned. Therefore we assume only the parameters, $\tilde{\beta}_i$,
describing $X_j$ are random, i.e.$^1$

$$\Delta \equiv \begin{pmatrix} 0 & 0 \\ 0 & \Delta_1 \end{pmatrix}$$

Finally, we assume that the intercept is a random variable, included in $X_j$. This assumption significantly simplifies the resulting algebra since it implies that the column space of $X_{\alpha_i}$, $M(X_{\alpha_i})$, is contained in the column space of $X_{1_i}$.

The standard generalized least squares estimator of the mean parameter vector is

$$\tilde{\beta} = (X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}\tilde{y} = \left(\sum_{j=1}^{N} X_j^t\Sigma_{jj}^{-1}X_j \right)^{-1} \sum_{i=1}^{N} X_i^t\Sigma_{ii}^{-1}\tilde{y}_i$$

We make one further generalization of Swamy's$^3$ results by assuming the individual covariance matrices $X_{1_j}^t X_1$, may be singular.

Due to the assumed block structure of $\Delta$, the covariance matrix for the $i$-th error structure, $\Sigma_{ii}$, simplifies to

$$\Sigma_{ii} = \sigma_i^2 I_i + X_1 \Delta_1 X_1^t$$

To simplify these expressions we note that Rao's equality (Ref. [4], p. 33) generalized to

$$(\sigma^2 I + X\Delta X^t)^{-1} = \sigma^{-2}[I - XEX^t] + XE(\sigma^2 E + \hat{\Delta})^{-1}EX^t$$

where $E$ is the Moore-Penrose generalized inverse of $X^t X$, $E = (X^t X)^{-1}$ and $\hat{\Delta}$ is the normalized projection of $\Delta$ onto the row space of $X$, $\hat{\Delta} = P_x \Delta P_x$.

We note that $XEX^t$ is the projection operator onto the column space of $X$. Thus the first two terms of eq. (7) are the projection perpendicular...
to the column space of $X, P_{\perp}$. Since $M(X_\alpha) \subset M(X_1), P_{\perp}X_\alpha = 0$, where $P_{\perp} = I_i - X_iE_i$. Thus

$$X_j^t\Sigma_j^{-1}X_j = \begin{pmatrix} X_{o, j}^t \\ X_{1, j}^t \end{pmatrix} \Sigma_j^{-1}(X_{o, j}, X_{1, j}) = \begin{pmatrix} F_j^t \\ I_{m_1} \end{pmatrix} (\sigma_{jj}^2E_j + \hat{\Delta}_j)^-(F_j, I_{m_1}) \quad (7)$$

where $E_j \equiv (X_{1, j}^tX_{1, j})^{-1}F_j \equiv k_jE_j(X_{1, j}, \bar{x}_{o, j})$. We have used $X_{o, j}^tX_{1, j} = k_j\bar{x}_{o, j}\bar{x}_{1, j}^t$, where $\bar{x}_{1, j}$ is the mean value of $\bar{x}_i$ for the $j$-th individual. Similarly, $X_j^t\Sigma_j^{-1}y_j$ simplifies to

$$\begin{pmatrix} F_j^t \\ I_{m_1} \end{pmatrix} (\sigma_{jj}^2E_j + \hat{\Delta}_j)^-b_j^- \quad (8)$$

where $b_j^-$ is the generalized least squares estimator obtained using the Moore-Penrose inverse.

Thus the generalized least squares estimator for $\beta$ simplifies to

$$\Beta = \left( \sum_{j=1}^{N} \begin{pmatrix} F_j^t \\ I_{m_1} \end{pmatrix} (\sigma_{jj}^2E_j + \hat{\Delta}_j)^-(F_j, I_{1, j}) \right)^{-1} \sum_{i=1}^{N} \begin{pmatrix} F_i^t \\ I_{m_1} \end{pmatrix} (\sigma_{ii}^2E_i + \hat{\Delta}_i)^-b_i^- \quad (9)$$

Swamy's estimator for the within individual variation remains unchanged

$$\hat{\sigma}_{ii}^2 = \frac{\bar{y}_i^tP_{\perp}\bar{y}_i}{k_i - m_i} \quad (10)$$

where $m_i$ is the actual number of degrees of freedom in the $i$ th individual regression. However the estimator for $\Delta_1$ becomes significantly more complicated when $X_{1, i}^tX_1, i$ is singular. The variance matrix is defined from

$$\sum_{i=1}^{N} Q_i\Delta_iQ_i = \frac{N}{N - 1} S_b - \sum_{i=1}^{N} \hat{\sigma}_{ii}^2(X_{1, i}^tX_1, i)^- \quad (11)$$
where $Q_i$ projects onto the column space of $X_i^T X_i$ and

$$ S_b = \sum_{i=1}^{N} (\hat{b}_i - \hat{b}_i^-)(\hat{b}_i^- - \hat{b}_i^-)' \quad (12) $$

Our definition of $S_b$ differs from Swamy’s because we have used our estimated $\hat{\beta}$ vector to iteratively define $\hat{b}_i^-$. This is not only a more efficient estimator, but also necessary to take into account the effects of different $\bar{x}_o$ for different individuals. The precise definition of $\hat{b}_i^-$ is

$$ \hat{b}_i^- = Q_i \left( \tilde{\beta} + \hat{\beta}_o \cdot \bar{x}_o \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} \right) \quad (13) $$

The second term corrects the constant term in $\hat{b}_i$ for $\bar{x}_o$.

If $X_i^T X_i$ is never singular, the LHS of (11) reduces to $N \Delta_1$. Unfortunately, for degenerate cases we must solve a $m_1^2 \times m_1^2$ matrix problem. We rewrite eq. (11) as

$$ \left( \sum_{i=1}^{N} Q_i \bigotimes Q_i \right) \tilde{\Delta}_1 = \tilde{R} \quad (11') $$

where $\bigotimes$ denotes the Kronecker product and

$$ \tilde{\Delta}_1' = (\Delta_{11}, \Delta_{21} \ldots \Delta_{m_11}, \Delta_{12} \ldots \Delta_{m_1m_1}) . $$

The resulting linear system may be solved to yield an estimate for $\Delta_1$.

**Acknowledgment**

This work was performed under U.S. Department of Energy Grant No. DE-FG02-86ER-53223.
REFERENCES


APPENDIX:

Principal Components Analysis for Degenerate Uncentered Matrices

We assume the $k$-th individual experiment is of the form

$$\tilde{x}_i^t (1, \tilde{x}^t + \tilde{x}_i^t)$$

where $\tilde{x}, \tilde{x}_i$ are $m_1$ vectors of independent variables which can be varied in the individual experiments. Let $X_T$ denote the $n \times (m_1 + 1)$ data matrix and $X_V$ be the $n \times m_1$ data matrix consisting of the last $m_1$ columns of $X_T$. We define the normalized, centered data matrix $\tilde{X}$ as

$$\tilde{X} = (X_V - 1_n \bar{x}^t)/\sqrt{n}$$

The data matrix has the form

$$X_T^t X_T = n \begin{pmatrix} 1 & \bar{x}^t \\ \bar{x} & \tilde{X}^t \tilde{X} + \bar{x} \bar{x}^t \end{pmatrix}$$

We now derive a simple expression for the Moore-Penrose generalized inverse of $X_T^t X_T$ when $\tilde{X}^t \tilde{X}$ is degenerate.

**Lemma.** Let $E$ be a $\ell_0 \times \ell_0$ symmetric non-negative definite matrix of rank $\ell_1$ and $\tilde{v}, \tilde{w}$ be $\ell$ vectors. Assume $v \in M(E)$ and $\bar{w}$ is orthogonal to $M(E)$. Let

$$\Omega \equiv E + (\tilde{v} + \bar{w})(\tilde{v} + \bar{w})^t,$$
the Moore-Penrose generalized inverse is

\[ \Omega_{MP} = E_{MP} - \frac{(\tilde{w}^\top E_{MP} + E_{MP} \tilde{v} \tilde{w}^\top)}{|w|^2} + (1 + \tilde{v} E_{MP} \tilde{v}) \frac{\tilde{w} \tilde{w}^\top}{|w|^4} \]

To determine the generalized inverse for non-centered matrices, we set

\[ E \equiv \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{X}^\top \tilde{X} \end{pmatrix}, \quad \tilde{v} + \tilde{w} = \begin{pmatrix} 1 \\ \ddots \\ \tilde{x} \end{pmatrix} \]

Let \( e_1, \ldots, e_n \) be the non-normalized null vectors of \( \tilde{X}^\top \tilde{X} \), we define

\[ \tilde{x}_f = \sum_{i=1}^{n_0} \tilde{e}_i \cdot \tilde{x}, \quad \tilde{w} = \begin{pmatrix} 1 \\ \ddots \\ \tilde{x}_f \end{pmatrix}. \]

If \( \tilde{X}^\top \tilde{X} \) is nearly degenerate, the principal components regression can be obtained by projecting out the nearly degenerate vectors.

Finally, the projection operator onto the row space of \( X_T \) is

\[ P_{X_T} = E_{MP} + \frac{\tilde{w} \tilde{w}^\top}{|w|^2} . \]
Riedel, Kurt S
Advanced statistics for tokamak transport colinearity and... c.1
Library
CIMS