

Shrinkage Estimation of Linear Regression Models with GARCH Errors

S. Hossain

*Department of Mathematics & Statistics, University of Winnipeg
Winnipeg, Manitoba R3B 2E9, Canada
sh.hossain@uwinnipeg.ca*

M. Ghahramani*

*Department of Mathematics & Statistics, University of Winnipeg
Winnipeg, Manitoba R3B 2E9, Canada
m.ghahramani@uwinnipeg.ca*

Received 29 May 2015

Accepted 27 March 2016

Abstract

This paper introduces shrinkage estimators for the parameter vector of a linear regression model with conditionally heteroscedastic errors such as the class of generalized autoregressive conditional heteroscedastic (GARCH) errors when some of the regression parameters are restricted to a subspace. We derive the asymptotic distributional biases and risks of the shrinkage estimators using a large sample theory. We show that if the shrinkage dimension exceeds two, the relative efficiency of the shrinkage estimator is strictly greater than that of the full model estimator. Furthermore, a Monte Carlo simulation study is conducted to examine the relative performance of the shrinkage estimators with the full model estimator. Our large sample theory and simulation study show that the shrinkage estimators dominate the full model estimator in the entire parameter space. We illustrate the proposed method using a real data set from econometrics.

Keywords: Stein-type shrinkage; likelihood ratio test; linear regression model; GARCH error; asymptotic bias; asymptotic risk.

2000 Mathematics Subject Classification: Primary 62H12, Secondary 62J07

1. Introduction

The linear regression model is a commonly used statistical tool for analysis of the relationship between the response and predictor variables. Econometricians often wish to build regression models for macroeconomic data. For example, they may want to predict returns on a stock based on interest rates, money supply, and the monthly treasury bill rate; these variables serve as proxies for macroeconomic announcements. For a comprehensive review of regression models for the conditional mean and conditional variance of equity returns, see [14] and, references therein. However,

*Corresponding author.

empirical studies of financial time series have shown that linear regression models involving economic variables fail to be adequate because the residuals are rarely homoscedastic. In the presence of heteroskedasticity, ordinary least squares estimates of the regression parameters will be unbiased, but the standard errors and consequently the confidence intervals will be too narrow, giving a false sense of precision. Taking this heteroskedasticity explicitly into account has a number of advantages: First, the loss in asymptotic efficiency from neglected heteroskedasticity may be arbitrarily large, which may lead to a lower power of statistical tests. Second, a correct specification of the heteroskedasticity provides unbiased estimates of the variance.

Many financial series such as returns on stock prices and foreign exchange rates, exhibit leptokurtosis and time-varying volatility (variance). These two features have been the subject of extensive research ever since ([29] and [7]) reported them. Autoregressive conditional heteroscedastic (ARCH) models (see, [7]; [9]) and generalized autoregressive conditional heteroscedastic (GARCH) models of [4], provide a convenient framework to study time-varying volatility in financial markets. An investigation of the literature reveals that despite growing attention in designing various regression models with GARCH errors, to the best of our knowledge, no interest has been shown in implementing the shrinkage estimation method for regression parameters. Recently, [13] studied simultaneous model selection and parameter estimation using LASSO and Shrinkage methods for the partially linear models with random coefficient autoregressive errors. An extensive survey of the theory and applications of these regression models is given by [6], [18], and [30]. Other surveys of GARCH(p, q) processes include [23], [21], [3], [19], [32] and [25], [34], and [22]. Introductory surveys include [12] and [11]. Intermediate and advanced textbooks that provide more details are [36], [35], and [20]. Some research monographs on the theoretical properties of GARCH models include [33], [17], [37]; see also nine chapters on GARCH modeling in [1].

The aim of this paper is to develop the James-Stein shrinkage estimation method for the linear regression model with GARCH errors when some of the macroeconomic predictors may not be active, i.e., not statistically significant for the response or when the prior information on inactive predictors is available. One common problem in regression modeling occurs frequently in selecting active predictors for the response, especially when a large number of predictors is under consideration. If many inactive predictors are selected, the statistical model loses its predictive power, and the results are difficult to interpret. The shrinkage method estimates the coefficients of active predictors by using the information contained in discarded inactive predictors. In this situation we can partition the regression parameter vector $\boldsymbol{\beta}$ into two sub-vectors as $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)'$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are assumed to have dimensions $k_1 \times 1$ and $k_2 \times 1$, respectively, such that $k = k_1 + k_2$. We are interested in estimating the parameter sub-vector $\boldsymbol{\beta}_1$ by incorporating the information of $\boldsymbol{\beta}_2$ into the estimation procedure. One may obtain more efficient estimators this way than the estimators obtained when the prior information is ignored. We set the values of $\boldsymbol{\beta}_2$ near some specified value which, without loss of generality, a null vector, $\boldsymbol{\beta}_2^0 = \mathbf{0}$.

The plan of the paper is as follows. We begin with the linear regression model with GARCH errors in Section 2 and proceed to the estimation strategies. Section 3 describes the asymptotic distributional biases and risks of the shrinkage estimators where the corresponding proofs are given in the Appendix. In Section 4, we use Monte Carlo simulations to evaluate the numerical performance of the shrinkage estimators with respect to the full model estimator. Section 5 applies the procedure to a real data example. Concluding remarks are given in Section 6.

2. Model description and estimation strategies

In this section, the model arising from the motivating examples is discussed and several estimation strategies for the regression parameters are outlined.

2.1. Linear Regression Model with GARCH(p, q) Errors

Consider a multiple linear regression model with GARCH errors. Let y_t be the response, $\mathbf{x}'_t = (x_{t1}, x_{t2}, \dots, x_{tk})$ be a $k \times 1$ predictor vector, and ε_t be the GARCH(p, q) errors. The model for the process y_t is:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (2.1)$$

where

$$\varepsilon_t = \sigma_t Z_t, \quad (2.2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \gamma_j \sigma_{t-j}^2, \quad (2.3)$$

Z_t is an i.i.d random variable with continuous density such that Z_t has mean zero and variance 1, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)'$ is a $k \times 1$ vector of unknown regression coefficients.

Let $\boldsymbol{\omega}' = (\alpha_0, \alpha_1, \dots, \alpha_p, \gamma_1, \dots, \gamma_q)$ be the $(1 + p + q) \times 1$ vector of unknown variance parameters where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\gamma_j \geq 0$ for $i = 1, \dots, p$ and $j = 1, \dots, q$. Define $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\omega}')$ be the $(k + 1 + p + q) \times 1$ vector of all unknown parameters with $\boldsymbol{\theta} \in \Theta$ and Θ is a compact subset of Euclidean space. When $q = 0$, we obtain the class of ARCH models introduced by Engle (1982) as an error term. A necessary and sufficient condition for ε_t to be covariance stationary with finite variance is $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \gamma_j < 1$, see [4].

2.1.1. The Full and Reduced Model Estimators

Earlier literature on inference from GARCH models is based on the maximum likelihood estimation (MLE) method. For example, [7] assumes conditional normality while [4] advocates the use of the t-distribution. As the distribution of the innovation term is rarely known, Gaussian quasi-maximum likelihood (QML) estimation is employed. The idea is to maximize the likelihood function written under the assumption that Z_t in (2.2) is Gaussian. The resulting estimators are known as QMLEs. The assumption of normality for Z_t turns out to be non-essential for the strong consistency and asymptotic normality of the QMLE provided that the innovation has a finite fourth moment. This result holds even if the true distribution is far from Gaussian, as shown by [23], [21], and [3]. Semiparametric estimation of the regression model with ARCH errors using estimating function theory has been studied by [24]. Similarly, [16] studied estimation of the pure GARCH model using the combined estimating function method. While no distributional assumptions are made about Z_t , knowledge the conditional skewness and conditional kurtosis of the innovation process is required. In this paper, we study shrinkage estimators and their asymptotic properties based on QMLE for regression models with GARCH errors. The details are as follows and are based on discussions in [15].

Let $\boldsymbol{\theta}_0 = (\beta_{01}, \dots, \beta_{0k}, \alpha_{00}, \alpha_{01}, \dots, \alpha_{0p}, \gamma_{01}, \dots, \gamma_{0q})'$ be the unknown true parameter value. Put $\mathbf{s}'_t = (1, \varepsilon_{t-1}^2, \dots, \varepsilon_{t-p}^2, \sigma_{t-1}^2, \dots, \sigma_{t-q}^2)$. Now conditioning on initial values

$\varepsilon_0^2, \dots, \varepsilon_{1-p}^2, \tilde{\sigma}_0^2, \dots, \tilde{\sigma}_{1-q}^2$, the natural logarithm of the quasi-likelihood function for the model (2.1) apart from some constant is given by:

$$L_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n l_t(\boldsymbol{\theta})$$

$$l_t(\boldsymbol{\theta}) = -\frac{1}{2} \log(\tilde{\sigma}_t^2) - \frac{1}{2} \frac{(y_t - \mathbf{x}'_t \boldsymbol{\beta})^2}{\tilde{\sigma}_t^2}, \tag{2.4}$$

where $\tilde{\sigma}_t^2$ is defined recursively, for $t \geq 1$, by

$$\tilde{\sigma}_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \gamma_j \tilde{\sigma}_{t-j}^2. \tag{2.5}$$

[15] showed that while the sequence $\tilde{\sigma}_t^2$ is not stationary, it can be viewed as an approximation of the strictly stationary solution of Equation (2.3) under the assumption that $\sum_{j=1}^q \gamma_j < 1$.

The information matrix for the variance parameter $\boldsymbol{\omega}$ is given by

$$I_{\boldsymbol{\omega}\boldsymbol{\omega}} = -\frac{1}{n} \sum_{t=1}^n E \left(\frac{\partial^2 l_t}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}'} \right) = \frac{1}{2n} \sum_{t=1}^n \left(\frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\omega}} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\omega}'} \right),$$

where

$$\frac{\partial \sigma_t^2}{\partial \boldsymbol{\omega}} = \mathbf{s}_t + \sum_{j=1}^q \gamma_j \frac{\partial \sigma_{t-j}^2}{\partial \boldsymbol{\omega}}.$$

The information matrix corresponding to $\boldsymbol{\beta}$ is given by

$$I_{\boldsymbol{\beta}\boldsymbol{\beta}} = -\frac{1}{n} \sum_{t=1}^n E \left(\frac{\partial^2 l_t}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right) = \frac{1}{n} \sum_{t=1}^n \left(\frac{\mathbf{x}'_t \mathbf{x}_t}{\sigma_t^2} - \frac{1}{2} \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\beta}} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\beta}'} \right),$$

where

$$\frac{\partial \sigma_t^2}{\partial \boldsymbol{\beta}} = -2 \sum_{i=1}^p \alpha_i x_{t-i} \varepsilon_{t-i} + \sum_{j=1}^q \gamma_j \frac{\partial \sigma_{t-j}^2}{\partial \boldsymbol{\beta}}.$$

The elements of the off-diagonal block of the information matrix are zero. That is,

$$I_{\boldsymbol{\omega}\boldsymbol{\beta}} = -\frac{1}{n} \sum_{t=1}^n E \left(\frac{\partial^2 l_t}{\partial \boldsymbol{\omega} \partial \boldsymbol{\beta}'} \right) = 0.$$

Thus, the variance-covariance matrix is the inverse of the information matrix

$$I = \begin{bmatrix} I_{\boldsymbol{\omega}\boldsymbol{\omega}} & \mathbf{0} \\ \mathbf{0} & I_{\boldsymbol{\beta}\boldsymbol{\beta}} \end{bmatrix}.$$

A QMLE of $\boldsymbol{\theta}$ is defined as the solution $\hat{\boldsymbol{\theta}}_n$ of

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} L_n(\boldsymbol{\theta}).$$

The natural logarithm of the quasi-likelihood function is maximized using the numerical optimization techniques for a given initial value of the parameter. Due to the asymptotic independence of

β and ω , β may be estimated without loss of efficiency based on a consistent estimate of ω and vice-versa. The successive conditional variances are computed using the GARCH recursion formula from (2.5) with updated estimates $\hat{\alpha}_0, \hat{\alpha}_i$ ($i = 1, 2, \dots, p$), and $\hat{\gamma}_j$ ($j = 1, 2, \dots, q$). The regression parameter β is estimated using generalized least squares where $\hat{\sigma}_t^2$ forms the diagonal entries of the weighting matrix. This iterative procedure is repeated until there is no change in $\hat{\beta}$. The resulting QML estimator is denoted by $\hat{\beta}^{FE}$, as the full model estimator (FE).

Let $A(L) = \sum_{i=1}^p \alpha_i L^i$ and $B(L) = 1 - \sum_{j=1}^q \gamma_j L^j$ be the characteristic polynomials, where L is the backshift operator with $A(L) = 0$ if $p = 0$ and $B(L) = 1$ if $q = 0$. In addition, let $\gamma(\mathbf{A}_0)$ be the top Lyapunov exponent of the sequence (A_{0t}) , where (A_{0t}) is a matrix defined on p.86 of Andersen *et al.* (2009) used in deriving the stationarity of the GARCH process. The following assumptions (see [3], [33] and [15]) are needed to ensure the consistency and asymptotic normality of the estimator $\hat{\beta}$:

A1: $\theta_0 \in \Theta$ and Θ is a compact subspace of the Euclidean space

A2: $\gamma(\mathbf{A}_0) < 0$ and $\forall \theta \in \Theta, \sum_{j=1}^q \gamma_j < 1$.

A3: Z_t^2 has a non-degenerate distribution with $E(Z_t) = 0$ and $E(Z_t^2) = 1$.

A4: if $q > 0$, $A(L)$ and $B(L)$ have no common roots, $A(1) \neq 0$, and $\alpha_{0p} + \gamma_{0q} \neq 0$.

A5: $E(Z_t^4) < \infty$.

A6: The information matrix $E\left(-\frac{\partial^2 l_t}{\partial \theta \partial \theta'}\right)$ is finite and positive definite.

Under the above assumptions, the full model estimator $\hat{\beta}^{FE}$ is consistent and asymptotically normally distributed with mean β and the variance-covariance matrix $\mathbf{I}_0(\hat{\beta}) = \lim_{n \rightarrow \infty} \mathbf{I}_{\beta\beta}/n$ and it is constructed from an estimate of the information matrix used in the last step of the optimization algorithm. The matrix $\mathbf{I}_0(\hat{\beta})$ with $\hat{\beta} = (\hat{\beta}'_1, \hat{\beta}'_2)'$ can be partitioned as

$$\begin{pmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{pmatrix},$$

The reduced model estimator (RE), $\hat{\beta}^{RE}$ of β can be obtained by maximizing the log-likelihood function (2.4) subject to the restriction $\beta_2 = \mathbf{0}$. If the restriction is true, the RE is expected to perform better than the FE, otherwise the RE tends to be biased, inefficient, and even inconsistent.

Remark. Under assumptions **A1-A6**, the error process (ε_t) will be strictly stationary. Applied time series analysts view strict stationarity as restrictive and focus their attention on covariance stationary series. However, second-order stationarity requires up to the fourth moment of the error term (and hence y_t) to be finite; this requirement is viewed as questionable for financial series.

2.2. Shrinkage and Positive Shrinkage Estimators

The shrinkage estimator which combines the full and reduced model estimators, can be defined as

$$\hat{\beta}^{SE} = \hat{\beta}^{RE} + (1 - \nu \hat{D}_n^{-1}) (\hat{\beta}^{FE} - \hat{\beta}^{RE}), \quad \nu = k_2 - 2 \quad \text{and} \quad \nu \geq 1,$$

where \hat{D}_n is defined below. In other words, the shrinkage estimator $\hat{\beta}^{SE}$ is the linear combination $(1 - \lambda) \hat{\beta}^{FE} + \lambda \hat{\beta}^{RE}$ of a full model estimator $\hat{\beta}^{FE}$ and a reduced model estimator $\hat{\beta}^{RE}$. The parameter λ determines the extent to which these estimates are pooled together. If $\lambda = 1$ then the reduced

model estimator dominates completely, whereas for $\lambda = 0$ no shrinkage occurs. The test statistic D_n is defined as

$$D_n = 2[l(\hat{\boldsymbol{\beta}}^{FE}; \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) - l(\hat{\boldsymbol{\beta}}^{RE}; \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)],$$

$$= n\hat{\boldsymbol{\beta}}_2' \mathbf{I}_{22.1} \hat{\boldsymbol{\beta}}_2.$$

where $\mathbf{I}_{22.1} = \mathbf{I}_{22} - \mathbf{I}_{21} \mathbf{I}_{11}^{-1} \mathbf{I}_{12}$. If $\mathbf{I}(\boldsymbol{\beta})$ is replaced by a consistent estimator $\mathbf{I}(\hat{\boldsymbol{\beta}})$, then

$$\hat{D}_n = n\hat{\boldsymbol{\beta}}_2^\top \hat{\mathbf{I}}_{22.1} \hat{\boldsymbol{\beta}}_2 + o_P(1),$$

which is asymptotically χ^2 -distributed with k_2 degrees of freedom $n \rightarrow \infty$ when the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ is true.

The difficulties with the shrinkage estimator is that it is not a convex combination of full and reduced model estimators, and it has the tendency to over-shrink the full model estimator and reverse the sign. This can occur when the test statistic \hat{D}_n is less than ν . To moderate this effect, it is better to define a truncated estimator called a positive-part shrinkage estimator (PSE). This estimator is defined as

$$\hat{\boldsymbol{\beta}}^{PSE} = \hat{\boldsymbol{\beta}}^{RE} + (1 - \nu \hat{D}_n^{-1})^+ (\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE}),$$

where $z^+ = \max(0, z)$.

3. Asymptotic Results

3.1. Asymptotic Distributional Results: Bias

In this subsection we present the asymptotic distributions of the estimators and the necessary test statistics. This will facilitate the derivation of the asymptotic distributional bias (ADB) and, later, the asymptotic distributional risk (ADR) of the estimators of $\boldsymbol{\beta}$.

Under nonlocal (fixed) alternatives, all the estimators are asymptotically equivalent to $\hat{\boldsymbol{\beta}}^{FE}$, while $\hat{\boldsymbol{\beta}}^{RE}$ has unbounded risk. To obtain the non-degenerate asymptotic distribution and the meaningful bias and risk comparisons among these estimators we consider the local Pitman alternatives, we consider a sequence of local alternatives,

$$K_{(n)} : \boldsymbol{\beta}_2 = \frac{\boldsymbol{\delta}}{\sqrt{n}}, \tag{3.1}$$

Note that $\boldsymbol{\delta} = \mathbf{0}$ implies $\boldsymbol{\beta}_2 = \mathbf{0}$. So (3.1) is a special case of $K_{(n)}$. Under local $K_{(n)}$, the following theorem facilitates the derivation and numerical computation of the ADB and the ADR of the estimators outlined below:

Theorem 3.1.1. Under the local alternatives $K_{(n)}$ in (3.1) and the usual regularity conditions,

- (1) $\sqrt{n}\hat{\boldsymbol{\beta}}_2 \xrightarrow{\mathcal{L}} N(\boldsymbol{\delta}, \mathbf{I}_{22.1})$ as $n \rightarrow \infty$
- (2) As $n \rightarrow \infty$, the distribution of \hat{D}_n converge to a non-central chi-squared distribution $H_{k_2}(x; \Delta)$ with k_2 degrees of freedom and non-centrality parameter $\Delta = \boldsymbol{\delta}' \mathbf{I}_{22.1} \boldsymbol{\delta}$, $\mathbf{I}_{22.1} = \mathbf{I}_{22} - \mathbf{I}_{21} \mathbf{I}_{11}^{-1} \mathbf{I}_{12}$ is a positive definite matrix, and $H_{k_2}(x; \Delta) = P(\chi_{k_2}^2(\Delta) \leq x)$, $x \in \mathbb{R}^+$.

The outline of a proof is available in [31], Chapter 1.

The shrinkage estimators are, in general, biased estimators, although bias is accompanied by reduction in variance. We define ADB of an estimator β^* as

$$ADB(\beta^*) = E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\beta^* - \beta) \right],$$

where β^* be a generic notation for any of $\hat{\beta}^{FE}$, $\hat{\beta}^{RE}$, $\hat{\beta}^{SE}$ or $\hat{\beta}^{PSE}$.

Theorem 3.1.2. Using the above definition of ADB and Theorem 3.1.1, under the local alternatives $K_{(n)}$ in (3.1), and the usual regularity conditions, as $n \rightarrow \infty$,

$$ADB(\hat{\beta}^{PSE}) = ADB(\hat{\beta}^{SE}) - \mathbf{I}_{11}^{-1} \mathbf{I}_{12} \boldsymbol{\delta} [H_{\nu+4}(\nu, \Delta) - \nu E(Z_1 I(\nu Z_1 > 1))],$$

where $Z_1 = \chi_{k_2+2}^{-2}(\Delta)$, $ADB(\hat{\beta}^{SE}) = -\nu \mathbf{I}_{11}^{-1} \mathbf{I}_{12} \boldsymbol{\delta} E(Z_1)$, $ADB(\hat{\beta}^{RE}) = -\mathbf{I}_{11}^{-1} \mathbf{I}_{12} \boldsymbol{\delta}$, and $H_g(x; \Delta) = P(\chi_g^2(\Delta) \leq x)$, $x \in \mathbb{R}^+$ is the cumulative distribution function of a non-central chi-square with g degrees of freedom and non-centrality parameter Δ , and

$$E(\chi_g^{-2j}(\Delta)) = \int_0^\infty x^{-2j} dH_g(x; \Delta).$$

Proof. See the Appendix.

The constant term $\boldsymbol{\delta}$ is common to the ADBs of $\hat{\beta}^{RE}$, $\hat{\beta}^{SE}$, and $\hat{\beta}^{PSE}$ and the ADBs differ only by a constant factor Δ only. It then, suffices to compare Δ only. It is clear that the ADB of the $\hat{\beta}^{RE}$ is an unbounded function of Δ . On the other hand, the ADBs of both $\hat{\beta}^{SE}$ and $\hat{\beta}^{PSE}$ are bounded in Δ . Since $E(\chi_{k_2+2}^{-2}(\Delta))$ is a decreasing function of Δ , the ADB of $\hat{\beta}^{SE}$ starts from the origin, increases to a maximum, and then decreases towards 0 as $\Delta > 0$. The characteristics of $\hat{\beta}^{PSE}$ are similar to that of $\hat{\beta}^{SE}$.

3.2. Asymptotic Distributional Results: Risk

To derive expressions for the ADRs of the estimators, we define a quadratic loss function

$$\mathcal{L}(\beta^*; \mathbf{W}) = [\sqrt{n}(\beta^* - \beta)]' \mathbf{W} [\sqrt{n}(\beta^* - \beta)],$$

where \mathbf{W} is a positive semidefinite weight matrix. If we choose $\mathbf{W} = \mathbf{I}$, the identity matrix, this is the usual quadratic loss. Using a general \mathbf{W} gives a loss function that weights different β 's differently. We will use this identity matrix in the simulation study. One may select a weight matrix with different arbitrary entries on its main diagonal but in this situation, shrinkage estimators are not guaranteed to outperform the full model estimators. As an example, see the discussion in [27]. The

expected loss function is defined as

$$E[\mathcal{L}(\boldsymbol{\beta}^*; \mathbf{W})] \equiv \text{ADR}(\boldsymbol{\beta}^*, \boldsymbol{\beta}; \mathbf{W}) \equiv \text{ADR}(\boldsymbol{\beta}^*, \boldsymbol{\beta})$$

which is called the risk function. The risk function can be rewritten as

$$\begin{aligned} \text{ADR}(\boldsymbol{\beta}^*, \boldsymbol{\beta}; \mathbf{W}) &= nE\{(\boldsymbol{\beta}^* - \boldsymbol{\beta})\mathbf{W}(\boldsymbol{\beta}^* - \boldsymbol{\beta})'\} \\ &= n\text{trace}[\mathbf{W}E\{(\boldsymbol{\beta}^* - \boldsymbol{\beta})(\boldsymbol{\beta}^* - \boldsymbol{\beta})'\}] \\ &= \text{trace}(\mathbf{W}\boldsymbol{\Gamma}), \end{aligned} \tag{3.2}$$

where $\boldsymbol{\Gamma}$ is the asymptotic covariance matrix of $\boldsymbol{\beta}^*$.

Under the sequence of local alternatives, we define the asymptotic distribution function of an estimator $\boldsymbol{\beta}^*$ as

$$G(\mathbf{y}) = \lim_{n \rightarrow \infty} P[\sqrt{n}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) \leq \mathbf{y} | K_{(n)}],$$

where $G(\mathbf{y})$ is nondegenerate distribution function for the estimators we consider. We define the asymptotic distributional quadratic risk (ADR) by

$$\begin{aligned} \text{ADR}(\boldsymbol{\beta}^*; \mathbf{W}) &= \int \dots \int \mathbf{y}'\mathbf{W}\mathbf{y}dG(\mathbf{y}) \\ &= \text{trace}(\mathbf{W}\boldsymbol{\Gamma}), \end{aligned}$$

where $\boldsymbol{\Gamma} = \int \dots \int \mathbf{y}\mathbf{y}'dG(\mathbf{y})$ is the dispersion matrix for the distribution $G(\mathbf{y})$.

An estimator $\boldsymbol{\beta}^*$ is then said to dominate an estimator $\boldsymbol{\beta}^0$ asymptotically if $\text{ADR}(\boldsymbol{\beta}^*; \boldsymbol{\beta}) \leq \text{ADR}(\boldsymbol{\beta}^0; \boldsymbol{\beta})$. If, in addition, $\text{ADR}(\boldsymbol{\beta}^*; \boldsymbol{\beta}) < \text{ADR}(\boldsymbol{\beta}^0; \boldsymbol{\beta})$ for at least some $(\boldsymbol{\beta}, \mathbf{W})$, then $\boldsymbol{\beta}^*$ strictly dominates $\boldsymbol{\beta}^0$.

Theorem 3.1.3. Under the local alternatives $K_{(n)}$ in (3.1) and the usual regularity conditions, as $n \rightarrow \infty$, we obtain the ADR functions of the proposed estimators by virtue of the following theorem:

$$\begin{aligned} \text{ADR}(\hat{\boldsymbol{\beta}}^{PSE}; \mathbf{W}) &= \text{ADR}(\hat{\boldsymbol{\beta}}^{SE}; \mathbf{W}) - \text{trace}(\mathbf{W}\mathbf{B})E[(1 - vZ_1)^2I(vZ_1 > 1)] \\ &\quad - \boldsymbol{\gamma}'\mathbf{W}\boldsymbol{\gamma}E[(1 - vZ_2)^2I(vZ_2 > 1)] \\ &\quad + 2\boldsymbol{\gamma}'\mathbf{W}\boldsymbol{\gamma}E[(1 - vZ_1)I(vZ_1 > 1)], \text{ where} \\ \text{ADR}(\hat{\boldsymbol{\beta}}^{SE}; \mathbf{W}) &= \text{ADR}(\hat{\boldsymbol{\beta}}^{FE}; \mathbf{W}) - 2v\text{trace}(\mathbf{W}\mathbf{B})[2\Delta E(Z_2^2) + E(Z_1)] \\ &\quad + v(v + 4)\boldsymbol{\gamma}'\mathbf{W}\boldsymbol{\gamma}E(Z_2^2), \end{aligned}$$

$$Z_2 = \chi_{k_2+4}^{-2}(\Delta), \text{ADR}(\hat{\boldsymbol{\beta}}^{RE}; \mathbf{W}) = \text{ADR}(\hat{\boldsymbol{\beta}}^{FE}; \mathbf{W}) - \text{trace}(\mathbf{W}\boldsymbol{\Phi}) + \boldsymbol{\gamma}'\mathbf{W}\boldsymbol{\gamma}, \text{ADR}(\hat{\boldsymbol{\beta}}^{FE}; \mathbf{W}) = \text{trace}[\mathbf{W}\mathbf{I}_{11.2}^{-1}], \boldsymbol{\gamma} = \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\boldsymbol{\delta}, \boldsymbol{\Phi} = \mathbf{I}_{11.2}^{-1} - \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\mathbf{I}_{22.1}^{-1}\mathbf{I}_{21}\mathbf{I}_{11}^{-1}, \mathbf{I}_{11.2} = \mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21}, \text{ and } \mathbf{B} = \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\mathbf{I}_{22.1}^{-1}\mathbf{I}_{21}\mathbf{I}_{11}^{-1}.$$

Proof. See the Appendix.

4. Simulation Results

In this section we have carried out a Monte Carlo simulation study to assess the relative performance of the proposed estimators with respect to the full model estimator. We generated response from the

following regression model with GARCH (1,1) errors:

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + \varepsilon_t, \tag{4.1}$$

where

$$\begin{aligned} \varepsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 \sigma_{t-1}^2, \end{aligned}$$

with $t = 1, 2, \dots, n$ where $n = 600, 700, \text{ and } 800$, and Z_t are i.i.d standard normal errors. The predictor values $\mathbf{x}'_t = (x_{t1}, x_{t2}, \dots, x_{tk})$ have been drawn from a multivariate standard normal distribution.

We partition the regression coefficients as $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)'$, where the coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $k_1 \times 1$ and $k_2 \times 1$ vectors, respectively with $k = k_1 + k_2$. Then the linear model (4.1) can be written as

$$y_t = \mathbf{X1}'_t \boldsymbol{\beta}_1 + \mathbf{X2}'_t \boldsymbol{\beta}_2 + \varepsilon_t,$$

where $\mathbf{X1}$ and $\mathbf{X2}$ are $k_1 \times n$ and $k_2 \times n$ dimensional partitioned design matrices, respectively. In this model, our objective is to construct a test for the hypothesis

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0} \text{ versus } H_a : \boldsymbol{\beta}_2 \neq \mathbf{0}.$$

We set the true values of $\boldsymbol{\beta}$ at $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2) = (\boldsymbol{\beta}'_1, \mathbf{0})'$ with $\boldsymbol{\beta}_1 = (1.6, 2.2, -1.2)$ and the weight matrix $\mathbf{W} = \mathbf{I}$. We also set the true value of GARCH parameters $\alpha_0 = 0.002$, $\alpha_1 = 0.2$, and $\gamma_1 = 0.7$. We define the parameter

$$\Delta = \|\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}\|^2,$$

where $\boldsymbol{\beta}^{(0)} = (\boldsymbol{\beta}'_1, \mathbf{0})'$, $\boldsymbol{\beta}_1$ is the true parameter in the model and $\|\cdot\|$ is the Euclidian norm. Samples were generated using Δ between 0 and 1. We provide detailed results for $(k_1, k_2) = (3, 3), (3, 6), (3, 12), (3, 17), (3, 20), \text{ and } (3, 25)$. The number of replications in the simulation was varied initially but it was determined that 1,000 was adequate for each combination of parameters because a further increase in the number did not change the results significantly.

The objective here is to investigate the behavior of the estimators for $\Delta \geq 0$. In order to do this, further samples are generated from the standard normal distribution (i.e. for different Δ between 0 and 1). Different values of $\boldsymbol{\beta}_2$ are chosen to obtain the different values of Δ . The criterion for comparing the performance of any estimator $\hat{\boldsymbol{\beta}}^*$ is based on the mean squared error (MSE). Based on the simulated data, we calculated the MSE of all the estimators studied in this paper. We consider the full model estimator $\hat{\boldsymbol{\beta}}^{FE}$ as the “benchmark” estimator, and thus the performance of the estimators is evaluated in terms of the MSE relative to the MSE of $\hat{\boldsymbol{\beta}}^{FE}$. For any estimator $\hat{\boldsymbol{\beta}}^*$, the simulated relative MSE (RelMSE) of $\hat{\boldsymbol{\beta}}^*$ to $\hat{\boldsymbol{\beta}}^{FE}$ is defined by

$$\text{RelMSE}(\hat{\boldsymbol{\beta}}^{FE} : \hat{\boldsymbol{\beta}}^*) = \frac{\text{Simulated MSE}(\hat{\boldsymbol{\beta}}^{FE})}{\text{Simulated MSE}(\hat{\boldsymbol{\beta}}^*)}.$$

Keep in mind that RelMSE greater than 1 indicates the degree of superiority of $\hat{\boldsymbol{\beta}}^*$ over $\hat{\boldsymbol{\beta}}^{FE}$.

The simulation results are presented in Figures 1, 2, & 3 ($\Delta \geq 0$) and Table 1 ($\Delta = 0$) when the active set of parameters $\boldsymbol{\beta}_1 = (1.6, 2.2, -1.2)$ and inactive set of parameters $\boldsymbol{\beta}_2 = (\boldsymbol{\beta}_3, \mathbf{m})$, where

Table 1. Simulated relative MSE of RE, SE, and PSE with respect to FE when the restricted parameter space is correct ($\Delta = 0$) for $n = 600, 700,$ and 700 .

$n = 600$						
Estimators	$k_2 = 3$	$k_2 = 6$	$k_2 = 12$	$k_2 = 17$	$k_2 = 20$	$k_2 = 25$
RE	2.07	3.54	5.10	7.10	7.90	8.98
SE	1.26	2.10	2.95	4.23	4.99	5.65
PSE	1.28	2.33	3.58	4.95	5.38	7.62
$n = 700$						
RE	1.93	2.98	4.79	6.69	7.72	8.90
SE	1.21	1.86	2.86	4.22	4.86	5.46
PSE	1.27	2.05	3.51	4.83	5.32	7.27
$n = 800$						
RE	1.88	2.92	4.78	6.68	7.70	8.73
SE	1.18	1.76	2.83	4.02	4.73	5.32
PSE	1.25	2.03	3.31	4.78	5.25	6.46

$\beta_3 = \sqrt{\Delta} = 0, 0.07, 0.10, 0.20, 0.32, 0.55, 0.71,$ and 1.00 and \mathbf{m} is a zero vector with varying dimensions, are presented in Figures 1 and 2 and Table 1. Note that there are $m = k_2 - 1$ inactive variables and that $\Delta = \beta_3^2$. We summarize our findings as follows:

- (i) The reduced model estimator $\hat{\beta}^{RE}$ dominates all the estimators at and near $\Delta = 0$. That is, if a reduced model is nearly correctly specified, then the RE is the optimal one. For large values of Δ , the RelMSE of $\hat{\beta}^{RE}$ decreases sharply, crosses the benchmark line at a certain level of Δ , and converges to zero as $\Delta \rightarrow \infty$ whereas the RelMSE of all the other estimators remains bounded and approaches 1.
- ii) Keeping n and Δ fixed, the RelMSE of the shrinkage estimator is a monotone increasing function of the number of inactive variables m . For $m = 16$, the RelMSE of the PSE can be as large as 4.95 (Table 1). This indicates a remarkable performance of PSE over FE. The figures also reveal that for small values of Δ , the positive shrinkage estimator offers a large improvement in RelMSE as the number m of inactive predictors is increased. Comparing the SE and the PSE shows that the PSE has a higher RelMSE than the SE, which shows the better performance of the PSE over the SE. Finally, our simulation finding is in agreement with the asymptotic results of Section 3.

5. Real data example

In this section, we apply the shrinkage estimation procedure to the [26] data set. The data consist of 14 macroeconomic variables sampled at an annual frequency. The sample varies for each series with a starting date between 1860 and 1909. However, each series ends in 1988. Data are available as a part of **R**'s **tseries** package ([28]). As two of the time series start from 1909, only the data from 1909 until the end of the series in 1988 will be used. The sample size is $n = 80$. The response variable is the log returns of stock prices as the price process is not weakly stationary. The predictor variables

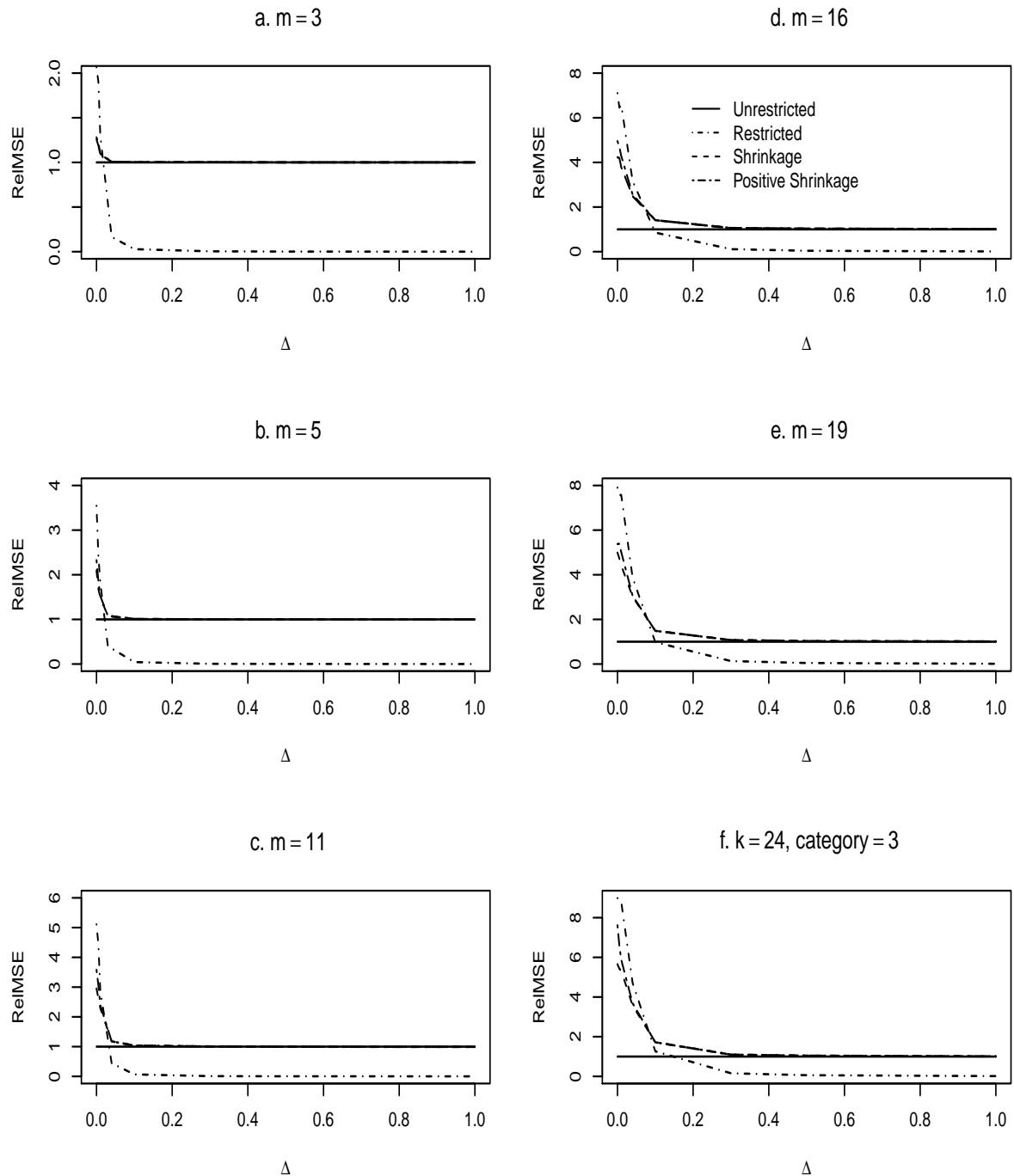


Fig. 1. Simulated relative efficiency with respect to $\hat{\beta}^{FE}$ of the estimators when the subspace misspecifies β_3 as zero as a function of $\Delta = \beta_3^2$. Here $m = k_2 - 1$ which is the number of inactive predictors and $n = 600$.

are real GNP (x_1), real per capita GNP (x_2), GNP deflator (x_3), money stock (x_4), real wages (x_5), industrial production index (x_6), bond yield (x_7), consumer price index (x_8), total employment (x_9),

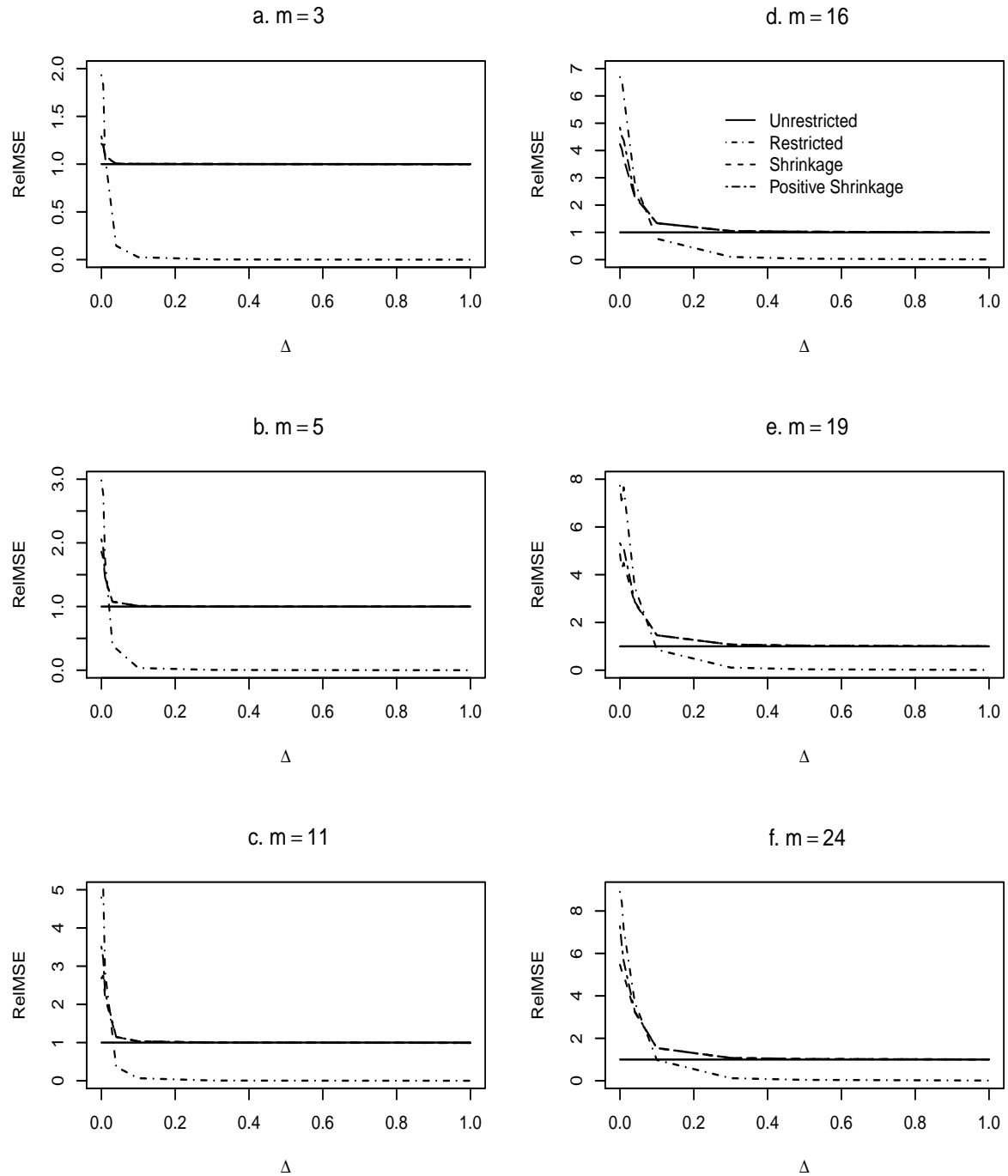


Fig. 2. Simulated relative efficiency with respect to $\hat{\beta}^{FE}$ of the estimators when the subspace misspecifies β_3 as zero as a function of $\Delta = \beta_3^2$. Here $m = k_2 - 1$ which is the number of inactive predictors and $n = 700$.

nominal GNP (x_{10}), velocity of money (x_{11}), nominal wages (x_{12}), and total unemployment rate

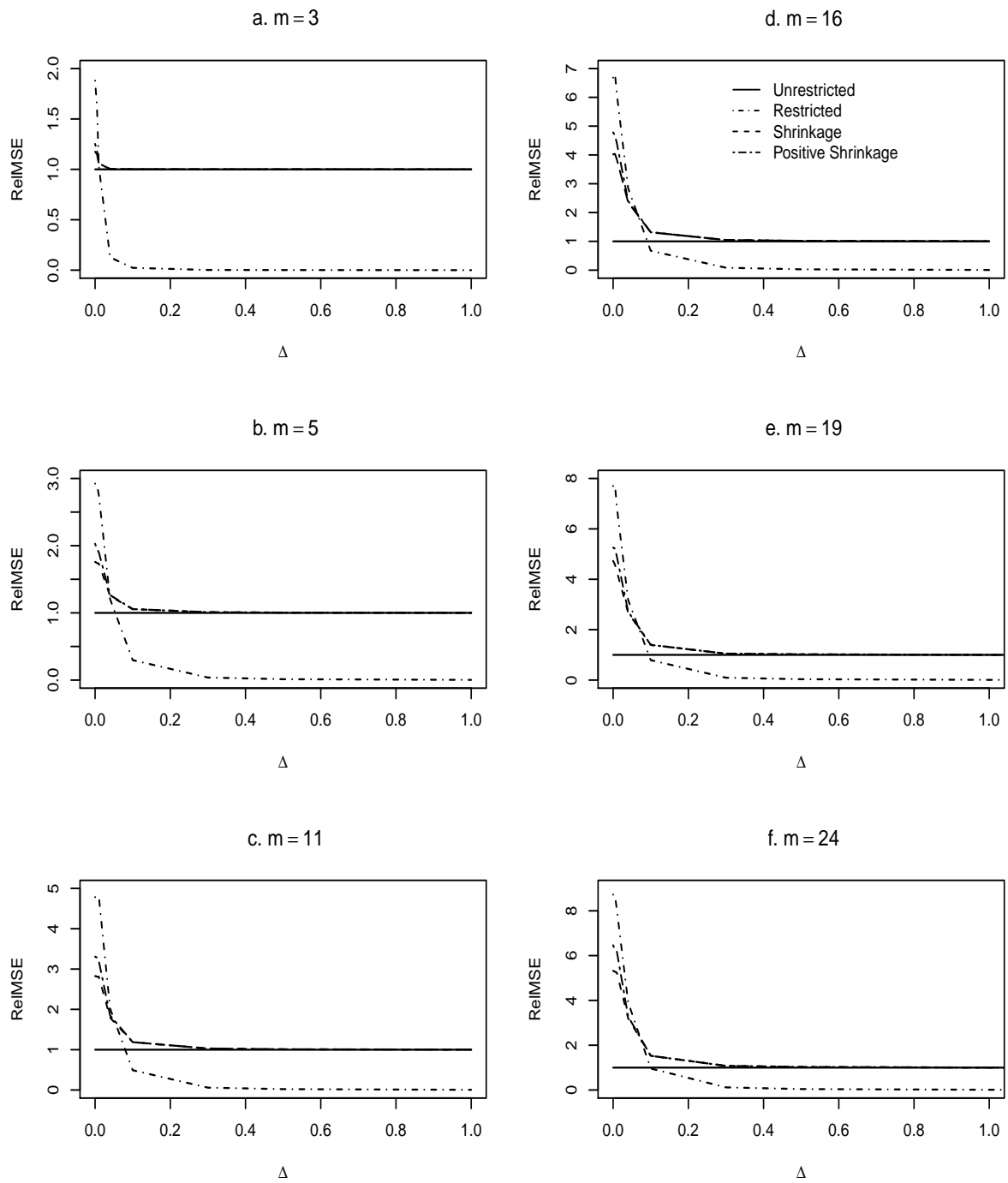


Fig. 3. Simulated relative efficiency with respect to $\hat{\beta}^{FE}$ of the estimators when the subspace misspecifies β_3 as zero as a function of $\Delta = \beta_3^2$. Here $m = k_2 - 1$ which is the number of inactive predictors and $n = 800$.

(x_{13}). We use the first difference of all predictors, where x_6 and x_8 being log-transformed before differencing in order, to stabilize the predictor variances.

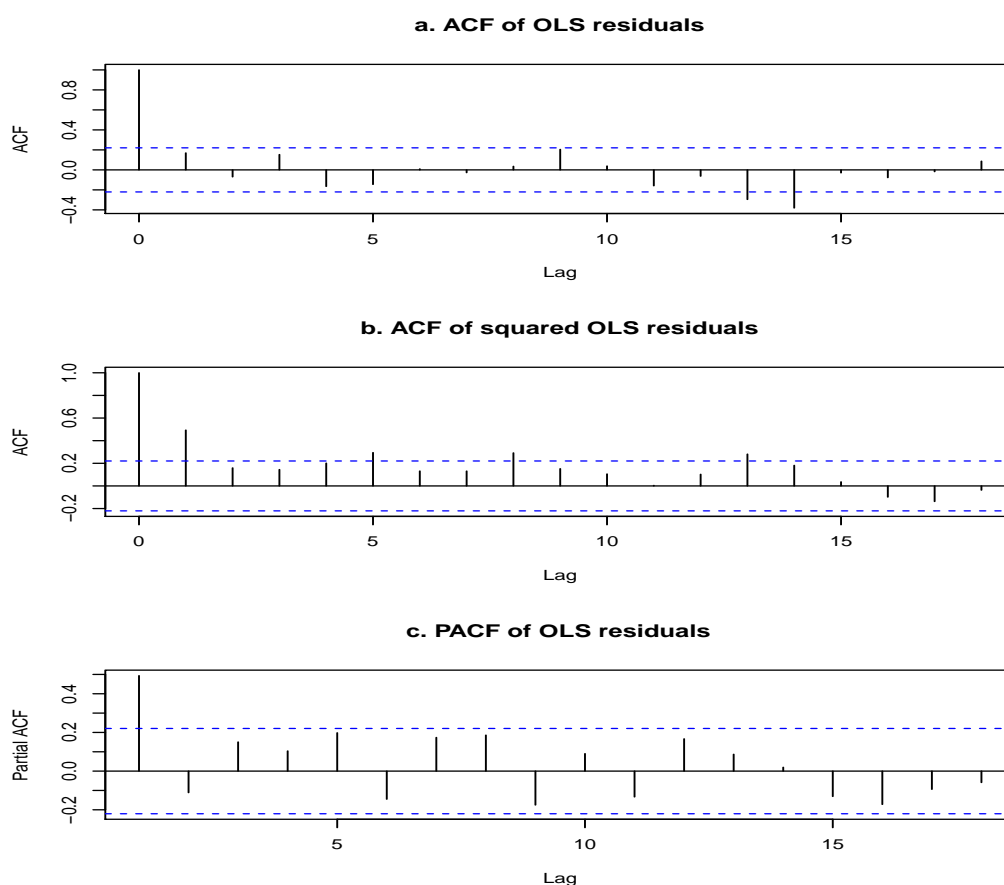


Fig. 4. ACF and PACF plots of OLS residuals for Nelson-Plosser data

We begin our analysis of the data with the ordinary least squares (OLS) method. The residuals from the OLS fit are uncorrelated according to the sample autocorrelation (ACF) and partial ACF (PACF) plots, see Figure 3(a, c). However, the sample ACF of squared residuals (see Figure 3(b)) shows that they are correlated having a significant spike at lag 1. This motivates us to fit a regression model with GARCH(1, 1) errors.

In order to form a restricted subspace of the total number of predictors, we used AIC or BIC criterion in the first step of the shrinkage method. It shows that $x_1, x_2, x_3, x_4,$ and x_5 are the active predictors, and the effects of the remaining six predictors may be ignored. We then form a constraint on the full model by using the inactive predictors. Here is the restricted subspace, $\beta_2 = (\beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})' = (0, 0, 0, 0, 0, 0, 0, 0), k = 13, k_1 = 5,$ and $k_2 = 8.$

We bootstrap from the data to examine the performance of the suggested estimators for estimating coefficients of the five active predictors. We draw bootstrap samples of size $n = 75$ by 10000 times with replacement from the data matrix (y_t, \mathbf{x}_t) to examine the point estimates, standard errors, and relative mean squared errors of the estimators. These results are reported in Table 2. Our numerical result reveals that the reduced and shrinkage estimators are superior to the full model estimator, which is in agreement with our theoretical and simulation results. As expected, the reduced model

Table 2. Estimate (first row) and standard error (second row) for real GNP (β_1), real per capita GNP (β_2), GNP deflator (β_3), money stock (β_4), real wages (β_5) on the log returns of stock prices. The RelMSE column gives the relative MSE based on bootstrap simulation of the estimators with respect to FE.

Estimators	β_1	β_2	β_3	β_4	β_5	RelMSE
FE	3.460	-2.876	-0.026	0.475	-1.072	1.0000
	2.233	1.455	0.012	0.093	0.045	
RE	3.368	-2.819	-0.251	0.364	-0.390	2.821
	1.446	1.383	0.008	0.027	0.023	
SE	3.455	-2.875	-0.049	0.462	-0.972	1.225
	2.041	1.373	0.010	0.086	0.037	
PSE	3.452	-2.871	-0.052	0.461	-0.971	1.225
	2.037	1.372	0.010	0.086	0.036	

estimator performs better than all the other estimators as the data is generated from an empirical distribution where the subset of predictors is nearly true.

6. Conclusion

We compared the performance of the reduced, shrinkage, and positive estimators with respect to the full model estimator in the context of a linear regression model with GARCH errors when some of the regression parameters lie in a subspace. We explored the risk properties of the shrinkage estimators via asymptotic distributional risk and a Monte Carlo experiment. We conclude that the risk improvement of the reduced model estimator over other estimators is substantial at and near the restriction $\beta_2 = \mathbf{0}$. However, the improvement starts diminishing as the restriction moves away from the restriction $\beta_2 = \mathbf{0}$. Furthermore, the shrinkage estimators with data based weights perform well if the restriction is true. In fact, the shrinkage estimators outperform the full model estimator $\hat{\beta}^{FE}$ in the entire parameter space for $\nu \geq 1$. Finally, a real data example and a simulation study support the idea that the proposed estimators are superior to the full model estimator.

Acknowledgments

We express our sincere thanks to the referees and the editor for their constructive and valuable suggestions, which led to an improvement of this paper. The research of Shakhawat Hossain and Melody Ghahramani was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada

References

- [1] T. G. Andersen, R. A. Davis, J.P. Kreiss and T. Mikosch, Handbook of Financial Time Series (Berlin: Springer Verlag, 2009).
- [2] G. G. Judge and M. E. Bock, The statistical implication of pretest and Stein-rule estimators in econometrics (Amsterdam: North-Holland, 1978).
- [3] I. Berkes, I. L. Horváth, and P. Kokoszka, GARCH processes: structure and estimation, *Bernoulli*, **9** (2003),201–127.
- [4] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**(1986), 307–327.

- [5] T. Bollerslev, A conditionally heteroskedastic time series model for speculative prices and rates of return, *The Review of Economics and Statistics*, **69**(1987), 542–547.
- [6] T. Bollerslev, R.Y. Chou, N. Jayaraman and K.F. Kroner, ARCH conditionally heteroskedastic time series model for speculative prices and rates of return, *Journal of Econometrics*, **52**(1992), 5–59.
- [7] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**(1982), 987–1007.
- [8] R. F. Engle and T. Bollerslev, Modeling the persistence of conditional variances, *Econometric Reviews*, **5**(1986), 1–50.
- [9] R. F. Engle and G. González-Rivera, Semi-parametric estimation of ARCH model, *Journal of Business and Economic Statistics*, **9**(1991), 345–358.
- [10] R. F. Engle, GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics, *Journal of Economic Perspectives*, **15**(2001), 157–168.
- [11] R. F. Engle, Risk and volatility: Econometric models and financial practice, *American Economic Review*, **94**(2004), 405–420.
- [12] R.F. Engle and A.J. Patton, What Good is a Volatility Model?, *Quantitative Finance*, **1**(2001), 237–245.
- [13] S. Fallahpour, S.E. Ahmed and K.A. Doksum, L_1 penalty and shrinkage estimation in partially linear models with random coefficient autoregressive errors, *Applied Stochastic Models in Business and Industry*, **28**(2012), 236–250.
- [14] M. J. Flannery and A.A. Protopapadakis, A. A., Macroeconomic factors do influence aggregate stock returns, *The Review of Financial Studies*, **15**(2002), 751–782.
- [15] Francq, C. and Zakoian, J. M., Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, *Bernoulli*, **10**(2004), 605–637.
- [16] M. Ghahramani and A. Thavaneswaran, Combining Estimating Functions for Volatility, *Journal of Statistical Planning and Inference*, **139**(2009), 1449–1461.
- [17] C. Francq, C. and J.M. Zakoian, GARCH Models: Structure, Statistical Inference and Financial Applications (New York: John Wiley, 2010).
- [18] C. Gouriéroux, C., ARCH Models and Financial Applications (New York: Springer Verlag, 1997).
- [19] L. Giraitis, R. Leipus and D. Surgailis, Recent advances in ARCH modelling. In A. Kirman and G. Teyssiere (eds), Long-Memory in Economics (Berlin: Springer, 2006), 3–38.
- [20] W. Greene, W., Econometric Analysis (7th ed.) (NJ: Prentice Hall, 2011).
- [21] P. Hall and Q. Yao, Inference in ARCH and GARCH models with heavy-tailed errors, *Econometrica*, **71**(2003), 285–317.
- [22] C. M. Hafner, GARCH modeling. In R. Meyers (Ed.), Encyclopedia of Complexity and Systems Science (New York: Springer Verlag 2008).
- [23] W.K. Li, S. Ling and M. McAleer, Recent theoretical results for time series models with GARCH errors, *Journal of Economic Surveys*, **16** (2002), 245–269.
- [24] D. X. Li and H. J. Turtle, Semiparametric ARCH Models: an estimating function approach, *Journal of Business Economics & Statistics*, **18** (2000), 174–186.
- [25] T. Mikosch and T. Straumann, Stable limits of martingale transforms with application to the estimation of GARCH parameters, *Annals of Statistics*, **34**(2006), 493–522.
- [26] C. R. Nelson and C. I. Plosser, Trends and random walks in macroeconomic time series, *Journal of Monetary Economics*, **10**(1982), 139–162.
- [27] A.K. Gupta, Saleh, A.K. Md. E. and P.K. Sen, Improved Estimation in a Contingency Table: Independence Structure, *Journal of the American Statistical Association*, **84**(1989), 525-532.
- [28] A. Trapletti, tseries: Time series analysis and computational finance. R package version 0.10–25 (2011), URL <http://CRAN.R-project.org/package=tseries>
- [29] D.F. Nicholls and B.G. Quinn, Random coefficient autoregressive models: An introduction, Lecture Notes in Statistics, 11.(New York: Springer Verlag, 1982).
- [30] C.J. O'Donnell and V. Rayner, Imposing stationarity constraints on the parameters of ARCH and GARCH models, in Siddhartha Chib, William Griffiths, Gary Koop, Dek Terrell (ed.) Bayesian Econometrics (Advances in Econometrics, Volume 23), 2008, Emerald Group Publishing Limited, pp.545-566.

- [31] A. K. Md. E. Saleh, Theory of preliminary test and stein-type estimation with applications, (New York: John Wiley, 2006).
- [32] D. Straumann and T. Mikosch, Quasi-MLE in heteroscedastic times series: a stochastic recurrence equations approach, *Annals of Statistics*, **34**(2006), 2449–95.
- [33] D. Straumann Estimation in Conditionally Heteroscedastic Time Series Models (Heidelberg: Springer-Verlag, 2005).
- [34] T. Teräsvirta, An introduction to univariate GARCH models. In Handbook of Financial Time Series, (Heidelberg: Springer-Verlag, 2009), pp. 17-42.
- [35] R. S. Tsay, Analysis of Financial Time Series, 3rd Edition (Hoboken, NJ: John Wiley & Sons, 2010).
- [36] M. Verbeek, A Guide to Modern Econometrics, 3rd Edition (Hoboken, NJ: John Wiley & Sons, 2008).
- [37] E. Xekalaki and S. Degiannakis, ARCH Models for Financial Applications (New York: John Wiley & Sons, 2010).

Appendix A. Appendices

The following Lemma will help us to derive the results of Theorems 3.1.2 and 3.1.3.

Lemma 1. Let \mathbf{z} be a k_2 vector that follows normal distribution with mean $\boldsymbol{\mu}_{k_2}$ vector and covariance matrix $\boldsymbol{\Sigma}_{k_2}$ as $\mathbf{z} \sim N_{k_2}(\boldsymbol{\mu}_{k_2}, \boldsymbol{\Sigma}_{k_2})$. Then, for any ϕ Borel measurable and real-valued integrable function, we have

$$E[\mathbf{z}\phi(\mathbf{z}'\mathbf{z})] = \boldsymbol{\mu}_{k_2}E[\phi(\chi_{k_2+2}^2(\Delta))]$$

$$E[\mathbf{z}\mathbf{z}'\phi(\mathbf{z}'\mathbf{z})] = \boldsymbol{\Sigma}_{k_2}E[\phi(\chi_{k_2+2}^2(\Delta))] + \boldsymbol{\mu}_{k_2}\boldsymbol{\mu}'_{k_2}E[\phi(\chi_{k_2+4}^2(\Delta))],$$

where $\Delta = \boldsymbol{\mu}'_{k_2}\boldsymbol{\Sigma}_{k_2}^{-1}\boldsymbol{\mu}_{k_2}$. The proof can be found in [2].

Proof of Theorems 3.1.2 and 3.1.3

To prove the theorems, we use $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \mathbf{0})'$. Under the local alternative $K_{(n)}$ and the assumed regularity conditions, we have as $n \rightarrow \infty$

$$\boldsymbol{\eta}_1 = \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_{11.2}^{-1})$$

$$\boldsymbol{\eta}_2 = \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE}) \xrightarrow{\mathcal{L}} N(\boldsymbol{\gamma}, \mathbf{B})$$

$$\boldsymbol{\eta}_3 = \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(-\boldsymbol{\gamma}, \boldsymbol{\Phi})$$

where $\boldsymbol{\gamma} = \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\boldsymbol{\delta}$, $\mathbf{B} = \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\mathbf{I}_{11.2}^{-1}\mathbf{I}_{21}\mathbf{I}_{11}^{-1}$, $\boldsymbol{\Phi} = \mathbf{I}_{22.1}^{-1} - \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\mathbf{I}_{22.1}^{-1}\mathbf{I}_{21}\mathbf{I}_{11}^{-1}$

Proof of Theorem 3.1.2. Here, we provide the proof of the bias expressions. It is obvious that $ADB(\hat{\boldsymbol{\beta}}^{FE}) = 0$. The ADB of the reduced, shrinkage, and positive shrinkage estimators are as follows:

$$ADB(\hat{\boldsymbol{\beta}}^{RE}) = E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta}) \right]$$

$$= E \left[\lim_{n \rightarrow \infty} n(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta} - \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\boldsymbol{\delta}/\sqrt{n}) \right] = -\mathbf{I}_{11}^{-1}\mathbf{I}_{12}\boldsymbol{\delta} = -\boldsymbol{\gamma}.$$

$$ADB(\hat{\boldsymbol{\beta}}^{SE}) = E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{SE} - \boldsymbol{\beta}) \right]$$

$$= E \left[\lim_{n \rightarrow \infty} \left(\sqrt{n}(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta}) - (\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})\mathbf{v}\hat{D}_n^{-1} \right) \right] = -\mathbf{v}\mathbf{I}_{11}^{-1}\mathbf{I}_{12}\boldsymbol{\delta}E(Z_1), \text{ by Lemma 1.}$$

$$\begin{aligned}
 ADB(\hat{\boldsymbol{\beta}}^{PSE}) &= E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{PSE} - \boldsymbol{\beta}) \right] \\
 &= E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{SE} - \boldsymbol{\beta}) - \lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})(1 - v\hat{D}_n^{-1})I(\hat{D}_n < v) \right] \\
 &= ABD(\hat{\boldsymbol{\beta}}^{SE}) - E \left[\boldsymbol{\eta}_2(1 - v\hat{D}_n^{-1})I(\hat{D}_n < v) \right] \\
 &= ABD(\hat{\boldsymbol{\beta}}^{SE}) - \boldsymbol{\gamma}E \left[(1 - vZ_1)I(vZ_1 > 1) \right], \text{ by Lemma 1} \\
 &= ABD(\hat{\boldsymbol{\beta}}^{SE}) - \boldsymbol{\gamma}[H_{v+4}(v, \Delta) - vE(Z_1I(Z_1 > 1))],
 \end{aligned}$$

where $\boldsymbol{\gamma} = \mathbf{I}_{11}^{-1}\mathbf{I}_{12}\boldsymbol{\delta}$, $v = k_2 - 2$, and $Z_1 = \chi_{k_2+2}^{-2}(\Delta)$.

Proof of Theorem 3.1.3. In order to prove this theorem, we first derive the asymptotic covariance matrices for the four estimators. The covariance matrix $\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}^*)$ of any estimator $\hat{\boldsymbol{\beta}}^*$ is defined as:

$$\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}^*) = E \left[\lim_{n \rightarrow \infty} n(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})' \right].$$

First, we derive the covariance matrices of FE and RE:

$$\begin{aligned}
 \boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}^{FE}) &= E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta})\sqrt{n}(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta})' \right] \\
 &= E(\boldsymbol{\eta}_1\boldsymbol{\eta}'_1) = \text{Var}(\boldsymbol{\eta}_1) + E(\boldsymbol{\eta}_1)E(\boldsymbol{\eta}'_1) = \text{Var}(\boldsymbol{\eta}_1) = \mathbf{I}_{11.2}^{-1}.
 \end{aligned}$$

$$\begin{aligned}
 \boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}^{RE}) &= E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta})\sqrt{n}(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta})' \right] \\
 &= E(\boldsymbol{\eta}_3\boldsymbol{\eta}'_3) = \text{Var}(\boldsymbol{\eta}_3) + E(\boldsymbol{\eta}_3)E(\boldsymbol{\eta}'_3) = \boldsymbol{\Phi} + \boldsymbol{\gamma}\boldsymbol{\gamma}'.
 \end{aligned}$$

Second, we derive the covariance matrices of the shrinkage estimators:

$$\begin{aligned}
 \boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}^{SE}) &= E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{SE} - \boldsymbol{\beta})\sqrt{n}(\hat{\boldsymbol{\beta}}^{SE} - \boldsymbol{\beta})' \right] \\
 &= E \left[\lim_{n \rightarrow \infty} \sqrt{n} \left(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta} + v\hat{D}_n^{-1}(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE}) \right) \right. \\
 &\quad \left. \sqrt{n} \left(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta} + v\hat{D}_n^{-1}(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE}) \right)' \right] \\
 &= E \left[\lim_{n \rightarrow \infty} (\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta})' - 2v\hat{D}_n^{-1}(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})(\hat{\boldsymbol{\beta}}^{FE} - \boldsymbol{\beta})' \right. \\
 &\quad \left. + v^2\hat{D}_n^{-2}(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})' \right] \\
 &= E \left[\boldsymbol{\eta}_1\boldsymbol{\eta}'_1 - 2v\hat{D}_n^{-1}\boldsymbol{\eta}_2\boldsymbol{\eta}'_1 + v^2\hat{D}_n^{-2}\boldsymbol{\eta}_2\boldsymbol{\eta}'_2 \right].
 \end{aligned}$$

Using the conditional mean of bivariate normal, the second term of $\boldsymbol{\Gamma}(\hat{\boldsymbol{\beta}}^{SE})$ without $-2v$ is equal to

$$\begin{aligned}
 E \left[\boldsymbol{\eta}_2\boldsymbol{\eta}'_1 \right] &= E \left[E(\boldsymbol{\eta}_2\boldsymbol{\eta}'_1\hat{D}_n^{-1} | \boldsymbol{\eta}_2) \right] \\
 &= E \left[\boldsymbol{\eta}_2 E(\boldsymbol{\eta}'_1\hat{D}_n^{-1} | \boldsymbol{\eta}_2) \right] = E \left[\boldsymbol{\eta}_2 (E(\boldsymbol{\eta}_1) + (\boldsymbol{\eta}_2 - \boldsymbol{\delta}))' \hat{D}_n^{-1} \right] \\
 &= E \left[\boldsymbol{\eta}_2(\boldsymbol{\eta}_2 - \boldsymbol{\delta})' \hat{D}_n^{-1} \right] = E \left[\boldsymbol{\eta}_2\boldsymbol{\eta}'_2\hat{D}_n^{-1} \right] - E \left[\boldsymbol{\eta}_2\boldsymbol{\delta}'\hat{D}_n^{-1} \right] \\
 &= \text{Var}(\boldsymbol{\eta}_2)E(Z_1) + E(\boldsymbol{\eta}_2)E(\boldsymbol{\eta}_2)'E(Z_2) - E(\boldsymbol{\eta}_2)\boldsymbol{\delta}E(Z_1), \text{ by Lemma 1} \\
 &= \mathbf{B}E(Z_1) + \boldsymbol{\gamma}\boldsymbol{\gamma}'E(Z_2) - \boldsymbol{\gamma}\boldsymbol{\delta}'E(Z_1)
 \end{aligned}$$

where $Z_2 = \chi_{k_2+4}^{-2}(\Delta)$. Therefore,

$$\begin{aligned} \Gamma(\hat{\boldsymbol{\beta}}^{SE}) &= \mathbf{I}_{11.2}^{-1} - 2\nu[\mathbf{B}E(Z_1) + \boldsymbol{\gamma}\boldsymbol{\gamma}'E(Z_2) - \boldsymbol{\gamma}\boldsymbol{\gamma}'E(Z_1)] \\ &\quad + \nu^2[\mathbf{B}E(Z_1^2) + \boldsymbol{\gamma}\boldsymbol{\gamma}'E(Z_2^2)] \\ &= \mathbf{I}_{11.2}^{-1} + \nu\mathbf{B}[\nu E(Z_1^2) - 2E(Z_1)] + \nu\boldsymbol{\gamma}\boldsymbol{\gamma}'[\nu E(Z_2^2) + 2E(Z_1) - 2E(Z_2)] \\ &= \Gamma(\hat{\boldsymbol{\beta}}^{FE}) - \nu\mathbf{B}[2\Delta E(Z_2^2) + E(Z_1)] + \nu(\nu + 4)\boldsymbol{\gamma}\boldsymbol{\gamma}'E(Z_2^2). \end{aligned}$$

$$\begin{aligned} \text{Again, } \Gamma(\hat{\boldsymbol{\beta}}^{PSE}) &= E \left[\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\boldsymbol{\beta}}^{PSE} - \boldsymbol{\beta}) \sqrt{n}(\hat{\boldsymbol{\beta}}^{PSE} - \boldsymbol{\beta})' \right] \\ &= \Gamma(\hat{\boldsymbol{\beta}}^{SE}) - 2E \left[\lim_{n \rightarrow \infty} n \left((\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta})'(1 - \nu\hat{D}_n^{-1})I(\hat{D}_n < \nu) \right) \right] \\ &\quad + E \left[\lim_{n \rightarrow \infty} n(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})'(1 - \nu\hat{D}_n^{-1})^2I(\hat{D}_n < \nu) \right] \\ &= \Gamma(\hat{\boldsymbol{\beta}}^{SE}) - 2E \left[\boldsymbol{\eta}_2\boldsymbol{\eta}_3'(1 - \nu\hat{D}_n^{-1})I(\hat{D}_n < \nu) \right] \\ &\quad - E \left[\boldsymbol{\eta}_2\boldsymbol{\eta}_2'(1 - \nu\hat{D}_n^{-1})^2I(\hat{D}_n < \nu) \right]. \end{aligned}$$

Consider the second term without -2 and use the rule of conditional expectation

$$\begin{aligned} &E \left[\boldsymbol{\eta}_2\boldsymbol{\eta}_3'(1 - \nu\hat{D}_n^{-1})I(\hat{D}_n < \nu) \right] \\ &= E \left[\boldsymbol{\eta}_2 E \left(\boldsymbol{\eta}_3'(1 - \nu\hat{D}_n^{-1})I(\hat{D}_n < \nu) \mid \boldsymbol{\eta}_2 \right) \right] \\ &= E \left[\boldsymbol{\eta}_2(-\boldsymbol{\gamma}' + \mathbf{0} \times \boldsymbol{\Phi}(\boldsymbol{\eta}_2 - \boldsymbol{\gamma})')(1 - \nu\hat{D}_n^{-1})I(\hat{D}_n < \nu) \right] \\ &= -E \left[\boldsymbol{\eta}_2\boldsymbol{\gamma}'(1 - \nu\hat{D}_n^{-1})I(\hat{D}_n < \nu) \right] \\ &= -\boldsymbol{\gamma}\boldsymbol{\gamma}'E \left[(1 - \nu Z_1)I(\nu Z_1 > 1) \right], \text{ by Lemma 1.} \end{aligned}$$

Therefore,

$$\begin{aligned} \Gamma(\hat{\boldsymbol{\beta}}^{PSE}) &= \Gamma(\hat{\boldsymbol{\beta}}^{SE}) + 2\boldsymbol{\gamma}\boldsymbol{\gamma}'E \left[(1 - \nu Z_1)I(\nu Z_1 > 1) \right] \\ &\quad - \mathbf{B}E \left[(1 - \nu Z_1)^2I(\nu Z_1 > 1) \right] - \boldsymbol{\gamma}\boldsymbol{\gamma}'E \left[(1 - \nu Z_2)^2I(\nu Z_2 > 1) \right]. \end{aligned}$$

The proof of Theorem 3.1.3 now follows using (3.2) and the above covariance matrices.