

Model-Based Covariance Estimation for Regression M - and GM -Estimators

Tobias Schoch

University of Applied Sciences Northwestern Switzerland FHNW
School of Business, Riggbachstrasse 16, CH-4600 Olten
tobias.schoch@fhnw.ch

September 14, 2022

1 Introduction

The population regression model is given by

$$\xi : \quad Y_i = \mathbf{x}_i^T \boldsymbol{\theta} + \sigma \sqrt{v_i} E_i, \quad \boldsymbol{\theta} \in \mathbb{R}^p, \quad \sigma > 0, \quad i \in U,$$

where the population U is of size N ; the parameters $\boldsymbol{\theta}$ and σ are unknown; the \mathbf{x}_i 's are known values (possibly containing outliers), $\mathbf{x}_i \in \mathbb{R}^p$, $1 \leq p < N$; the v_i 's are known positive (heteroscedasticity) constants; the errors E_i are independent and identically distributed (i.i.d.) random variables with zero expectation and unit variance; it is assumed that $\sum_{i \in U} \mathbf{x}_i \mathbf{x}_i^T / v_i$ is a non-singular ($p \times p$) matrix.

It is assumed that a sample s is drawn from U with sampling design $p(s)$ such that the independence structure of model ξ is maintained. The sample regression GM -estimator of $\boldsymbol{\theta}$ is defined as the root to the estimating equation $\hat{\Psi}_n(\boldsymbol{\theta}, \sigma) = \mathbf{0}$ (for all $\sigma > 0$), where

$$\hat{\Psi}_n(\boldsymbol{\theta}, \sigma) = \sum_{i \in s} w_i \Psi_i(\boldsymbol{\theta}, \sigma) \quad \text{with} \quad \Psi_i(\boldsymbol{\theta}, \sigma) = \eta \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\theta}}{\sigma \sqrt{v_i}}, \mathbf{x}_i \right) \frac{\mathbf{x}_i}{\sigma \sqrt{v_i}},$$

where the function $\eta : \mathbb{R} \times \mathbb{R}^p \rightarrow \mathbb{R}$ parametrizes the following estimators

$\eta(r, \mathbf{x}) = \psi(r)$	M -estimator,
$\eta(r, \mathbf{x}) = \psi(r) \cdot h(\mathbf{x})$	Mallows GM -estimator,
$\eta(r, \mathbf{x}) = \psi \left(\frac{r}{h(\mathbf{x})} \right) \cdot h(\mathbf{x})$	Schweppe GM -estimator,

where $\psi : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous, bounded, and odd (possibly redescending) function, and $h : \mathbb{R}^p \rightarrow \mathbb{R}_+$ is a weight function.

2 Covariance estimation

The model-based covariance matrix of θ is (Hampel, Ronchetti, Rousseeuw, and Stahel, 1986, Chapter 6.3)

$$\text{cov}_\xi(\theta, \sigma) = M^{-1}(\theta, \sigma) \cdot Q(\theta, \sigma) \cdot M^{-T}(\theta, \sigma) \quad \text{for known } \sigma > 0, \quad (1)$$

where

$$M(\theta, \sigma) = \sum_{i=1}^N E_\xi \{ \Psi'_i(\theta, \sigma) \}, \quad \text{where} \quad \Psi'_i(\theta, \sigma) = -\frac{\partial}{\partial \theta^*} \Psi_i(Y_i, \mathbf{x}_i; \theta^*, \sigma) \Big|_{\theta^* = \theta},$$

and

$$Q(\theta, \sigma) = \frac{1}{N} \sum_{i=1}^N E_\xi \{ \Psi_i(Y_i, \mathbf{x}_i; \theta, \sigma) \Psi_i(Y_i, \mathbf{x}_i; \theta, \sigma)^T \},$$

and E_ξ denotes expectation with respect to model ξ . For the sample regression GM -estimator $\hat{\theta}_n$, the matrices M and Q must be estimated. Expressions of the generic matrices M and Q in (1) are given as follows.

$$\begin{array}{lll} \widehat{M}_M = -\overline{\psi}' \cdot X^T W X & \widehat{Q}_M = \overline{\psi}^2 \cdot X^T W X & M\text{-est.} \\ \widehat{M}_{Mal} = -\overline{\psi}' \cdot X^T W H X & \widehat{Q}_{Mal} = \overline{\psi}^2 \cdot X^T W H^2 X & GM\text{-est. (Mallows)} \\ \widehat{M}_{Sch} = -X^T W S_1 X & \widehat{Q}_{Sch} = X^T W S_2 X & GM\text{-est. (Schweppe)} \end{array}$$

where

$$\begin{array}{ll} W = \text{diag}_{i=1, \dots, n} \{w_i\}, & H = \text{diag}_{i=1, \dots, n} \{h(\mathbf{x}_i)\}, \\ \overline{\psi}' = \frac{1}{\widehat{N}} \sum_{i \in s} w_i \psi' \left(\frac{r_i}{\widehat{\sigma} \sqrt{v_i}} \right), & \overline{\psi}^2 = \frac{1}{\widehat{N}} \sum_{i \in s} w_i \psi^2 \left(\frac{r_i}{\widehat{\sigma} \sqrt{v_i}} \right), \\ S_1 = \text{diag}_{i=1, \dots, n} \{s_1^i\}, & s_1^i = \frac{1}{\widehat{N}} \sum_{j \in s} w_j \psi' \left(\frac{r_j}{h(\mathbf{x}_i) \widehat{\sigma} \sqrt{v_j}} \right), \end{array}$$

and

$$S_2 = \text{diag}_{i=1, \dots, n} \{s_2^i\}, \quad s_2^i = \frac{1}{\widehat{N}} \sum_{j \in s} w_j \psi^2 \left(\frac{r_j}{h(\mathbf{x}_i) \widehat{\sigma} \sqrt{v_j}} \right).$$

Remarks.

- The i -th diagonal element of S_1 and S_2 depends on $h(\mathbf{x}_i)$, but the summation is over $j \in s$; see also (Marazzi, 1987, Chapter 6).
- When W is equal to the identity matrix I , the asymptotic covariance of $\hat{\theta}_M$ is equal to the expression in Huber (1981, Eq. 6.5), which is implemented in the R packages MASS (Venables and Ripley, 2002) and robeth (Marazzi, 2020).

- For the Mallows and Schweppe type *GM*-estimators and given that $\mathbf{W} = \mathbf{I}$, the asymptotic covariance coincides with the one implemented in package/ library `robeth` for the option “averaged”; see [Marazzi \(1993, Chapter 4\)](#) and [Marazzi \(1987, Chapter 2.6\)](#) on the earlier ROBETH-85 implementation.

3 Implementation

The main function – which is only a wrapper function – is `cov_reg_model`. The following display shows pseudo code of the main function.

```
cov_reg_model()
{
  get_psi_function()           // get psi function (fun ptr)
  get_psi_prime_function()     // get psi-prime function (fun ptr)
  switch(type) {
    case 0: cov_m_est()         // M-estimator
    case 1: cov_mallows_gm_est() // Mallows GM-estimator
    case 2: cov_schweppe_gm_est() // Schweppe GM-estimator
  }
  robsurvey_error()           // signal error in case of failure
}
```

The functions `cov_m_est()`, `cov_mallows_gm_est()`, and `cov_schweppe_gm_est()` implement the covariance estimators; see below. All functions are based on the subroutines in BLAS ([Blackford et al., 2002](#)) and LAPACK ([Anderson et al., 1999](#)).

To fix notation, denote the Hadamard product of the matrices \mathbf{A} and \mathbf{B} by $\mathbf{A} \circ \mathbf{B}$ and suppose that $\sqrt{\cdot}$ is applied element by element.

3.1 *M*-estimator

The covariance matrix is (up to $\hat{\sigma}$) equal to (see `cov_m_est`)

$$(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \quad (2)$$

and is computed as follows:

- Compute the factorization $\sqrt{\mathbf{w}} \circ \mathbf{X} := \mathbf{Q}\mathbf{R}$ (LAPACK: `dgeqrf`).
- Invert the upper triangular matrix \mathbf{R} by backward substitution to get \mathbf{R}^{-1} (LAPACK: `dtrtri`).
- Compute $\mathbf{R}^{-1} \mathbf{R}^{-T}$, which is equal to (2); taking advantage of the triangular shape of \mathbf{R}^{-1} and \mathbf{R}^{-T} (LAPACK: `dtrmm`).

3.2 Mallows GM -estimator

The covariance matrix is (up to $\hat{\sigma}$) equal to (see `cov_mallows_gm_est`)

$$(\mathbf{X}^T \mathbf{W} \mathbf{H} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{H}^2 \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{H} \mathbf{X})^{-1} \quad (3)$$

and is computed as follows:

- Compute the QR factorization: $\sqrt{\mathbf{w} \cdot \mathbf{h}} \circ \mathbf{X} := \mathbf{Q} \mathbf{R}$ (LAPACK: `dgeqrf`).
- Invert the upper triangular matrix \mathbf{R} by backward substitution to get \mathbf{R}^{-1} (LAPACK: `dtrtri`).
- Define a new matrix: $\mathbf{A} \leftarrow \sqrt{\mathbf{h}} \circ \mathbf{Q}$ (extraction of \mathbf{Q} matrix with LAPACK: `dorgqr`).
- Update the matrix: $\mathbf{A} \leftarrow \mathbf{A} \mathbf{R}^{-T}$ (taking advantage of the triangular shape of \mathbf{R}^{-1} ; LAPACK: `dtrmm`).
- Compute $\mathbf{A} \mathbf{A}^T$, which corresponds to the expression in (3); (LAPACK: `dgemm`).

3.3 Schweppe GM -estimator

The covariance matrix is (up to $\hat{\sigma}$) equal to (see `cov_schweppe_gm_est`)

$$(\mathbf{X}^T \mathbf{W} \mathbf{S}_1 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{S}_2 \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{S}_1 \mathbf{X})^{-1}. \quad (4)$$

Put $\mathbf{s}_1 = \text{diag}(\mathbf{S}_1)$, $\mathbf{s}_2 = \text{diag}(\mathbf{S}_2)$, and let \cdot/\cdot denote elemental division (i.e., the inverse of the Hadamard product). The covariance matrix in (4) is computed as follows

- Compute the factorization $\sqrt{\mathbf{w} \circ \mathbf{s}_1} \circ \mathbf{X} := \mathbf{Q} \mathbf{R}$ (LAPACK: `dgeqrf`).
- Invert the upper triangular matrix \mathbf{R} by backward substitution to get \mathbf{R}^{-1} (LAPACK: `dtrtri`).
- Define a new matrix: $\mathbf{A} \leftarrow \sqrt{\mathbf{s}_2/\mathbf{s}_1} \circ \mathbf{Q}$ (extraction of \mathbf{Q} matrix with LAPACK: `dorgqr`).
- Update the matrix: $\mathbf{A} \leftarrow \mathbf{A} \mathbf{R}^{-T}$ (taking advantage of the triangular shape of \mathbf{R}^{-1} ; LAPACK: `dtrmm`).
- Compute $\mathbf{A} \mathbf{A}^T$, which corresponds to the expression in (4); (LAPACK: `dgemm`).

Remark. Marazzi (1987) uses the Cholesky factorization (see his subroutines `RTASKV` and `RTASKW`) which is computationally a bit cheaper than our QR factorization.

References

- ANDERSON, E., Z. BAI, C. BISCHOF, L. S. BLACKFORD, J. DEMMEL, J. DONGARRA, J. D. CROZ, A. GREENHAUM, S. HAMMARLING, A. MCKENNEY, AND D. SORENSEN (1999). *LAPACK Users' Guide*, Philadelphia: Society for Industrial and Applied Mathematics (SIAM), 3rd ed.
- BLACKFORD, L. S., A. PETITET, R. POZO, K. REMINGTON, R. C. WHALEY, J. DEMMEL, J. DONGARRA, I. DUFF, S. HAMMARLING, G. HENRY, M. HEROUX, L. KAUFMAN, AND A. LUMSDAINE (2002). An updated set of basic linear algebra subprograms (BLAS), *ACM Transactions on Mathematical Software* **28**, 135–151.
- HAMPEL, F. R., E. M. RONCHETTI, P. J. ROUSSEEUW, AND W. A. STAHEL (1986). *Robust Statistics: The Approach Based on Influence Functions*, New York: John Wiley and Sons.
- HUBER, P. J. (1981). *Robust Statistics*, New York: John Wiley and Sons.
- MARAZZI, A. (1987). *Subroutines for Robust and Bounded Influence Regression in ROBETH*, Cahiers de Recherches et de Documentation, 3 ROBETH 2, Division de Statistique et Informatique, Institut Universitaire de Médecine Sociale et Préventive, Lausanne, ROBETH-85 Document No. 2, August 1985, revised April 1987.
- MARAZZI, A. (1993). *Algorithms, Routines, and S Functions for Robust Statistics: The FORTRAN Library ROBETH with an interface to S-PLUS*, New York: Chapman & Hall.
- MARAZZI, A. (2020). *robeth: R Functions for Robust Statistics*, r package version 2.7-6.
- VENABLES, W. N. AND B. D. RIPLEY (2002). *Modern Applied Statistics with S*, New York: Springer, 4th ed.