CESIS Electronic Working Paper Series

Paper No. 183

Developing Median Regression for SURE models

– with application to 3-generation immigrants' data in Sweden

Zangin Zeebari* and Ghazi Shukur**

(Centre for Labour Market Policy (CAFO),
Växjö University*,Jönköping International Business School**)

August 2009

The Royal Institute of Technology
Centre of Excellence for Science and Innovation Studies (CESIS)
http://www.cesis.se

Developing Median Regression for SURE Models

with Application to 3-Generation Immigrants' data in Sweden

By

[1]Zangin Zeebari and Ghazi Shukur[2]

[1&2]Centre for Labour Market Policy (CAFO) Department of Economics and Statistics, Växjö University, Sweden and

[2] Jönköping International Business School, Sweden

Abstract

In this paper we generalize the median regression method in order to make it applicable to systems of regression equations. Given the existence of proper systemwise medians of the errors from different equations, we apply the weighted median regression with the weights obtained from the covariance matrix of errors from different equations calculated by conventional SURE method. The Seemingly Unrelated Median Regression Equations (SUMRE) method produces results that are more robust than the usual SURE or single equations OLS estimations when the distributions of the dependent variables are not symmetric. Moreover, the estimations of the SUMRE method are also more efficient than those of the cases of single equation median regressions when the cross equations errors are correlated. More precisely, the aim of our SUMRE method is to produce a harmony of existing skewness and correlations of errors in systems of regression equations. A theorem is derived and indicates that even with the lack of statistically significant correlations between the equations, using the SMRE method instead of the SURE method will not damage the estimation of parameters.

A Monte Carlo experiment was conducted to investigate the properties of the SUMRE method in situations where the number of equations in the system, number of observations, strength of the correlations of cross equations errors and the departure from the normality distribution of the errors were varied. The results show that, when the cross equations correlations are medium or high and the level of skewness of the errors of the equations are also medium or high, the SUMRE method produces estimators that are more efficient and less biased than the ordinary SURE GLS estimators. Moreover, the estimates of applying the SUMRE method are also more efficient and less biased than the estimates obtained when applying the OLS or single equation median regressions. In addition, our results from an empirical application are in accordance with what we discovered from the simulation study, with respect to the relative gain in efficiency of SUMRE estimators compared to SURE estimators, in the presence of Skewness of error terms.

Key words: Median regression, SURE models, robustness, efficiency

1. Introduction

Regression analysis is often used to explain the behaviour of an explained variable for fixed values of the explanatory variables. Traditionally, this kind of analysis is focused on the mean, i.e., by using a conditional mean function we try to summarise the relationship between the explained variable and the explanatory variables. The Ordinary Least Squares (OLS) method is a typical estimation method for this purpose. Intuitively, the OLS estimation method describes the relationship between these variables when the distribution of the dependent variable is symmetric. Otherwise, when this symmetry does not exist the mean will not be the most proper measure of central tendency for calculating the conditional function. In practical studies, there exist numerous cases where the data of interest, in one way or another, are not symmetric. The distributions of for example earning variables are often highly skewed. This, of course, may render inferences invalid when using standard estimation methodology such as OLS. In such cases, other measures of central tendency, like median, might be more appropriate for this purpose.

The median regression is a statistical technique intended to estimate and draw inferences about conditional median functions. Just as the classical linear regression method based on minimising sums of squared errors enables one to estimate models for conditional mean functions, the median regression method offers a mechanism for estimating models for the conditional median function. Moreover, the median regression is less sensitive to outliers and departure from the normality assumption than the ordinary linear regression method is.

Originally, median regression was suggested by Koenker and Bassett (1978) as a robust regression technique, so called L_1 or Least Absolute Deviation (LAD) regression, as an alternative to the OLS for a case where the errors are not normally distributed. For these reasons, this method and other robust estimation methods have been used in many empirical works instead of the traditional OLS method. Practically, the median regression is more difficult to apply than the standard OLS method since it requires special algorithms that previously were not readily available in standard statistical software packages. However, recent versions of STATA do include routines for estimating the median regression.

The median regression has been mainly applied strictly to single equation environments and disappointingly to multivariate regression (excluding SURE models). Many models are expressed in terms of multivariate models (sometimes referred to as systems of equations), due to the fact that the different marginal models are connected to each other. Treating each equation separately, may lead to the loss of efficiency and to the reduction of the validity of the conclusions. In general, the use of median regression is quite uncommon in multivariate models, which may partly be due to the lack of availability of a standard methodology, or even a standard definition of multivariate median.

The purpose of this study is to generalize the median regression and make it applicable to systems of regression equations. Given the existence of proper systemwise medians of the errors from different equations, we apply the weighted median regression with the weights obtained from the covariance matrix of cross equations errors calculated by the ordinary SURE method. The SURE method is considered as one of the most successful and efficient methods for estimating seemingly unrelated regressions with the assumption of symmetric regression errors. The resulting SURE model has stimulated countless theoretical and empirical results in econometrics and other areas, (see Zellner, 1962; Srivastava and Giles, 1987; Chib and Greenberg, 1995). The benefit of SURE models in our case is that the SURE estimators utilise the information present in the correlations of the cross equations errors and hence are more efficient than other estimation methods such as the OLS method.

The paper is arranged as follows. In Section 2, we discuss the methodology. First, we give a formal definition of median regression which is widely used in the literature, and discuss how its notion is connected with the OLS method. Next, we give a formal definition of the SURE model and its development from the OLS method, which was introduced by Zellner. Finally, we introduce a new method, which we call the SUMRE method, and discuss how it is connected with both the median regression and the SURE method. In Section 3, we present the design of our Monte Carlo experiment and discuss the criteria used to evaluate the efficiency of the SUMRE model. In Section 4, we interpret the output of the Monte Carlo experiment and make a comparison between the SUMRE method, the SURE method and the method of separate median regressions of single equations. Section 5 contains an empirical application of the introduced SUMRE method to some data taken from Multi-Generation Register at Statistics Sweden on three generations of male immigrants from Finland to Sweden. The conclusions of the paper are presented in Section 6. The outputs of the Monte Carlo experiment are arranged in tables and presented in the appendix.

2. Methodology

In this section, we present the methodology with regard to the median regression, the SURE method and our introduced SUMRE method.

2.1. Median Regression:

Traditional regression analysis places heavy reliance on the conditional mean function, that is, it fits a model based on the relationship between the mean of the response given a fixed value of predictors. This approach suffices when the data have a symmetric distribution— such as with the Gaussian distribution. In this case, the median coincides with the mean, and even all other quintiles could be approximately predicted by the use of further information about the dispersion of errors. Even for symmetrically distributed errors with longer or shorter tails than those of a Gaussian distribution, some adjustments of the conditional mean function like robust methods could be used.

For asymmetric distributions, the mean seems less desirable, but other measures like the median could be taken as more suitable alternatives for the study of the locational behaviour of a random variable. If we seek the mean of a distribution through a statistical decision theoretic problem, represented as an optimization problem of a loss function, a suitable loss function for this purpose is a quadratic loss function, as shown below,

$$L(Y,\ \theta) = c\|Y-\theta\|^2 \qquad (2.1.1)$$

where c is a positive real constant and θ is a function used to predict the mean of the random variable Y. With this loss function, the risk (expectation of the loss function) is the same whatever positive constant c is chosen, so, for more convenience it is usually chosen to be 1.

$$R(Y;\theta) = E\left(L(Y,\ \theta)\right) = \int_{-\infty}^{+\infty} \|y-\theta\|^2\, dF(y)$$

$$\therefore\ \ 0 = \nabla_{\hat{\theta}} R\left(\hat{\theta}\right) = \nabla_{\hat{\theta}} \int_{-\infty}^{+\infty} \left\|y-\hat{\theta}\right\|^2 dF(y)$$

$$= -2\left(E(Y)-\hat{\theta}\right)$$

$$\therefore\ \ \hat{\theta} = E(Y) \qquad (2.1.2)$$

This means that the expectation of loss function (2.1.1) is minimized if $\hat{\theta}$ is chosen to be the mean of Y, or in other words, the solution to the minimization problem gives the mean of Y. Using a random sample of n independently and identically distribute (iid) random variables, and consequently, replacing the unknown distribution function of the random variables by the empirical distribution function, we move from the realm of mathematics into the theory of statistics, and obtain the following risk function.

$$
\begin{aligned}
R(Y;\theta) &= \int_{-\infty}^{+\infty} \|y-\theta\|^2 \, dF_n(y) \\
&= \int_{-\infty}^{+\infty} \|y-\theta\|^2 \, d\left(\frac{1}{n}\sum_{i=1}^{n} I(y_i \le y)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \|y_i-\theta\|^2 ,
\end{aligned}
\qquad (2.1.3)
$$

which is the sum of squared errors divided by n. Minimizing the risk (2.1.3) gives the sample estimate of the mean.

Suppose θ is a function of vector $\mathbf{x}$ (or the conditional mean of y given $\mathbf{x}$ is required) through the relationship $\theta = \mu_y(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$, then minimizing the risk function with respect to $\boldsymbol{\beta}$, gives the following solution to the optimization problem (2.1.3):

$$
\begin{aligned}
\therefore \quad 0 &= \nabla_{\hat{\boldsymbol{\beta}}} R\left(Y;\mathbf{x}'\hat{\boldsymbol{\beta}}\right) \\
&= \nabla_{\hat{\boldsymbol{\beta}}} \frac{1}{n}\sum_{i=1}^{n} \left\|y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}\right\|^2 \\
&= \nabla_{\hat{\boldsymbol{\beta}}} \frac{1}{n}\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)
\end{aligned}
$$

$$
\therefore \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} , \qquad (2.1.4)
$$

which is known as the least squared errors estimation of the parameter vector $\boldsymbol{\beta}$.

However, in decision-making, we can try to predict other parameters of a distribution function, like unique mode, or a mode in a specific interval of parameter space, median or any other quantiles of the distribution function of Y. The median of a distribution is obtained if the quadratic loss function in (2.1.1) is replaced by the absolute deviation loss function,

$$L(Y,\ \xi) = c|Y-\xi|, \tag{2.1.5}$$

where c is a positive real constant and ξ is a function used to predict the median of Y. Each positive value of c gives the same risk, and for more convenience it is usually chosen to be 1. The probability distribution of a real-valued random variable Y is a right-continuous left-limit (not necessarily) monotone increasing function:

$$F(y) = P(Y \le y). \tag{2.1.6}$$

The quantile function is the inverse of the distribution function

$$Q_Y(\tau) = F^{-1}(\tau) = \inf\{y:\ F(y) \ge \tau\} \tag{2.1.7}$$

for $0 < \tau < 1$. The value of the quantile function for each τ is called τ th quantile of Y. It is obvious that the median is the 0.5th quantile of Y.

By minimizing the risk of the loss function (2.1.5) we get the median of Y, as shown below.

$$\begin{aligned}
\therefore\ \ 0 &= \frac{\partial R\left(Y;\hat{\xi}\right)}{\partial \hat{\xi}} \\
&= \frac{\partial}{\partial \hat{\xi}} \int_{-\infty}^{+\infty} |y-\xi|\ \ dF(y) \\
&= \frac{\partial}{\partial \hat{\xi}} \left\{ \int_{-\infty}^{\hat{\xi}} (y-\hat{\xi})\ \ dF(y) - \int_{\hat{\xi}}^{+\infty} (y-\hat{\xi})\ \ dF(y) \right\} \\
&= 1 - 2F(\hat{\xi})
\end{aligned}$$

$$\therefore\ \ \hat{\xi} = F^{-1}(1/2) = \text{median}\ . \tag{2.1.8}$$

This solution may not be unique, but an interval of values may satisfy the minimization, since F is monotone (but not necessarily a strict monotone). In this case, the smallest value in the interval is chosen. Replacing the unknown F by the empirical distribution function F_n, when we have a sample of n iid random variables, gives the following risk function:

$$R(Y;\xi) = E\big(L(Y,\ \xi)\big)$$
$$= \int_{-\infty}^{+\infty} |y-\xi| \ dF(y)$$
$$= \int_{-\infty}^{+\infty} |y-\xi| \ d\left(\frac{1}{n}\sum_{i=1}^{n} I\left(y_i \le \xi\right)\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n}|y_i - \xi|. \qquad (2.1.9)$$

Minimizing the risk (2.1.9) gives the sample median. A minimizer that minimizes a function divided by n minimizes the function, as well. Or, a simpler argument is that we can choose $c = n$ in the loss function (2.1.5). Thus, we take the risk function (2.1.9) as the base of an objective function for finding the sample least absolute deviations fitting function, in a simpler form as follows:

$$R(\xi) = \sum_{i=1}^{n}|y_i - \xi|. \qquad (2.1.10)$$

Here, the problem is to find a value for ξ that minimises the objective function (2.1.10), which also minimises the risk function (2.1.9). This optimization problem is expressed as follows (see Bassett & Koenker, 1978):

$$\min_{\xi \in \mathrm{R}} R(\xi) = \min_{\xi \in \mathrm{R}} \sum_{i=1}^{n}|y_i - \xi|. \qquad (2.1.11)$$

The optimization problem (2.1.11) is a linear programming problem, which after adding $2n$ artificial variables $\{u_i, v_i\}, i = 1, \ldots, n$, is formulated as (see Charnes, Cooper & Ferguson, 1955; Wagner 1959):

$$\min_{(\xi,\mathbf{u},\mathbf{v}) \in \mathrm{R}\times\mathrm{R}_+^n\times\mathrm{R}_+^n} \left\{\mathbf{1}_n'\mathbf{u} + \mathbf{1}_n'\mathbf{v} \ \middle| \ \mathbf{1}_n\xi + \mathbf{u} + \mathbf{v} = \mathbf{y}\right\}. \qquad (2.1.12)$$

If ξ is a function of $\mathbf{x}$ through a linear relationship (or when the conditional absolute deviation function of y given $\mathbf{x}$ is required), $\xi = Q_y(\boldsymbol{\beta} \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$, then we want a minimizer $\hat{\boldsymbol{\beta}}$ that minimizes the objective function, as shown below,

$$\min_{\boldsymbol{\beta}\in \mathbb{R}^p} \sum_{i=1}^{n} \left| y_i - \mathbf{x}_i'\boldsymbol{\beta} \right| . \tag{2.1.13}$$

Using the linear programming reformulation of the problem as

$$\min_{(\boldsymbol{\beta},\mathbf{u},\mathbf{v})\in \mathbb{R}^p\times\mathbb{R}_+^n\times\mathbb{R}_+^n} \left\{ \mathbf{1}_n'\mathbf{u} + \mathbf{1}_n'\mathbf{v} \;\middle|\; \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{v} = \mathbf{y} \right\}, \tag{2.1.14}$$

gives the solution (minimizer) to the objective function (2.1.13), where the rows of $n \times p$ matrix $\mathbf{X}$ are transposes of vectors $\mathbf{x}_i$, for $i = 1, \ldots, n$.

Optimality Conditions:

On one hand, if $\hat{\boldsymbol{\beta}}$, the solution to the linear programming (2.1.14), is the minimizer of the objective function (2.1.13), it must be equivalent to the solution of differentiating the risk function of the objective function (2.1.13) and then letting it be equal to zero, or to that of a similar method. On the other hand, the risk function $R(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left| y_i - \mathbf{x}_i'\boldsymbol{\beta} \right|$ is not a smooth, but a piecewise continuous convex linear function, which is differentiable with respect to $\boldsymbol{\beta}$ except at those points at which one or more errors $Y_i - \mathbf{x}_i'\boldsymbol{\beta}$ are zero (see Karst, 1958). For this reason, instead of an ordinary derivative, we use the directional derivative with respect to $\boldsymbol{\beta}$ in all directions $\mathbf{w}$, with $\|\mathbf{w}\| = 1$, and define the function $\gamma(t) = \boldsymbol{\beta} + t\mathbf{w}$, for $t \in [-1,1]$, where $\gamma(0) = \boldsymbol{\beta}$ and $\gamma'(0) = \mathbf{w}$, as described below (See Bassett & Koenker, 1978):

$$\begin{aligned}
\nabla_{\mathbf{w}} R(\boldsymbol{\beta}) \equiv \nabla_{\vec{\mathbf{w}}} R(\boldsymbol{\beta}, \mathbf{w}) &= \frac{d}{dt}\left(R \circ \gamma(t) \right) \Big|_{t=0} \\
&= \frac{d}{dt} \sum_{i=1}^{n} \left| y_i - \mathbf{x}_i'\boldsymbol{\beta} - t\mathbf{x}_i'\mathbf{w} \right| \Big|_{t=0} \\
&= - \sum_{i=1}^{n} \psi\left(y_i - \mathbf{x}_i'\boldsymbol{\beta},\ -\mathbf{x}_i'\mathbf{w} \right) \mathbf{x}_i'\mathbf{w}
\end{aligned} \tag{2.1.15}$$

where

$$\psi(u,v) = \begin{cases} \operatorname{sgn}(u) , & \text{if } \; u \neq 0 \\ \operatorname{sgn}(v) , & \text{if } \; u = 0 \end{cases} . \tag{2.1.16}$$

and the function sgn(.) is the sign of its argument.

Here, the necessary minimization condition of a smooth function $\nabla R(\hat{\boldsymbol{\beta}})=0$ is met by the condition that the directional derivative at $\hat{\boldsymbol{\beta}}$ is nonnegative in all directions. Each p-tuple solution $\hat{\boldsymbol{\beta}}$, named a basic solution, interpolates at least p observations. But of course not each p-element set of observations might give a basic solution, since for some of observations there might be a linear relationship between their $\mathbf{x}$'s, i.e., the matrix of their $\mathbf{x}$'s might be singular. Let $N=\{1,\ 2,\ \ldots,\ n\}$ and h be any combination of p elements of N whose corresponding rows of $\mathbf{X}$ are not linearly related to each other. Also, let $\mathbf{X}(h)$ be the rows of matrix $\mathbf{X}$ corresponding to the elements of h, and similarly, $\mathbf{y}(h)$ the vector of elements of $\mathbf{y}$ corresponding to the elements of h and associated with $\mathbf{X}(h)$. A basic solution, which is obtained by the observations indexed by the elements of h, is as follows:

$$\mathbf{b}(h)=\mathbf{X}(h)^{-1}\mathbf{y}(h)\,. \tag{2.1.17}$$

If $\mathbf{b}(h)$ is to be a minimizer of the objective function (2.1.14), i.e., when the optimality holds in $\mathbf{b}(h)$, the directional derivative of objective function (2.1.13) at $\mathbf{b}(h)$ must be nonnegative in all directions $\mathbf{w}$, as shown below:

$$0\le\nabla_{\vec{\mathbf{w}}}R\big(\mathbf{b}(h),\mathbf{w}\big)=-\sum_{i=1}^{n}\psi\big(y_i-\mathbf{x}_i'\mathbf{b}(h),-\mathbf{x}_i'\mathbf{w}\big)\mathbf{x}_i'\mathbf{w}\,. \tag{2.1.18}$$

Let $\mathbf{v}=\mathbf{X}(h)\,\mathbf{w}$, then the optimality condition is simplified as:

$$0\le-\sum_{i\in h}\psi\big(0,-v_i\big)v_i-\sum_{j\notin h}\psi\Big(y_j-\mathbf{x}_j'\mathbf{b}(h),-\mathbf{x}_j'\mathbf{X}(h)^{-1}\mathbf{v}\Big)\mathbf{x}_j'\mathbf{X}(h)^{-1}\mathbf{v}$$

$$\therefore\quad 0\le\sum_{i\in h}|v_i|-\sum_{j\notin h}\psi\Big(y_j-\mathbf{x}_j'\mathbf{b}(h),-\mathbf{x}_j'\mathbf{X}(h)^{-1}\mathbf{v}\Big)\mathbf{x}_j'\mathbf{X}(h)^{-1}\mathbf{v}\,, \tag{2.1.19}$$

for all directions $\mathbf{v} \in \mathrm{R}^p$. All the directions of R^p are spanned by basis vectors $\mathbf{e}_i$, $i = 1,\ldots,p$. Thus, choosing $\mathbf{v} = \pm\mathbf{e}_i$, for $i = 1,\ldots.,p$, gives the following $2p$ inequalities.

$$\begin{cases} 0 \le 1 - \sum_{j \notin h} \psi\left(y_j - \mathbf{x}'_j\mathbf{b}(h), -\mathbf{x}'_j\mathbf{X}(h)^{-1}\mathbf{e}_i\right)\mathbf{x}'_j\mathbf{X}(h)^{-1}\mathbf{e}_i & ,i = 1,\ldots.,p \\ 0 \le 1 + \sum_{j \notin h} \psi\left(y_j - \mathbf{x}'_j\mathbf{b}(h), \mathbf{x}'_j\mathbf{X}(h)^{-1}\mathbf{e}_i\right)\mathbf{x}'_j\mathbf{X}(h)^{-1}\mathbf{e}_i & i = 1,\ldots.,p \end{cases} \quad (2.1.20)$$

If the distribution of Y is continuous, then the probability of $y_j - \mathbf{x}_j\mathbf{b}(h) = 0$ is zero, for any $j \notin h$. This means that the errors of the observations not indexed by the elements of h could not be zero. Consequently, the first argument of each of $\psi\left(y_j - \mathbf{x}'_j\mathbf{b}(h), -\mathbf{x}'_j\mathbf{X}(h)^{-1}\mathbf{e}_i\right)$ and $\psi\left(y_j - \mathbf{x}'_j\mathbf{b}(h), \mathbf{x}'_j\mathbf{X}(h)^{-1}\mathbf{e}_i\right)$ could not be zero and both of them are reduced to the simpler form $\operatorname{sgn}\left(y_j - \mathbf{x}'_j\mathbf{b}(h)\right)$, for $j \notin h$. Then, combining $2p$ inequalities in (2.1.20), the optimality condition at $\mathbf{b}(h)$ becomes the following simpler inequality:

$$-\mathbf{1}_p \le \left(\mathbf{X}(h)'\right)^{-1} \sum_{j \notin h} \operatorname{sgn}\left(y_j - \mathbf{x}'_j\mathbf{b}(h)\right)\mathbf{x}_j \le \mathbf{1}_p \quad (2.1.21)$$

The solution is unique if the inequalities are strict (See Bassett & Koenker, 1978).

2.2. SURE Models:

Consider a general system of m linear regression equations given by

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{e}_i, \qquad i = 1,\ldots,M \quad (2.2.1)$$

where, $\mathbf{Y}_i$ is a $T \times 1$ vector of the dependent variables, $\mathbf{e}_i$ is a $T \times 1$ vector of random errors with $E\left(\mathbf{e}_i\right) = \mathbf{0}$, and $\mathbf{X}_i$ is a $T \times k_i$ matrix of observations on k_i independent variables including a constant term, and $\boldsymbol{\beta}_i$ an $k_i \times 1$ vector of coefficients to be estimated. The number of equations in the system is M, where T is the number of observations per equation. k_i is the number of rows in the vector $\boldsymbol{\beta}_i$ which is equal to the number of independent variables in the i^{th} equation (including the intercept). Those M equations in the system can be written as

$$\mathbf{Y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{e}_1$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$\mathbf{Y}_M = \mathbf{X}_M\boldsymbol{\beta}_M + \mathbf{e}_M \tag{2.2.2}$$

and then they can be combined into a comprehensive model written as

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_M \end{pmatrix}.$$

This model can be rewritten compactly as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{e} \tag{2.2.3}$$

where, $\mathbf{Y}$ and $\mathbf{e}$ are of dimension $TM \times 1$, $\mathbf{X}$ is of dimension $TM \times k$, and finally $\mathbf{B}$ is of the dimension $k \times 1$, with $k = \sum_{i=1}^{M} k_i$.

Assumptions:

At this stage we have to make the following assumptions:

a) $\mathbf{X}_i$ is fixed with rank k_i.

b) $\text{plim}\frac{1}{T}\mathbf{X}_i'\mathbf{X}_i = \mathbf{Q}_{ii}$ is non-singular with finite and fixed elements, i.e., invertible.

c) In addition, we assume that $\text{plim}\frac{1}{T}\mathbf{X}_i'\mathbf{X}_j = \mathbf{Q}_{ij}$ also has finite and fixed elements.

d) $E\left(\mathbf{e}_i\mathbf{e}_j'\right) = \sigma_{ij}\mathbf{I}_T$, where σ_{ij} is the covariance between the i^{th} and the j^{th} equations.

e) $E\left(\mathbf{e}\right) = \mathbf{0}$, and

f) $E\left(\mathbf{ee}'\right) = \boldsymbol{\Psi} = \boldsymbol{\Sigma} \otimes \mathbf{I}_T$, where $\boldsymbol{\Sigma} = \left[\sigma_{ij}\right]_{M\times M}$ is a positive definite matrix and $\otimes$ is the Kronecker product. Thus, the errors in each equation are assumed to be homoscedastic and non-autocorrelated, but that there exists contemporaneous correlation between corresponding errors in different equations.

The OLS estimator of $\mathbf{B}$ is

$$\hat{\mathbf{B}}_{OLS} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \quad (2.2.4)$$

with the covariance matrix

$$\operatorname{var}\left(\hat{\mathbf{B}}_{OLS}\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Psi}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}. \quad (2.2.5)$$

Also, the Generalized Least Squares (GLS) estimator of $\mathbf{B}$ is given by:

$$\hat{\mathbf{B}}_{GLS} = \left(\mathbf{X}'\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{I}_T\right)\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{I}_T\right)\mathbf{Y} \quad (2.2.6)$$

with the covariance matrix

$$\operatorname{var}\left(\hat{\mathbf{B}}_{GLS}\right) = \left(\mathbf{X}'\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{I}_T\right)\mathbf{X}\right)^{-1}. \quad (2.2.7)$$

Generally, $\mathbf{\Sigma}$ is not an observable matrix and must be estimated from a sample of T observations from each equation. The estimated $M\times M$ matrix is denoted by $\mathbf{S}$, and replaces $\mathbf{\Sigma}$ in (2.2.6) and in all other places where $\mathbf{\Sigma}$ is used. The "feasible generalized least squares" (FGLS) estimator of $\mathbf{B}$ in (2.2.3), is computed as below.

$$\hat{\mathbf{B}}_{F} = \left(\mathbf{X}'\left(\mathbf{S}^{-1}\otimes\mathbf{I}_T\right)\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{S}^{-1}\otimes\mathbf{I}_T\right)\mathbf{Y} \quad (2.2.8)$$

The components of the matrix $\mathbf{S}=\left[s_{ij}\right]$ are estimated, as follows. First, we calculate

$$\begin{aligned}\hat{\mathbf{u}}_i &= \mathbf{Y}_i - \mathbf{X}_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i\mathbf{Y}_i \\ &= \left(\mathbf{I}_T - \mathbf{X}_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i\right)\mathbf{Y}_i = \mathbf{P}_i\mathbf{Y}_i. \quad (2.2.9)\end{aligned}$$

Then, the error vectors $\hat{\mathbf{u}}_i$ obtained from the OLS estimates of separate equations and the matrices $\mathbf{P}_i$, defined in (2.2.9), are used to get an unbiased estimation of s_{ij}, as follows:

$$s_{ij} = \frac{1}{trace\left(\mathbf{P}_i\mathbf{P}_j\right)}.\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j, \qquad \text{for } i, j = 1, \ldots, M \tag{2.2.10}$$

2.3. Seemingly Unrelated Median Regression Equations (SUMRE) Models:

When the covariance matrix of the error terms of the regression model is not a scalar covariance matrix of the form $\sigma^2\mathbf{I}$, the OLS estimator of the parameter $\boldsymbol{\beta}$ is still unbiased, but there is no guarantee that it is the best linear unbiased estimate (BLUE). In this case, when the covariance matrix of error terms is a positive definite matrix $\boldsymbol{\Sigma} \neq \sigma^2\mathbf{I}$, Aitken's generalized least squares method is preferable (See Aitken, 1934). The idea behind the Aitken's generalized least squares method is described below. With the lack of scalar covariance matrix of a regression model, using a proper transformation, we can obtain a new model with a scalar covariance matrix of error terms. Let $\mathbf{G}$ be an $n \times n$ full rank non-stochastic matrix, and then define the following new model;

$$\mathbf{Y}^* = \mathbf{GY}$$

$\mathbf{X}^* = \mathbf{GX}$, still is a non-stochastic matrix,

$\mathbf{e}^* = \mathbf{Ge}$,

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{e}^* \tag{2.3.1}$$

$\mathrm{var}\left(\mathbf{e}^*\right) = \mathbf{G\Sigma G}' = \sigma^2\mathbf{I}$, ($\mathbf{G}$ must suitably be chosen for this restriction).

The matrix $\sigma\boldsymbol{\Sigma}^{-1/2}$ is a good candidate for $\mathbf{G}$, by which the new estimation of $\boldsymbol{\beta}$ becomes

$$\mathbf{b_G} = \left(\left(\mathbf{X}^*\right)'\mathbf{X}^*\right)^{-1}\left(\mathbf{X}^*\right)'\mathbf{Y} = \left(\mathbf{X}'\mathbf{G}'\mathbf{GX}\right)^{-1}\mathbf{X}'\mathbf{G}'\mathbf{GY} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \tag{2.3.2}$$

The estimation is the same whatever value for σ^2 is assumed. Therefore, for simplicity and convenience we let $\sigma^2 = 1$, and consequently, $\mathbf{G} = \boldsymbol{\Sigma}^{-1/2}$. The new estimator $\mathbf{b_G}$ is called the generalised least squares estimator, denoted as $\hat{\boldsymbol{\beta}}_{\mathbf{GLS}}$.

The transformation (2.3.1) is in fact a multivariate standardization of the variables $\mathbf{Y}$ and $\mathbf{X}$, and the GLS method is then a simple application of the OLS method on those standardized variables. In this case, the original squared errors, which are determined in terms of squared Euclidian distances, are replaced by generalized squared errors, which are defined in terms of Mahalanobis distances. In 1962, Zellner used the notion of generalized least squares method, which at the time was proposed by Aitken for regression methods with one independent variable Y, through a smart change in the form of the design matrix of the data.

In our method, we use the same Zellner's design matrix and use the transformation (2.3.1) but instead of using the OLS method, which gives Aitken's GLS estimates, we use median regression. The difference is in the use of the norms used for calculating the errors. In the OLS method, the 2-norm (squared Euclidian distance) is used to calculate the errors and after transformation the distances are transformed to Mahalanobis distances, whereas in our method, the 1-norm (taxicab or city-block) distance is used to calculate the errors and after the transformation the distances are transformed to a new form of distance. The new method might be called the Generalized Least Absolute Deviations (GLAD) method. This enables us to estimate all the parameters of a system of seemingly unrelated median regression equations, where the correlations between the equations are also taken into account.

The relationship between the SURE method and the OLS method in some aspects is reflected in the relationship between the SURE method and ordinary median regression, and that is due to the structure of the design matrix of SURE models. For instance, if we use the OLS method to estimate the parameters of a SURE model, it results in separate OLS parameter estimations of SURE equations. The same argument holds for applying ordinary median regression on a SURE model. In this case, median regression estimation of the SURE models is the same as applying the median regression on each equation of the SURE model separately, as shown in Theorem 2.1, below.

Theorem 2.1: The median regression estimation of the SURE model in (2.2.3) takes the form

$$\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_1' \quad \hat{\boldsymbol{\beta}}_2' \quad \cdots \quad \hat{\boldsymbol{\beta}}_M']',$$

where $\hat{\boldsymbol{\beta}}_i$ is the estimation of the median regression model, $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{e}_i$, for $i = 1,\ldots,M$.

Proof: Any solution to the median regression estimation of parameters must satisfy the optimality condition (see Bassett & Koenker, 1978)

$$-\sum_{i=1}^{TM} \psi\left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}, -\mathbf{x}_i'\mathbf{w}\right) \mathbf{x}_i'\mathbf{w} \geq 0,$$

in all directions $\mathbf{w} \in \mathrm{R}^k$, where, $\mathbf{w} = \left(\mathbf{w}_1', \mathbf{w}_2', \ldots, \mathbf{w}_M'\right)'$, and $\hat{\boldsymbol{\beta}} = \left(\mathbf{b}_1', \mathbf{b}_2', \ldots, \mathbf{b}_M'\right)'$.

Let $\mathbf{w} = \left(\mathbf{0}', \mathbf{0}', \ldots, \mathbf{w}_j', \ldots, \mathbf{0}'\right)'$, for $j = 1, \ldots, M$, the optimality condition becomes

$$-\sum_{i=1}^{T} \psi\left(y_{ji} - \mathbf{x}_{ji}' \mathbf{b}_j, -\mathbf{x}_{ji}'\mathbf{w}_j\right) \mathbf{x}_{ji}'\mathbf{w}_j \geq 0. \qquad (2.3.3)$$

Therefore, according to the optimality condition (2.1.18) of the median regression estimation of parameters for the *j*th regression equation and the inequality (2.3.3), $\mathbf{b}_j$ must be a solution to the *j*th equation. This means $\hat{\boldsymbol{\beta}}_j = \mathbf{b}_j$, when $\hat{\boldsymbol{\beta}}_j$ is unique. █

According to this theorem, applying ordinary median regression on SURE models gives the notion of applying OLS instead of GLS on SURE models, since also the OLS estimation of parameters of the SURE model collapses to OLS estimation of separate equations. This means that applying ordinary median regression on SURE models abandons the information imbedded in the correlation matrix of cross equations errors. Our goal is to search for a method in which the median regression is applicable on SURE models and at the same time the information imbedded in the correlation matrix of cross equations errors is maintained.

In this paper we use the same notion of Aitken's GLS method to deal with the correlations between the equations of the SURE models, but this time for median regression, i.e., we use generalized least absolute deviation (GLAD) method. Then, the same transformation (2.3.1), which is used for GLS parameter estimation of SURE models in (2.2.3), is used, as follows.

$$\mathbf{Y}^* = \mathbf{\Psi}^{-1/2}\mathbf{Y} = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_T)\mathbf{Y}$$

$$\mathbf{X}^* = \mathbf{\Psi}^{-1/2}\mathbf{X} = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_T)\mathbf{X}$$

$$\mathbf{e}^* = \mathbf{\Psi}^{-1/2}\mathbf{e} = (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_T)\mathbf{e},$$

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{e}^*$$

$$\text{var}\left(\mathbf{e}^*\right) = \mathbf{\Psi}^{-1/2}\,\mathbf{\Psi}\,\mathbf{\Psi}^{-1/2} = \left(\mathbf{I}_M \otimes \mathbf{I}_T\right) = \mathbf{I}_{TM}. \tag{2.3.4}$$

Here, just like the objective function (2.1.13), we want to find a minimizer $\hat{\boldsymbol{\beta}}$ that minimizes the new objective function as shown below.

$$\min_{\boldsymbol{\beta}\in\mathsf{R}^k} \sum_{i=1}^{TM} \left| y_i^* - \mathbf{x}_i'^* \boldsymbol{\beta} \right|. \tag{2.3.5}$$

Using the linear programming reformulation of the problem as

$$\min_{(\boldsymbol{\beta},\mathbf{u},\mathbf{v})\in\mathsf{R}^k\times\mathsf{R}_+^{TM}\times\mathsf{R}_+^{TM}} \left\{ \mathbf{1}_{TM}'\,\mathbf{u} + \mathbf{1}_{TM}'\,\mathbf{v} \;\middle|\; \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u} + \mathbf{v} = \mathbf{y}^* \right\}, \tag{2.3.6}$$

gives the solution, where $k = \sum_{i=1}^{M} k_i$, and k_i is the dimension of $\mathbf{X}_i$, for $i = 1,\ldots,M$.

The objective function could be simplified as follows.

$$\min_{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_M)\in\mathsf{R}^{k_1}\times\mathsf{R}^{k_2}\times\ldots\times\mathsf{R}^{k_M}} \sum_{i=1}^{M}\sum_{j=1}^{T}\left|\sum_{m=1}^{M}\gamma^{im}\left(y_{mj} - \mathbf{x}_{mj}'\,\boldsymbol{\beta}_m\right)\right| \tag{2.3.7}$$

where k_i is the dimension of $\mathbf{X}_i$, for $i = 1,\ldots,M$, and γ^{im} is the im^{th} element of the matrix $\mathbf{\Sigma}^{-1/2}$, for $i,m = 1,\ldots,M$.

Bassett and Koenker (1978) have shown that the single equation median estimators asymptotically follow a multivariate normal distribution. Also, in our model $\dfrac{(\mathbf{X}^*)'\mathbf{X}^*}{TM}$ defined in (2.3.4) is a positive definite matrix, and the distribution of $\mathbf{e}^*$ is continuous. This means that the required assumptions mentioned in Bassett and Koenker (1978) are fulfilled, and hence the SUMRE estimators are also asymptotically normally distributed.

Theorem 2.2: With finite variances of equation errors, the SUMRE method estimation in (2.3.7) tends to the median regression estimations of separate equations in (2.2.3), as the correlations between the equations tend to zero.

Proof: When the correlations between the equations tend to zero, the limits of γ^{ik} for $i \neq k$ tend to zero, as well. In this case, concerning the objective function (2.3.7) and the fact that $\gamma^{ii} > 0$, we have

$$\lim_{\substack{\gamma^{ik} \to 0 \\ i \neq k}} \left(\min_{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M) \in \mathrm{R}^{k_1} \times \mathrm{R}^{k_2} \times \ldots \times \mathrm{R}^{k_M}} \sum_{i=1}^{M} \sum_{j=1}^{T} \left| \sum_{k=1}^{M} \gamma^{ik} \left(y_{kj} - \mathbf{x}'_{kj} \boldsymbol{\beta}_k \right) \right| \right)$$

$$= \min_{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M) \in \mathrm{R}^{k_1} \times \mathrm{R}^{k_2} \times \ldots \times \mathrm{R}^{k_M}} \lim_{\substack{\gamma^{ik} \to 0 \\ i \neq k}} \left(\sum_{i=1}^{M} \sum_{j=1}^{T} \left| \sum_{k=1}^{M} \gamma^{ik} \left(y_{kj} - \mathbf{x}'_{kj} \boldsymbol{\beta}_k \right) \right| \right)$$

$$= \min_{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M) \in \mathrm{R}^{k_1} \times \mathrm{R}^{k_2} \times \ldots \times \mathrm{R}^{k_M}} \sum_{i=1}^{M} \left(\gamma^{ii} \sum_{j=1}^{T} \left| y_{ij} - \mathbf{x}'_{ij} \boldsymbol{\beta}_i \right| \right)$$

$$= \sum_{i=1}^{M} \left(\min_{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M) \in \mathrm{R}^{k_1} \times \mathrm{R}^{k_2} \times \ldots \times \mathrm{R}^{k_M}} \gamma^{ii} \sum_{j=1}^{T} \left| y_{ij} - \mathbf{x}'_{ij} \boldsymbol{\beta}_i \right| \right)$$

$$\cong \sum_{i=1}^{M} \left(\min_{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M) \in \mathrm{R}^{k_1} \times \mathrm{R}^{k_2} \times \ldots \times \mathrm{R}^{k_M}} \sum_{j=1}^{T} \left| y_{ij} - \mathbf{x}'_{ij} \boldsymbol{\beta}_i \right| \right). \qquad (2.3.8)$$

This means that the objective function of SUMRE model is equivalent to the summation of objective functions of median regressions for separate equations. Mathematically, complete equivalence holds when the correlations between the equations are exactly zero. ∎

The above theorem indicates that even if there is no statistically significant correlation between the equations, using SUMRE method instead of SURE method will not damage the estimation of parameters.

We know that GLS estimation of SURE models will collapse to OLS estimations of separate regression equations, when $\mathbf{X}_i$ of all equations are identical, whereas this is not the case with the estimation of SUMRE models. This could be verified from the objective function (2.3.7). In other words, having the same variables $\mathbf{X}$ in common between all equations does not cause the collapse of SUMRE estimations to the estimations of separate median regression equations. An obvious reason of that is due to the problematic estimation of multivariate median regression, since having $\mathbf{X}_i$ variables identical in all equations changes the SURE model to an ordinary multivariate regression model.

3. Monte Carlo Design and Experiment

In the following two subsections, we discuss the way we designed our Monte Carlo experiment and the idea behind it. The idea behind the design is due to some criteria we have used to assess the efficiency of our method and the factors that, intuitively and based on the theory, may change the efficiency.

3.1. Criteria for performance evaluation

In a Monte Carlo study, set up to look at the good properties of the estimators for the purpose of comparison between them, we calculate the *mean squared error* (*MSE*) of the estimators through *generalized sample variance*, *total sample variance* and *squared bias* of the estimators. Those calculations are done by simply calculating the estimators in repeated samples under fixed combinations of conditions (factors). More precisely, the mean vectors and the covariance matrices of the parameter estimates are estimated from 2000 replications of the Monte Carlo experiment, for each combination of imposed factors. The idea behind choosing these statistics arises from the following argument.

For a single parameter θ, the efficiency of an estimator $\hat{\theta}$ is defined as,

$$Eff(\hat{\theta}) = 1 \big/ MSE(\hat{\theta}), \qquad (3.1.1)$$

where,

$$MSE(\hat{\theta}) = E\left(\hat{\theta}-\theta\right)^2 = E\left(\hat{\theta}-E(\hat{\theta})\right)^2 + E\left(E(\hat{\theta})-\theta\right)^2 = \operatorname{var}\left(\hat{\theta}\right) + bias^2\left(\hat{\theta}\right). \qquad (3.1.2)$$

In our simulation, we look at the efficiency of the estimators from two different points of view. First, if we take the multivariate MSE, we need to estimate the total sample variance and squared bias of $\hat{\boldsymbol{\theta}}$, as shown below.

$$MSE(\hat{\boldsymbol{\theta}}) = E\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)'\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)$$

$$= E\left(\hat{\boldsymbol{\theta}} - E(\hat{\theta})\right)'\left(\hat{\boldsymbol{\theta}} - E(\hat{\theta})\right) + E\left(E(\hat{\theta}) - \boldsymbol{\theta}\right)'\left(E(\hat{\theta}) - \boldsymbol{\theta}\right)$$

$$= trace\left(E\left(\hat{\boldsymbol{\theta}} - E(\hat{\theta})\right)\left(\hat{\boldsymbol{\theta}} - E(\hat{\theta})\right)'\right) + \left(E(\hat{\theta}) - \boldsymbol{\theta}\right)'\left(E(\hat{\theta}) - \boldsymbol{\theta}\right)$$

$$= trace\left(\text{cov}(\hat{\boldsymbol{\theta}})\right) + \left(bias(\hat{\boldsymbol{\theta}})\right)'\left(bias(\hat{\boldsymbol{\theta}})\right) \qquad (3.1.3)$$

Second, for a multidimensional parameter $\boldsymbol{\theta}$, if we take the ellipsoid of estimators $\hat{\boldsymbol{\theta}}$, which is centred at $\boldsymbol{\theta}$, the efficiency is considered as the inverse of the volume of that ellipsoid. In this case, we define a matrix of mean squared error, as follows.

$$MSE(\hat{\boldsymbol{\theta}}) = E\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)'$$

$$= E\left(\hat{\boldsymbol{\theta}} - E(\hat{\theta})\right)\left(\hat{\boldsymbol{\theta}} - E(\hat{\theta})\right)' + E\left(E(\hat{\theta}) - \boldsymbol{\theta}\right)\left(E(\hat{\theta}) - \boldsymbol{\theta}\right)'$$

$$= \text{cov}\left(\hat{\boldsymbol{\theta}}\right) + \left(E(\hat{\theta}) - \boldsymbol{\theta}\right)\left(E(\hat{\theta}) - \boldsymbol{\theta}\right)'. \qquad (3.1.4)$$

As the squared volume of ellipsoid of estimators centred at $\boldsymbol{\theta}$ is proportional to the determinant of $MSE(\hat{\boldsymbol{\theta}})$, we will have

$$\det\left(MSE(\hat{\boldsymbol{\theta}})\right) \leq \det\left(\text{cov}(\hat{\boldsymbol{\theta}})\right) + \det\left(\left(E(\hat{\theta}) - \boldsymbol{\theta}\right)\left(E(\hat{\theta}) - \boldsymbol{\theta}\right)'\right)$$

$$\therefore \det\left(MSE(\hat{\boldsymbol{\theta}})\right) \leq \det\left(\text{cov}(\hat{\boldsymbol{\theta}})\right), \qquad (3.1.5)$$

since the second matrix on the right hand side is not of full rank but of rank one. Thus, a reduction in the determinant of the covariance matrix (generalized variance) of $\hat{\boldsymbol{\theta}}$ increases its efficiency.

Among two estimators, the one with smaller MSE (defined in 3.1.3 but not in 3.1.4) is usually considered to be more efficient than the other. To compare the efficiency of an estimator with the efficiency of another estimator, we usually use the relative efficiency, as shown below.

$$\text{relative efficiency of } \hat{\theta}_2 \text{ to } \hat{\theta}_1 = \frac{\text{efficiency of } \hat{\theta}_2}{\text{efficiency of } \hat{\theta}_1} = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}. \tag{3.1.6}$$

We use the same formula to compute the relative efficiency of each of the SURE GLS estimators and the median regression estimators of single equations to the efficiency of SUMRE estimators. A value greater than 1 means relative inefficiency of SUMRE estimators to the estimators of one of the two other methods, whereas a value less than 1 is an indicator of relative efficiency of SUMRE estimators.

All mathematical expectations are computed based on the empirical distribution function, i.e. the method of moments for estimating, due to the fact that the number of repeated runs of the Monte Carlo experiment was relatively large (2000 replications). Also, asymptotically normal distribution could be used for any statistical inference about the estimators.

3.2. Factors that vary in the experiment

A number of factors can affect estimation properties of the SUMRE method. These are: the number of equations (M), the sample size (T), the skewness of the distribution of the errors from each equation, and the correlations between errors from different equations. In the rest of this section, we will explain these factors, in some more detail.

The Monte Carlo experiment was performed by generating the data as follows:

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{1}_T & \mathbf{X}_i^* \end{bmatrix}_{T\times k_i}, \qquad (3.2.1)$$

Where $\mathbf{X}_i^*$ is a $T\times(k_i-1)$ matrix, whose rows are vectors with the multivariate normal distribution $N_{k_i-1}\left(\boldsymbol{\mu}_i, \sigma_{\mathbf{X}_i}^2 \mathbf{I}_{k_i-1}\right)$,

$$\mathbf{Y}_i(\tau) = \mathbf{X}_i'\boldsymbol{\beta}_i(\tau) + \boldsymbol{\varepsilon}_i, \qquad (3.2.2)$$

$$\boldsymbol{\beta}_i(\tau) = \begin{bmatrix} 1 & \boldsymbol{\beta}_i^*(\tau) \end{bmatrix}, \qquad (3.2.3)$$

$$\boldsymbol{\beta}_i^*(\tau) = \alpha(\tau,\gamma_i)\boldsymbol{\beta}_i, \qquad (3.2.4)$$

for a constant vector $\boldsymbol{\beta}_i$, and $\alpha(\tau,\gamma_i)$ an exponential function of skewness level γ_i and the quantiles τ, for $0<\tau<1$, and $\boldsymbol{\varepsilon}_i \sim N_T\left(\mathbf{0}, \sigma_i^2\mathbf{I}_T\right)$, for $i = 1,2,\ldots,M$.

Our primary interest lies in the analysis of systemwise estimation, and thus the number of equations to be estimated is of central importance. As the number of equations grows the computation time becomes longer, and we took a system with five equations as our largest model. This represents a fairly medium-sized model of the type that is used in, for example, agriculture, economics or labour markets, while a three-equation system is a typical small model. Moreover, different levels of strength of correlations between these equations have been imposed.

To get the desired strength of correlations we impose a monotonically decreasing functional relationship between $\sigma_{X_i}^2$ and the correlation coefficient as follows:

$$\sigma_{\mathbf{X}_i}^2 = g(\rho). \qquad (3.2.5)$$

Here, we suppose that the correlation coefficients are almost the same (which are supposed to be ρ) between all the equations. The correlation coefficient is Pearson's product-moment coefficient of sample correlation, computed as follows.

$$r_{Y_i,Y_k} = \frac{\sum_{j=1}^{T} Y_{ij}Y_{kj} - T\overline{\mathbf{Y}}_i\overline{\mathbf{Y}}_k}{\sqrt{\sum_{j=1}^{T} Y_{ij}^2 - T\overline{\mathbf{Y}}_i^2}\sqrt{\sum_{j=1}^{T} Y_{kj}^2 - T\overline{\mathbf{Y}}_k^2}} \quad , \; i,k = 1,\ldots,M \, . \qquad (3.2.6)$$

Using different functions $g_{ij}(\rho_{ik})$ instead of $g(\rho)$, gives different correlation coefficients between pairs of equations. For simplicity, we took almost the same level of correlation between all the equations, as we mentioned just above.

When generating the data, we also imposed different distributional properties in terms of the degrees of skewness of the errors. This goal was achieved by multiplying the vector $\boldsymbol{\beta}_i$ by an exponential function of both the level of skewness γ_i, and the quantiles $\tau \in (0,1)$, during the processes of generating the data.

The skewness is computed according to the following formula.

$$\gamma_i = \frac{T.\sqrt{T-1}}{T-2} . \frac{\sum_{j=1}^{T}\left(y_{ij} - \overline{y}_i\right)^3}{\left(\sum_{j=1}^{T}\left(y_{ij} - \overline{y}_i\right)^2\right)^{3/2}} \quad , \; i = 1,\ldots,M \, . \qquad (3.2.7)$$

Using proper coefficients in the functions $g(\rho)$ and $\alpha(\tau,\gamma_i)$ it is possible to get a data set with desired level of correlations between the equations and level of skewness of the errors, respectively. However, because of the interaction between these two functions, the data set probably would not have the required properties. Therefore, we regularly checked the data set, through computing the coefficients of correlations between the equations and the coefficients of skewness of errors using the formulas (3.2.6) and (3.2.7), respectively, and we then removed those data sets that violated the desired properties. Thus, what were important for us in this study were not the exact well-known probability distributions of errors, but the desired properties of the distributions, represented in level of skewness of the distributions and level of correlations between the distributions.

Another prime factor that affects the performance of the SUMRE method is the number of observations. We have investigated sample sizes of 50, 250 and 1000 observations that will cover small, medium and large samples. Also, for each replication of the experiment with the desired combination of factors we generated a set of data, estimated the model using that data set, and computed the generalized sample variance, total sample variance, squared bias and the mean squared errors of the estimators.

To make a comparison between the estimators of the SUMRE method and of each of the ordinary SURE method and separate ordinary median regressions of equations, we performed conventional median regression on each equation and the ordinary SURE method on all equations together. Once again, we computed the generalized sample variance, total sample variance, squared bias and the MSE of the estimators obtained in each of these two methods. Then, relative efficiencies of SURE method and the method of separate median regression equations to the efficiency of SUMRE method were computed through the ratios of the generalized sample variance, total sample variance and MSE of their estimators.

The factors that vary for different models are presented in Table 3.2.1, and the results of the simulation, for systems of 3 and 5 equations, are summarized in Tables 1 to Table 12 in the appendix. All the calculations were performed using the GAUSS, version 8.0.6 program.

Table 3.2.1 Values of Factors that vary for Different Models

Factor	Symbol	Design
No. of Equations in the Model	M	3, 5
No. of Observations in the Simulated Sample	T	50, 250, 1000
Level of Correlations between Equations	ρ	Low (0.0 - 0.2), Medium (0.4 – 0.6) High (over 0.8)
Level of Skewness	γ	Low (0.0 - 0.5), Medium (1.5 – 2.0) High (over 3.0)

4. Results

The results of the simulation, represented in squared bias, MSE and each of the determinant ratio and the trace ratio of the covariance matrix of SUMRE estimators to those of each of ordinary SURE GLS estimators and the estimators of the separate median regressions of single equations, are presented in Table 1 to Table 12 in the appendix.

One can look at the results of the simulation from different perspectives, in order to draw comparisons between efficiencies of the SUMRE method and the two other methods mentioned just before. The efficiencies of any two methods (SUMRE with SURE or SUMRE with separate median regressions) could be compared at the presence of a specific level of skewness, level of correlation(s), number of equations and sample size, and/or any combination of these factors. At first glance, one may think that the changes in ratios are regular and parallel to the changes in these factors, but by delving into the columns of the tables, it would be discovered that the changes could be interpreted differently, when the results are looked at from different perspectives.

To reduce the terminology, we use the abbreviations *L*, *M*, *H*, *S* and *C* to stand for *Low*, *Medium*, *High*, *Skewness* and *Correlation*, respectively. Then, for instance, *MSHC* stands for a model with medium level of skewness of all equations errors and high level of correlations of cross equations errors. Or, for instance, *LC* stands for low correlations between errors from different equations, and so on.

Another attempt to simplify the terminology is using the term determinant ratio to mean the ratio of the determinant of the covariance matrix of SUMRE estimators to the determinant of the covariance matrix of SURE GLS estimators or to that of separate median regressions of single equations. Also, trace ratio, by analogy, is the ratio of the traces of the corresponding covariance matrices. Each of MSE ratio and squared bias ratio is meant in a way analogous to that in which trace ratio and determinant ratio are labelled.

Finally, to further simplify the terminology, we use the term level of skewness to mean the level of skewness of errors from each equation, and by level of correlations the level of correlations between cross equations errors. Also, the method of separate median regressions is used as the shorthand for the method in which we deal with each of the equations of SURE model separately and apply conventional median regressions on single equations. And, by SURE estimators we mean feasible SURE GLS estimators.

4.1 Determinant Ratio Comparison

The changes in determinant ratio do not follow the same pattern when we compare the SUMRE method to the SURE method and when it is to be compared to the method of separate median regressions. The difference is due to the fact that strong correlations between cross equations errors are beneficial to the SURE method and detrimental to separate median regressions of single equations whereas the converse is true for the high levels of skewness. These differences are explained in the following two subsections. In the entire subsection 4.1, we use Table 1 and Table 2, to make comparisons.

4.1.1 SUMRE versus SURE GLS

From the tables, assuming the low skewness as fixed, by which the cases *LSLC*, *LSMC* and *LSHC* are included, the determinant ratio (of SUMRE to SURE), without exception, increases as the level of correlations increases. This ratio, which is, in a sense, inefficiency of the SUMRE method related to the SURE method, increases for increasing number of equations, as well.

Surprisingly, no asymptote of the determinant ratio could be revealed from the simulation tables for the case of *LS* models. In other words, in the case of the *LS* models the sample size even asymptotically does not have any effect on the gap between SUMRE method and SURE method. However, the determinant ratios tend to zero asymptotically, in the presence of a high level of skewness (non-*LS* cases), no matter what the level of correlations is. The more equations in the model and the higher the level of skewness, the more rapid the reduction in the determinant ratios occurs, for non-*LS* cases.

For fixed levels of correlation, the gap between SUMRE and SURE methods reduces rapidly as the level of skewness increases. The reduction of this gap continues in a way that after relatively small rises in the level of skewness, SUMRE method estimators become more efficient than those of SURE method, based on the value of the determinant ratio which becomes less than 1, and further on very close to zero.

As we mentioned above, for LS models, the determinant ratio becomes larger and larger as the model moves from the case LSLC towards LSMC and LSHC cases. In other words, with a fixed low level of skewness, the determinant ratio rises with incremental levels of correlation. Generalising this notion for all levels of skewness, may lead to getting the wrong idea. As it's obvious from the tables, in the presence of a level of skewness, the determinant ratio is the maximum at medium levels of correlation. By looking at the tables in more detail, we observe that with a medium level of skewness and a medium level of correlation, the SURE method is more efficient than the SUMRE method, whereas in all other non-LS models, the SUMRE method is more efficient.

4.1.2 SUMRE versus Separate Median Regressions

As Theorem 2.2 (on page 15) indicates, for very low correlations between errors from different equations, the SUMRE method almost collapses into separate median regression models of single equations. The collapse is complete for mathematically zero correlations between cross equations errors. Far from mathematical models, statistical models or simulated models do not yield exact zero correlations between cross equations errors. However, statistical models or simulated models with very low levels of correlations could asymptotically resemble mathematical models with zero correlations between cross equations errors. This resemblance is very sensitive to the level of skewness. The less the level of skewness, the stronger the resemblance is.

Overall, for low levels of correlation, the gap between the SUMRE method and the method of separate median regressions is not very large. This gap asymptotically tends to zero, looking at the value of determinant ratio which increases to 1. The rise of the value of the determinant ratio becomes slower as the number of equations included in the model increases. This means that in spite of the sensitivity of asymptotically resembling a SURE model with zero correlations to high levels of skewness, the sensitivity increases even more as the number of equations included in the model increases.

What is expected intuitively and could be seen from the tables is the increasing relative efficiency of the SUMRE method compared to the method of separate median regressions, when the level of correlations increases. In the presence of correlations but no levels of skewness, no asymptote of the determinant ratio could be guessed at. This means that the relative efficiency of the SUMRE method compared to the method of separate median regressions, which is in order in the presence of correlations and the lack of skewness, remains fixed and free from the effect of the sample size, no matter what the level of skewness is. However, the determinant ratio asymptotically tends to zero in the presence of correlations and high levels of skewness.

4.1.2 SURE versus Separate Median Regressions

It is not the aim of this paper to compare the SURE method with the method of separate median regression equations, but we do want to take note of something here. Of the matching determinant ratios in Table 1 and Table 2, and associated with each of the two methods of SURE and separate median regressions, the greater ratio means its associated method is more efficient. With lower levels of skewness and higher levels of correlations, for instance *MSMC*, *MSHC* and *LS* models, the determinant ratios corresponding to the SURE method are greater than their matching determinant ratios corresponding to the method of separate median regressions, which again means the SURE method is more efficient. The converse is true with higher levels of skewness and lower levels of correlations, e.g., all *HS* models.

4.2 Trace Ratio Comparison

The idea behind selecting trace or determinant of the covariance matrix of parameters for checking the efficiency of a vector parameter estimator arises from two different viewpoints of looking at the MSE of a vector parameter, as described in subsection 3.1. With the trace of covariance matrix of the parameters, we maintain the focus on only the variances of the parameters, whereas with the determinant of the covariance matrix we change this focus and look at the covariances between the parameters, as well. Therefore, it will not be surprising that generally the trace ratio and determinant ratio are not equal or even not consistent, under the same factors imposed on (or present in) a SURE model, when we compare the SUMRE method to each of the SURE method and/or the method of separate median regression equations.

However, we must notice that the trace ratio by itself (without adding it to the squared bias) does not give a correct result when we check the efficiency of a vector parameter estimator. Nevertheless, for the sake of simplicity someone may take only the trace to check the efficiency of a vector parameter estimator. In this paper, apart from trace and squared bias, we will look at the MSE of the vector parameter estimators and MSE ratio, as well. In the following two subsections we are going to explain the differences between the trace ratios of the SUMRE method to each of the SURE method and the method of separate median regressions. We devote the entire subsection 4.2 to a discussion of the results in Tables 3 and Table 4.

4.2.1 SUMRE versus SURE GLS

Most of the facts discussed in the subsection (4.1.1) are consistent with the facts that could be discovered from a comparison of the SUMRE method to the SURE method based on trace ratio, except for some slight differences.

Keeping the level of correlations fixed, the trace ratio increases as the level of skewness increases. With very low levels of skewness, the trace ratio is greater than 1, which is an indicator of inefficiency of the SUMRE method relative to the SURE method. With the level of skewness fixed (excluding the case *MSMC*, just like subsection 4.1.1), incremental levels of correlations do not reduce the trace ratio a great deal but only slightly. A reason for the peculiar behaviour of *MSMC* case may be the higher bias of SURE parameters in the presence of a medium level of skewness and a medium level of correlations, since with small changes in MSE, according to the equation (3.1.3), when the squared bias increases, the trace decreases. Finally, neither the sample size nor the number of equations taken into the model has any remarkable effect on the trace ratio.

4.2.2 SUMRE versus Separate Median Regressions

Most of the facts that could be discovered from Table 3 and Table 4, concerning the trace ratio of the SUMRE method compared to the method of separate median regressions, are very nearly consistent with the facts that are discussed in subsection 4.1.2.

With a fixed level of correlations, the trace ratio does not change remarkably with increases in the level of skewness, whereas with a fixed level of skewness, a considerable reduction of trace ratio occurs with each rise of the level of correlations. Trace ratio is neither remarkably changed by the sample size nor by the number of equations of the model.

4.3 Squared Bias Ratio Comparison

In the following two subsections, we are going to compare the efficiency of the SUMRE method and the efficiencies of each of SURE method and the method of separate median regressions, based on the squared bias ratio. As we mentioned before, according to the equation (3.1.3) none of the squared bias ratio and trace ratio alone could yield a complete and valid conclusion about the efficiency or even the relative efficiency of some estimators. However, first, for the sake of simplicity, one may desire to compare the efficiency of some parameters, based on the squared bias ratio. Second, one may not wish to look at the efficiency or relative efficiency of some estimators, but merely wish to look at their bias. In both of these cases, the squared bias ratio can give a representation of some properties of an estimator. Throughout subsection 4.3, we use Table 5 and Table 6 to make comparisons. We compare the SUMRE method separately with each of SURE method and the method of separate median regression equations, in the following two subsections.

4.3.1 SUMRE versus SURE GLS

In all cases, except the *LSHC* case and asymptotically each of *LSMC* and *MSHC* cases, the squared bias ratio is less than 1. Furthermore, this ratio is very small in *HS* cases. However, in the absence of skewness the ratio becomes greater than 1, which is an indicator of a smaller bias of SURE parameters in those cases (*LS* cases). In the presence of correlations, the ratio grows in the magnitude asymptotically. This is because of the asymptotically unbiasedness of the parameters of the SUMRE method.

Another fact concerning the squared bias ratio of the SUMRE method to the SURE method is that keeping whatever level of correlations as fixed, the ratio increases with any increase in the level of skewness, and taking whatever level of skewness as fixed, the ratio decreases with almost any rise in the level of correlation.

4.3.2 SUMRE versus Separate Median Regressions

In the case *LSLC*, the squared bias of separate median regressions estimators is less than the squared bias of SUMRE estimators. Moreover, even in this case *LSLC*, the two squared biases are asymptotically equal, since the ratio reduces to tend to 1 asymptotically. This is due to the fact that estimators of both methods are asymptotically unbiased. Furthermore, in all non-*LC* cases, the ratio is less than 1. Finally, it is worth noticing that there seems to be the effect of the interaction between the skewness and the correlations on the squared bias ratio.

4.4 MSE Ratio Comparison

The MSE ratio is computed using equation (3.1.6), and this is true only for the MSEs defined in equation (3.1.3). For MSEs defined in equation (3.1.4), using the inequality (3.1.5), the MSE ratio is replaced by the determinant ratio. Therefore, depending on the squared bias ratio whether it is small or large, the trend of changes in MSE ratio is somehow consistent and parallel to the trend of changes in trace ratio, for each pair of methods—SUMRE to SURE, or SUMRE to the method of separate median regressions. In the following two subsections we shed light on those changes, in more detail. Throughout subsection 4.4, we use only the results in Table 7 and Table 8.

4.4.1 SUMRE versus SURE GLS

With the exception of *LS* cases, the MSE ratio is less than 1 in all the cases, which indicates the relative efficiency of SUMRE estimators to the estimators of the SURE method. Taking each level of correlation as fixed, the ratio decreases as the level of skewness increases. If we interchange these two factors, i.e., fix the level of skewness and increase the level of correlations, the ratio seems to be odd in *MSMC* case, due to a decrease in the squared bias in that case. Also, the gap between SUMRE and SURE methods is not reduced asymptotically. Overall, the results in MSE ratio are almost consistent with the results in trace ratio.

4.4.2 SUMRE versus Separate Median Regressions

In the presence of correlations, the relative efficiency of SUMRE estimators over the estimators of separate median regressions is improved, according to the value of the MSE ratio, which is less than 1 in most of non-*LC* cases. However, a very high level of skewness disturbs the delicate balance between the levels of skewness and correlations in *HSLC* case. It is worth noticing that the results based on the MSE ratio are almost consistent with the results based on the trace ratio. Finally, the MSE ratio is not asymptotically changed in a considerable amount.

4.5. Squared Bias Changes

In this section, we are more interested in focusing on each method by itself—and we are not going to compare any methods with each other. Because of that, the results in Table 9 and Table 10, that are the only results used in this subsection, are not ratios (scale free) but the squared bias of each method (scale squared). As we mentioned before, the results of squared bias are more meaningful for the efficiency of some estimators if the trace of the covariance matrix of the estimators is also taken into account. However, by looking at the biasness of the estimators one can get a useful idea about the affect of any combination of factors present in the model on the estimations.

For SURE estimators, we realize that, without exception, the squared bias rises at each level of correlation, when the level of skewness increases. The squared bias reduces as the sample size increases. An interesting fact is that in *MS* cases, the squared bias of SURE estimators rises up as the level of correlations rises. For other cases, the changes are not regular, especially in *MC* cases.

Concerning estimators of separate median regressions, the squared bias increases with each rise in the level of skewness, at all levels of correlations, and the squared bias increases with each rise in the level of correlations, at all levels of skewness. In other words, the changes in squared bias are parallel to the changes in each of the level of skewness and the level of correlations. However, the changes are reduced as the sample size increases.

The changes in squared bias of SUMRE estimators are not exactly parallel with the changes in squared bias of the two other methods. But, some similarities are present in the changes occurring to the squared bias of all the three methods. One such is that for all of the three methods, at a fixed level of correlation, the squared bias increases as the level of skewness increases. The other one is that the squared bias decreases as the sample size increases.

The changes in squared bias for all three methods are almost parallel, except for the case of *LSMC*, where the changes in squared bias of SUMRE estimators and SURE estimators are not as regular as the change of squared bias of the estimators of separate median regressions is.

4.6 MSE Changes

In this subsection, analogous to subsection 4.5, we look at the changes that occur in the MSE (defined in 1.3.3) of the estimators of each of the methods SUMRE, SURE and separate median regressions, using the results presented in Table 11 and Table 12 exclusively. When we do not look at the MSE of estimators comparatively, the results based on the MSE defined in (3.1.4) are almost consistent with those based on the MSE defined in (3.1.3). Using the linear algebra theory, for the MSE defined in (3.1.4) we have

$$\begin{aligned}
\det\left(MSE(\hat{\boldsymbol{\theta}})\right) &= \det\left(E\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)' \right) \\
&\le \left(\frac{1}{k}\right)^k trace\left(E\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)' \right)^k \\
&\le trace\left(E\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)' \right)^k \\
&= \left(E\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)'\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \right)^k
\end{aligned}$$

This means any reduction in MSE defined in (3.1.3) automatically implies a reduction in the MSE defined in (3.1.4).

Concerning SURE estimators, for the non-*MC* cases the MSE increases as the level of skewness increases. Among the *MC* cases, the MSE reduces for the *MSMC* case. Keeping the level of skewness as fixed, the behaviour of MSE is odd for the *MC* cases.

MSE of the estimators of separate median regressions is reduced for the *MS* cases, whatever the level of correlations is, but their MSE increases at fixed levels of skewness, as the level of correlations increases, except for the case *MSHC*.

For SUMRE estimators, the MSE almost (but not quite) increases as the level of skewness increases. Depending on the balance between the level of skewness and the level of correlations the MSE increases with very high levels of skewness and very high levels of correlation. What is common between the changes in MSE of the estimators of all the above methods is that the MSE decreases as the sample size increases.

5. Empirical Application

To assess the performance of our SUMRE method, we consider an empirical example and compare its estimators with the feasible SURE GLS estimators. In our empirical example we use some data taken from Multi-Generation Register at Statistics Sweden on three generations of male immigrants from Finland to Sweden. The male immigrants have been selected based on their grandfather's (the father's father) place of birth which in our case is Finland. A more thorough explanation of the variables in the data set can be found in (Hammarstedt 2009; Ekberg, Hammarstedt and Shukur, 2009).

The first group in the study contains all the male individuals who were born in Finland and living in Sweden in the year 1960. This group constitutes the first-generation immigrants in the study. For all individuals in the first generation we have data on yearly earnings and their background variables like: age, educational attainment, civil status and region of residence in Sweden in the year 1968. All the individuals in that group were between 25 and 64 years of age at the observation time (in year 1968).

The second group in the study contains the second generation male immigrants, which are the first group individual's biological sons. Yearly earnings, as well as data on the background variables like: age, educational attainment, civil status and region of residence in Sweden, are observed for the second generation in the year 1980. All the individuals in that group were between 25 and 64 years of age at the observation time (in year 1980).

Finally, the third group in the study contains the third-generation male immigrants, which are the second group individual's biological sons. For this group, we have the same data as we mentioned above for the first and second generation of male immigrants, in the year 2003. All the individuals in that group were between 25 and 64 years of age at the observation time (in year 2003).

For each individual, earnings are defined as yearly taxable income from work which includes income from wage-employment, self-employment, sickness pay and parents' allowances. We only include individuals who are in their working ages (i.e. 25-64 years of age) and, furthermore, active on the labour market (i.e. have positive earnings) at the observation time.

To link the generations together, we first identify individuals from the first generation, and then their sons and the sons of their sons (their grandsons) that have earnings. Since there is the possibility that individuals from the first generation might have more than one son having earnings, and their sons, in turn, might have more than one son with earnings, information about the first and second generations might appear more than one time in the data. More precisely, since the correspondence between the first, the second and the third generation of male immigrants having earnings is not a one-to-one correspondence (bijection), for the second and especially the first generations we will have some replicated observations. Otherwise, to get a one-to-one correspondence between the individuals of different generations, we had to remove some individuals from the second and especially the third generations, and among two or more sons having earning select only one of them to be taken into the sample. In this case, we would have two problems. First, which son should be selected? Second, the samples would be much smaller. This means that the removal of the individuals would not be without bias and a vast body of information would be lost. Though this is not happening very often elsewhere, we decided to include replicated individuals from the first and the second generations in the data when they have more than one son having earnings, in an attempt to construct balanced SURE models.

The system of three equations is a model like (2.2.1), as shown below:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i , \qquad i = 1,2,3 \qquad (5.1)$$

where, what the symbols in the model stand for are described as follows. The vector $\mathbf{Y}_i$ is a $T \times 1$ vector of observations on the dependent variable representing the natural logarithm of the yearly earnings of the individuals of the *i*th generation. The matrix $\mathbf{X}_i$ is an $T \times k_i$ matrix of presumed non-stochastic explanatory variables representing each of intercept, age, square of age, dummy variables for each of civil status, living in metropoles, living in northern part of Sweden and an ordinal variable indicating the level of educational attainments. The vector $\boldsymbol{\beta}_i$ is a $k_i \times 1$ vector of unknown parameters in the model. The vector $\boldsymbol{\varepsilon}_i$ is an $T \times 1$ vector of random error term with $E(\boldsymbol{\varepsilon}_i) = \mathbf{0}$. The symbol T stands for the number of observations per equation. Since the sample size is 647 then $T = 647$. Finally, k_i is the number of columns (number of explanatory variables with intercept) of $\mathbf{X}_i$, for $i = 1,2,3$. In this example, all the equations have the same number of explanatory variables, i.e., $k_i = 7$, for $i = 1,2,3$.

If we expand upon each equation of the proposed model, we will have the following regression equations.

$$\ln(earnings_i) = \beta_{io} + \beta_{i1}Age_i + \beta_{i2}Age_i^2 + \beta_{i3}CivilStaus_i$$
$$+ \beta_{i4}LiveInMetropoles_i + \beta_{i5}LiveInNorth_i + \beta_{i6}Education_i \,,$$
$$\text{for } i = 1,2,3 \qquad (5.2)$$

We used the bootstrap method, based on 2000 times of resampling of observations (and not of estimated errors), to compute the standard error of the estimated parameters of SUMRE and SURE methods. At each time of resampling, feasible SURE parameters are estimated, using the equations (2.2.8), (2.2.9) and (2.2.10), and the parameters of the SUMRE method are estimated using the equation (2.3.7).

In Table 5.1, we see that the correlation coefficients, which are computed using the formula (3.2.6), are rather low, particularly between the first and the third generation immigrants and between the second and the third generation immigrants.

Table 5.1. Correlation Coefficients of the Cross-Equation Errors

Generations	First Generation	Second Generation	Third Generation
First Generation	1	0.1600	-0.0151
Second Generation		1	0.0660
Third Generation			1

Using the formula (3.2.7), Table 5.2 shows that the distribution of the logarithm of yearly earnings for the first generation immigrants is not very skewed, whereas for the second and the third generation immigrants it is highly skewed to the left.

Table 5.2. Coefficients of Skewness of Errors

Generation	First Generation	Second Generation	Third Generation
Skewness	-0.6464	-2.9863	-2.4189

As it appears from Table 5.3, the differences between SURE and SUMRE estimates are not considerable in the first equation, due to the reason of very low skewness of the errors in the first equation (see Table 5.2), whereas the differences are remarkable in the second and third equations for the opposite reason. A similar argument holds for the standard errors. But which estimates are more efficient?

Table 5.3. SURE & SUMRE Parameter Estimators and Standard Deviations of Estimators

Equation (of)	Variables	Parameter Estimator		Standard Deviation	
		SURE	SUMRE	SURE	SUMRE
First Generation	Age	-0.0379	-0.0922 *	0.0391	0.0445
	Age Squared	0.0004	0.0009 *	0.0004	0.0004
	Civil Status	0.3177 *	0.2420 *	0.0911	0.1071
	Living in Metropoles	0.1155	0.1463 *	0.0602	0.0611
	Living in North	-0.1873 *	-0.1019	0.0624	0.0655
	Education	0.1782 *	0.1808 *	0.0238	0.0241
	Constant	4.8909 *	6.3040 *	0.9063	0.9516
Second Generation	Age	0.0848	0.0472	0.0448	0.0296
	Age Squared	-0.0010	-0.0005	0.0007	0.0004
	Civil Status	0.0750	0.0659 *	0.0421	0.0221
	Living in Metropoles	0.0358	0.0530	0.0435	0.0305
	Living in North	-0.0852	-0.0265	0.0562	0.0292
	Education	0.0276 *	0.0321 *	0.0074	0.0062
	Constant	4.5550 *	5.2396 *	0.7318	0.4899
Third Generation	Age	0.3495 *	0.1819 *	0.0931	0.0516
	Age Squared	-0.0046 *	-0.0024 *	0.0015	0.0008
	Civil Status	0.2505 *	0.0835 *	0.0930	0.0408
	Living in Metropoles	0.2634 *	0.1499 *	0.0806	0.0477
	Living in North	-0.2000	-0.1122 *	0.1106	0.0524
	Education	0.0188	0.0380 *	0.0201	0.0114
	Constant	0.8626	4.0287 *	1.5280	0.8886

(.)* means that the estimate is significant at 5% level.

However, the SUMRE estimates are more efficient at the presence of skewness, due to the fact that their standard errors are lower than the standard errors of their matching SURE estimates. This argument is consistent with the results of the Monte Carlo simulation experiment.

As we previously mentioned, we are only interested in explaining the benefits that we gain from using our SUMRE method, under specific conditions. However, if we delve into the results that we have obtained, and thence look at the parameter estimator of SUMRE method separately or compare them with the parameter estimators of SURE method, we can realize more interesting benefits from using our SUMRE method. Therefore, giving a brief explanation of the results from this empirical exercise might be necessary.

Taking the theory of economics into account, we realize some odd results in the Table 5.3. In a quadratic functional relationship between the natural logarithm of earnings and the age, a negative coefficient for age squared and a positive coefficient for age are obtained, as it is the case with the corresponding parameter estimates of the second and the third generations. The opposite is obtained for the first generation, i.e., the relationship between the natural logarithm of earnings and the age for the first generation is negative.

Figure 5.1 Quadratic fit of ln(earnings) vs. Age

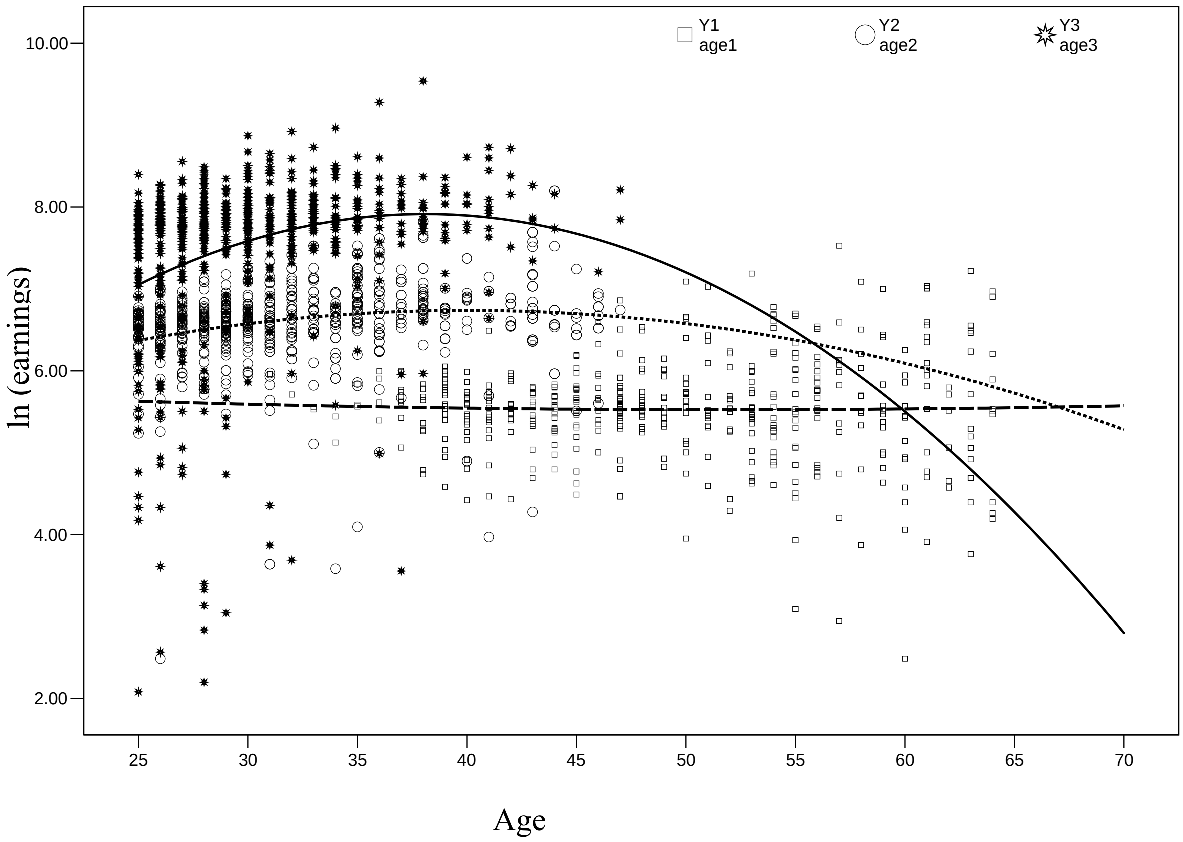

Table 5.4. Summary Statistics of Age and Education

Case Summaries	Generation's Age			Generation's Education		
	First	Second	Third	First	Second	Third
Mean	50.3	32.6	30.2	7.2	10.7	12.4
Median	51	31	29	7	11	12
Minimum	31	25	25	7	7	7
Maximum	64	47	47	12	18	18
Std. Deviation	7.75	5.42	4.52	0.89	2.76	1.89

An obvious reason for that is the age interval taken into the sample for each group of immigrants, as shown in Table 5.4 and Figure 5.1. Individuals of the first generation immigrants are relatively much older (and also less educated) than the individuals of the second and third generations. Consequently, for the first generation immigrants the age was not beneficial to their earnings while the opposite is true for the second and especially the third generation immigrants, since most of the individuals of these two groups taken into the sample are of an age less than 40. Therefore, the quadratic fit lines for each generation will be different, especially for the third generation, which is a sharp quadratic line having a very low intercept (see Table 5.4 and Figure 5.1).

If we look at the SUMRE results for the first generation male immigrants, we find that the variable age has an unexpected negative sign for its estimated coefficient (although non significant) while the age-square has a positive estimate for its coefficient. The variables civil (married) and metro (big cities) have positive significant effects while northern has a negative effect. The number of years of education (school) has also shown to have positive effect (although non significant). Since the skewness is small in this equation, these results are fairly similar to those of the SURE model.

Another fascinating result of applying the SUMRE method in this exercise is the significance of almost all of SUMRE estimators while many of their corresponding SURE estimators are not significant at the 5% level of significance, as indicated in Table 5.3.

On the other hand, since the skewnesses are higher in the equations of the second and the third generations, we find that the results from the SUMRE differ from those of the SURE. This implies that results from the SUMRE are more accurate and representative than those from the SURE. The estimated parameters of the independent variables in these two equations have the expected signs.

Another fact that could be abstracted from the results of Table 5.3 is the relative efficiency of SUMRE estimators in the first and second generation equations, where the data of these two generations are highly skewed. This could be evaluated through the smaller standard error of the estimated parameters. But, on the other hand, for the first generation equation, the SURE estimators are more efficient due to the very low level of skewness of the data in that group. This agrees with the results that we obtained from the Monte Carlo simulation experiment, which indicate that in the presence of skewness of the data, the SUMRE estimators are more efficient than SURE estimators.

6. Summary and Conclusions

In this paper, we generalize the median regression method and make it applicable to systems of regression equations. Given the existence of proper systemwise medians of the errors from different equations, we apply the weighted median regression with the weights obtained from the covariance matrix of errors from different equations calculated by the ordinary SURE method. The SURE method is considered to be one of the most successful and efficient methods for estimating seemingly unrelated regressions with the assumption of symmetric regression errors of each equation. The benefit of SURE models in our case is that the SURE estimators utilise the information present in the correlations of the cross equations errors and hence are more efficient than other estimation methods like the OLS method. The Seemingly Unrelated Median Regression Equations (SUMRE) Models produce results that are more robust than the usual SURE or single equations OLS estimation when the distributions of the dependent variables are not symmetric. Moreover, the results are also more efficient than for the cases of single equations median regressions whose cross equations errors are correlated. More precisely, the aim of our SUMRE method is to produce a harmony of existing skewness and correlations of errors in systems of regression equations. A theorem is derived and indicates that even with the lack of statistically significant correlations between the equations, using SUMRE method instead of SURE method will not damage the estimation of parameters.

A Monte Carlo experiment with 2000 replications has been conducted to investigate the properties of the SUMRE method in situations where the number of equations in the system, number of observations, strength of the correlations of cross equations errors, and the departure from the normality distribution of the errors, have been varied. The results have shown that, when the cross equations correlations are medium or high and the level of skewness of the errors of the equations are also medium or high, the SUMRE methods produces estimators that are more efficient and less biased than the ordinary SURE GLS estimators. Moreover, the results are also more efficient and less biased than in the cases where OLS or single equation median regressions are applied.

Our results from the empirical application are in accordance with what we discovered from the simulation study, with respect to the relative gain in efficiency of SUMRE estimators compared to SURE estimators, in the presence of Skewness of error terms.

References

Aitken, A. C. (1934-35), “On least-Squares and Linear Combination of Observations”, *Proceeding of the Royal Society of Edinburgh*, **55**, 42-8.

Bassett, Jr. & Koenker, R. (1978). “Asymptotic Theory of Least Absolute Error Regression”, *J. Am. Statist. Assoc.*, **73**, pp. 618-622.

Charnes, A., Cooper W.W., & Ferguson, R.O. (1955), “Optimal Estimation of Executive Compensation by Linear Programming”, *Management Science*, Vol **1**, No. 2, pp. 138-151.

Chib, S., and Greenberg, E. (1995). “Hierarchical Analysis of SUR model with Extensions to Correlated Serial Errors and Time Varying Parameter Models”, *Journal of Econometrics*, **68**, 339-360.

Ekberg, J., Hammarstedt, M., and Shukur, G. (2009). ”SUR estimation of earnings differentials between three generations of immigrants and natives”, Submitted for publication in the *Annals of Regional Sciences*.

Hammarstedt, M. (2009). Intergenerational mobility and the earnings position among first-, second- and third-generation immigrants in Sweden. *Kyklos*, **62**, 275-292.

Karst, O.J., (1958), “Linear Curve Fitting Using Least Deviations”, *Journal of the American Statistical Association*, Vol **53**, No. 281, pp. 118-132.

Koenker, R. & Bassett G. (1978), Regression quantiles, *Econometrica*, **46**(1), 33–50.

Otto J. Karst. (1958). “Linear Curve Fitting Using Least Deviations”, *J. Am. Statist. Assoc.*, 53, pp. 118-132.

Srivastava, V., and Giles, D. (1987). “*Seemingly Unrelated Regression Equations Models*”, New York: Marcel Dekker.

Wagner, H. (1959), “Linear Programming Techniques for Regression Analysis”, Journal of the American Statistical Association, Vol **54**, No. 285, pp. 206-212.

Zellner, A. (1962). “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias”, *J. Am. Statist. Assoc.*, **57**, pp. 348-68.

Appendix

Table 1: Determinant Ratio (3 Equations)

Sample Size		Correlation Low			Medium			High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	5.09e+02	2.17e-15	6.54e-67	1.81e+03	2.31e-04	6.50e-31	1.88e+04	2.14e-14	6.50e-58
	Separate Median Regression	8.01e-01	6.60e-01	1.53e-02	2.63e-03	1.88e-06	3.80e-06	2.94e-23	5.51e-35	1.57e-21
250	SURE GLS	4.53e+02	4.91e-21	7.68e-98	1.44e+03	1.02e-07	6.33e-42	1.42e+04	4.70e-19	5.75e-86
	Separate Median Regression	8.77e-01	8.20e-01	2.76e-01	4.15e-03	1.26e-06	2.00e-05	1.63e-24	4.27e-35	5.21e-26
1000	SURE GLS	4.27e+02	4.24e-22	1.63e-106	2.37e+03	2.60e-09	7.89e-45	1.25e+04	8.59e-20	4.68e-91
	Separate Median Regression	8.35e-01	8.94e-01	3.98e-01	5.72e-03	1.83e-06	6.44e-05	2.95e-24	2.70e-36	2.59e-27

Table 2: Determinant Ratio (5 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	5.06e+04	1.03e-22	7.90e-66	3.19e+05	6.94e-10	1.74e-68	1.08e+06	2.44e-15	1.36e-94
	Separate Median Regression	1.45e+00	3.57e-01	4.66e-02	2.87e-10	8.88e-11	7.34e-12	5.09e-34	5.36e-29	2.70e-30
250	SURE GLS	4.95e+04	2.07e-34	7.49e-92	2.59e+05	4.19e-16	5.60e-93	7.23e+05	1.90e-23	4.24e-156
	Separate Median Regression	1.47e+00	4.12e-01	1.94e-01	3.08e-10	5.52e-10	4.35e-10	5.12e-34	2.09e-30	9.08e-40
1000	SURE GLS	4.20e+04	8.07e-39	6.03e-99	2.62e+05	3.57e-19	1.09e-99	7.17e+05	9.40e-26	5.57e-166
	Separate Median Regression	1.26e+00	3.93e-01	2.32e-01	3.43e-10	2.72e-09	2.64e-09	8.68e-34	7.50e-30	1.10e-39

Table 3: Trace Ratio (3 Equations)

Sample Size		Correlation: Low, Skewness: Low	Low / Medium	Low / High	Medium / Low	Medium / Medium	Medium / High	High / Low	High / Medium	High / High
50	SURE GLS	1.54e+00	1.13e-01	5.24e-05	1.60e+00	5.55e-01	1.23e-02	1.62e+00	8.30e-02	1.68e-04
50	Separate Median Regression	1.00e+00	1.01e+00	8.34e-01	6.16e-01	5.78e-01	8.47e-01	2.27e-02	4.46e-03	6.52e-02
250	SURE GLS	1.51e+00	4.82e-02	3.24e-07	1.57e+00	2.84e-01	1.81e-03	1.34e+00	3.84e-02	1.70e-06
250	Separate Median Regression	1.03e+00	1.04e+00	1.05e+00	6.55e-01	7.36e-01	1.19e+00	1.58e-02	3.99e-03	2.60e-02
1000	SURE GLS	1.44e+00	3.98e-02	7.28e-08	1.62e+00	2.19e-01	1.06e-03	1.38e+00	3.34e-02	6.84e-07
1000	Separate Median Regression	1.00e+00	1.04e+00	1.05e+00	6.71e-01	8.68e-01	1.39e+00	1.48e-02	2.54e-03	1.85e-02

Table 4: Trace Ratio (5 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	1.55e+00	1.37e-01	3.26e-03	1.58e+00	4.87e-01	3.26e-03	1.51e+00	2.68e-01	1.88e-04
	Separate Median Regression	1.02e+00	1.04e+00	1.00e+00	3.77e-01	7.06e-01	8.63e-01	3.45e-02	1.03e-01	1.06e-01
250	SURE GLS	1.55e+00	4.30e-02	2.24e-04	1.58e+00	2.52e-01	2.84e-04	1.46e+00	1.45e-01	7.36e-07
	Separate Median Regression	1.01e+00	1.07e+00	1.09e+00	3.88e-01	1.05e+00	1.40e+00	3.27e-02	1.02e-01	5.79e-02
1000	SURE GLS	1.52e+00	2.64e-02	1.02e-04	1.55e+00	1.86e-01	1.38e-04	1.48e+00	1.12e-01	2.78e-07
	Separate Median Regression	1.01e+00	1.07e+00	1.14e+00	3.97e-01	1.21e+00	1.62e+00	3.21e-02	1.15e-01	5.63e-02

Table 5: Squared Bias Ratio (3 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	1.95e-01	2.86e-02	4.68e-05	5.38e-01	9.53e-02	1.24e-02	6.41e+00	1.07e-01	5.07e-05
	Separate Median Regression	4.07e-01	6.18e-01	1.69e+00	1.18e-01	2.26e-01	1.33e+01	3.33e-02	9.28e-02	4.96e-02
250	SURE GLS	2.73e-02	1.02e-02	7.43e-07	1.16e+00	4.88e-02	2.30e-04	1.15e+00	1.10e+00	1.91e-05
	Separate Median Regression	3.20e-01	2.17e-01	1.37e+00	6.71e-01	3.62e-02	5.07e-01	6.48e-02	7.35e-02	1.58e-02
1000	SURE GLS	4.61e-02	2.64e-02	8.19e-08	1.06e+01	2.51e-01	1.02e-04	1.08e+00	1.83e+00	1.39e-04
	Separate Median Regression	5.43e-01	3.24e-01	9.72e-01	2.19e-01	8.49e-02	2.00e-01	4.92e-02	1.07e-01	3.47e-02

Table 6: Squared Bias Ratio (5 Equations)

Sample Size		Correlation Low			Medium			High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	1.33e+00	3.81e-03	1.20e-04	3.16e-01	1.08e-02	4.09e-03	7.01e-01	8.58e-03	6.51e-08
	Separate Median Regression	8.89e-01	1.45e+00	5.47e+00	4.58e-01	4.28e-01	1.20e+01	4.04e-01	7.91e-02	4.31e-03
250	SURE GLS	1.72e+00	2.54e-04	8.79e-06	1.64e-01	1.97e-02	7.83e-06	6.74e-01	8.05e-03	1.32e-08
	Separate Median Regression	1.42e+00	5.48e-01	2.67e+00	2.15e-01	5.26e+00	3.66e+00	1.16e-01	5.02e-02	2.30e-03
1000	SURE GLS	1.06e+00	2.99e-04	7.08e-07	2.56e-01	2.21e-02	1.43e-06	6.06e-01	1.02e-02	5.15e-09
	Separate Median Regression	9.29e-01	1.05e+00	1.92e+00	8.91e-02	6.98e+00	1.67e+00	6.86e-02	1.73e-01	6.24e-04

Table 7: MSE Ratio (3 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	1.36e+00	7.55e-01	1.86e-04	1.61e+00	9.78e-01	3.08e-03	1.61e+00	5.82e-01	7.55e-04
	Separate Median Regression	1.00e+00	1.02e+00	1.10e+00	6.30e-01	6.87e-01	8.68e-01	1.44e-01	2.04e-01	1.27e-01
250	SURE GLS	1.32e+00	5.91e-01	3.02e-06	1.62e+00	8.62e-01	4.89e-04	1.56e+00	5.53e-01	1.29e-04
	Separate Median Regression	1.01e+00	1.04e+00	1.05e+00	6.44e-01	6.74e-01	1.21e+00	1.25e-01	2.10e-01	1.66e-01
1000	SURE GLS	1.21e+00	5.44e-01	1.05e-06	1.54e+00	8.44e-01	3.01e-04	1.55e+00	6.25e-01	1.23e-04
	Separate Median Regression	1.01e+00	1.05e+00	1.08e+00	6.38e-01	5.24e-01	1.14e+00	1.14e-01	1.75e-01	1.44e-01

Table 8: MSE Ratio (5 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	1.55e+00	1.19e-01	2.74e-03	1.51e+00	3.86e-01	3.32e-03	1.35e+00	1.23e-01	3.51e-05
	Separate Median Regression	1.02e+00	1.04e+00	1.01e+00	3.77e-01	7.03e-01	9.35e-01	3.82e-02	1.02e-01	1.03e-01
250	SURE GLS	1.55e+00	3.82e-02	1.91e-04	1.47e+00	1.61e-01	2.01e-04	1.34e+00	8.74e-02	4.52e-07
	Separate Median Regression	1.01e+00	1.07e+00	1.09e+00	3.85e-01	1.09e+00	1.41e+00	3.46e-02	9.81e-02	4.53e-02
1000	SURE GLS	1.52e+00	2.26e-02	8.17e-05	1.44e+00	1.21e-01	8.08e-05	1.33e+00	6.69e-02	2.29e-07
	Separate Median Regression	1.01e+00	1.07e+00	1.14e+00	3.77e-01	1.29e+00	1.62e+00	3.35e-02	1.18e-01	4.14e-02

Table 9: Squared Bias (3 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	2.22e+01	2.66e+03	1.26e+10	7.20e+02	5.70e+03	1.10e+07	6.89e+01	5.73e+04	2.32e+09
	Separate Median Regression	1.06e+01	1.24e+02	3.47e+05	3.28e+03	2.41e+03	1.02e+04	1.33e+04	6.60e+04	2.37e+06
	SUMRE	4.34e+00	7.63e+01	5.89e+05	3.87e+02	5.43e+02	1.36e+05	4.42e+02	6.12e+03	1.18e+05
250	SURE GLS	7.38e+01	4.80e+02	1.99e+09	1.18e+02	1.58e+03	4.16e+06	3.59e+01	3.35e+03	1.75e+08
	Separate Median Regression	6.29e+00	2.26e+01	1.08e+03	2.04e+02	2.13e+03	1.89e+03	6.38e+02	5.00e+04	2.12e+05
	SUMRE	2.01e+00	4.90e+00	1.48e+03	1.37e+02	7.70e+01	9.58e+02	4.13e+01	3.67e+03	3.34e+03
1000	SURE GLS	8.85e+01	2.66e+02	1.83e+09	2.45e+00	6.88e+02	3.65e+06	1.72e+01	2.32e+03	2.67e+07
	Separate Median Regression	7.52e+00	2.17e+01	1.54e+02	1.19e+02	2.04e+03	1.86e+03	3.77e+02	3.95e+04	1.07e+05
	SUMRE	4.08e+00	7.03e+00	1.50e+02	2.61e+01	1.73e+02	3.71e+02	1.85e+01	4.23e+03	3.72e+03

Table 10: Squared Bias (5 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	9.19e-02	5.86e+05	3.92e+08	7.39e+02	2.57e+04	5.60e+07	4.95e+03	3.87e+05	1.20e+12
	Separate Median Regression	1.38e-01	1.54e+03	8.62e+03	5.10e+02	6.47e+02	1.90e+04	8.58e+03	4.19e+04	1.81e+07
	SUMRE	1.23e-01	2.23e+03	4.72e+04	2.34e+02	2.77e+02	2.29e+05	3.47e+03	3.32e+03	7.82e+04
250	SURE GLS	6.75e-03	9.27e+04	8.14e+07	2.35e+02	1.39e+04	6.65e+07	6.99e+02	4.50e+04	7.42e+10
	Separate Median Regression	8.21e-03	4.31e+01	2.68e+02	1.79e+02	5.21e+01	1.42e+02	4.07e+03	7.21e+03	4.26e+05
	SUMRE	1.16e-02	2.36e+01	7.15e+02	3.85e+01	2.74e+02	5.21e+02	4.71e+02	3.62e+02	9.77e+02
1000	SURE GLS	6.12e-03	3.31e+04	3.04e+07	6.27e+01	3.88e+03	3.31e+07	1.92e+02	1.39e+04	7.70e+09
	Separate Median Regression	6.96e-03	9.42e+00	1.12e+01	1.81e+02	1.23e+01	2.84e+01	1.69e+03	8.24e+02	6.36e+04
	SUMRE	6.46e-03	9.90e+00	2.15e+01	1.61e+01	8.59e+01	4.73e+01	1.16e+02	1.42e+02	3.97e+01

Table 11: MSE (3 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	1.76e+04	5.17e+04	1.68e+11	4.86e+05	5.50e+04	5.15e+08	2.09e+05	4.32e+05	5.22e+09
	Separate Median Regression	2.40e+04	3.84e+04	2.84e+07	1.24e+06	7.83e+04	1.83e+06	2.33e+06	1.23e+06	3.12e+07
	SUMRE	2.41e+04	3.90e+04	3.13e+07	7.81e+05	5.38e+04	1.59e+06	3.36e+05	2.52e+05	3.94e+06
250	SURE GLS	3.31e+03	1.01e+04	3.26e+10	9.01e+04	1.09e+04	9.67e+07	3.72e+04	7.89e+04	8.23e+08
	Separate Median Regression	4.34e+03	5.78e+03	9.42e+04	2.27e+05	1.39e+04	3.91e+04	4.64e+05	2.08e+05	6.43e+05
	SUMRE	4.37e+03	5.99e+03	9.86e+04	1.46e+05	9.38e+03	4.73e+04	5.81e+04	4.37e+04	1.06e+05
1000	SURE GLS	8.69e+02	2.62e+03	1.05e+10	2.31e+04	3.00e+03	2.96e+07	8.87e+03	2.20e+04	2.08e+08
	Separate Median Regression	1.05e+03	1.36e+03	1.02e+04	5.57e+04	4.84e+03	7.84e+03	1.20e+05	7.84e+04	1.77e+05
	SUMRE	1.05e+03	1.43e+03	1.10e+04	3.56e+04	2.53e+03	8.91e+03	1.37e+04	1.37e+04	2.55e+04

Table 12: MSE (5 Equations)

Sample Size		Correlation: Low			Correlation: Medium			Correlation: High		
		Skewness			Skewness			Skewness		
		Low	Medium	High	Low	Medium	High	Low	Medium	High
50	SURE GLS	8.72e+01	4.33e+06	2.36e+09	1.50e+04	1.22e+05	8.34e+08	2.47e+04	6.91e+05	1.48e+12
	Separate Median Regression	1.32e+02	4.96e+05	6.39e+06	6.01e+04	6.70e+04	2.96e+06	8.75e+05	8.36e+05	5.04e+08
	SUMRE	1.35e+02	5.16e+05	6.46e+06	2.27e+04	4.71e+04	2.77e+06	3.34e+04	8.50e+04	5.18e+07
250	SURE GLS	1.61e+01	8.23e+05	5.32e+08	2.93e+03	3.55e+04	2.20e+08	4.50e+03	1.07e+05	1.89e+11
	Separate Median Regression	2.47e+01	2.93e+04	9.30e+04	1.12e+04	5.23e+03	3.13e+04	1.74e+05	9.53e+04	1.88e+06
	SUMRE	2.50e+01	3.14e+04	1.02e+05	4.31e+03	5.72e+03	4.42e+04	6.03e+03	9.35e+03	8.54e+04
1000	SURE GLS	3.99e+00	2.26e+05	1.51e+08	7.39e+02	9.81e+03	7.93e+07	1.11e+03	3.13e+04	4.30e+10
	Separate Median Regression	6.02e+00	4.78e+03	1.08e+04	2.82e+03	9.20e+02	3.97e+03	4.41e+04	1.77e+04	2.38e+05
	SUMRE	6.06e+00	5.10e+03	1.23e+04	1.06e+03	1.19e+03	6.41e+03	1.47e+03	2.09e+03	9.86e+03

* * *