Computing a Test Statistic

Chris Paige

Joint work with   Xiao-Wen Chang

Computer Science,   McGill University

Montreal, Quebec, Canada

ERCIM WG MATRIX COMPUTATIONS
AND STATISTICS WORKSHOP

Prague August 27-29, 2004.

Brief Summary

- Statistical testing can warn us of disturbances in measurements.
- The generalized likelihood ratio (GLR) test statistic δ_{TS} is a good indicator.
- Standard ways of computing

$$\delta_{\text{TS}} = \sigma^{-2}(r_0^T V^{-1} r_0 - r_a^T V^{-1} r_a)$$

  are extremely numerically unstable.
- We give a numerically stable method for this statistic, & the estimates of the parameter vectors.
- This method works when V is singular, & has other uses.

Notation

- GLLS = "Generalized Linear Least Squares".

- GLR = "Generalized Likelihood Ratio".

- $\mathcal{E}\{\cdot\}$ the expected value, $\mathrm{cov}\{\cdot\}$ the covariance,
$$\mathrm{cov}\{x\} \equiv \mathcal{E}\{(x - \mathcal{E}\{x\})(x - \mathcal{E}\{x\})^T\}.$$

- $v \sim \mathcal{N}(\bar{v}, \sigma^2 V)$: v is a random vector,
  normally distributed,
  with mean $\bar{v}$ and covariance $\sigma^2 V$.

Linear model under H_0

Linear model under the null-hypothesis H_0:

$$\mathrm{H}_0 : \quad y = Ax + v, \quad v \sim \mathcal{N}(0, \sigma^2 V),$$

where

- $y \in \Re^m$ random measurement vector,
- $A \in \Re^{m \times n}$, $m \geq n$, known design matrix,
- $x \in \Re^n$ unknown parameter vector,
- $v \in \Re^m$ random noise vector,
- $V \in \Re^{m \times m}$ known symm. pos. def. matrix.

Possible outliers may invalidate estimation results.

Linear model under H_a

Restrict misspecification to the mean of y,
i.e. an error of additive nature.

The alternative hypothesis H_a then reads

$$\mathrm{H}_a: \quad y = Ax + Cd + v, \quad v \sim \mathcal{N}(0, \sigma^2 V),$$

where

- known matrix $C \in \Re^{m \times q}$ specifies the type of model error that can occur,
- $[A, C]$ has full column rank (fcr),
- $d \in \Re^q$ is an unknown constant vector.

Special Case

$$y = Ax + Cd + v, \quad v \sim \mathcal{N}(0, \sigma^2 V).$$

- $V = I$, and a possible outlier in only one measurement (which one is unknown). The case leads to the "w-test statistic".
- Taking $\quad C = e_i, \quad i = 1, \ldots, m,$
  $e_i \equiv (0, \ldots, 0, 1, 0, \ldots, 0)^T,$
  gives m alternative hypotheses:

$$\mathrm{H}_i : \quad y = Ax + e_i \delta_i + v, \quad v \sim \mathcal{N}(0, \sigma^2 I).$$

MLE and BLUE

The Maximum Likelihood Estimates (MLE)
(and Best Linear Unbiased Estimates (BLUE))
x_0 of x under H_0 ,
& $\{x_a, d_a\}$ of $\{x, d\}$ under H_a ,
solve respectively:

$$\mathrm{GLLS}_0 : \min_x \{(y - Ax)^T V^{-1}(y - Ax) = r^T V^{-1} r\},$$

where $r \equiv y - Ax,$ and

$$\mathrm{GLLS}_a : \min_{x,d} (y - Ax - Cd)^T V^{-1}(y - Ax - Cd).$$

Test Statistic

Write $r_0 \equiv y - Ax_0$, $r_a \equiv y - Ax_a - Cd_a$.

GLR *test statistic*, testing H_0 against H_a, is:

$$\delta_{TS} \equiv \sigma^{-2}(r_0^T V^{-1} r_0 - r_a^T V^{-1} r_a) \geq 0.$$

The extra term Cd in $y = Ax + Cd + v$ *decreases* $r_0^T V^{-1} r_0$ to become $r_a^T V^{-1} r_a$. A large change shows Cd is *significant*.

Given a threshold θ (determined by the requirements of the specific application).

- When $\delta_{TS} > \theta$, reject H_0 in favour of H_a.
- Otherwise accept H_0.

Well Known Fact:

A test statistic doesn't need high accuracy. So

"Any reasonable method can be used for

$$\delta_{\mathrm{TS}} = \sigma^{-2}(r_0^T V^{-1} r_0 - r_a^T V^{-1} r_a) \quad \text{" ??}$$

Well Known Fact:

A test statistic doesn't need high accuracy. So

"Any reasonable method can be used for

$$\delta_{\mathrm{TS}} = \sigma^{-2}(r_0^T V^{-1} r_0 - r_a^T V^{-1} r_a) \quad \text{" ??}$$

NOT SO.

Harmless Example

For $\epsilon \approx 2 * 10^{-16}$ the following example looks harmless, with $\kappa_2(A) \approx 4.44$ and $\kappa_2(V) \approx 33,000$:

$$y = \begin{bmatrix} 5.48223618514353 \\ 0.90878847962427 \\ 25.94493985828999 \\ 5.91432884267696 \end{bmatrix},$$

$$A = \begin{bmatrix} 0.73591311187945 & 1.98690117078759 \\ 0.01599725305719 & 1.85723508466859 \\ 0.15632753551635 & 2.35754116764473 \\ 0.65858764131884 & 0.21189908130823 \end{bmatrix}, \quad C = \begin{bmatrix} 2.72281454895206 \\ 1.87323437713309 \\ 2.51387627488834 \\ 0.87049065185808 \end{bmatrix},$$

$$\sigma = 1, \qquad V =$$

$$\begin{bmatrix} 9.140496886810 & -5.179920639550 & 22.018803142087 & -2.448166448348 \\ -5.179920639550 & 31.269615846900 & -38.726345506531 & 1.768700005165 \\ 22.018803142087 & -38.726345506531 & 244.102164709880 & 43.463631186108 \\ -2.448166448348 & 1.768700005165 & 43.463631186108 & 15.497722556410 \end{bmatrix}$$

Harmless Example, ctd.

Exact solution and test statistic:

$$x_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \delta_{\mathrm{TS}} = \delta_0 - \delta_a = 2 - 1 = 1.$$

From $\quad x_0 = (A^T V^{-1} A)^{-1} A^T V^{-1} y, \quad r_0 = y - A x_0,$
mathematically (after some cancellation):

$$\delta_0 \equiv r_0^T V^{-1} r_0 = y^T V^{-1} y - y^T V^{-1} A (A^T V^{-1} A)^{-1} A^T V^{-1} y.$$

We computed this using the Matlab code:

```
V1=inv(V);   G=A'*V1;   W=G*A;
d=y'*(V1*y)-y'*(G'*(inv(W)*(G*y)));
```

We computed $\delta_a \equiv r_a^T V^{-1} r_a$ similarly.

Harmless Example, ctd

We saw in theory $\delta_{\mathrm{TS}} = \delta_0 - \delta_a = 1 \geq 0.$

But the Matlab result with $\epsilon \approx 2 * 10^{-16}$ was

$$\delta_{\mathrm{TS}} = \delta_0 - \delta_a \approx -14,$$

(instead of 1), an obviously nonsensical result.

A simple reminder:
combining a sequence of individually reliable computations does not necessarily lead to an overall numerically acceptable computation.

Harmless Example, ctd.

Another "obvious" approach:

Computing $\quad x_0 = (A^T V^{-1} A)^{-1} A^T V^{-1} y,$
then $\quad r_0 = y - A x_0, \quad$ then $\quad \delta_0 \equiv r_0^T V^{-1} r_0,$
(and similarly for δ_a), gave

$$\delta_{\mathrm{TS}} = \delta_0 - \delta_a \approx 0.44.$$

Our method (to be given later) gave (to 15 dec. dig.)

$$\delta_{\mathrm{TS}} = 1.00000000078345,$$

$$x_0 = [1.00000000000001, 2.00000000000001]^T.$$

What We Learnt:

- It is important to use a numerically stable algorithm for computing δ_{TS}.
- All the more so in real time applications when IEEE standard double precision floating point arithmetic is not available.
- It is probably worthwhile making a numerically stable code available.

Paige's 1978 GLLS Formulation

Factor the symmetric positive definite V

$$V = BB^T, \qquad B \in \Re^{m \times m}.$$

E.g. the Cholesky factorization of V gives a B.

Then for $v \sim \mathcal{N}(0, \sigma^2 V)$ we can write

$$v \equiv Bu, \qquad u \sim \mathcal{N}(0, \sigma^2 I).$$

The linear models can be replaced by

$$\mathrm{H}_0 : \; y = Ax + Bu, \qquad u \sim \mathcal{N}(0, \sigma^2 I),$$
$$\mathrm{H}_a : \; y = Ax + Cd + Bu, \qquad u \sim \mathcal{N}(0, \sigma^2 I).$$

GLLS **Formulations, ctd.**

With $V^{-1} = B^{-T}B^{-1}$, the previous GLLS_0 is:

$$\min_x \{(y-Ax)^T V^{-1}(y-Ax) = \| \underbrace{B^{-1}(y-Ax)}_{u} \|_2^2\},$$

& problems GLLS_0, GLLS_a can be rewritten:

$$\mathrm{GLLS}_0 : \min_{u,x} \|u\|_2^2 \quad \text{subject to} \;\; y = Ax + Bu;$$

$$\mathrm{GLLS}_a : \min_{u,x,d} \|u\|_2^2 \quad \text{s.t.} \;\; y = Ax + Cd + Bu.$$

GLLS version of $\delta_{\rm TS}$

$$\mathrm{GLLS}_0: \min_{u,x} \|u\|_2^2 \quad \text{s.t. } y = Ax + Bu;$$
$$\mathrm{GLLS}_a: \min_{u,x,d} \|u\|_2^2 \quad \text{s.t. } y = Ax + Cd + Bu.$$

Let u_0 & u_a be the optimal u for GLLS_0 & GLLS_a, so

$$u_0 = B^{-1}(y - Ax_0) = B^{-1}r_0,$$
$$u_a = B^{-1}(y - Ax_a - Cd_a) = B^{-1}r_a.$$

These with $V^{-1} = B^{-T}B^{-1}$ show

$$\delta_{\rm TS} = \sigma^{-2}(r_0^T V^{-1} r_0 - r_a^T V^{-1} r_a) = \sigma^{-2}(\|u_0\|_2^2 - \|u_a\|_2^2).$$

No inverse of B or V appears in the above GLLS_0, GLLS_a, or this last expression—1 key for stability.

Solution Derivation

Use the Generalized QR (GQR) of $[A, C]$ & B:

The QR factorization of fcr $m \times (n{+}q)$ $[A, C]$:

$$\begin{array}{c} P^T[A,C] \\ \quad\; n \quad q \end{array} = \begin{array}{cl} \left[\begin{array}{cc} U_A & U_{AC} \\ 0 & U_C \\ 0 & 0 \end{array}\right] & \begin{array}{l} n \\ q \\ m-n-q \end{array} \\ \begin{array}{cc} n & \quad q \end{array} & \end{array} , \quad P^{-1} = P^T;$$

and the RQ factorization of $m \times m$ $P^T B$:

$$P^T B Q = \begin{array}{cl} \left[\begin{array}{ccc} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{array}\right] & \begin{array}{l} n \\ q \\ m-n-q \end{array} \\ \begin{array}{ccc} n & \quad q & \quad m-n-q \end{array} & \end{array} , \quad Q^{-1} = Q^T.$$

Solution Derivation, ctd.

GQR transforms GLLS_a (and GLLS_0 as well):

$$\min \|u\|_2^2 \quad \text{s.t.} \;\; y = Ax + Cd + Bu.$$

Solution Derivation, ctd.

GQR transforms GLLS_a (and GLLS_0 as well):

$$\min \|u\|_2^2 \quad \text{s.t.} \; y = Ax + Cd + Bu.$$

$$\Downarrow$$

$$\min \|u\|_2^2 \; \text{ s.t. } \; \underbrace{P^T y}_{z} = P^T Ax + P^T Cd + P^T BQ \underbrace{Q^T u}_{w}$$

Solution Derivation, ctd.

GQR transforms GLLS_a (and GLLS_0 as well):

$$\min \|u\|_2^2 \quad \text{s.t.} \;\; y = Ax + Cd + Bu.$$

$$\Downarrow$$

$$\min \|u\|_2^2 \;\; \text{s.t.} \quad \underbrace{P^T y}_{z} = P^T Ax + P^T Cd + P^T BQ\, \underbrace{Q^T u}_{w}$$

$$\Downarrow$$

$$\min\{\|u\|_2^2 = \|Q^T u\|_2^2 \equiv \|w\|_2^2\} \quad \text{s.t.}$$

$$\begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x + \underbrace{\begin{bmatrix} U_{AC} \\ U_C \\ 0 \end{bmatrix} d}_{\text{omit for } \mathrm{H}_0} + \underbrace{\begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix}}_{w_3 \text{ fully determined}}.$$

Solution Derivation, under H_a

$$u_a^T u_a = r_a^T V^{-1} r_a = \min(\|w_A\|_2^2 + \|w_C\|_2^2 + \|w_3\|_2^2) \quad \text{s.t.}$$

$$\begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} U_{AC} \\ U_C \\ 0 \end{bmatrix} d + \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix}.$$

Under H_a: the optimal solution satisfies:

$$w_A^a = 0, \quad w_C^a = 0, \quad r_a^T V^{-1} r_a = \|w_3^a\|_2^2, \quad \begin{bmatrix} U_A & U_{AC} & R_{A3} \\ 0 & U_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_a \\ d_a \\ w_3^a \end{bmatrix} = \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix}.$$

w_3^a the generalization of Styan's LUSH residuals.

Solution Derivation, under H_0

$$u_0^T u_0 = r_0^T V^{-1} r_0 = \min(\ \|w_A\|_2^2 + \|w_C\|_2^2 + \|w_3\|_2^2 \ ) \quad \text{s.t.}$$

$$\begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x + \underbrace{\begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix}}_{w_C \text{ determined exactly}} .$$

Under H_0: the optimal solution satisfies:

$$w_A^0 = 0, \qquad \begin{bmatrix} U_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_0 \\ w_C^0 \\ w_3^0 \end{bmatrix} = \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} .$$

Solution Derivation, Final.

For the GLR test statistic, we know

$$\delta_{\text{TS}} = \sigma^{-2}(r_0^T V^{-1} r_0 - r_a^T V^{-1} r_a),$$

$$r_0^T V^{-1} r_0 = \|w_C^0\|_2^2 + \|w_3^0\|_2^2,$$

$$r_a^T V^{-1} r_a = \|w_3^a\|_2^2,$$

but $\quad R_3 w_3^0 = R_3 w_3^a = z_3, \quad$ so

$$\delta_{\text{TS}} = \sigma^{-2}\|w_C^0\|_2^2.$$

A simple, directly computable result.

Summary: Computer Solution of:

$$\mathrm{GLLS}_0 : \min_{u,x} \|u\|_2^2 \quad \text{s.t.} \ y = Ax + Bu \ ;$$

$$\mathrm{GLLS}_a : \min_{u,x,d} \|u\|_2^2 \quad \text{s.t.} \ y = Ax + Cd + Bu \ .$$

GQR of $[A, C]$ and B gives:

$$\begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x + \underbrace{\begin{bmatrix} U_{AC} \\ U_C \\ 0 \end{bmatrix} d}_{\text{omit for } \mathrm{H}_0} + \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix} .$$

Computer Solution, ctd.

Under H$_a$: we obtain $\{x_a, d_a\}$ by solving:

$$\begin{bmatrix} U_A & U_{AC} & R_{A3} \\ 0 & U_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_a \\ d_a \\ w_3^a \end{bmatrix} = \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix}.$$

Under H$_0$: we obtain x_0 by solving:

$$\begin{bmatrix} U_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_0 \\ w_C^0 \\ w_3^0 \end{bmatrix} = \begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix}.$$

GLR test statistic : $\boxed{\delta_{\mathrm{TS}} = \sigma^{-2}\|w_C^0\|_2^2}$

Numerical Stability of Algorithm

Computed $\hat{\delta}_{\mathrm{TS}}$ & $\hat{x}_0$ are the exact test statistic & MLE under H_0 for data:

$$
\begin{aligned}
\tilde{y} &\equiv y + \Delta y, & \|\Delta y\|_2 &= O(\epsilon)\|y\|_2, \\
\tilde{A} &\equiv A + \Delta A, & \|\Delta A\|_F &= O(\epsilon)\|A\|_F, \\
\tilde{B} &\equiv B + \Delta B, & \|\Delta B\|_F &= O(\epsilon)\|B\|_F, \\
\tilde{C} &\equiv C + \Delta C, & \|\Delta C\|_F &= O(\epsilon)\|C\|_F, \\
\tilde{\sigma} &\equiv \sigma + \Delta \sigma, & |\Delta \sigma| &= O(\epsilon)|\sigma|.
\end{aligned}
$$

The computations of δ_{TS} & x_0 are numerically stable! Similarly the computation of the MLE $\{x_a, d_a\}$ under H_a are numerically stable.

Covariance Matrix representation

What is $\mathrm{cov}\{x_0\}$ under H_0?
Under H_0 we have the model & estimate :

$$\begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x + \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} w_A \\ w_C \\ w_3 \end{bmatrix}$$

$$\begin{bmatrix} z_A \\ z_C \\ z_3 \end{bmatrix} = \begin{bmatrix} U_A \\ 0 \\ 0 \end{bmatrix} x_0 + \begin{bmatrix} R_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} 0 \\ w_C^0 \\ w_3^0 \end{bmatrix}$$

Subtracting the 1st equation from the 2nd leads to

$$\begin{bmatrix} U_A & R_{AC} & R_{A3} \\ 0 & R_C & R_{C3} \\ 0 & 0 & R_3 \end{bmatrix} \begin{bmatrix} x_0 - x \\ w_C^0 - w_C \\ w_3^0 - w_3 \end{bmatrix} = \begin{bmatrix} R_A w_A \\ 0 \\ 0 \end{bmatrix}.$$

This shows that $w_3^0 - w_3 = 0$ & $w_C^0 - w_C = 0$, so

$$U_A(x_0 - x) = R_A w_A.$$

Since $w = Q^T u \sim \mathcal{N}(0, \sigma^2 I)$, we have

$$U_A \cdot \mathrm{cov}\{x_0\} \cdot U_A^T = \sigma^2 R_A R_A^T.$$

The most reliable & useful representation of $\mathrm{cov}\{x_0\}$:
it covers all cases,
& can be updated in a numerically stable way.

An Example of Singular V.

New theory & algorithm handle singular V.
for example Linear Equality Constraints:
For the null-hypothesis H_0:

$$y = Ax + v, \quad v \sim \mathcal{N}(0, \sigma^2 V),$$

$$\text{subject to} \quad Ex = f.$$

If $V = BB^T$, with B fcr, so $v = Bu$, $u \sim \mathcal{N}(0, \sigma^2 I)$,
apply our algorithm directly to GLLS_0 problem:

$$\min \|u\|_2^2 \quad \text{subject to} \quad \begin{bmatrix} y \\ f \end{bmatrix} = \begin{bmatrix} A \\ E \end{bmatrix} x + \begin{bmatrix} B \\ 0 \end{bmatrix} u.$$

Similarly for H_a. Gives both test statistic & estimates.

Summary: Theory

- The standard formula for the GLR test statistic δ_{TS} is not defined when V is singular.
- We gave a new formulation for δ_{TS} (by reformulating the two problems for estimating the parameter vectors x & $\{x, d\}$).
- We gave a representation of the covariance matrices for the MLEs x_0 & x_a.

  The new formulations are well defined even when V is singular.

  The theory trivially handles the case where there are linear constraints $Ex = f$.

Summary: Practice

- The standard formula for the GLR test statistic δ_{TS} is not good for computation if any of A, $[A, C]$, or V is ill-conditioned.

- A numerically stable algorithm based on the GLLS method was proposed for computing δ_{TS} and the MLEs x_0 & x_a.

- We showed how to compute the covariance matrix representations for the MLEs x_0 & x_a.

- The algorithm handles the singular case, & where there are linear constraints $Ex = f$.

Some References

- Chang, X-W., Paige, C. C. and Tiberius, C. C. J. M.. *Computation of a Test Statistic in Data Quality Control*. SIAM Journal on Scientific Computing (SISC), accepted June 2004.

- Kourouklis, S. and Paige, C. C. (1981). *A constrained least squares approach to the general Gauss-Markov linear model*. Journal of the American Statistical Association, **76**, 620 – 625.

- Paige, C. C. (1978). *Numerically stable computations for general univariate linear models*. Communications in Statistics Series B, **B7**, 437 – 453.