# Fast robust SUR with economical and actuarial applications

Mia Hubert, Tim Verdonck and Ozlem Yorulmaz

November 30, 2015

**Abstract**

The seemingly unrelated regression (SUR) model is a generalization of a linear regression model consisting of more than one equation, where the error terms of these equations are contemporaneously correlated. The standard Feasible Generalized Linear Squares (FGLS) estimator is efficient as it takes into account the covariance structure of the errors, but it is also very sensitive to outliers. The robust SUR estimator of Bilodeau & Duchesne (2000) can accommodate outliers, but it is hard to compute. First we propose a fast algorithm, FastSUR, for its computation and show its good performance in a simulation study. We then provide diagnostics for outlier detection and illustrate them on a real data set from economics. Next we apply our FastSUR algorithm in the framework of stochastic loss reserving for general insurance. We focus on the General Multivariate Chain Ladder (GMCL) model that employs SUR to estimate its parameters. Consequently, this multivariate stochastic reserving method takes into account the contemporaneous correlations among run-off triangles and allows structural connections between these triangles. We plug in our FastSUR algorithm into the GMCL model to obtain a robust version.

**Keywords:** Seemingly Unrelated Regression; Feasible Generalized Least Squares; S-estimator; Outlier; Claims Reserving.

# 1 Introduction

Many studies in econometrics, insurance and finance are based on regression models containing more than one equation. Unconsidered factors that influence the error term in one equation often also influence the error terms in other equations. Ignoring this dependence structure of the error terms and estimating these equations separately using ordinary least squares (OLS) leads to inefficient estimates. Therefore, Zellner (1962) proposed the Seemingly Unrelated Regressions (SUR) model that is composed of several regression equations that are linked by the fact that their error terms are contemporaneously correlated. This system of structurally related equations is simultaneously estimated with a generalized least squares (GLS) estimator that takes the covariance structure of the error terms into account. An extensive summary of the literature dealing with the SUR model and its various extensions can be found in the book of Srivastava & Giles (1987) and the chapter by Fiebig (2001).

The SUR models have found considerable use in many applications in econometrics, finance and insurance. For example, Carrieri & Majerbi (1996), Angbazo & Narayanan (2006) and Williams (2013) studied the effects of financial crises and shocks on the insurance industry, banks and stock markets whereas Kaul (1987) and Dincer & Wang (2011) examined relationships between income and consumption, inflation and stock markets, ethnic diversity and economic growth. In this paper we illustrate the SUR model on studying the relation between the Foreign Direct Investment (FDI) by multinational corporations and several macroeconomic variables. Data are available for six countries over the period 1981-2012. This yields $m = 6$ regression blocks, each with $n = 32$ observations. Since there is a possible existence of common factors (such as economy-wide or worldwide shocks) that influence all countries at the same time, cross-correlations between the error terms may exist. This motivates the use of the SUR model for fitting this system of equations. As another application we consider claims reserving in general insurance, which is a major actuarial issue. Recently the multivariate reserving approach has received extensive attention, see e.g. Merz & Wüthrich (2009), Shi & Frees (2011), De Jong (2012), Happ & Wüthrich (2013), Shi (2014). We will focus on the general multivariate chain ladder model of Zhang (2010), where the parameters are estimated using SUR.

Since the common GLS estimation procedure for the SUR model is based on the classical covariance matrix and OLS estimation, the method as a whole is very sensitive to outliers. Outliers are observations that differ from the majority of the data and it is well known that they can have a large impact on classical statistical methods. Therefore, robust alternatives have already been presented in the literature. Koenker & Portnoy

(1990) proposed a robust version of the SUR model based on M-estimators. Since this procedure is not affine equivariant and does not take full account of the multivariate nature of the problem, Bilodeau & Duchesne (2000) introduced a method based on S-estimators. S-estimators were introduced for regression by Rousseeuw & Yohai (1984), whereas Davies (1987) and Lopuhaä (1989) studied S-estimators for multivariate location and scatter. The robust SUR method of Bilodeau & Duchesne (2000) combines both types of S-estimators, which results in an estimator that is regression and affine equivariant and is able to detect multivariate outliers.

First we resume the SUR model and the standard GLS estimators in Section 2. Section 3 describes the robust SUR method of Bilodeau & Duchesne (2000) and proposes a new algorithm for its computation. We then show its good performance in a simulation study (Section 4). In Section 5 we provide a diagnostic tool to detect outlying observations, and illustrate it on a data set from macroeconomics. Section 6 elaborates on the use of the robust SUR method in the context of stochastic loss reserving, whereas Section 7 provides some directions for further research.

# 2 Classical SUR

In general the SUR model comprises $m$ linear equations (also called blocks), each of which is assumed to satisfy the Gauss-Markov conditions. Each block contains an equal number $n$ of observations, hence the system can be written as

$$\begin{cases} y_1 &= \boldsymbol{X}_1\beta_1 + \varepsilon_1 \\ & \vdots \\ y_m &= \boldsymbol{X}_m\beta_m + \varepsilon_m \end{cases} \tag{1}$$

where $y_j = (y_{1j}, y_{2j}, \ldots, y_{nj})'$ is the $n \times 1$ response vector of the $j$th block, $\boldsymbol{X}_j$ is the $n \times p_j$ matrix of explanatory variables, $\beta_j$ is the $p_j \times 1$ vector of regression parameters while $\varepsilon_j$ corresponds to the $n \times 1$ error vector which satisfies $E(\varepsilon_j) = 0$ and $\text{Cov}(\varepsilon_j) = \sigma_{jj}\boldsymbol{I}_n$ (for all $j = 1, ..., m$). Here, $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. Note that each block has its own dependent variable and potentially different sets of exogenous explanatory variables. We further assume for each $j$ that $\text{rank}(\boldsymbol{X}_j) = p_j \leqslant n$ to avoid singular solutions. These seemingly unrelated regression blocks are linked through their zero mean error structure. It is namely assumed that the error vectors are contemporaneously but not serially correlated herein. This means that for given observations $i$ and $l$, across the regression equations $j$ and $k$, it holds that

$$E(\varepsilon_{ij}\varepsilon_{ik}) = \sigma_{jk} \text{ for all} \quad i = 1, 2, \ldots, n$$
$$E(\varepsilon_{ij}\varepsilon_{lj}) = 0 \quad \text{when} \quad i \neq l$$
$$E(\varepsilon_{ij}\varepsilon_{lk}) = 0 \quad \text{when} \quad j \neq k, \quad i \neq l$$

The system of $m$ seemingly unrelated regression equations (1) can be stacked in two equivalent compact matrix forms. First, we can express it as a multiple linear regression model:

$$y = \boldsymbol{X}\beta + \varepsilon \tag{2}$$

where $y = (y_1', \ldots, y_m')'$ is the $nm \times 1$ response vector,

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{X}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{X}_m \end{bmatrix}$$

is the $nm \times p$ structured design matrix, with $p = \sum_{j=1}^{m} p_j$, $\beta = (\beta_1', \ldots, \beta_m')'$ is the $p \times 1$ parameter vector and $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_m')'$ is the error vector with

$$\text{Cov}(\varepsilon) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix} \otimes \boldsymbol{I}_n = \begin{bmatrix} \sigma_{11}\boldsymbol{I}_n & \sigma_{12}\boldsymbol{I}_n & \cdots & \sigma_{1m}\boldsymbol{I}_n \\ \sigma_{21}\boldsymbol{I}_n & \sigma_{22}\boldsymbol{I}_n & \cdots & \sigma_{2m}\boldsymbol{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}\boldsymbol{I}_n & \sigma_{m2}\boldsymbol{I}_n & \cdots & \sigma_{mm}\boldsymbol{I}_n \end{bmatrix}.$$

Note that $\sigma_{jj}$ is the variance of the error term in the $j$th equation, whereas $\sigma_{jk}$ is the covariance between the error terms in equation $j$ and the error terms in equation $k$.

Another formulation of the SUR model uses the multivariate linear regression model:

$$\boldsymbol{Y} = \tilde{\boldsymbol{X}}\mathcal{B} + \mathcal{E} \tag{3}$$

where $\boldsymbol{Y} = (y_1, \ldots, y_m)$ is the $n \times m$ response matrix, $\tilde{\boldsymbol{X}} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m]$ the $n \times p$ design matrix,

$$\mathcal{B} = \begin{bmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_m \end{bmatrix} = \text{diag}(\beta_1, \beta_2, \ldots, \beta_m) \tag{4}$$

the structured $p \times m$ parameter matrix and $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_m)$ the error matrix with $\text{Cov}(\mathcal{E}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$. Equivalently we can write the error matrix as $\mathcal{E} = \boldsymbol{Y} - \tilde{\boldsymbol{X}}\mathcal{B} =$

$(e_1, \ldots, e_n)'$ with $e_i$ the $m$-dimensional vector containing the errors of the $i$th observation in each block.

Each equation in (1) could be estimated separately using the OLS estimator but this would ignore the covariance structure of the errors. A more efficient estimator is obtained as the GLS estimator with weight matrix $\boldsymbol{W} = \text{Cov}(\varepsilon)$. As $\boldsymbol{\Sigma}$ is typically unknown, a Feasible GLS (FGLS) estimator is preferred that replaces the unknown $\boldsymbol{W}$ with a consistent estimate. The FGLS estimator is an iterative two-step procedure that uses estimates for $\beta$ to estimate $\boldsymbol{\Sigma}$, which is then used to improve the regression estimates $\hat{\beta}$. To start, each equation is estimated by OLS, yielding $\hat{\beta}_j$. Then iteratively:

- The residuals $\hat{\varepsilon}_j = y_j - \boldsymbol{X}_j\hat{\beta}_j$ from the $m$ equations are used to estimate the error covariance matrix $\hat{\boldsymbol{W}} = \hat{\boldsymbol{\Sigma}} \otimes \boldsymbol{I}_n$ with $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}(\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m)'(\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_m)$.

- New estimates of $\beta$ are obtained as

$$\hat{\beta} = (\boldsymbol{X}'\hat{\boldsymbol{W}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{W}}^{-1}y = (\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X})^{-1}\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_n)y.$$

The estimated covariance matrix of $\hat{\beta}$ is given by $\text{Cov}(\hat{\beta}) = (\boldsymbol{X}'\hat{\boldsymbol{W}}^{-1}\boldsymbol{X})^{-1}$.

# 3 Robust SUR

## 3.1 The robust SUR method

We first recall the definition of the S-estimator of the SUR model, as introduced in Bilodeau & Duchesne (2000): it is defined as the couple $(\hat{\boldsymbol{\mathcal{B}}}, \hat{\boldsymbol{\Sigma}})$ which minimizes $|\boldsymbol{S}|$ under the condition

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\sqrt{e_i(\boldsymbol{B})'\boldsymbol{S}^{-1}e_i(\boldsymbol{B})}\right) = b \tag{5}$$

over all $(\boldsymbol{B}, \boldsymbol{S})$ with $\boldsymbol{B} = \text{diag}(b_1, \ldots, b_m) \in \mathbb{R}^{p \times m}$, $b_j \in \mathbb{R}^{p_j \times 1}$ for all $j = 1, \ldots, m$, $e_i(\boldsymbol{B})'$ the $i$th row of $\boldsymbol{Y} - \tilde{\boldsymbol{X}}\boldsymbol{B}$, and $\boldsymbol{S}$ an $m \times m$ symmetric positive definite (SPD) matrix. In order to obtain a robust estimator which is consistent and asymptotically normal, $\rho$ should satisfy the following conditions:

(C1) $\rho$ is symmetric around zero and twice continuously differentiable

(C2) $\rho(0) = 0$ and $\rho$ is strictly increasing on $[0, c_0]$ and constant on $[c_0, \infty[$ for some $c_0 > 0$.

The constant $b$ can be computed as $E_{F_0}[\rho(|\boldsymbol{e}|)]$ where $\boldsymbol{e} \sim F_0$. As such, $F_0 = N_m(\mathbf{0}, \boldsymbol{I}_m)$ ensures consistency at the model with normal errors. For $\rho$ one often chooses the function

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4} & \text{for} \quad |x| \leqslant c \\ \frac{c^2}{6} & \text{for} \quad |x| > c \end{cases}$$

where $c$ is an appropriate tuning constant (Rousseeuw & Yohai 1984). The derivative of this function is known as Tukey's bisquare function:

$$\rho'(x) = \psi(x) = \begin{cases} x\left(1 - \left(\frac{x}{c}\right)^2\right)^2 & \text{for } |x| \leqslant c \\ 0 & \text{for } |x| > c. \end{cases}$$

We will use this $\rho$-function throughout the paper. For fixed $b$, the value of the tuning constant $c$ determines the breakdown value, see Van Aelst & Willems (2005). For high-dimensional regressors, one could also consider other $\rho$-functions that downweight outliers more appropriately (Rocke 1996).

In addition to the minimization condition mentioned above the robust SUR estimators of $\beta$ and $\boldsymbol{\Sigma}$ also satisfy the following equations (Bilodeau & Duchesne 2000):

$$\beta = (\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{D})\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{D})y \tag{6}$$

$$\boldsymbol{\Sigma} = m(\boldsymbol{Y} - \tilde{\boldsymbol{X}}\boldsymbol{B})'\boldsymbol{D}(\boldsymbol{Y} - \tilde{\boldsymbol{X}}\boldsymbol{B})/\sum_{i=1}^{n} v(d_i) \tag{7}$$

where $d_i^2 = e_i'\boldsymbol{\Sigma}^{-1}e_i$ and $\boldsymbol{D} = \text{diag}(w(d_i))$, for $w(u) = \rho'(u)/u$ and $v(u) = \rho'(u)u - \rho(u) + b$.

## 3.2 The FastSUR algorithm

S-estimators have good robustness properties, but they are computationally expensive. The original resampling algorithm of Rousseeuw & Yohai (1984) was first improved by the SURREAL algorithm of Ruppert (1992). Next, a better performance was achieved by the FastS algorithm of Salibian-Barrera & Yohai (2006) for regression and Salibian-Barrera et al. (2006) for multivariate location and scatter. This is currently the most popular algorithm and is e.g. included in the R packages `robustbase` and `rrcov`, and the FSDA Matlab toolbox (Riani et al. 2012).

For the computation of the robust SUR method (5), Bilodeau & Duchesne (2000) adapted the SURREAL algorithm of Ruppert (1992). Here we propose the FastSUR algorithm, which implements the ideas of the FastS algorithm into the robust SUR estimator.

First, the $\boldsymbol{S}$ in (5) is written as $\sigma^2\boldsymbol{\Gamma}$ with $|\boldsymbol{\Gamma}| = 1$ and $\sigma = |\boldsymbol{S}|^{1/2m}$, so that the equivalent objective is to find the triplet $(\hat{\mathcal{B}}, \hat{\boldsymbol{\Gamma}}, \hat{\sigma})$ that minimizes $\sigma$ under the restriction

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{\sqrt{e_i(\boldsymbol{B})'\boldsymbol{\Gamma}^{-1}e_i(\boldsymbol{B})}}{\sigma}\right) = b$$

over all $(\boldsymbol{B}, \boldsymbol{\Gamma}, \sigma)$ where $\boldsymbol{B} = \text{diag}(\beta_1, \ldots, \beta_m) \in \mathbb{R}^{p\times m}$, $\boldsymbol{\Gamma}$ is an $m \times m$ SPD matrix with $|\boldsymbol{\Gamma}| = 1$ and $\sigma$ is a positive scalar. The robust SUR estimates are then given by $(\hat{\mathcal{B}}, \hat{\boldsymbol{\Sigma}} = \hat{\sigma}^2\hat{\boldsymbol{\Gamma}})$.

The algorithm starts with $N$ initial estimates $(\hat{\boldsymbol{B}}_1^{(0)}, \hat{\boldsymbol{\Gamma}}_1^{(0)}, \hat{\sigma}_1^{(0)}), \ldots, (\hat{\boldsymbol{B}}_N^{(0)}, \hat{\boldsymbol{\Gamma}}_N^{(0)}, \hat{\sigma}_N^{(0)})$ obtained as follows:

(a) Choose a random subsample of $\max(p_1, \ldots, p_m)$ integers from the first $n$ integers.

(b) Calculate OLS on the corresponding rows of $\boldsymbol{Y}$ and $\tilde{\boldsymbol{X}}$, giving $\hat{\boldsymbol{B}}_l^{(0)}$. If this subset yields a singular solution in a block, randomly increase the number of observations in that block, until a nonsingular OLS solution is obtained.

(c) Compute the robust covariance matrix of the residuals by their one-step M-estimator (Maronna et al. 2006, pag. 197):

- Set the initial covariance matrix $\hat{\boldsymbol{\Sigma}}^{(0)} = \text{diag}(\text{mad}(\boldsymbol{Y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{B}}_l^{(0)}))^2$ where the median absolute deviation (mad) is computed on each column of the residual matrix.
- Compute the robust distances $d_i = \sqrt{e_i(\hat{\boldsymbol{B}}_l^{(0)})'(\hat{\boldsymbol{\Sigma}}^{(0)})^{-1}e_i(\hat{\boldsymbol{B}}_l^{(0)})}$.
- Update the initial covariance matrix: $\hat{\boldsymbol{\Sigma}}_l^{(0)} = \frac{1}{n}\sum_{i=1}^{n} w(d_i)e_i(\hat{\boldsymbol{B}}_l^{(0)})e_i(\hat{\boldsymbol{B}}_l^{(0)})'$.

Then we set $\hat{\boldsymbol{\Gamma}}_l^{(0)} = |\hat{\boldsymbol{\Sigma}}_l^{(0)}|^{-1/m}\hat{\boldsymbol{\Sigma}}_l^{(0)}$ and $\hat{\sigma}_l^{(0)} = \text{med}_{i=1}^{n}\sqrt{e_i(\hat{\boldsymbol{B}}_l^{(0)})'(\hat{\boldsymbol{\Gamma}}_l^{(0)})^{-1}e_i(\hat{\boldsymbol{B}}_l^{(0)})}$ for all $l = 1, \ldots, N$. Next, those estimates are refined by performing $k$ so-called $I$-steps, resulting in

$$(\hat{\boldsymbol{B}}_1^{(k)}, \hat{\boldsymbol{\Gamma}}_1^{(k)}, \hat{\sigma}_1^{(k)}), \ldots, (\hat{\boldsymbol{B}}_N^{(k)}, \hat{\boldsymbol{\Gamma}}_N^{(k)}, \hat{\sigma}_N^{(k)}).$$

The $j$th $I$-step to refine the estimate $(\hat{\boldsymbol{B}}_l^{(j-1)}, \hat{\boldsymbol{\Gamma}}_l^{(j-1)}, \hat{\sigma}_l^{(j-1)})$ goes as follows:

1. Refine the scale: $\hat{\sigma}_l^{(j)} = \hat{\sigma}_l^{(j-1)}\sqrt{\frac{1}{nb}\sum_{i=1}^{n}\rho\left(\frac{\sqrt{e_i(\hat{\boldsymbol{B}}_l^{(j-1)})'(\hat{\boldsymbol{\Gamma}}_l^{(j-1)})^{-1}e_i(\hat{\boldsymbol{B}}_l^{(j-1)})}}{\hat{\sigma}_l^{(j-1)}}\right)}$.

2. Use $\hat{\sigma}_l^{(j)}$ to compute weights $w_i^{(j)} = \frac{\rho'(u_i)}{u_i}$ with $u_i = \frac{\sqrt{e_i(\hat{\boldsymbol{B}}_l^{(j-1)})'(\hat{\boldsymbol{\Gamma}}_l^{(j-1)})^{-1}e_i(\hat{\boldsymbol{B}}_l^{(j-1)})}}{\hat{\sigma}_l^{(j)}}$.

3. Update $\hat{\boldsymbol{B}}_l^{(j-1)}$ following equation (6): Let $\boldsymbol{D} = \text{diag}(w_i^{(j)})$ and $\boldsymbol{W} = (\sigma_l^{(j)})^{-2}(\hat{\boldsymbol{\Gamma}}_l^{(j-1)})^{-1} \otimes \boldsymbol{D}$, then $\hat{\beta}^{(j)} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}y$ and $\hat{\boldsymbol{B}}_l^{(j)} = \text{diag}(\hat{\beta}_1^{(j)}, \ldots, \hat{\beta}_m^{(j)})$.

4. Update $\hat{\boldsymbol{\Sigma}}$ following equation (7): $\hat{\boldsymbol{\Sigma}}_l^{(j)} = m(\boldsymbol{Y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{B}}_l^{(j)})'\boldsymbol{D}(\boldsymbol{Y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{B}}_l^{(j)})/\sum_{i=1}^n v(u_i)$, which leads to the refinement $\hat{\boldsymbol{\Gamma}}_l^{(j)} = |\hat{\boldsymbol{\Sigma}}_l^{(j)}|^{-1/m}\hat{\boldsymbol{\Sigma}}_l^{(j)}$.

After performing $k$ $I$-steps, the scale $\hat{\sigma}_l^{(k)}$ is improved for each $(\hat{\boldsymbol{B}}_l^{(k)}, \hat{\boldsymbol{\Gamma}}_l^{(k)}, \hat{\sigma}_l^{(k)})$ by iteratively solving

$$\hat{\sigma}_l^{(k+1)} = \hat{\sigma}_l^{(k)} \sqrt{\frac{1}{nb}\sum_{i=1}^n \rho\left(\frac{\sqrt{e_i(\hat{\boldsymbol{B}}_l^{(k)})'(\hat{\boldsymbol{\Gamma}}_l^{(k)})^{-1}e_i(\hat{\boldsymbol{B}}_l^{(k)})}}{\hat{\sigma}_l^{(k)}}\right)} \tag{8}$$

until convergence while keeping $\hat{\boldsymbol{B}}_l^{(k)}$ and $\hat{\boldsymbol{\Gamma}}_l^{(k)}$ fixed. We keep the $v$ refined estimates $(\hat{\boldsymbol{B}}_1^{(*)}, \hat{\boldsymbol{\Gamma}}_1^{(*)}, \hat{\sigma}_1^{(*)})$, …, $(\hat{\boldsymbol{B}}_v^{(*)}, \hat{\boldsymbol{\Gamma}}_v^{(*)}, \hat{\sigma}_v^{(*)})$ with the smallest fully iterated scales. Note that not all scales $\hat{\sigma}_l^{(k)}$, $l = 1, \ldots, N$ need to be computed by solving (8). The first $v$ scales $\hat{\sigma}_l^{(k)}$, $l = 1, \ldots, v$ are always computed, but for $l > v$ the $l$th scale is only computed if

$$\frac{1}{n}\sum_{i=1}^n \rho\left(\frac{\sqrt{e_i(\hat{\boldsymbol{B}}_l^{(k)})'(\hat{\boldsymbol{\Gamma}}_l^{(k)})^{-1}e_i(\hat{\boldsymbol{B}}_l^{(k)})}}{A}\right) < b$$

where $A$ is the maximum of the $v$ best scales that were fully iterated so far. The $v$ estimates $(\hat{\boldsymbol{B}}_1^{(*)}, \hat{\boldsymbol{\Gamma}}_1^{(*)}, \hat{\sigma}_1^{(*)})$, …, $(\hat{\boldsymbol{B}}_v^{(*)}, \hat{\boldsymbol{\Gamma}}_v^{(*)}, \hat{\sigma}_v^{(*)})$ with the smallest scales need to be refined until convergence using $I$-steps as described above, and the final estimate $(\hat{\boldsymbol{B}}^{(F)}, \hat{\boldsymbol{\Gamma}}^{(F)}, \hat{\sigma}^{(F)})$ is the one with the smallest scale after full refinement. The final robust SUR estimates are then $\hat{\boldsymbol{\mathcal{B}}} = \hat{\boldsymbol{B}}^{(F)}$ (or equivalently $\hat{\beta} = (\hat{\beta}_1', \ldots, \hat{\beta}_m')'$) and $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}^{(F)})^2\hat{\boldsymbol{\Gamma}}^{(F)}$.

Note that the number of subsets $N$, the number of I-steps $k$ and the number of refined estimates $v$ can be chosen by the user. In our experience the settings $N = 500, k = 2$ and $v = 5$ (as in Salibian-Barrera & Yohai (2006)) work well for many data sizes and contamination patterns, but using larger values for these parameters might be useful at large data sets with potentially a high contamination level.

# 4 Simulation study

In this section we study the performance of our FastSUR algorithm on artificial data sets. As a benchmark, we always compare our results with the FGLS algorithm, as computed within the R package `systemfit` (Henningsen & Hamann 2007).

We carried out an extensive simulation study on data sets of different dimension and here we report the results of two settings, namely, $A : n = 100, p_j = 5, m = 8$ and $B : n = 30, p_j = 3, m = 4$. For each simulation setting, we generated $K = 100$ data sets, with fixed $\beta$ and fixed $\mathbf{\Sigma}$. Each block contains the same number of explanatory variables $p_j$. The values of $\beta_j$ are randomly drawn from a uniform distribution $U([0, 10])$ (although we could equally well fix them to zero due to the regression equivariance of the estimators). For generating the covariance matrix $\mathbf{\Sigma}$ we followed the methodology of Joe (2006), taking $[1, 4]$ as the range for the variances. For each data set, the independent variables were generated from a $(p_j - 1)$-variate standard normal distribution $N_{p_j-1}(0, \boldsymbol{I}_{p_j-1})$. We considered two different distributions for the error terms $\varepsilon = (\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_m')'$. First, we studied normal errors generated from $N_{nm}(0, \mathbf{\Sigma} \otimes \boldsymbol{I}_n)$ and secondly we considered heavy tailed errors following a Student distribution with 3 degrees of freedom, $t_3(0, \mathbf{\Sigma} \otimes \boldsymbol{I}_n)$. The response variables were then computed according to model (1).

In order to contaminate the data, we replaced the first 5%, 10% and 30% of observations in each block by bad leverage points, by replacing some predictor variables with a random value from $U([20, 30])$ while keeping the response value unchanged. It is well-known that these type of outliers, being both outlying in the space of the predictor variables as in the errors, are considered the most influential type of outliers that often cause the regression estimates to be highly biased (Rousseeuw & Leroy 1987). For each simulation setting and each data set, we applied both FGLS and robust FastSUR. We always used a breakdown value of 25%, except at those data sets with 30% contamination where the breakdown value was set to 50%.

In order to measure the performance of the estimators, we evaluated their bias and mean squared error:

$$Bias = \left\| \frac{1}{K} \sum_{k=1}^{K} \hat{\beta}^{(k)} - \beta \right\|$$

$$MSE = \frac{1}{K} \sum_{k=1}^{K} \left\| \hat{\beta}^{(k)} - \beta \right\|^2$$

where $\beta$ is the true regression vector, and $\hat{\beta}^{(k)}$ is the FGLS or robust SUR estimate on the $k$th sample.

The results are given in Table 1 for data setting A and in Table 2 for setting B. It can be observed that for uncontaminated data sets, the FastSUR results are close to FGLS. When there is contamination, the bias and MSE of FGLS explode, whereas the robust FastSUR algorithm yields very satisfactory results that do not deviate much from the uncontaminated case. Only with a very large amount of outliers in each block, FastSUR has a slightly increased bias and MSE.

Table 1: Simulation results for data setting A with 0%, 5%, 10% and 30% contamination, and a normal or heavy-tailed error distribution.

| *Outliers* | 0% | | 5% | | 10% | | 30% | |
|---|---|---|---|---|---|---|---|---|
| | FGLS | FastSUR | FGLS | FastSUR | FGLS | FastSUR | FGLS | FastSUR |
| *Normal* | | | | | | | | |
| Bias | 0.212 | 0.212 | 15.203 | 0.241 | 23.084 | 0.223 | 33.599 | 0.259 |
| MSE | 3.145 | 3.176 | 358.724 | 3.565 | 884.582 | 3.529 | 1193.012 | 5.788 |
| $t_3$ | | | | | | | | |
| Bias | 0.305 | 0.301 | 16.537 | 0.292 | 24.247 | 0.313 | 32.463 | 0.782 |
| MSE | 9.405 | 9.545 | 359.924 | 6.918 | 906.993 | 7.577 | 1016.514 | 10.283 |

Table 2: Simulation results for data setting B with 0%, 5%, 10% and 30% contamination, and a normal or heavy-tailed error distribution.

| *Outliers* | 0% | | 5% | | 10% | | 30% | |
|---|---|---|---|---|---|---|---|---|
| | FGLS | FastSUR | FGLS | FastSUR | FGLS | FastSUR | FGLS | FastSUR |
| *Normal* | | | | | | | | |
| Bias | 0.213 | 0.204 | 11.076 | 0.170 | 11.433 | 0.179 | 16.104 | 0.232 |
| MSE | 3.442 | 3.591 | 143.736 | 3.305 | 154.008 | 3.423 | 288.558 | 8.167 |
| $t_3$ | | | | | | | | |
| Bias | 0.376 | 0.393 | 11.084 | 0.311 | 11.415 | 0.348 | 16.116 | 2.176 |
| MSE | 8.642 | 8.744 | 144.803 | 7.669 | 156.985 | 8.198 | 274.167 | 10.892 |

We also ran more simulations in which the outliers were differently positioned in each block, but the outcomes were always comparable to the results of Table 1 and 2. Whereas FGLS collapses at data sets with outliers, FastSUR is more resistant to them.

# 5 Outlier detection

Applying FastSUR on a real data set does not only result in robust parameter estimates, it also provides diagnostics for outlier detection. First, based on the multiple regression model (2), we can quantify the outlyingness of observations within one block by their *standardized residual*. For the $i$th observation of block $j$, it is equal to $r_{ij}$ being the $i$th value of the standardized residual vector $r_j = (y_j - \boldsymbol{X}_j \hat{\beta}_j)/\hat{\sigma}_{jj} = (r_{1j}, \ldots, r_{nj})'$. Under gaussian errors, an observation is typically considered to be outlying if its absolute standardized residual exceeds 2.5.

Next, we can consider the multivariate regression model (3). The $i$th row in the data matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ then corresponds with the measurements of the $i$th observation in each block ($i = 1, \ldots, n$). Consequently, the $i$th *residual distance*

$$\text{ResD}_i = \sqrt{e_i(\hat{\mathcal{B}})' \hat{\boldsymbol{\Sigma}}^{-1} e_i(\hat{\mathcal{B}})} \tag{9}$$

can be used to detect outlying behavior of one of the $n$ rows. Under normal errors, a residual distance larger than $\sqrt{\chi^2_{m,0.975}}$ (the square root of the 0.975 quantile of the $\chi^2_m$ distribution) is flagged as being unusually large.

To illustrate these diagnostics, we study the Foreign Direct Investment (FDI) of six countries (India, Indonesia, Columbia, Mexico, Turkey and Chile) over the period 1981-2012. The countries constitute the $m = 6$ blocks in our SUR model, with measurements of $n = 32$ years per country. FDI is considered to be the main source of economic growth. Agiomirgianakis et al. (2003) referred that FDI is mostly defined as capital flows resulting from the behavior of multinational companies (MNCs) and the factors to affect the behavior of MNCs may also affect the magnitude and the direction of FDI. As predictor variables we include several macroeconomic variables: the growth rate per capita GPD, the rate of inflation measured by annual percentage change of consumer prices, and the degree of openness which is computed as the sum of nominal export and import divided by the nominal GDP. The data are collected from The World Bank, World Development Indicators [1].

The relationship between FDI and its potential determinant variables has been a prominent topic in the last two decades. In several studies SUR and panel models (Kok & Ersoy 2009) were applied. The reason for using the SUR model is the possible existence of common factors that influence all the countries at the same time and bring about the cross correlations between the error terms.

[1] website:`http://data.worldbank.org/data-catalog/world-development-indicators`

We apply both FLGS and FastSUR (with 50% breakdown value) on our data set and compare the residuals by means of several outlier maps (Hubert et al. 2008). First, for each country $j$ we plot the FGLS standardized residuals $r_{ij}$ versus the Mahalanobis distance of the observed predictor variables $\mathrm{MD}_{ij}$ (for $i = 1, \ldots, n$). The latter are computed as

$$\mathrm{MD}_{ij} = \sqrt{(x_{ij} - \bar{x}_j)' \boldsymbol{S}_j^{-1} (x_{ij} - \bar{x}_j)} \tag{10}$$

with $\bar{x}_j$ and $\boldsymbol{S}_j$ the mean and covariance matrix of $\boldsymbol{X}_j = (x_{1j}, \ldots, x_{nj})'$. If the predictor variables are normally distributed, the squared Mahalanobis distances approximately have a $\chi^2_{p_j}$ distribution, hence we set the cutoff to $\sqrt{\chi^2_{p_j,0.975}}$. For Indonesia this yields the residual plot in Figure 1(a). The lines indicate the cutoff values for the residuals and the Mahalanobis distances. Observations exceeding these cutoff values are considered to be outliers according to the FGLS estimates. We see that for the years 1997, 2000 and 2012 the standardized residuals are slightly larger than expected, but the macroeconomic predictor variables do not have abnormal values. These type of observations are called vertical outliers. In 1998 the FDI can be well predicted, but the larger MD indicates that the explanatory variables deviate from the other years. This observation is called a good leverage point. Looking at the raw data for Indonesia we notice that all predictor variables have an extreme value in 1998 (a negative growth rate of -14%, a huge inflation rate of 75% and a degree of openness which is over 98%). Figure 1(b) shows the FastSUR results. Here the vertical axis shows the standardized residuals based on the FastSUR estimates, whereas the horizontal axis depicts the robust distances

$$\mathrm{RD}_{ij} = \sqrt{(x_{ij} - \hat{\mu}_j)' \hat{\boldsymbol{\Sigma}}_j^{-1} (x_{ij} - \hat{\mu}_j)}$$

with $\hat{\mu}_j$ and $\hat{\boldsymbol{\Sigma}}_j$ the FastS estimates of the center and scatter of $\boldsymbol{X}_j$. Now, the years 2001 and 2003 show up as very clear vertical outliers, and also in 2000 the fit is not very good. In all these three years the FDI was negative. Furthermore, 1998 is indicated as a very prominent good leverage point. Note that when we analyze solely the data from Indonesia by means of OLS and robust S-regression, we obtain the residual plots of Figure 1(c) and (d). In both cases the year 2003 is not detected as a vertical outlier.

The residual distance plot, showing all $n$ residual distances (9), is depicted in Figure 2(a) for FGLS and (b) for robust SUR. Here, we notice a huge difference between both estimates. Whereas the SUR estimates flag none of the years as outlying (although there is an increasing trend), robust SUR finds that most of the recent years have a very different behavior. This corresponds with the global financial and economic crisis period which has a substantial effect on the emerging countries.

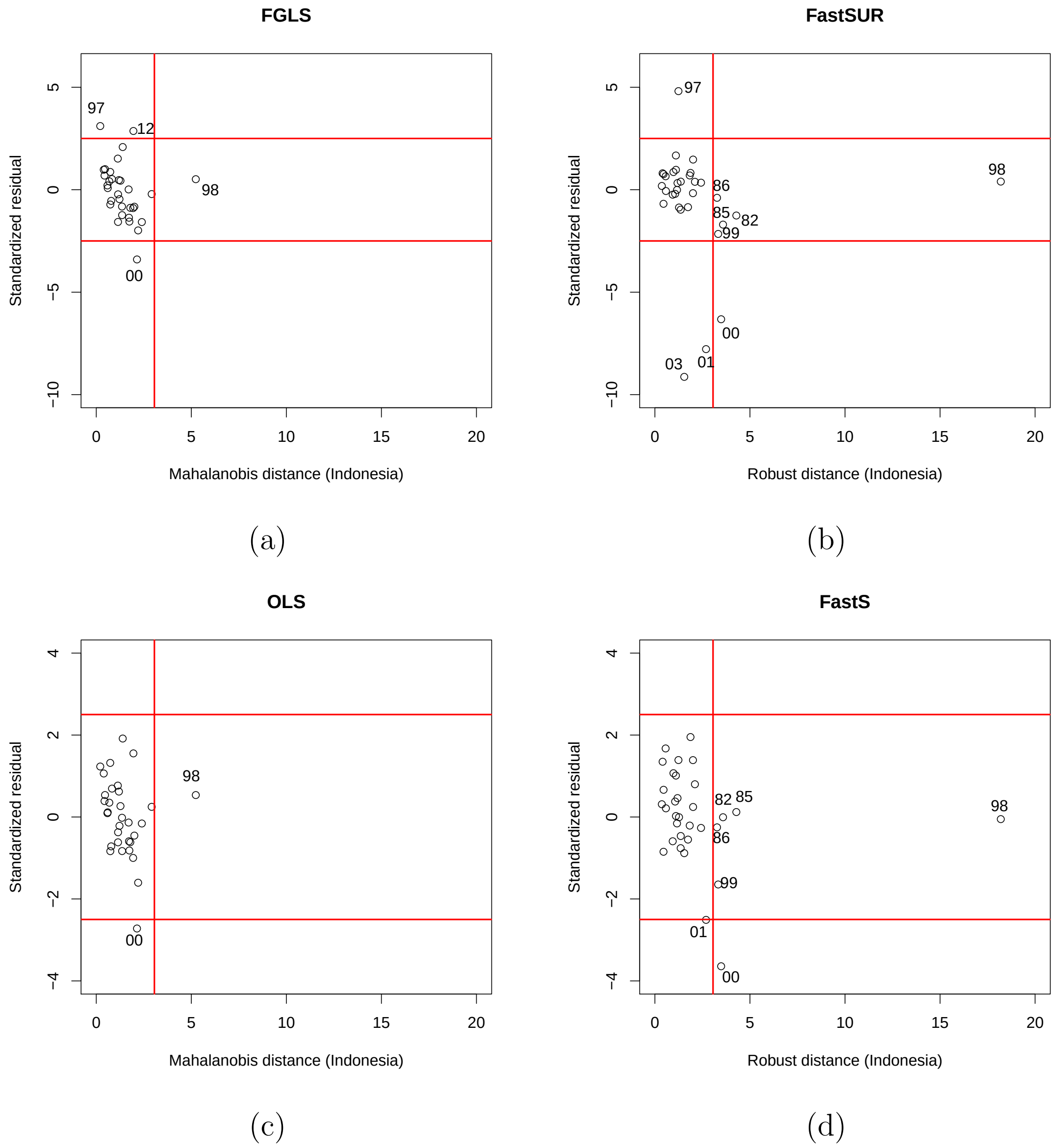
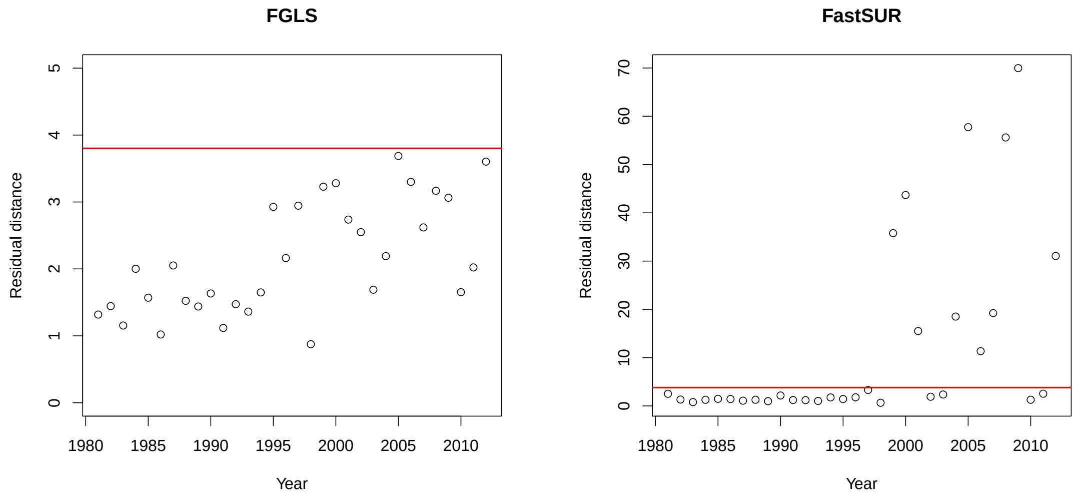



Figure 1: Regression diagnostic plots of Indonesia based on (a) FGLS; (b) robust SUR; (c) OLS and (d) robust S-regression.

# 6 Actuarial application

Stochastic claims reserving is a major actuarial problem in general insurance and with the introduction of new regulatory guidelines for the insurance business there is a growing awareness that modern statistical techniques should be used. Claims reserves are often the largest position on the liability side of the balance sheet of a general insurance company. We study claims that take months or years to emerge depending on the complexity of




(a) (b)

Figure 2: Residual distance plot based on (a) classical estimates; (b) robust estimates.

the damage. The delay in payment is, for example, due to long legal procedures or difficulties in determining the size of the claims. Therefore, insurers have to build up reserves enabling them to pay the outstanding claims and to meet claims arising in the future on the written contracts. In this section, we describe how the reserve estimates can be obtained using the SUR technique. Our FastSUR algorithm will then be plugged in in this methodology and its good performance will be illustrated on a real data set.

## 6.1 Chain Ladder method

We assume that $C_{ik}$ (for $1 \leqslant i \leqslant I$ and $1 \leqslant k \leqslant K$) are the cumulative claims amount of accident year $i$ and development year $k$. For representation of the data it is common to use a run-off triangle as in Table 3.

The ultimate goal of claims reserving boils down to completing the triangle into a square (or a rectangle if estimates are required pertaining to development years of which no data are recorded at hand) since the total of the values found in the lower right triangle equals the overall reserve $R$ that will need to be paid in future. The Chain Ladder (CL) method is the most popular method for estimating $R$. Many problems related with the CL method have already been solved in literature, for example Mack (1999) included a tail factor, whereas Kunkler (2006) and De Alba (2006) dealt with the presence of negative values in the run-off triangle.

In practice, a general insurance company subdivides portfolios into several correlated

Table 3: Run-off triangle.

| development year / accident year | 1 | 2 | $\cdots$ | $k$ | $\cdots$ | $K-1$ | $K$ |
|---|---|---|---|---|---|---|---|
| 1 | $C_{1,1}$ | $C_{1,2}$ | $\cdots$ | $C_{1,k}$ | $\cdots$ | $C_{1,K-1}$ | $C_{1,K}$ |
| 2 | $C_{2,1}$ | $C_{2,2}$ | $\cdots$ | $C_{2,k}$ | $\cdots$ | $C_{2,K-1}$ | |
| $\vdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | | |
| $i$ | $C_{i,1}$ | $C_{i,2}$ | $\cdots$ | $C_{i,k}$ | | | |
| $\vdots$ | $\cdots$ | $\cdots$ | $\cdots$ | | | | |
| $I$ | $C_{I,1}$ | | | | | | |

subportfolios, such that each subportfolio (which is represented by a run-off triangle) satisfies certain homogeneity properties. However, the CL method applies to a single run-off triangle and therefore neglects the contemporaneous correlations existing between subportfolios. Since it is well-known that the CL predictions for the sum of several run-off triangles in general differ from the sum of the Chain Ladder predictions for the single run-off triangles (Aine 1994), the claims reserving problem is recently studied in a multivariate context (Braun 2004, Merz & Wüthrich 2007). Pröhl & Schmidt (2005) and Schmidt (2006) introduced a Multivariate Chain Ladder (MCL) model, where the multivariate estimators take into account the dependence structure between the subportfolios and which are optimal in terms of a classical optimality criterion. Merz & Wüthrich (2008) provide a conditional mean squared error of prediction (MSEP) estimator for this multivariate version, which is useful to quantify the uncertainties in the reserve estimates.

## 6.2 MCL in SUR framework

Zhang (2010) recently showed that the estimators in the MCL model can find their equivalents in the SUR framework. We give a brief description of this General Multivariate Chain Ladder (GMCL) model and refer to Zhang (2010) for more details. Assume that we have $m$ correlated run-off triangles with $I$ accident and $K$ development years (for simplicity, we assume $I = K$) and that the claims from different accident years are independent. Denote $\tilde{C}_{i,k} = \left(C_{i,k}^{(1)}, \ldots, C_{i,k}^{(m)}\right)'$ as the vector of cumulative claims at accident year $i$ and development year $k$ and consider the following model structure for development period $k$ (i.e. from development year $k$ to $k+1$):

$$\tilde{C}_{i,k+1} = \mathcal{B}_k \tilde{C}_{i,k} + \varepsilon_{i,k} \qquad \text{for } i = 1, \ldots, I-k. \tag{11}$$

Here $\mathcal{B}_k$ is the corresponding $m \times m$ development matrix that contains the development parameters $\beta_j = (\beta_{j1}, \ldots, \beta_{jm})'$ for run-off triangle $j \leqslant m$ in the $j$th row and $\varepsilon_{i,k}$ are symmetrically distributed errors. Therefore, the development of one run-off triangle in development period $k$ can depend on the claims in the other run-off triangles at development year $k$. Moreover, it is assumed that

$$E(\varepsilon_{i,k}|\mathcal{D}_{i,k}) = 0 \tag{12}$$

$$\text{Cov}(\varepsilon_{i,k}|\mathcal{D}_{i,k}) = \text{diag}(\tilde{C}_{i,k})^{1/2}\, \boldsymbol{\Sigma}_k \,\text{diag}(\tilde{C}_{i,k})^{1/2} \tag{13}$$

where $\mathcal{D}_{i,k} = \left\{ C_{i,k}^{(j)} | i \leqslant k, j = 1, \ldots, m \right\}$ is the set of claims for accident year $i$ up to and including development year $k$ and $\boldsymbol{\Sigma}_k$ is a symmetric positive definite $m \times m$ matrix.

Zhang (2010) has rewritten the model structure for development period $k$ (11) as the following system of equations

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 & 0 & \ldots & 0 \\ 0 & \boldsymbol{X}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{X}_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \tag{14}$$

where for $j = 1, 2, \ldots, m$ and $n = I - k$ it holds that

- $y_j = C_{\leq,k+1}^{(j)}$ is the $n \times 1$ vector of all observed losses at development year $k+1$ from the $j$th triangle
- $\boldsymbol{X}_j = \left( C_{<,k}^{(1)}, \ldots, C_{<,k}^{(m)} \right)$ is the $n \times m$ matrix of the first $n$ observations at development year $k$ from each triangle (hence $\boldsymbol{X}_1 = \ldots = \boldsymbol{X}_m$)
- $\varepsilon_j$ is the $n \times 1$ vector of error terms in the $j$th equation.

From (12) and (13) it follows that

$$\text{Cov}(\varepsilon) = E(\varepsilon\varepsilon') = \text{diag}(V)^{1/2}(\boldsymbol{\Sigma}_k \otimes \boldsymbol{I}_n)\text{diag}(V)^{1/2}$$

where $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_m')'$ and $V = \left( (C_{<,k}^{(1)})', \ldots, (C_{<,k}^{(m)})' \right)'$ is the $nm \times 1$ vector of the first $n$ observed claims at development year $k$. Pre-multiplying both sides of model (14) by $\text{diag}(V)^{-1/2}$ leads to a regression model whose error covariance matrix $\text{Cov}(\varepsilon^*)$ is consistent with the SUR assumption:

$$\text{Cov}(\varepsilon^*) = \text{diag}(V)^{-1/2}\,\text{Cov}(\varepsilon)\text{diag}(V)^{-1/2} = \boldsymbol{\Sigma}_k \otimes \boldsymbol{I}_n.$$

After estimating the development parameters $\beta_j$ ($j = 1, \ldots, m$) using the FGLS estimation procedure consecutively for all development periods $k = 1, \ldots, K-1$, the overall reserve estimate $\hat{R}$ (for $m$ triangles simultaneously) can be obtained.

In the univariate setting ($m = 1$) Verdonck et al. (2009) and Verdonck & Debruyne (2011) have already demonstrated the sensitivity of the CL method to outliers. Even one outlier can lead to a huge over- or underestimation of the overall reserve estimate. Since the traditional SUR estimator is not robust, the corresponding GMCL estimates are also unreliable in the presence of outliers. Note that robustness now even plays a more important role, since an outlier in one of the run-off triangles may now also affect the estimates of outstanding claim amounts of the other run-off triangles. When we plug in our robust SUR algorithm in the GMCL model, we obtain robust reserve estimates and diagnostics for outlier detection.

## 6.3 Real example

To study the performance of the robust GMCL estimator, we focus on the real example that is presented in Zhang (2010). The studied portfolio from Schedule P of General Accident Insurance Company (published by NAIC) consists of three run-off triangles that are given in Table 4, 5 and 6.

Table 4: Cumulative paid triangle from Personal Auto.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 101125 | 209921 | 266618 | 305107 | 327850 | 340669 | 348430 | 351193 | 353353 | 353584 |
| 2 | 102541 | 203213 | 260677 | 303182 | 328932 | 340948 | 347333 | 349813 | 350523 | |
| 3 | 114932 | 227704 | 298120 | 345542 | 367760 | 377999 | 383611 | 385224 | | |
| 4 | 114452 | 227761 | 301072 | 340669 | 359979 | 369248 | 373325 | | | |
| 5 | 115597 | 243611 | 315215 | 354490 | 372376 | 382738 | | | | |
| 6 | 127760 | 259416 | 326975 | 365780 | 386725 | | | | | |
| 7 | 135616 | 262294 | 327086 | 367357 | | | | | | |
| 8 | 127177 | 244249 | 317972 | | | | | | | |
| 9 | 128631 | 246803 | | | | | | | | |
| 10 | 126288 | | | | | | | | | |

Similar as in Zhang (2010), we fit the multivariate GMCL model for development years 1-7 and the univariate CL model for development years 8-10 (since the gain of increasing model complexity after year 7 is minor). Applying the classical and the robust methods on the original data yields respectively an overall reserve estimate of 1 049 664 and 1 052 546, which is very close to each other. When we multiply for example claim $C^{(1)}_{2,2}$ by 10, then the classical and robust overall reserve estimate equals respectively 825 530 and 1 048 768. When there is a significant difference between the classical and robust overall

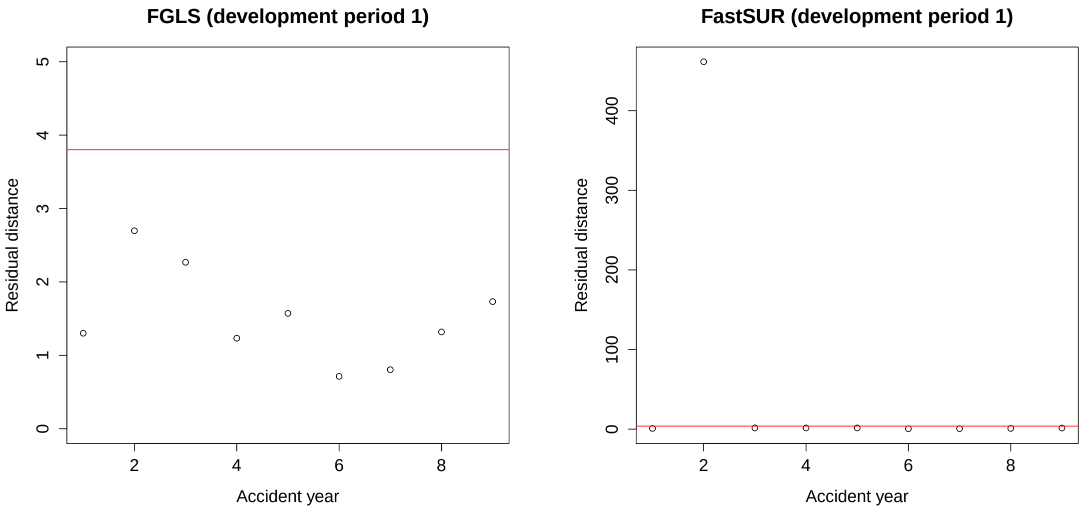
Table 5: Cumulative incurred triangle from Personal Auto.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 325423 | 336426 | 346061 | 347726 | 350995 | 353598 | 354797 | 355025 | 354986 | 355363 |
| 2 | 323627 | 339267 | 344507 | 349295 | 351038 | 351583 | 352050 | 352231 | 352193 | |
| 3 | 358410 | 386330 | 385684 | 384699 | 387678 | 387954 | 388540 | 389436 | | |
| 4 | 405319 | 396641 | 391833 | 384819 | 380914 | 380163 | 379706 | | | |
| 5 | 434065 | 429311 | 422181 | 409322 | 394154 | 392802 | | | | |
| 6 | 417178 | 422307 | 413486 | 406711 | 406503 | | | | | |
| 7 | 398929 | 398787 | 398020 | 400540 | | | | | | |
| 8 | 378754 | 361097 | 369328 | | | | | | | |
| 9 | 351081 | 335507 | | | | | | | | |
| 10 | 329236 | | | | | | | | | |

Table 6: Cumulative paid triangle from Commercial Auto.

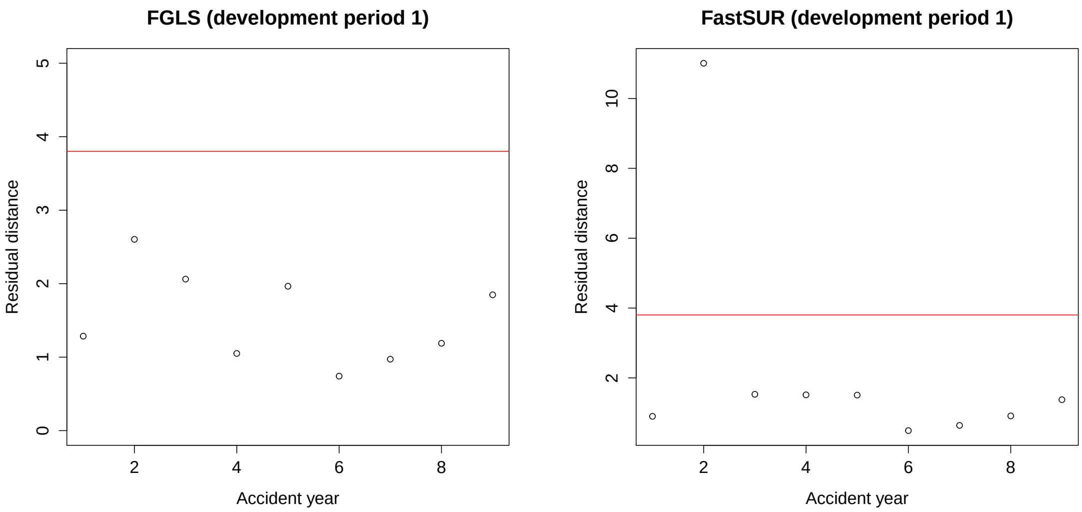
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 19827 | 44449 | 61205 | 77398 | 88079 | 95695 | 99853 | 104789 | 105427 | 106690 |
| 2 | 22331 | 48480 | 68789 | 92356 | 104958 | 112399 | 115638 | 117415 | 118571 | |
| 3 | 22533 | 44484 | 65691 | 88435 | 102044 | 112672 | 115973 | 118359 | | |
| 4 | 23128 | 51328 | 81542 | 98063 | 113149 | 121515 | 124347 | | | |
| 5 | 25053 | 57220 | 84607 | 104936 | 117663 | 126180 | | | | |
| 6 | 30136 | 64767 | 92288 | 108835 | 121326 | | | | | |
| 7 | 34764 | 69125 | 91354 | 111987 | | | | | | |
| 8 | 31803 | 63471 | 92439 | | | | | | | |
| 9 | 40559 | 77667 | | | | | | | | |
| 10 | 46285 | | | | | | | | | |

reserve estimate, then it is highly recommended to examine the data more closely and to study the most influential observations. When plotting the residual distances obtained with the classical and robust GMCL, we see from Figure 3 that only the robust version succeeds in detecting the outlier. Let us now multiply $C_{2,2}^{(1)}$ and $C_{2,2}^{(2)}$ by 1.2 and divide $C_{2,2}^{(3)}$ by 1.2. The overall reserve estimates for the classical and robust GMCL method equals respectively 1 036 407 and 1 048 768. Although this (multivariate) outlier configuration does not have a very large impact on the overall reserve estimates, the robust GMCL method still succeeds in detecting the atypical observation as we can see in Figure 4. Also here the non-robust GMCL does not indicate any anomaly.




(a) (b)

Figure 3: Residual distance plot for the first contamination based on (a) classical GMCL estimates; (b) robust GMCL estimates.




(a) (b)

Figure 4: Residual distance plot for the second contamination based on (a) classical GMCL estimates; (b) robust GMCL estimates.

# 7 Conclusion and outlook

Our new algorithm for robust SUR provides robust parameter estimates and useful outlier diagnostics, as illustrated on two real data sets. Several extensions are interesting to study. Robust inference might be obtained using the robust bootstrap as in Van Aelst & Willems (2005). This would provide standard errors of the robust point estimates, and hence a fast approach towards robust model selection. High collinearity among the predictor variables could be alleviated by applying a robust PCA method to the predictor variables as in Hubert & Verboven (2003) or by generalizing the robust PLS method of Hubert & Vanden Branden (2003). When the set of predictor variables includes categorical ones, the algorithm might need many more random subsets than the current $N = 500$. Techniques as in Hubert & Rousseeuw (1996) or Maronna & Yohai (2000) could be extended towards this setting. Also the use of multivariate $\tau$-estimators could be considered, by combining ideas from Maronna & Yohai (1997) and Salibian-Barrera et al. (2008).

Applied to the framework of stochastic loss reserving for general insurance we obtain robust estimates in the GMCL model. The proposed methodology is helpful to build up a more realistic reserve, certainly when used in addition to the classical GMCL estimates. Another advantage of the SUR model in the claims reserving context is that is also allows structural connections among triangles and hence allows to construct flexible models. In the classical approach, this is already described and illustrated in Zhang (2010). Studying this using the robust GMCL model could also be an interesting topic for future research.

# References


Agiomirgianakis, G., Asteriou, D. & Papathoma, K. (2003), 'The determinants of foreign direct investment: a panel data study for the OECD countries', *Working Papers 03/06, City University London*.

Aine, B. (1994), 'Additivity of chain-ladder projections', *Astin Bulletin* **24**, 311–318.

Angbazo, L. & Narayanan, R. (2006), 'Catastrophic shocks in the property-liability insurance industry: evidence on regulatory and contagion effects', *Economics Letters* **37**, 372–391.

Bilodeau, M. & Duchesne, P. (2000), 'Robust estimation of the SUR model', *Canadian Journal of Statistics* **28**, 277–288.

Braun, C. (2004), 'The prediction error of the chain ladder method applied to correlated run-off triangles', *Astin Bulletin* **34**, 399–424.

Carrieri, F. & Majerbi, B. (1996), 'The pricing of exchange risk in emerging stock markets', *Journal of International Business Studies* **63**, 619–637.

Davies, L. (1987), 'Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices', *The Annals of Statistics* **15**, 1269–1292.

De Alba, E. (2006), 'Claims reserving when there are negative values in the runoff triangle: Bayesian analysis using the three-parameter log-normal distribution', *North American Actuarial Journal* **10**, 45–59.

De Jong, P. (2012), 'Modeling dependence between loss triangles', *North American Actuarial Journal* **16**, 74–86.

Dincer, O. & Wang, F. (2011), 'Ethnic diversity and economic growth in China', *Journal of Economic Policy Reform* **14**, 1–10.

Fiebig, D. (2001), 'Seemingly unrelated regression' in *A Companion to Theoretical Econometrics*, Blackwell Publishing Ltd, pp. 101–121.

Happ, S. & Wüthrich, M. V. (2013), 'Paid-incurred chain reserving method with dependence modeling', *Astin Bulletin* **43**, 1–20.

Henningsen, A. & Hamann, J. (2007), 'systemfit: A package for estimating systems of simultaneous equations in R', *Journal of Statistical Software* **23**, 1–40.

Hubert, M. & Rousseeuw, P. (1996), 'Robust regression with both continuous and binary regressors', *Journal of Statistical Planning and Inference* **57**, 153–163.

Hubert, M., Rousseeuw, P. & Van Aelst, S. (2008), 'High breakdown robust multivariate methods', *Statistical Science* **23**, 92–119.

Hubert, M. & Vanden Branden, K. (2003), 'Robust methods for Partial Least Squares Regression', *Journal of Chemometrics* **17**, 537–549.

Hubert, M. & Verboven, S. (2003), 'A robust PCR method for high-dimensional regressors', *Journal of Chemometrics* **17**, 438–452.

Joe, H. (2006), 'Generating random correlation matrices based on partial correlations', *Journal of Multivariate Analysis* **97**, 2177–2189.

Kaul, G. (1987), ‘Stock returns and inflation: The role of the monetary sector’, *Journal of Financial Economics* **18**, 253–276.

Koenker, R. & Portnoy, S. (1990), ‘M-estimation of multivariate regressions’, *Journal of the American Statistical Association* **85**, 1060–1068.

Kok, R. & Ersoy, B. (2009), ‘Analyses of FDI determinants in developing countries’, *International Journal of Social Economics* **36**, 105–123.

Kunkler, M. (2006), ‘Modelling negatives in stochastic reserving models’, *Insurance: Mathematics and Economics* **38**, 540–555.

Lopuhaä, H. (1989), ‘On the relation between S-estimators and M-estimators of multivariate location and covariance’, *The Annals of Statistics* **17**, 1662–1683.

Mack, T. (1999), ‘The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor’, *Astin Bulletin* **29**, 361–366.

Maronna, R., Martin, D. & Yohai, V. (2006), *Robust Statistics: Theory and Methods*, Wiley, New York.

Maronna, R. & Yohai, V. (1997), ‘Robust estimation in simultaneous equations models’, *Journal of Statistical Planning and Inference* **57**, 233–244.

Maronna, R. & Yohai, V. (2000), ‘Robust regression with both continuous and categorical predictors’, *Journal of Statistical Planning and Inference* **89**, 197–214.

Merz, M. & Wüthrich, M. (2007), ‘Prediction error of the chain ladder reserving method applied to correlated run off triangles’, *Annals of Actuarial Science* **2**, 25–50.

Merz, M. & Wüthrich, M. (2008), ‘Prediction error of the multivariate chain ladder reserving method’, *North American Actuarial Journal* **12**, 175–197.

Merz, M. & Wüthrich, M. V. (2009), ‘Combining chain-ladder and additive loss reserving method for dependent lines of business’, *Variance* **3**, 270–291.

Pröhl, C. & Schmidt, K. (2005), *Multivariate chain ladder*, Techn. Univ., Inst. für Mathematische Stochastik.

Riani, M., Perrotta, D. & Torti, F. (2012), ‘FSDA: A MATLAB toolbox for robust analysis and interactive data exploration’, *Chemometrics and Intelligent Laboratory Systems* **116**, 17–32.

Rocke, D. M. (1996), 'Robustness properties of $S$-estimators of multivariate location and shape in high dimension', *The Annals of Statistics* **24**, 1327–1345.

Rousseeuw, P. & Leroy, A. (1987), *Robust Regression and Outlier Detection*, Wiley-Interscience, New York.

Rousseeuw, P. & Yohai, V. (1984), Robust regression by means of S-estimators, *in* J. Franke, W. Härdle & R. Martin, eds, 'Robust and Nonlinear Time Series Analysis', Lecture Notes in Statistics No. 26, Springer-Verlag, New York, pp. 256–272.

Ruppert, D. (1992), 'Computing S-estimators for regression and multivariate location/dispersion', *Journal of Computational and Graphical Statistics* **1**, 253–270.

Salibian-Barrera, M., Van Aelst, S. & Willems, G. (2006), 'PCA based on multivariate MM-estimators with fast and robust bootstrap', *Journal of the American Statistical Association* **101**, 1198–1211.

Salibian-Barrera, M., Willems, G. & Zamar, R. (2008), 'The fast-$\tau$ estimator for regression', *Journal of Computational and Graphical Statistics* **17**, 659–682.

Salibian-Barrera, M. & Yohai, V. (2006), 'A fast algorithm for S-regression estimates', *Journal of Computational and Graphical Statistics* **15**, 414–427.

Schmidt, K. (2006), 'Optimal and additive loss reserving for dependent lines of business', *CAS forum (fall)* pp. 319–351.

Shi, P. (2014), 'A copula regression for modeling multivariate loss triangles and quantifying reserving variability', *Astin Bulletin* **44**, 85–102.

Shi, P. & Frees, E. W. (2011), 'Dependent loss reserving using copulas', *Astin Bulletin* **41**, 449–486.

Srivastava, V. & Giles, D. (1987), *Seemingly Unrelated Regression Models: Estimation and Inference*, Marcel Dekker, New York.

Van Aelst, S. & Willems, G. (2005), 'Multivariate regression S-estimators for robust estimation and inference', *Statistica Sinica* **15**, 981–1001.

Verdonck, T. & Debruyne, M. (2011), 'The influence of individual claims on the chain-ladder estimates: analysis and diagnostic tool', *Insurance: Mathematics and Economics* **48**, 85–98.

Verdonck, T., Van Wouwe, M. & Dhaene, J. (2009), 'A robustification of the chain-ladder method', *North American Actuarial Journal* **13**, 280–298.

Williams, B. (2013), 'Income volatility of Indonesian banks after the Asian financial crisis', *Journal of the Asia Pacific Economy* **18**, 333–358.

Zellner, A. (1962), 'Robust estimation of the SUR model', *Canadian Journal of Statistics* **57**, 348–368.

Zhang, Y. (2010), 'A general multivariate chain ladder model', *Insurance: Mathematics and Economics* **46**, 588–599.