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

Tobias Schoch

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

September 14, 2022

1 Introduction

The population regression model is given by

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

where the population U is of size N; the parameters $\boldsymbol{\theta}$ and σ are unknown; the $\boldsymbol{x}_i$'s are known values (possibly containing outliers), $\boldsymbol{x}_i \in \mathbb{R}^p$, $1 \leq p < N$; 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 $\sum_{i \in U} \boldsymbol{x}_i \boldsymbol{x}_i^T / v_i$ 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 $\widehat{\boldsymbol{\Psi}}_n(\boldsymbol{\theta}, \sigma) = \boldsymbol{0}$ (for all $\sigma > 0$), where

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

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

$$\begin{aligned}
&\eta(r, \boldsymbol{x}) = \psi(r) && M\text{-estimator}, \\
&\eta(r, \boldsymbol{x}) = \psi(r) \cdot h(\boldsymbol{x}) && \text{Mallows } GM\text{-estimator}, \\
&\eta(r, \boldsymbol{x}) = \psi\left(\frac{r}{h(\boldsymbol{x})}\right) \cdot h(\boldsymbol{x}) && \text{Schweppe } GM\text{-estimator},
\end{aligned}$$

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

2 Covariance estimation

The model-based covariance matrix of $\boldsymbol{\theta}$ is (Hampel, Ronchetti, Rousseeuw, and Stahel, 1986, Chapter 6.3)

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

where

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

and

$$\boldsymbol{Q}(\boldsymbol{\theta}, \sigma) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}_\xi\big\{\boldsymbol{\Psi}_i(Y_i, \boldsymbol{x}_i; \boldsymbol{\theta}, \sigma) \boldsymbol{\Psi}_i(Y_i, \boldsymbol{x}_i; \boldsymbol{\theta}, \sigma)^T\big\},$$

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

$$\begin{aligned}
\widehat{\boldsymbol{M}}_M &= -\overline{\psi'} \cdot \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} & \widehat{\boldsymbol{Q}}_M &= \overline{\psi^2} \cdot \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} & &M\text{-est.}\\
\widehat{\boldsymbol{M}}_{Mal} &= -\overline{\psi'} \cdot \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{H} \boldsymbol{X} & \widehat{\boldsymbol{Q}}_{Mal} &= \overline{\psi^2} \cdot \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{H}^2 \boldsymbol{X} & &GM\text{-est. (Mallows)}\\
\widehat{\boldsymbol{M}}_{Sch} &= -\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{S}_1 \boldsymbol{X} & \widehat{\boldsymbol{Q}}_{Sch} &= \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{S}_2 \boldsymbol{X} & &GM\text{-est. (Schweppe)}
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{W} &= \operatorname{diag}_{i=1,\ldots,n}\{w_i\}, & \boldsymbol{H} &= \operatorname{diag}_{i=1,\ldots,n}\{h(\boldsymbol{x}_i)\},\\
\overline{\psi'} &= \frac{1}{\widehat{N}} \sum_{i \in s} w_i \psi'\left(\frac{r_i}{\widehat{\sigma}\sqrt{v_i}}\right), & \overline{\psi^2} &= \frac{1}{\widehat{N}} \sum_{i \in s} w_i \psi^2\left(\frac{r_i}{\widehat{\sigma}\sqrt{v_i}}\right),\\
\boldsymbol{S}_1 &= \operatorname{diag}_{i=1,\ldots,n}\{s_1^i\}, & s_1^i &= \frac{1}{\widehat{N}} \sum_{j \in s} w_j \psi'\left(\frac{r_j}{h(\boldsymbol{x}_i)\widehat{\sigma}\sqrt{v_j}}\right),
\end{aligned}$$

and

$$\boldsymbol{S}_2 = \operatorname{diag}_{i=1,\ldots,n}\{s_2^i\}, \qquad s_2^i = \frac{1}{\widehat{N}} \sum_{j \in s} w_j \psi^2\left(\frac{r_j}{h(\boldsymbol{x}_i)\widehat{\sigma}\sqrt{v_j}}\right).$$

Remarks.

- The i-th diagonal element of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ depends on $h(\boldsymbol{x}_i)$, but the summation is over $j \in s$; see also (Marazzi, 1987, Chapter 6).
- When $\boldsymbol{W}$ is equal to the identity matrix $\boldsymbol{I}$, the asymptotic covariance of $\widehat{\boldsymbol{\theta}}_M$ 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 $\boldsymbol{W} = \boldsymbol{I}$, 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 $\boldsymbol{A}$ and $\boldsymbol{B}$ by $\boldsymbol{A} \circ \boldsymbol{B}$ and suppose that $\sqrt{\cdot}$ is applied element by element.

3.1 M-estimator

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

$$(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \tag{2}$$

and is computed as follows:

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

3.2 Mallows GM-estimator

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

$$\left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{H} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{H}^2 \boldsymbol{X} \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{H} \boldsymbol{X}\right)^{-1} \tag{3}$$

and is computed as follows:

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

3.3 Schweppe GM-estimator

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

$$\left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{S}_1 \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{S}_2 \boldsymbol{X} \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{S}_1 \boldsymbol{X}\right)^{-1}. \tag{4}$$

Put $\boldsymbol{s}_1 = \mathrm{diag}(\boldsymbol{S}_1)$, $\boldsymbol{s}_2 = \mathrm{diag}(\boldsymbol{S}_2)$, 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 $\sqrt{\boldsymbol{w} \circ \boldsymbol{s}_1} \circ \boldsymbol{X} := \boldsymbol{QR}$ (LAPACK: `dgeqrf`).
- Invert the upper triangular matrix $\boldsymbol{R}$ by backward substitution to get $\boldsymbol{R}^{-1}$ (LAPACK: `dtrtri`).
- Define a new matrix: $\boldsymbol{A} \leftarrow \sqrt{\boldsymbol{s}_2 / \boldsymbol{s}_1} \circ \boldsymbol{Q}$ (extraction of $\boldsymbol{Q}$ matrix with LAPACK: `dorgqr`).
- Update the matrix: $\boldsymbol{A} \leftarrow \boldsymbol{A}\boldsymbol{R}^{-T}$ (taking advantage of the triangular shape of $\boldsymbol{R}^{-1}$; LAPACK: `dtrmm`).
- Compute $\boldsymbol{A}\boldsymbol{A}^T$, 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.