

Article

Ridge Type Shrinkage Estimation of Seemingly Unrelated Regressions And Analytics of Economic and Financial Data from “Fragile Five” Countries

Bahadır Yüzbaşı ^{1,*}  and S. Ejaz Ahmed ²¹ Department of Econometrics, Inonu University, 44280 Malatya, Turkey² Department of Mathematics and Statistics, Brock University, St. Catharines, ON L2S 3A1, Canada; sahmed5@brocku.ca

* Correspondence: b.yzb@hotmail.com

Received: 26 April 2020; Accepted: 16 June 2020; Published: 18 June 2020



Abstract: In this paper, we suggest improved estimation strategies based on preliminarily test and shrinkage principles in a seemingly unrelated regression model when explanatory variables are affected by multicollinearity. To that end, we split the vector regression coefficient of each equation into two parts: one includes the coefficient vector for the main effects, and the other is a vector for nuisance effects, which could be close to zero. Therefore, two competing models per equation of the system regression model are obtained: one includes all the regression of coefficients (full model); the other (sub model) includes only the coefficients of the main effects based on the auxiliary information. The preliminarily test estimation improves the estimation procedure if there is evidence that the vector of nuisance parameters does not provide a useful contribution to the model. The shrinkage estimation method shrinks the full model estimator in the direction of the sub-model estimator. We conduct a Monte Carlo simulation study in order to examine the relative performance of the suggested estimation strategies. More importantly, we apply our methodology based on the preliminarily test and the shrinkage estimations to analyse economic data by investigating the relationship between foreign direct investment and several economic variables in the “Fragile Five” countries between 1983 and 2018.

Keywords: shrinkage estimator; seemingly unrelated regression model; multicollinearity; ridge regression

1. Introduction

A seemingly unrelated regression (SUR) system, originally proposed by Zellner (1962), comprises multiple individual regression equations that are correlated with each other. Zellner’s idea was to improve estimation efficiency by combining several equations into a single system. Contrary to SUR estimation, the ordinary least squares (OLS) estimation loses its efficiency and will not produce best linear unbiased estimates (BLUE) when the error terms between the equations in the system are correlated. This method has a wide range of applications in economic and financial data and other similar areas (Shukur 2002; Srivastava and Giles 1987; Zellner 1962). For example, Dincer and Wang (2011) investigated the effects of ethnic diversity on economic growth. Williams (2013) studied the effects of financial crises on banks. Since it considers multiple related equations simultaneously, a generalized least squares (GLS) estimator is used to take into account the effect of errors in these equations. Barari and Kundu (2019) reexamined the role of the Federal Reserve in triggering the recent housing crisis with a vector autoregression (VAR) model, which is a special case of the SUR model with lagged variables and deterministic terms as common regressors. One might also consider the correlations of explanatory variables in SUR models. Alkhamisi and Shukur (2008)

and Zeebari et al. (2012, 2018) considered a modified version of the ridge estimation proposed by Hoerl and Kennard (1970) for these models. Alkhamisi (2010) proposed two SUR-type estimators by combining the SUR ridge regression and the restricted least squares methods. These recent studies demonstrated that the ridge SUR estimation is superior to classical estimation methods in the presence of multicollinearity. Srivastava and Wan (2002) considered the Stein-rule estimators from James and Stein (1961) in SUR models with two equations.

In our study, we consider preliminarily test and shrinkage estimation, more information on which can be found in Ahmed (2014), in ridge-type SUR models when the explanatory variables are affected by multicollinearity. In a previous paper, we combined penalized estimations in an optimal way to define shrinkage estimation (Ahmed and Yüzbaşı 2016). Gao et al. (2017) suggested the use of the weighted ridge regression model for post-selection shrinkage estimation. Yüzbaşı et al. (2020) gave detailed information about generalized ridge regression for a number of shrinkage estimation methods. Srivastava and Wan (2002) and Arashi and Roozbeh (2015) considered Stein-rule estimation for SUR models. Erdugan and Akdeniz (2016) proposed a restricted feasible SUR estimate of the regression coefficients.

The organization of this paper is as follows: In Section 2, we briefly review the SUR model and some estimation techniques, including the ridge type. In Section 3, we introduce our new estimation methodology. A Monte Carlo simulation is conducted in Section 4, and our economic data are analysed in Section 5. Finally, some concluding remarks are given in Section 6.

2. Methodology

Consider the following model:

$$Y_i = X_i\beta_i + \varepsilon_i, i = 1, 2, \dots, M, \tag{1}$$

the i^{th} equation of an M seemingly unrelated regression equation with T number of observations per equation. Y_i is a $T \times 1$ vector of T observations; X_i is a $T \times p_i$ full column rank matrix of T observations on p_i regressors; and β_i is a $p_i \times 1$ vector of unknown parameters.

Equation (1) can be rewritten as follows:

$$Y = X\beta + \varepsilon, \tag{2}$$

where $Y = (Y'_1, Y'_2, \dots, Y'_M)'$ is the vector of responses and $\varepsilon = (\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_M)'$ is the vector of disturbances with dimension $TM \times 1$, $X = \text{diag}(X_1, X_2, \dots, X_M)$ of dimension $TM \times p$, and $\beta = (\beta'_1, \beta'_2, \dots, \beta'_M)'$ of dimension $p \times 1$, for $p = \sum_{i=1}^M p_i$.

The disturbances vector ε satisfies the properties:

$$E(\varepsilon) = \mathbf{0}$$

and:

$$E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma_{11}\mathbf{I} & \dots & \sigma_{1M}\mathbf{I} \\ \vdots & \ddots & \\ \sigma_{M1}\mathbf{I} & \dots & \sigma_{MM}\mathbf{I} \end{bmatrix} = \Sigma \otimes \mathbf{I},$$

where $\Sigma = [\sigma_{ij}], i, j = 1, 2, \dots, M$ is an $M \times M$ positive definite symmetric matrix, \otimes stands for the Kronecker product, and \mathbf{I} is an identity matrix of order of $T \times T$. Following Greene (2019), we assume strict exogeneity of X_i ,

$$E[\varepsilon|X_1, X_2, \dots, X_M] = \mathbf{0},$$

and homoscedasticity:

$$E[\varepsilon_i\varepsilon'_i|X_1, X_2, \dots, X_M] = \sigma_{ii}\mathbf{I}.$$

Therefore, it is assumed that disturbances are uncorrelated across observations, that is,

$$E [\varepsilon_{it}\varepsilon_{js}|\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \sigma_{ij}, \text{ if } t = s \text{ and } 0 \text{ otherwise,}$$

and it is assumed that disturbances are correlated across equations, that is,

$$E [\varepsilon_i\varepsilon_j'|\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \sigma_{ij}\mathbf{I}.$$

The OLS and GLS estimator of model (2) are thus given as:

$$\hat{\beta}^{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

and:

$$\hat{\beta}^{\text{GLS}} = (\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Y}.$$

$\hat{\beta}^{\text{OLS}}$ simply consists of the OLS estimators computed separately from each equation and omits the correlations between equation, as can be seen in [Kuan \(2004\)](#). Hence, it should use the GLS estimator when correlations exist among equations. However, the true covariance matrix $\boldsymbol{\Sigma}$ is generally unknown. The solution for this problem is a feasible generalized least squares (FGLS) estimation, which uses covariance matrix $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ in the estimation of GLS. In many cases, the residual covariance matrix is calculated by:

$$\sigma_{ij} = \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{T - \max(p_i, p_j)}, \quad i, j = 1, \dots, M,$$

where $\hat{\varepsilon}_i = \mathbf{Y}_i - \mathbf{X}_i\hat{\beta}_i$ represents residuals from the i^{th} equation and $\hat{\beta}_i$ may be the OLS or ridge regression (RR) estimation such that $(\mathbf{X}_i'\mathbf{X}_i + \lambda\mathbf{I})^{-1}\mathbf{X}_i'\mathbf{Y}_i$ with the tuning parameter $\lambda \geq 0$. Note that we use the RR solution to estimate $\boldsymbol{\Sigma}$ in our numerical studies because we assume that two or more explanatory variables in each equation are linearly related. Therefore, $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Sigma}} \otimes \mathbf{I}$, the FGLS of the SUR system, is:

$$\hat{\beta}^{\text{FGLS}} = (\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Y}.$$

By following [Srivastava and Giles \(1987\)](#) and [Zeebari et al. \(2012\)](#), we first transform Equation (2) by using the following transformations, in order to retain the information included in the correlation matrix of cross equation errors:

$$\mathbf{Y}_* = (\hat{\boldsymbol{\Sigma}}^{-1/2} \otimes \mathbf{I})\mathbf{Y}, \mathbf{X}_* = (\hat{\boldsymbol{\Sigma}}^{-1/2} \otimes \mathbf{I})\mathbf{X} \text{ and } \boldsymbol{\varepsilon}_* = (\hat{\boldsymbol{\Sigma}}^{-1/2} \otimes \mathbf{I})\boldsymbol{\varepsilon}.$$

Hence, Model (2) turns into:

$$\mathbf{Y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*. \tag{3}$$

The spectral decomposition of the symmetric matrix $\mathbf{X}_*'\mathbf{X}_*$ is $\mathbf{X}_*'\mathbf{X}_* = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ with $\mathbf{P}\mathbf{P}' = \mathbf{I}$. Model (3) can then be written as:

$$\begin{aligned} \mathbf{Y}_* &= \mathbf{X}_*\mathbf{P}\mathbf{P}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_* \\ &= \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}_*, \end{aligned} \tag{4}$$

with $\mathbf{Z} = \mathbf{X}_*\mathbf{P}$, $\boldsymbol{\alpha} = \mathbf{P}'\boldsymbol{\beta}$ and $\mathbf{Z}'\mathbf{Z} = \mathbf{P}'\mathbf{X}_*'\mathbf{X}_*\mathbf{P} = \boldsymbol{\Lambda}$, so that $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues and \mathbf{P} is a matrix whose columns are eigenvectors of $\mathbf{X}_*'\mathbf{X}_*$.

The OLS estimator of model (4) is:

$$\hat{\boldsymbol{\alpha}}^{\text{OLS}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_*.$$

The least squares estimates of β in model (2) can be obtained by an inverse linear transformation as:

$$\hat{\beta}^{OLS} = (\mathbf{P}')^{-1} \hat{\alpha}^{OLS} = \mathbf{P} \hat{\alpha}^{OLS}. \tag{5}$$

Furthermore, by following Alkhamisi and Shukur (2008), the full model ridge SUR regression parameter estimation is:

$$\hat{\alpha}^{RR} = (\mathbf{Z}'\mathbf{Z} + \mathbf{K})^{-1} \mathbf{Z}'\mathbf{Y}_*, \tag{6}$$

where $\mathbf{K} = \text{diag}(\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_M)$, $\mathbf{K}_i = \text{diag}(k_{i1}, k_{i2}, \dots, k_{ip_i})$ and $k_{ij} = \frac{1}{(\hat{\alpha}^{OLS})_{ij}^2} > 0$ for $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, p_i$.

Now let us assume that uncertain non-sample prior information (UNPI) on the vector of β parameters is available, either from previous studies, expert knowledge, or researcher’s experience. This information might be of use for the estimation of parameters, in order to improve the quality of the estimators when the sample data have a low quality or may not be reliable Ahmed (2014). It is assumed that the UNPI on the vector of parameters will be restricted by the equation for Model (2),

$$\mathbf{R}\beta = \mathbf{r}, \tag{7}$$

where $\mathbf{R} = \text{diag}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M)$, $\mathbf{R}_i, i = 1, \dots, M$ is a known $m_i \times p_i$ matrix of rank $m_i < p_i$ and \mathbf{r} is a known $\sum_i^M m_i \times 1$ vector. In order to use restriction (7) in Equation (2), we transform it as follows:

$$\mathbf{RPP}'\beta = \mathbf{H}\alpha = \mathbf{r}, \tag{8}$$

where $\mathbf{H} = \mathbf{RP}$ and $\alpha = \mathbf{P}'\beta$, which is defined above. Hence, the restricted ridge SUR regression estimation is obtained from the following objective function:

$$\begin{aligned} \tilde{\alpha}^{RR} &= \arg \min_{\alpha} \{(\mathbf{Y}_* - \mathbf{Z}\alpha)'(\mathbf{Y}_* - \mathbf{Z}\alpha)\} \text{ w.r.t } \mathbf{H}\alpha = \mathbf{r} \text{ and } \alpha\mathbf{K}\alpha' \leq \tau^2, \\ &= \hat{\alpha}^{RR} - \mathbf{Z}_K^{-1} \mathbf{H}' (\mathbf{H}\mathbf{Z}_K^{-1} \mathbf{H}')^{-1} (\mathbf{H}\hat{\alpha}^{RR} - \mathbf{r}), \end{aligned} \tag{9}$$

where $\mathbf{Z}_K = (\mathbf{Z}'\mathbf{Z} + \mathbf{K})$.

Theorem 1. The risks of $\hat{\alpha}^{RR}$ and $\tilde{\alpha}^{RR}$ are given by:

$$\begin{aligned} R(\hat{\alpha}^{RR}; \alpha) &= \text{tr} \left[(\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} (\mathbf{\Lambda} + \mathbf{K})^{-1} \right] + \alpha' \mathbf{K}' (\mathbf{\Lambda} + \mathbf{K})^{-2} \mathbf{K}\alpha, \\ R(\tilde{\alpha}^{RR}; \alpha) &= \text{tr} \left[(\mathbf{\Lambda} + \mathbf{K})^{-1} \left(\mathbf{\Lambda} - \mathbf{H}' (\mathbf{H}\mathbf{\Lambda}^{-1} \mathbf{H}')^{-1} \mathbf{H} \right) (\mathbf{\Lambda} + \mathbf{K})^{-1} \right] \\ &\quad + \alpha' \mathbf{K}' (\mathbf{\Lambda} + \mathbf{K})^{-2} \mathbf{K}\alpha + \delta' \mathbf{\Lambda} (\mathbf{\Lambda} + \mathbf{K})^{-2} \mathbf{\Lambda}\delta \\ &\quad + 2\delta' \mathbf{\Lambda} (\mathbf{\Lambda} + \mathbf{K})^{-2} \mathbf{K}\alpha, \end{aligned}$$

where $\delta = \mathbf{\Lambda}^{-1} \mathbf{H}' (\mathbf{H}\mathbf{\Lambda}^{-1} \mathbf{H}')^{-1} (\mathbf{H}\alpha - \mathbf{r})$.

Proof. For the risk of the estimators $\hat{\alpha}^{RR}$ and $\tilde{\alpha}^{RR}$, we consider:

$$R(\alpha^*; \alpha) = E \left[(\alpha^* - \alpha)' (\alpha^* - \alpha) \right] = \text{tr} [M(\alpha^*)],$$

where α^* is the one of the estimators $\hat{\alpha}^{RR}$ and $\tilde{\alpha}^{RR}$ and $M(\alpha^*) = E \left[(\alpha^* - \alpha) (\alpha^* - \alpha)'\right]$. Since:

$$\begin{aligned} \hat{\alpha}^{RR} &= (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{Z}'\mathbf{Y}_* \\ &= (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} \hat{\alpha}^{OLS} \\ &= \left[\mathbf{\Lambda}^{-1} (\mathbf{\Lambda} + \mathbf{K}) \right]^{-1} \hat{\alpha}^{OLS} \end{aligned}$$

$$\begin{aligned} &= [\mathbf{I} + \Lambda^{-1}\mathbf{K}]^{-1} \hat{\alpha}^{\text{OLS}} \\ &= \Lambda(\mathbf{K})\hat{\alpha}^{\text{OLS}} \text{ and} \\ \hat{\alpha}^{\text{OLS}} &= \Lambda^{-1}\mathbf{Z}'\mathbf{Y}_* \\ &= \alpha + \Lambda^{-1}\mathbf{Z}'\varepsilon_*, \end{aligned}$$

where $\Lambda = \mathbf{Z}'\mathbf{Z}$.

$$\begin{aligned} E(\hat{\alpha}^{\text{RR}} - \alpha) &= E(\Lambda(\mathbf{K})\hat{\alpha}^{\text{OLS}} - \alpha) \\ &= [\Lambda(\mathbf{K}) - \mathbf{I}]\alpha. \end{aligned}$$

Using $\Lambda(\mathbf{K}) = [\mathbf{I} + \Lambda^{-1}\mathbf{K}]^{-1}$, $k_{ij} \geq 0$, we get:

$$\begin{aligned} \Lambda^{-1}(\mathbf{K}) &= \mathbf{I} + \Lambda^{-1}\mathbf{K} \\ \mathbf{I} &= \Lambda(\mathbf{K}) + \Lambda(\mathbf{K})\Lambda^{-1}\mathbf{K} \\ \Lambda(\mathbf{K}) - \mathbf{I} &= -\Lambda(\mathbf{K})\Lambda^{-1}\mathbf{K} \\ &= -(\Lambda + \mathbf{K})^{-1}\mathbf{K}. \end{aligned}$$

Hence,

$$\begin{aligned} E(\hat{\alpha}^{\text{RR}} - \alpha) &= -(\Lambda + \mathbf{K})^{-1}\mathbf{K}\alpha \\ \text{Var}(\hat{\alpha}^{\text{RR}} - \alpha) &= \text{Var}(\Lambda(\mathbf{K})\hat{\alpha}^{\text{OLS}}) \\ &= (\Lambda + \mathbf{K})^{-1}\Lambda(\Lambda + \mathbf{K})^{-1}. \end{aligned}$$

Therefore, the risk of $\hat{\alpha}^{\text{RR}}$ is directly obtained by definition. Similarly,

$$\begin{aligned} \tilde{\alpha}^{\text{RR}} &= \Lambda(\mathbf{K})\tilde{\alpha}^{\text{OLS}} \\ &= \Lambda(\mathbf{K})\left(\hat{\alpha}^{\text{OLS}} - \Lambda^{-1}\mathbf{H}'(\mathbf{H}\Lambda^{-1}\mathbf{H}')^{-1}(\mathbf{H}\hat{\alpha}^{\text{OLS}} - \mathbf{r})\right) \\ &= \Lambda(\mathbf{K})\hat{\alpha}^{\text{OLS}} - (\Lambda + \mathbf{K})^{-1}\mathbf{H}'(\mathbf{H}\Lambda^{-1}\mathbf{H}')^{-1}(\mathbf{H}\hat{\alpha}^{\text{OLS}} - \mathbf{r}) \\ E(\tilde{\alpha}^{\text{RR}} - \alpha) &= E(\Lambda(\mathbf{K})\hat{\alpha}^{\text{OLS}} - \alpha) - E\left((\Lambda + \mathbf{K})^{-1}\mathbf{H}'(\mathbf{H}\Lambda^{-1}\mathbf{H}')^{-1}(\mathbf{H}\hat{\alpha}^{\text{OLS}} - \mathbf{r})\right) \\ &= -(\Lambda + \mathbf{K})^{-1}\mathbf{K}\alpha - \Lambda(\mathbf{K})\delta, \end{aligned}$$

and,

$$\begin{aligned} \text{Var}(\tilde{\alpha}^{\text{RR}} - \alpha) &= \text{Var}\left(\Lambda(\mathbf{K})\left(\hat{\alpha}^{\text{OLS}} - \Lambda^{-1}\mathbf{H}'(\mathbf{H}\Lambda^{-1}\mathbf{H}')^{-1}(\mathbf{H}\hat{\alpha}^{\text{OLS}} - \mathbf{r})\right)\right) \\ &= \Lambda(\mathbf{K})\left(\Lambda^{-1} - \Lambda^{-1}\mathbf{H}'(\mathbf{H}\Lambda^{-1}\mathbf{H}')^{-1}\mathbf{H}\Lambda^{-1}\mathbf{H}'(\mathbf{H}\Lambda^{-1}\mathbf{H}')^{-1}\mathbf{H}\Lambda^{-1}\right)\Lambda'(\mathbf{K}) \\ &= (\Lambda + \mathbf{K})^{-1}\left(\Lambda - \mathbf{H}'(\mathbf{H}\Lambda^{-1}\mathbf{H}')^{-1}\mathbf{H}\right)(\Lambda + \mathbf{K})^{-1}. \end{aligned}$$

Thus, the risk of $\tilde{\alpha}^{\text{RR}}$ is directly obtained by definition. \square

3. Preliminary Test and Shrinkage Estimation

Researchers have determined that restricted estimation (RE) generally performs better than the full model estimator (FME) and leads to smaller sampling variance than the FME when the UNPI is correct. However, the RE might be a noteworthy competitor of FME even though the restrictions may, in fact, not be valid; we refer to [Groß \(2003\)](#) and [Kaçiranlar et al. \(2011\)](#). It is important that the consequences of incorporating UNPI in the estimation process depends on the usefulness of the

information. The preliminary test estimator (PTE) uses UNPI, as well as the sample information. The PTE chooses between the RE and the FME through a pretest. We consider the SUR-PTE of α as follows:

$$\hat{\alpha}^{\text{PTE}} = \tilde{\alpha}^{\text{RR}} I(F_n < F_{m, M \cdot T - p}(\alpha)) + \hat{\alpha}^{\text{RR}} I(F_n \geq F_{m, M \cdot T - p}(\alpha)), \tag{10}$$

where $F_{m, M \cdot T - p}(\alpha)$ is the upper α -level critical value from the central F -distribution, $I(A)$ stands for the indicator function of the set A , and F_n is the F test for testing the null hypothesis of (8), given by:

$$F_n = \frac{(\mathbf{H}\hat{\alpha}^{\text{OLS}} - \mathbf{r})' (\mathbf{H}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{H}')^{-1} (\mathbf{H}\hat{\alpha}^{\text{OLS}} - \mathbf{r}) / m}{\hat{\varepsilon}_*'\hat{\varepsilon}_* / (M \cdot T - p)}, \tag{11}$$

where m is the number of restrictions and p is the total number of estimated coefficients. Under the null hypothesis (8), F_n is F distributed with n and $M \cdot T - p$ degrees of freedom (Henningsen et al. 2007). The PTE selects strictly between FME and RE and depends strongly on the level of significance. Later, we will define the Stein-type regression estimator (SE) of α . This estimator is the smooth version of PTE, given by,

$$\hat{\alpha}^{\text{SE}} = \hat{\alpha}^{\text{RR}} - d (\hat{\alpha}^{\text{RR}} - \tilde{\alpha}^{\text{RR}}) F_n^{-1}, \tag{12}$$

where $d = (m - 2)(T - p) / m(T - p + 2)$ is the optimum shrinkage constant. It is possible that the SE may have the opposite sign of the FME due to small values of F_n . To alleviate this problem, we consider the positive-rule Stein-type estimator (PSE) defined by:

$$\hat{\alpha}^{\text{PSE}} = \tilde{\alpha}^{\text{RR}} + (1 - dF_n^{-1}) I(F_n > d) (\hat{\alpha}^{\text{RR}} - \tilde{\alpha}^{\text{RR}}). \tag{13}$$

4. Simulation

In this section, the performance of the preliminary test and shrinkage SUR ridge estimators of β are investigated via Monte Carlo simulations. We generate the response from the following model:

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \vdots & \mathbf{X}_M \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{pmatrix}.$$

The explanatory variables are generated from a multivariate normal distribution $MVN_{p_i}(\mathbf{0}, \Sigma_x)$, and the random errors are generated from $MVN_M(\mathbf{0}, \Sigma_\varepsilon)$. We summarize the simulation details as follows:

1. Generate the i^{th} explanatory matrix \mathbf{X}_i from $MVN_{p_i}(\mathbf{0}, \Sigma_x)$ so that \mathbf{X}_i is different from \mathbf{X}_j for all $i, j = 1, 2, \dots, M$, where $diag(\Sigma_x) = 1$ and $off - diag(\Sigma_x) = \rho_x$. The ρ_x regulates the strength of collinearity among explanatory variables per equation. In this study, we consider $\rho_x = 0.5, 0.9$. Further, the response is centred, and the predictors are standardized for each equation.
2. The variance-covariance matrix of errors for interdependency among equations is defined by $diag(\Sigma_\varepsilon) = 1$ and $off - diag(\Sigma_\varepsilon) = \rho_\varepsilon = 0.5, 0.9$, and the errors are generated from $MVN_M(\mathbf{0}, \Sigma_\varepsilon)$ and $M = 2, 3$ for each replication.
3. We consider that an SUR regression model is assumed to be sparse. Hence, the vector of coefficients can be partitioned as $\beta' = (\beta'_1, \beta'_2)'$ where β_1 is the the coefficient vector of the main effects, while β_2 is the vector of nuisance, which means that it does not contribute to the model significantly. We set $\beta_1 = (1, -3, 2)'$ and $\beta_2 = \mathbf{0}$. For the suggested estimations, we consider the restriction of $\beta_2 = \mathbf{0}$ and test it. We also investigate the behaviours of the estimators when the restriction is not true. To this end, we add a Δ value to one component of β_2 so that it violates the null hypothesis.

Here, we use Δ values between zero and two and use $\alpha = 0.05$. We also consider that the lengths of the nuisance parameter β_2 are two and four, respectively. Therefore, the restricted matrices are:

$$\mathbf{R}_i = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{r}_i = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ if } p_i = 5 \text{ and}$$

$$\mathbf{R}_i = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{r}_i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \text{ if } p_i = 7,$$

where $i = 2, 3$ are the number of equation. Hence, \mathbf{R} will be $diag(\mathbf{R}_1, \mathbf{R}_2)$ and $diag(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3)$. Furthermore, \mathbf{r} will be $(\mathbf{r}'_1, \mathbf{r}'_2)'$ and $(\mathbf{r}'_1, \mathbf{r}'_2, \mathbf{r}'_3)'$.

- The performance of an estimator is evaluated by using the relative mean squared error (RMSE) criterion. The RMSE of an estimator $\hat{\alpha}^*$ with respect to $\hat{\alpha}^{RR}$ is defined as follows:

$$RMSE(\hat{\alpha}^*) = \frac{MSE(\hat{\alpha}^{RR})}{MSE(\hat{\alpha}^*)},$$

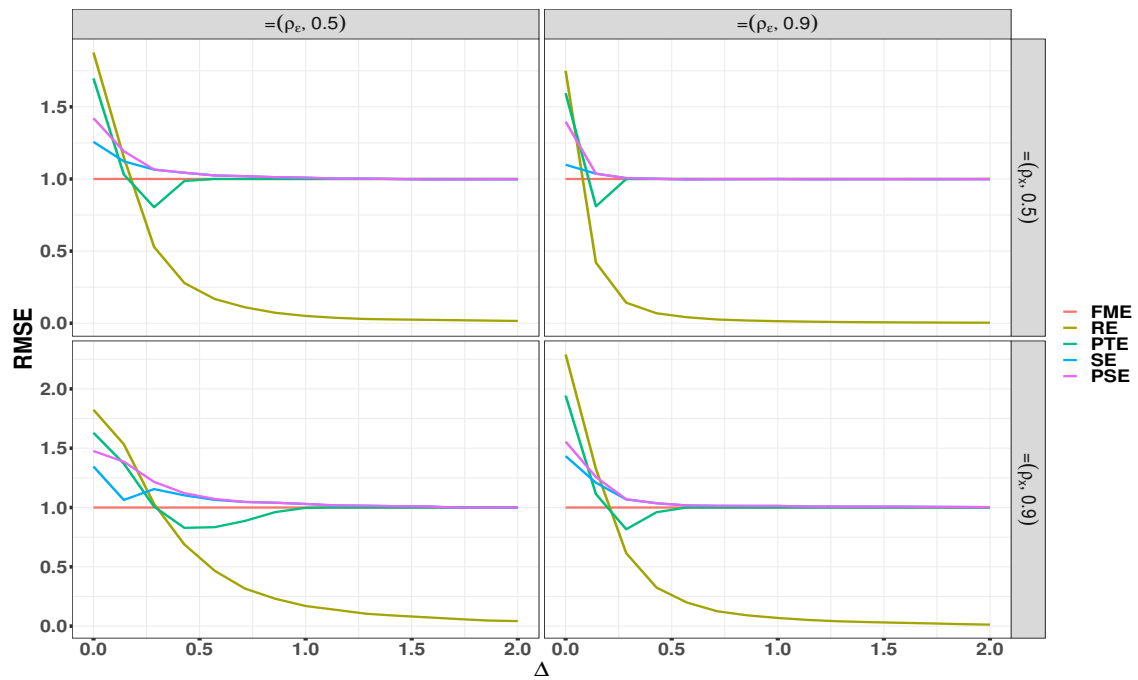
where $\hat{\alpha}^*$ is one of the listed estimators. If the RMSE of an estimator is larger than one, it indicates that it is superior to $\hat{\alpha}^{RR}$.

Table 1 provides notations and a symbol key for the benefit of the reader.

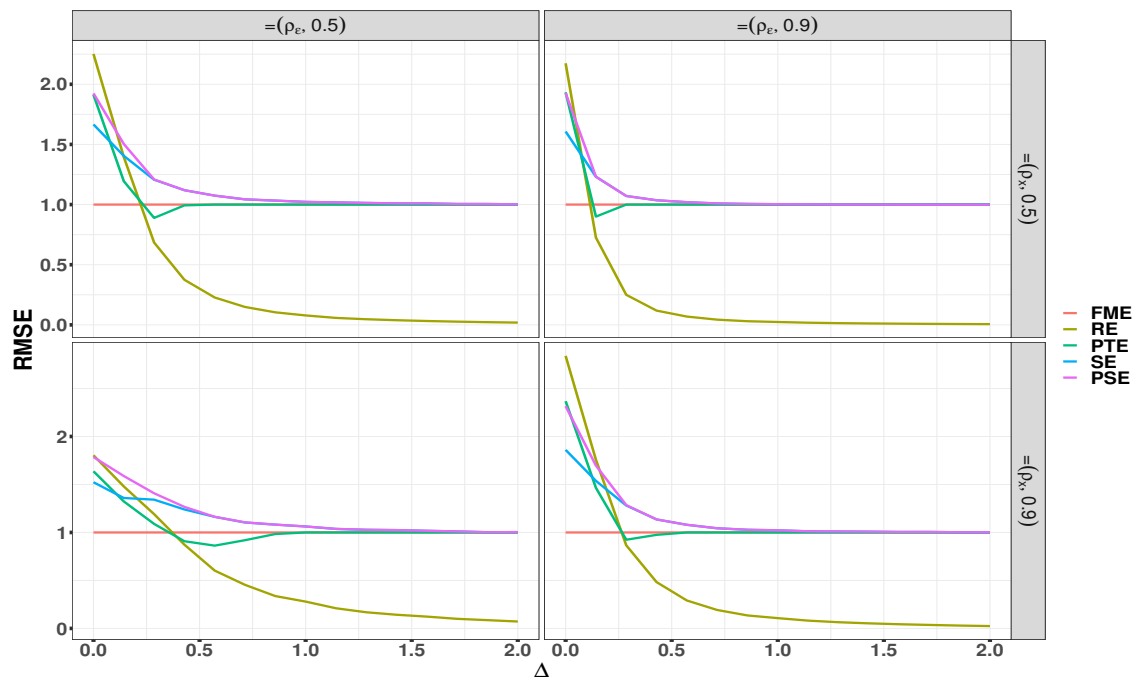
Table 1. Values and explanations of the symbols.

Symbol	Description	Design
M	the number of equations	2,3
T	the number of observations per equation	100
p_i	the number of covariates per equation	5,7
Σ_ϵ	the variance–covariance matrix of errors	$diag(\Sigma_\epsilon) = 1$ and off – $diag(\Sigma_\epsilon) = \rho_\epsilon$
ρ_ϵ	off-diagonal elements of Σ_ϵ	0.5,0.9
Σ_x	the variance–covariance matrix of covariates	$diag(\Sigma_x) = 1$ and off – $diag(\Sigma_x) = \rho_x$
ρ_x	off-diagonal elements of Σ_x	0.5,0.9
Δ	the magnitude of violation of the null hypothesis	[0, 2]

We plot the simulation results in Figures 1 and 2. The simulation results for some other parameter configurations were also obtained, but are not included here for the sake of brevity.

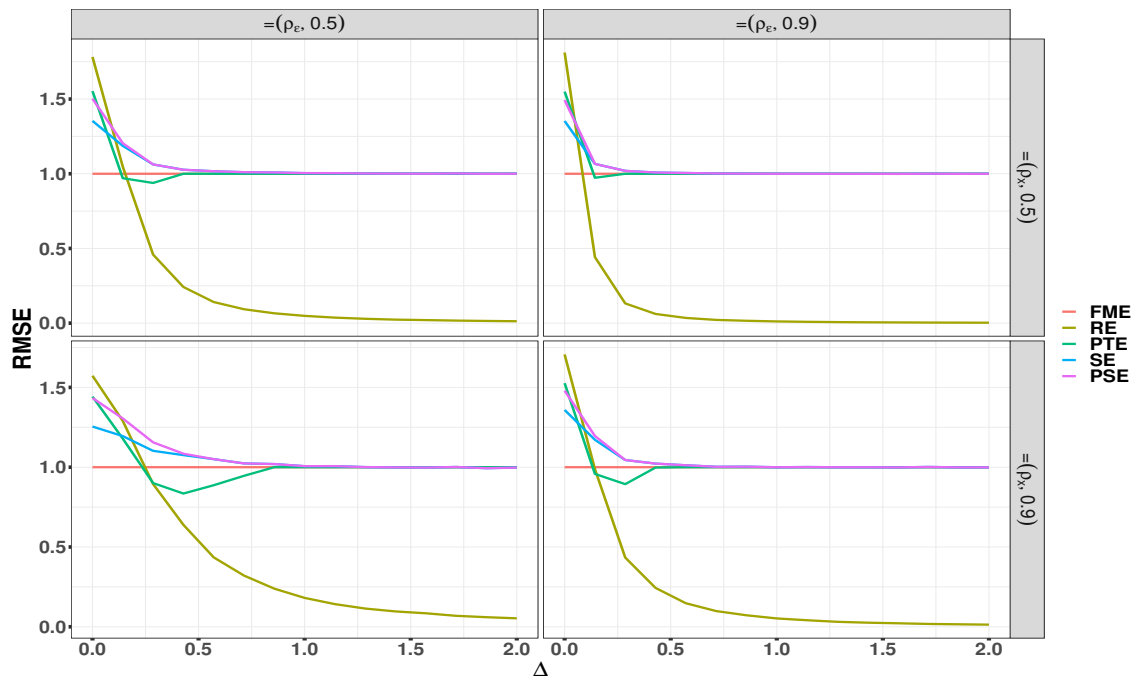


(a) As $p_i = 5, i = 1, 2$

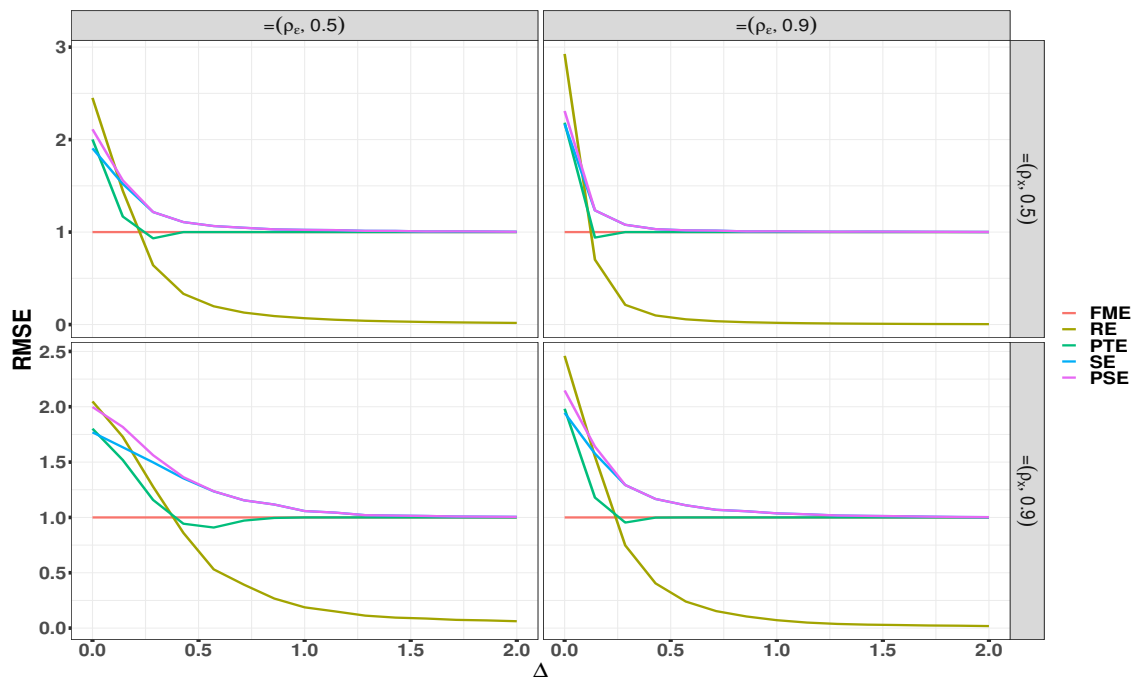


(b) As $p_i = 7, i = 1, 2$

Figure 1. RMSE of the estimators as a function of Δ when $M = 2, T = 100, \rho_x = 0.5, 0.9,$ and $\rho_\epsilon = 0.5, 0.9.$ FME, full model estimator; RE, restricted estimation; PTE, preliminary test estimator; PSE, positive-rule Stein-type estimator.



(a) As $p_i = 5, i = 1, 2$



(b) As $p_i = 7, i = 1, 2$

Figure 2. RMSE of the estimators as a function of Δ when $M = 3, T = 100, \rho_x = 0.5, 0.9,$ and $\rho_\epsilon = 0.5, 0.9.$

According to these results:

1. When $\Delta = 0$, which means that the null hypothesis is true or that the restrictions are consistent, the RE estimator always performs competitively when compared to other estimators. The PTE mostly outperforms the SE and PSE when $p_i = 5$, while it loses its efficiency when $p_i = 7$ when compared to PSE. The SE may perform worse than the FME due its sign problem, as is indicated in Section 3.

2. When $\Delta > 0$, which means that the null hypothesis is violated or the restrictions are invalid, the RE loses its efficiency, and its RMSE goes to zero, meaning that it becomes inconsistent. The RMSE of PTE decreases and remains below one for some values of Δ , but approaches one for larger values of Δ . The performance of PSE decreases, but its efficiency remains above the FME, for intermediate values of Δ , while it acts as the FME for larger values of Δ . It can be concluded that the PSE is a robust estimator even if the restriction is not true.
3. We examined both medium and high correlation between disturbance terms. The results showed that the performance of suggested estimators was consistent with its theory; see [Ahmed \(2014\)](#).
4. We examined both medium and high correlation between regressors across different equations. The results showed that the performance of suggested estimators was consistent with its theory; see [Yüzbaşı et al. \(2017\)](#).

5. Application

In the following section, we will apply the proposed estimation strategies to a financial dataset to examine the relative performance of the listed estimators. To illustrate and compare the listed estimators, we will study the effect of several economic and financial variables on the performance of the “Fragile Five” countries (coined by [Stanley 2013](#)) in terms of their attraction of direct foreign investment (FDI) over the period between 1983 and 2018. The “Fragile Five” include Turkey (TUR), South Africa (ZAF), Brazil (BRA), India (IND), and Indonesia (IDN). [Agiomirgianakis et al. \(2003\)](#), [Hubert et al. \(2017\)](#), and [Akn \(2019\)](#) used the FDI as the dependent variable across countries. With five countries, we have $M = 5$ blocks in our SUR model, with measurements of $T = 36$ years per equation. Table 2 provides information about prediction variables, and the raw data are available from the World Bank¹.

Table 2. Descriptions of variables.

Variables	Descriptions
Dependent Variables	
FDI	Foreign direct investment, net inflows (% of GDP)
Covariates	
GROWTH	GDP per capita growth (annual %)
DEFLATOR	Inflation, GDP deflator (annual %)
EXPORTS	Exports of goods and services (% of GDP)
IMPORTS	Imports of goods and services (% of GDP)
GGFCE	General government final consumption expenditure (% of GDP)
RESERVES	Total reserves (includes gold, current US\$)/GDP (current US\$)
PREM	Personal remittances, received (% of GDP)
BALANCA	Current account balance (% of GDP)

We suggest the following model:

$$FDI_{it} = \beta_{0i} + \beta_{1i}GROWTH + \beta_{2i}DEFLATOR + \beta_{3i}EXPORTS + \beta_{4i}IMPORTS + \beta_{5i}GGFCE + \beta_{6i}RESERVES + \beta_{7i}PREM + \beta_{8i}BALANCE + \epsilon_{it}, \tag{14}$$

where i denotes countries ($i = TUR, ZAF, BRA, IND, IDN$) and t is time ($t = 1, 2, \dots, T$). Following [Salman \(2011\)](#), the errors of each equation are assumed to be normally distributed with mean zero, homoscedastic, and serially not autocorrelated. Furthermore, there is contemporaneous correlation between corresponding errors in different equations. We test these assumptions along with the assumptions in Section 2. We first check the following assumptions of each equation:

¹ <https://data.worldbank.org>.

Nonautocorrelation of errors: There are a number of viable tests in the reviewed literature for testing the autocorrelation. For example, the Ljung–Box test is widely used in applications of time series analysis, and a similar assessment may be obtained via the Breusch–Godfrey test and the Durbin–Watson test. We apply the Ljung–Box test of (Ljung and Box 1978). The null hypothesis of the Ljung–Box Test, H_0 , is that the errors are random and independent. A significant p -value in this test rejects the null hypothesis that the time series is not autocorrelated. Results reported in Table 3 suggest a rejection of H_0 for the equations of both TUR and IND at any conventional significance level. Thus, the estimation results will be clearly unsatisfactory for these two equation models. To tackle this problem, we performed the first differences procedure to transform the variables. After transformation, the test statistics and p -values of the equation TUR and IND were $\chi^2_{(1)} = 1.379$, $p = 0.240$ and $\chi^2_{(1)} = 0.067$, $p = 0.794$, respectively. Hence, each equation satisfied the assumption of nonautocorrelation. We confirmed our result using the Durbin–Watson test.

Table 3. Ljung-Box test.

Equation	Test Statistic	p -Value
TUR	$\chi^2_{(1)} = 6.853$	0.008
ZAF	$\chi^2_{(1)} = 0.704$	0.401
BRA	$\chi^2_{(1)} = 0.489$	0.483
IND	$\chi^2_{(1)} = 6.301$	0.012
IDN	$\chi^2_{(1)} = 1.061$	0.302

Homoscedasticity of errors: To test for heteroscedasticity, we used the Breusch–Pagan test (Breusch and Pagan 1979). The results in Table 4 failed to reject the null hypothesis in each equation.

Table 4. Breusch–Pagan test.

Equation	Test Statistic	p -Value
TUR	$\chi^2_{(8)} = 3.686$	0.884
ZAF	$\chi^2_{(8)} = 10.003$	0.264
BRA	$\chi^2_{(8)} = 7.544$	0.479
IND	$\chi^2_{(8)} = 6.455$	0.596
IDN	$\chi^2_{(8)} = 8.328$	0.402

The assumption homoscedasticity in each equation was thus met.

Normality of errors: To test for normality, there are various tests such as Shapiro–Wilk, Anderson–Darling, Cramer–von Mises, Kolmogorov–Smirnov, and Jarque–Bera. In this study, we performed the Jarque–Bera goodness-of-fit test (Jarque and Bera 1980).

Table 5. Jarque–Bera test.

Equation	Test Statistic	p -Value
TUR	$\chi^2_{(2)} = 3.969$	0.137
ZAF	$\chi^2_{(2)} = 72.852$	0.000
BRA	$\chi^2_{(2)} = 2.355$	0.308
IND	$\chi^2_{(2)} = 1.815$	0.403
IDN	$\chi^2_{(2)} = 2.794$	0.247

The null hypothesis for the test is that the data are normally distributed. The results reported in Table 5 suggested a rejection of H_0 only for ZAF. We also performed the Kolmogorov–Smirnov test for

ZAF, and the results showed that the errors were normally distributed. Thus, each equation satisfied the assumption of normality.

Cross-sectional dependence: To test whether the estimated correlation between the sections was statistically significant, we applied the [Breusch and Pagan \(1980\)](#) Lagrange multiplier (LM) statistic and the [Pesaran \(2004\)](#) cross-section dependence (CD) tests. The null hypothesis of these tests claims there is no cross-section dependence. Both tests in [Table 6](#) suggested a rejection of the null hypothesis that the residuals from each equation were significantly correlated with each other. Consequently, the SUR model would be the preferred technique, since this model assumed contemporaneous correlation across equations. Therefore, the joint estimation of all parameters rather than OLS, on each equation, was more efficient ([Kleiber and Zeileis 2008](#)).

Table 6. Cross-section dependence test results. LM, Lagrange multiplier; CD, cross-section dependence.

Correlation Matrix of Residuals				
	TUR	ZAF	BRA	IND
ZAF	−0.207			
BRA	0.066	−0.187		
IND	0.414	−0.107	−0.016	
IDN	0.128	−0.334	−0.064	0.235

Breusch and Pagan LM and Pesaran CD tests		
Test	Test Statistic	p-Value
LM	$\chi^2_{(2)} = 29.516$	0.001
CD	$Z = -4.353$	0.000

Specification test: The regression equation specification error test (RESET) designed by [Ramsey \(1969\)](#) is a general specification test for the linear regression model. It tests the exogeneity of the independent variables, that is the null hypothesis is $E[\varepsilon_i|X_i] = 0$. Thus, rejecting the null hypothesis indicates that there is a correlation between the error term and the regressors or that nonlinearities exist in the functional form of the regression. The results reported in [Table 7](#) suggested a rejection of H_0 only for IDN.

Table 7. The regression equation specification error test (RESET) test.

Equation	Test Statistic	p-Value
TUR	$F(8, 18) = 0.458$	0.869
ZAF	$F(8, 19) = 1.185$	0.357
BRA	$F(8, 19) = 1.062$	0.428
IND	$F(8, 18) = 1.648$	0.180
IDN	$F(8, 19) = 7.788$	0.000

Multicollinearity: We calculated the variance inflation factor (VIF) values among the predictors. A VIF value provides the user with a measure of how many times larger the $\text{Var}(\beta_j)$ will be for multicollinear data than for orthogonal data. Usually, the multicollinearity is not a problem, as the VIFs are generally not significantly larger than one ([Mansfield and Helms 1982](#)). In the literature, values of VIF that exceed 10 are often regarded as indicating multicollinearity, but in weaker models, values above 2.5 may be a cause for concern. Another measure of multicollinearity is to calculate the condition number (CN) of $X_i'X_i$, which is the square root of the ratio of the largest characteristic root of $X_i'X_i$ to the smallest. [Belsley et al. \(2005\)](#) suggested that a CN greater than fifteen poses a concern, a CN in excess of 20 is indicative of a problem, and a CN close to 30 represents a severe problem. [Table 8](#) displays the results from a series of multicollinearity diagnostics. In general, EXPORTS, IMPORTS, and BALANCE were found to be problematic with regard to VIF values, while the others may be a little concerning. On the other hand, the results from the CN test suggested that there was a very serious

concern about multicollinearity for the equations of ZAF, BRA, and IDN. In light of these results, it was clear that the problem of multicollinearity existed in the equations. According to Greene (2019), the SUR estimation is more efficient when the less correlation exists between covariates. Therefore, the ridge-type SUR estimation will be a good solution of this problem.

Table 8. Variance inflation factor (VIF) and CN values.

Equation	GROWTH	DEFLATOR	EXPORTS	IMPORTS	GGFCE	RESERVES	PREM	BALANCE	CN
TUR	4.039	2.638	16.289	13.959	2.055	1.791	1.462	19.073	11.122
ZAF	3.070	7.354	69.891	191.248	3.891	12.918	5.847	60.728	248.221
BRA	1.204	1.614	18.324	32.159	6.329	3.131	3.938	13.301	85.336
IND	1.745	1.757	6.545	6.653	1.517	1.527	1.378	2.712	6.535
IDN	7.835	7.786	44.842	34.152	5.564	8.274	3.072	15.022	127.099

Structural change: To investigate the stability of the coefficients in each equation, we used the CUSUM (cumulative sum) test of Brown et al. (1975) that checks for structural changes. The null hypothesis is that of coefficient constancy, while the alternative suggests inconsistent structural change in the model over time. The results in Table 9 suggested the stability of coefficients over time.

Table 9. CUSUM test.

Equation	Test Statistic	p-Value
TUR	$T = 0.734$	0.653
ZAF	$T = 0.417$	0.995
BRA	$T = 0.496$	0.966
IND	$T = 0.413$	0.995
IDN	$T = 0.401$	0.997

Following Lawal et al. (2019), we selected important variables in each equation of the SUR model and implemented the stepwise AIC forward regression by using the function `ols_step_forward_aic` from the `olsrr` package in the R project. The statistically significant variables are shown in Table 10. After that, the sub-models were constituted by using these variables per equation.

Table 10. Important variables per equation.

Equation	GROWTH	DEFLATOR	EXPORTS	IMPORTS	GGFCE	RESERVES	PREM	BALANCE
TUR			✓					
ZAF							✓	
BRA		✓		✓				✓
IND				✓			✓	✓
IDN			✓	✓			✓	

In light of the selected variables in Table 10, we construct the matrices of restrictions as follows:

$$\mathbf{R}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{R}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{r}_1 = \mathbf{r}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$\begin{aligned}
 \mathbf{R}_3 &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \mathbf{R}_4 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\
 \mathbf{R}_5 &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{r}_3 = \mathbf{r}_4 = \mathbf{r}_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix};
 \end{aligned}$$

thus, the reduced models are given by:

$$\text{TUR} : \text{FDI}_t = \beta_0 + \beta_3 \text{EXPORTS} + \epsilon_t, \tag{15}$$

$$\text{ZAF} : \text{FDI}_t = \beta_0 + \beta_7 \text{PREM} + \epsilon_t, \tag{16}$$

$$\text{BRA} : \text{FDI}_t = \beta_0 + \beta_2 \text{DEFLATOR} + \beta_4 \text{IMPORTS} + \beta_8 \text{BALANCE} + \epsilon_t, \tag{17}$$

$$\text{IND} : \text{FDI}_t = \beta_0 + \beta_4 \text{IMPORTS} + \beta_7 \text{PREM} + \beta_8 \text{BALANCE} + \epsilon_t, \tag{18}$$

$$\text{IDN} : \text{FDI}_t = \beta_0 + \beta_3 \text{EXPORTS} + \beta_4 \text{IMPORTS} + \beta_7 \text{PREM} + \epsilon_t, \tag{19}$$

Next, we combined Model (14) and Models (15)–(19) using the shrinkage and preliminarily test strategies outlined in Section 3. Before we performed our analysis, the response was centred, and the predictors were standardized for each equation so that the intercept term was omitted. We then split the data by using the time series cross-validation technique of Hyndman and Athanasopoulos (2018) into a series of training sets and a series of testing sets. Each test set consisted of a single observation for the models that produced one step-ahead forecasts. In this procedure, the observations in the corresponding training sets occurred prior to the observation of the test sets. Hence, it was ensured that no future observations could be used in constructing the forecast. We used the function `createTimeSlices` from the `caret` package in the R project here. The listed models were applied to the data, and predictions were made based on the divided training and test sets. The process was repeated 15 times, and for each subset’s prediction, the mean squared error (MSE) and the mean absolute error (MAE) were calculated. The means of the 15 MSEs and MAEs were then used to evaluate the performance for each method. We also report the relative performances (RMAE and RMSE) with respect to the full model estimator for easier comparison. If a relative value of an estimator is larger than one, it is superior to the full model estimator.

In Table 11, we report the MSE and MAE values and their standard errors to see the stability of the algorithm. Based on this table, as expected, the RE had the smallest measurement values since the insignificant variables were selected as close to correct as possible. We saw that the performance of the PSE after the RE was best by following the SE and the PTE. Moreover, the performance of the OLS was the worst due to the problem of multicollinearity.

Table 11. Comparison of forecasting performance.

	RE	FME	PTE	SE	PSE	OLS
MAE	0.572 (0.114)	1.076 (0.148)	0.656 (0.139)	0.649 (0.124)	0.646 (0.122)	1.166 (0.165)
RMAE	1.879	1	1.639	1.656	1.664	0.922
MSE	0.598 (0.061)	0.800 (0.062)	0.624 (0.067)	0.624 (0.069)	0.622 (0.068)	0.831 (0.064)
RMSE	1.338	1	1.283	1.282	1.287	0.963

The numbers in parenthesis are the corresponding standard errors of the MAE and MSE.

In order to test whether the two competing models had the same forecasting accuracy, we used the two-sided statistical Diebold–Mariano (DM) test (Diebold and Mariano 1995) when the forecasting horizon was extended to one year, and the loss functions were both squared errors and absolute errors. A significant *p*-value in this test rejected the null hypothesis that the models had different forecasting accuracy. The results based on the absolute-error loss in Table 12 suggested that the FME had different prediction accuracy with all methods except RE. Additionally, the forecasting accuracy of the OLS differed from the listed estimators. On the other hand, the results of the DM test based on the squared error loss suggested that the observed differences between the RE and shrinkage estimators were significant.

Table 12. Diebold–Mariano test for the forecasting results.

LF		FME	RE	PTE	SE	PSE
MAE	RE	−1.308 (0.191)				
	PTE	−2.601 (0.009 ***)	−0.608 (0.543)			
	SE	−2.146 (0.032 **)	−0.733 (0.463)	0.276 (0.783)		
	PSE	−2.163 (0.031 **)	−0.702 (0.483)	0.33 (0.741)	0.551 (0.582)	
	OLS	3.734 (0.000 ***)	1.700 (0.089 *)	2.972 (0.003 ***)	2.543 (0.011 **)	2.56 (0.010 **)
MSE	SM	−0.187 (0.852)				
	PTE	−1.968 (0.049 **)	−2.165 (0.030 **)			
	SE	−1.444 (0.149)	−2.392 (0.017 **)	1.443 (0.149)		
	PSE	−1.474 (0.140)	−2.374 (0.018 **)	1.436 (0.151)	−1.496 (0.135)	
	OLS	3.528 (0.000 ***)	0.691 (0.490)	2.379 (0.017 **)	1.904 (0.057 *)	1.933 (0.053 *)

The numbers in parenthesis are the corresponding *p*-values; LS is the “loss function” of the method to compute; * *p* < 0.1, ** *p* < 0.05, *** *p* < 0.01.

Finally, the estimates of coefficients of all countries are given in Table 13.

Table 13. Estimated coefficients.

Estimation	Country	GROWTH	DEFLATOR	EXPORTS	IMPORTS	GGFCE	RESERVES	PREM	BALANCE
RE	TUR	0 (0)	0 (0)	0.102 (0.003)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	ZAF	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0.509 (0.007)	0 (0)
	BRA	0 (0)	−0.231 (0.005)	0 (0)	0.912 (0.005)	0 (0)	0 (0)	0 (0)	−0.377 (0.006)
	IND	−0.113 (0.002)	0 (0)	0 (0)	0.122 (0.003)	0 (0)	0.052 (0.002)	0 (0)	0 (0)
	IDN	0 (0)	0 (0)	−1.903 (0.015)	1.341 (0.014)	0 (0)	0 (0)	0.427 (0.004)	0 (0)
FME	TUR	−0.369 (0)	−0.064 (0)	0.327 (0.003)	−0.187 (0)	−0.140 (0)	0.036 (0)	0.061 (0)	−0.543 (0)
	ZAF	−0.120 (0)	−0.397 (0)	0.521 (0)	−0.139 (0)	0.012 (0)	−0.810 (0)	0.551 (0.007)	−0.420 (0)
	BRA	0.059 (0)	−0.18 (0.005)	0.999 (0)	−0.265 (0.005)	0.859 (0)	0.004 (0)	−0.210 (0)	−1.042 (0.006)
	IND	−0.071 (0.002)	0.018 (0)	−0.091 (0)	0.253 (0.003)	−0.002 (0)	−0.024 (0.002)	0.149 (0)	0.150 (0)
	IDN	0.280 (0)	0.177 (0)	−2.348 (0.015)	1.689 (0.014)	−0.009 (0)	−0.233 (0)	0.475 (0.004)	0.242 (0)
PTE	TUR	−0.151 (0)	−0.026 (0)	0.188 (0.003)	−0.072 (0)	−0.058 (0)	0.014 (0)	0.026 (0)	−0.219 (0)
	ZAF	−0.048 (0)	−0.140 (0)	0.184 (0)	−0.014 (0)	0.012 (0)	−0.318 (0)	0.514 (0.007)	−0.160 (0)
	BRA	0.024 (0)	−0.215 (0.005)	0.420 (0)	0.404 (0.005)	0.350 (0)	0.011 (0)	−0.073 (0)	−0.664 (0.006)
	IND	−0.094 (0.002)	0.006 (0)	−0.032 (0)	0.174 (0.003)	0.003 (0)	0.021 (0.002)	0.059 (0)	0.062 (0)
	IDN	0.126 (0)	0.064 (0)	−2.072 (0.015)	1.475 (0.014)	−0.002 (0)	−0.087 (0)	0.445 (0.004)	0.094 (0)
SE	TUR	−0.110 (0)	−0.019 (0)	0.166 (0.003)	−0.053 (0)	−0.042 (0)	0.010 (0)	0.018 (0)	−0.160 (0)
	ZAF	−0.036 (0)	−0.108 (0)	0.142 (0)	−0.022 (0)	0.006 (0)	−0.237 (0)	0.517 (0.007)	−0.119 (0)
	BRA	0.017 (0)	−0.218 (0.005)	0.299 (0)	0.554 (0.005)	0.253 (0)	0.005 (0)	−0.057 (0)	−0.579 (0.006)
	IND	−0.100 (0.002)	0.005 (0)	−0.025 (0)	0.160 (0.003)	0.001 (0)	0.029 (0.002)	0.043 (0)	0.045 (0)
	IDN	0.088 (0)	0.048 (0)	−2.031 (0.015)	1.442 (0.014)	−0.002 (0)	−0.065 (0)	0.440 (0.004)	0.070 (0)
PSE	TUR	−0.112 (0)	−0.019 (0)	0.168 (0.003)	−0.055 (0)	−0.043 (0)	0.010 (0)	0.019 (0)	−0.164 (0)
	ZAF	−0.037 (0)	−0.112 (0)	0.146 (0)	−0.023 (0)	0.006 (0)	−0.243 (0)	0.517 (0.007)	−0.122 (0)
	BRA	0.018 (0)	−0.217 (0.005)	0.307 (0)	0.546 (0.005)	0.260 (0)	0.005 (0)	−0.059 (0)	−0.584 (0.006)
	IND	−0.099 (0.002)	0.005 (0)	−0.026 (0)	0.161 (0.003)	0.001 (0)	0.029 (0.002)	0.044 (0)	0.046 (0)
	IDN	0.090 (0)	0.050 (0)	−2.033 (0.015)	1.444 (0.014)	−0.002 (0)	−0.067 (0)	0.441 (0.004)	0.071 (0)
OLS	TUR	−0.400 (0)	−0.071 (0)	0.379 (0.003)	−0.230 (0)	−0.147 (0)	0.042 (0)	0.066 (0)	−0.618 (0)
	ZAF	−0.142 (0)	−0.443 (0)	0.648 (0)	−0.291 (0)	0.007 (0)	−0.916 (0)	0.613 (0.007)	−0.515 (0)
	BRA	0.067 (0)	−0.171 (0.005)	1.078 (0)	−0.327 (0.005)	0.909 (0)	−0.018 (0)	−0.225 (0)	−1.110 (0.006)
	IND	−0.074 (0.002)	0.020 (0)	−0.105 (0)	0.275 (0.003)	−0.004 (0)	−0.029 (0.002)	0.156 (0)	0.163 (0)
	IDN	0.270 (0)	0.179 (0)	−2.468 (0.015)	1.768 (0.014)	−0.016 (0)	−0.246 (0)	0.478 (0.004)	0.281 (0)

The numbers in parenthesis are the corresponding standard errors.

6. Conclusions

In this paper, we proposed the shrinkage and preliminary test estimation methods in a system of regression models when the disturbances were dependent and correlations existed among regressors in each equation. To build the model, we first multiplied both sides of Model (1) by the inverse variance–covariance matrix of the disturbances and transformed the values using spectral decomposition. We defined the full model estimator by following Alkhamisi and Shukur (2008) and the restricted estimator by assuming a UNPI on the vector of parameters. Finally, we combined them in an optimal way by applying the shrinkage and preliminary test strategies. To illustrate and compare the relative performance of these methods, we conducted a Monte Carlo simulation. The simulated results demonstrated that the RE outperformed all other estimators when there was sufficient evidence that the vector nuisance parameters were a zero vector, that is $\Delta = 0$. However, the RE lost its efficiency as Δ increased and became unbounded when Δ was large. The PSE dominated the FME at the small values of Δ , while the SE and PSE outshone the FME in the entire parametric space. However, the PSE was better than the SE because it controlled for the over-shrinking problem in SE. We also investigated the performance of the suggested estimations via a real-world example using financial data for the “Fragile Five” countries. The results of our data analysis were consistent with the simulated results.

For further research, one can use the other penalized techniques for the SUR model such as the smoothly clipped absolute deviation (SCAD) by Fan and Li (2001), the least absolute shrinkage and selection operator (LASSO) by Tibshirani (1996), and the adaptive LASSO estimators by Zou (2006), as well as our preliminary and shrinkage estimations.

Author Contributions: Conceptualization, B.Y. and S.E.A.; methodology, S.E.A.; software, B.Y.; validation, B.Y. and S.E.A.; formal analysis, B.Y.; investigation, S.E.A.; resources, S.E.A.; data curation, B.Y.; writing—original draft preparation, B.Y.; writing—review and editing, B.Y. and S.E.A.; visualization, B.Y.; supervision, S.E.A.; project administration, B.Y.; funding acquisition, S.E.A. All authors have read and agreed to the published version of the manuscript.

Funding: The research of S. Ejaz Ahmed is supported by the Natural Sciences and the Engineering Research Council of Canada (NSERC).

Acknowledgments: The authors thank Guest Editors Shuangzhe Liu and Milind Sathye and the three reviewers for their detailed reading of the manuscript and their valuable comments and suggestions that led to a considerable improvement of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Agiomirgianakis, George Myron, Dimitrios Asteriou, and Kalliroi Papathoma. 2003. *The Determinants of Foreign Direct Investment: A Panel Data Study for the Oecd Countries*. Monograph (Discussion Paper). London: Department of Economics, City University London.
- Ahmed, S. Ejaz. 2014. *Penalty, Shrinkage and Pretest Strategies: Variable Selection and Estimation*. New York: Springer.
- Ahmed, S. Ejaz, and Bahadır Yüzbaşı. 2016. Big data analytics: Integrating penalty strategies. *International Journal of Management Science and Engineering Management* 11: 105–15. [\[CrossRef\]](#)
- Akın, Tuğba. 2019. The effects of political stability on foreign direct investment in fragile five countries. *Central European Journal of Economic Modelling and Econometrics* 11: 237–55.
- Alkhamisi, M. A. 2010. Simulation study of new estimators combining the sur ridge regression and the restricted least squares methodologies. *Statistical Papers* 51: 651–72. [\[CrossRef\]](#)
- Alkhamisi, M. A., and Ghazi Shukur. 2008. Developing ridge parameters for sur model. *Communications in Statistics—Theory and Methods* 37: 544–64. [\[CrossRef\]](#)
- Arashi, Mohammad, and Mahdi Roozbeh. 2015. Shrinkage estimation in system regression model. *Computational Statistics* 30: 359–76. [\[CrossRef\]](#)
- Barari, Mahua, and Srikanta Kundu. 2019. The role of the federal reserve in the us housing crisis: A var analysis with endogenous structural breaks. *Journal of Risk and Financial Management* 12: 125. [\[CrossRef\]](#)
- Belsley, David A., Edwin Kuh, and Roy E. Welsch. 2005. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Hoboken: John Wiley & Sons, vol. 571.

- Breusch, Trevor S., and Adrian R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society* 47: 1287–94. [CrossRef]
- Breusch, Trevor S., and Adrian R. Pagan. 1980. The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies* 47: 239–53. [CrossRef]
- Brown, Robert L., James Durbin, and James M. Evans. 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological)* 37: 149–63. [CrossRef]
- Diebold, Francis X., and Robert S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 20: 134–44.
- Dincer, Oguzhan C., and Fan Wang. 2011. Ethnic diversity and economic growth in China. *Journal of Economic Policy Reform* 14: 1–10. [CrossRef]
- Erdugan, Funda, and Fikri Akdeniz. 2016. Restricted estimator in two seemingly unrelated regression model. *Pakistan Journal of Statistics and Operation Research* 12: 579–88. [CrossRef]
- Fan, Jianqing, and Runze Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96: 1348–60. [CrossRef]
- Gao, Xiaoli, S. E. Ahmed, and Yang Feng. 2017. Post selection shrinkage estimation for high-dimensional data analysis. *Applied Stochastic Models in Business and Industry* 33: 97–120. [CrossRef]
- Greene, William H. 2019. *Econometric Analysis*. Harlow: Pearson.
- Groß, Jürgen. 2003. Restricted ridge estimation. *Statistics & Probability Letters* 65: 57–64.
- Henningens, Arne, and Jeff D. Hamann. 2007. Systemfit: A package for estimating systems of simultaneous equations in r. *Journal of Statistical Software* 23: 1–40. [CrossRef]
- Hoerl, Arthur E., and Robert W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12: 55–67. [CrossRef]
- Hubert, Mia, Tim Verdonck, and Özlem Yorulmaz. 2017. Fast robust sur with economical and actuarial applications. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 10: 77–88. [CrossRef]
- Hyndman, Rob J., and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. Melbourne: OTexts.
- James, W., and C. Stein. 1961. Proc. fourth berkeley symp. math. statist. probab. In *Estimation with Quadratic Loss*. Berkeley: University California Press, vol. 1, pp. 361–79.
- Jarque, Carlos M., and Anil K. Bera. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6: 255–59. [CrossRef]
- Kaçıranlar, Selahattin, Sadullah Sakallıoğlu, M. Revan Özkale, and Hüseyin Güler. 2011. More on the restricted ridge regression estimation. *Journal of Statistical Computation and Simulation* 81: 1433–48. [CrossRef]
- Kleiber, Christian, and Achim Zeileis. 2008. *Applied Econometrics with R*. Heidelberg: Springer Science & Business Media.
- Kuan, C.-M. 2004. *Introduction to Econometric Theory*. Lecture Notes. Available online: [ftp://nozdr.ru/biblio/kolxoz/G/GL/Kuan%20C.-M.%20Introduction%20to%20econometric%20theory%20\(LN,%20Taipei,%202002\)\(202s\)_GL_.pdf](ftp://nozdr.ru/biblio/kolxoz/G/GL/Kuan%20C.-M.%20Introduction%20to%20econometric%20theory%20(LN,%20Taipei,%202002)(202s)_GL_.pdf) (accessed on 18 June 2020).
- Lawal, Afeez Abolaji, and Oluwayemesi Oyeronke ALABA. 2019. Exploratory analysis of some sectors of the economy: A seemingly unrelated regression approach. *African Journal of Applied Statistics* 6: 649–61. [CrossRef]
- Ljung, Greta M., and George E. P. Box. 1978. On a measure of lack of fit in time series models. *Biometrika* 65: 297–303. [CrossRef]
- Mansfield, Edward R., and Billy P. Helms. 1982. Detecting multicollinearity. *The American Statistician* 36: 158–60.
- Pesaran, M. Hashem. 2004. *General Diagnostic Tests for Cross Section Dependence in Panels*. CESifo Working Papers. Cambridge: University of Cambridge, Faculty of Economics. .
- Ramsey, James Bernard. 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)* 31: 350–71. [CrossRef]
- Salman, A. Khalik. 2011. Using the sur model of tourism demand for neighbouring regions in sweden and norway. In *Advances in Econometrics-Theory and Applications*. London: IntechOpen, 98p.
- Shukur, Ghazi. 2002. Dynamic specification and misspecification in systems of demand equations: A testing strategy for model selection. *Applied Economics* 34: 709–25. [CrossRef]
- Srivastava, Virendera K., and David E. A. Giles. 1987. *Seemingly Unrelated Regression Equations Models: Estimation and Inference*. New York: CRC Press, vol. 80.

- Srivastava, Viren K., and Alan T. K. Wan. 2002. Separate versus system methods of stein-rule estimation in seemingly unrelated regression models. *Communications in Statistics—Theory and Methods* 31: 2077–99. [[CrossRef](#)]
- Stanley, Morgan. 2013. Fx pulse preparing for volatility. *Global Outlook* 1: 1–37.
- Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58: 267–88. [[CrossRef](#)]
- Williams, Barry. 2013. Income volatility of Indonesian banks after the Asian financial crisis. *Journal of the Asia Pacific Economy* 18: 333–58. [[CrossRef](#)]
- Yüzbaşı, Bahadır, S. Ejaz Ahmed, and Mehmet Güngör. 2017. Improved penalty strategies in linear regression models. *REVSTAT—Statistical Journal* 15: 251–76.
- Yüzbaşı, Bahadır, Mohammad Arashi, and S. Ejaz Ahmed. 2020. Shrinkage estimation strategies in generalised ridge regression models: Low/high-dimension regime. *International Statistical Review* 88: 229–51. [[CrossRef](#)]
- Zeebari, Zangin, B. M. Golam Kibria, and Ghazi Shukur. 2018. Seemingly unrelated regressions with covariance matrix of cross-equation ridge regression residuals. *Communications in Statistics-Theory and Methods* 47: 5029–53. [[CrossRef](#)]
- Zeebari, Zangin, Ghazi Shukur, and B. M. G. Kibria. 2012. Modified ridge parameters for seemingly unrelated regression model. *Communications in Statistics-Theory and Methods* 41: 1675–91. [[CrossRef](#)]
- Zellner, Arnold. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57: 348–68. [[CrossRef](#)]
- Zou, Hui. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101: 1418–29. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).