

Structural Breaks in Seemingly Unrelated Regression Models

Shahnaz Parsaeian*

Department of Economics, University of Kansas, USA

E-mail: sh.parsaeian@ku.edu

July 18, 2023

Abstract

This paper develops an efficient Stein-like shrinkage estimator for estimating slope parameters under structural breaks in seemingly unrelated regression models, which is then used for forecasting. The proposed method is a weighted average of two estimators: a restricted estimator that estimates the parameters under the restriction of no break in the coefficients, and an unrestricted estimator that considers break points and estimates the parameters using the observations within each regime. It is established that the asymptotic risk of the Stein-like shrinkage estimator is smaller than that of the unrestricted estimator, which is the method typically used to estimate the slope coefficients under structural breaks. Furthermore, this paper proposes an averaging minimal mean squared error estimator in which the averaging weight is derived by minimizing its asymptotic risk. The superiority of the two proposed estimators over the unrestricted estimator in terms of the mean squared forecast errors are also derived. Further, analytical comparison between the asymptotic risks of the proposed estimators is provided. Insights from the theoretical analysis are demonstrated in Monte Carlo simulations, and through an empirical example of forecasting output growth rates of G7 countries.

Keywords: Forecasting; Seemingly unrelated regression; Structural breaks; Stein-like shrinkage estimator; Minimal mean squared error estimator

JEL Classification: C13, C23, C52, C53

*Address correspondence to: Shahnaz Parsaeian, sh.parsaeian@ku.edu.

1 Introduction

Seemingly unrelated regression (SUR) models include several individual units that are linked by the fact that their disturbances are correlated. The correlation among the equation disturbances could come from several sources, such as correlated technological shocks, tax policy changes, or credit crunches, which may affect all individual units together. Consequently, such models have been broadly used in econometrics and applied works. Using the correlation among the equations in SUR models, one can improve the efficiency of estimates compared to those obtained by an equation-by-equation least squares, see [Zellner \(1962\)](#). However, standard estimators in SUR are arguably restrictive, as they assume that the slope coefficients are constant over time. As an economy may experience an unexpected shock across time (such as oil price shocks, financial crises, or technological shocks), and such a shock is likely to have impact on economic variables simultaneously, it is important to consider structural breaks in SUR models. This is a critical issue to consider because structural breaks are an important source of forecast failures in macroeconomics and finance as documented by [Pesaran and Timmermann \(2002, 2007\)](#), [Pesaran et al. \(2006\)](#), [Clements and Hendry \(2006, 2011\)](#), [Giacomini and Rossi \(2009\)](#), [Inoue and Rossi \(2011\)](#), [Pesaran and Pick \(2011\)](#), [Pesaran et al. \(2013\)](#), [Rossi \(2013\)](#), [Barnett et al. \(2016\)](#), and [Lee et al. \(2022\)](#), among others. To deal with this challenge, this paper considers structural breaks in SUR models, and develops estimation and forecasting methods in this framework.

There has been a recent increase in literature concerning the estimation and tests of common breaks in panel data models with a main focus on the detection of break points and their asymptotic properties. Large N and T panels involves nuisance parameters that increase at a quadratic rate because the cross-section dimension of the panel is allowed to rise. One solution to deal with this issue is to restrict the covariance matrix of the errors using a common factor specification with a fixed number of unobserved factors, as discussed in [Bai and Ng \(2002\)](#), [Coakley et al. \(2002\)](#), and [Phillips and Sul \(2003\)](#), to mention a few. Despite the considerable attention to the detection of break points, only a few studies have attempted to forecast panel data models, see [Smith and Timmermann \(2018\)](#), [Smith \(2018\)](#), and [Liu \(2022\)](#) which use a Bayesian framework. Our paper differs from their work in that it considers the so-called frequentist model averaging framework, which relies on the data on hand. Further, the main purpose of our paper is to find a more accurate

estimate of the slope parameters in the sense of mean squared errors, and also optimal forecast in the sense of mean squared forecast errors (MSFE) in SUR models.¹

This paper develops two averaging estimators in SUR models for estimating the slope coefficients and forecasting under structural breaks when the cross-section dimension is fixed while the time dimension is allowed to increase without bounds. It is important to allow for large T because, for example, technological changes or policy implementations are likely to happen over long time horizons. Moreover, the model allows for the cross-sectional dependence to gain from the correlation across individuals, which is not applicable in a univariate time-series model. Ignoring cross-sectional dependence of errors can have serious consequences and can result in misleading inference and even inconsistent estimators, depending on the extent of the cross-sectional dependence. This cross-correlations could arise due to omitted common effects, spatial effects, or as a result of interactions within socioeconomic networks, as discussed in [Chudick and Pesaran \(2015\)](#).

Our first proposed estimator is a weighted average of an unrestricted estimator and a restricted estimator in which the averaging weight takes the form of the James-Stein weight, cf. [Stein \(1956\)](#) and [James and Stein \(1961\)](#).² The restricted estimator is built under the restriction of no break in the coefficients, that is, the coefficients across different regimes are restricted to be the same as if there were no structural break. Therefore, the restricted estimator will be biased when there is a break, but is the most efficient one. However, the unrestricted estimator considers the break points and only uses the observations within each regime to estimate the coefficients. Thus, the unrestricted estimator is consistent but less efficient. The weighted average of the unrestricted estimator and the restricted estimator therefore balances the trade-off between the bias and the variance efficiency depending on the magnitude of the break. We establish the asymptotic risk for the Stein-like shrinkage estimator and show that its asymptotic risk is smaller than that of the unrestricted estimator, which is the common method for estimating the slope coefficients under structural breaks. We note that this out-performance comes from the efficiency of the restricted estimator in which it exploits the observations in neighboring regime(s). Therefore, the proposed estimator provides a better estimate and ultimately a better forecast in the sense of MSFE.

¹A SUR model can be seen as a panel data model in which the cross-section dimension (N) is small and the time series dimension (T) is large.

²See also [Massoumi \(1978\)](#), [Hansen \(2016, 2017\)](#) and [Mehrabani and Ullah \(2020\)](#) for applying the Stein-type shrinkage weight in different contexts.

The second proposed estimator is a weighted average of the unrestricted estimator and the restricted estimator in which the averaging weight is derived by minimizing the asymptotic risk. This estimator is called the minimal mean squared error estimator. We derive the asymptotic risk of the minimal mean squared error estimator and show that it is smaller than the asymptotic risk of the unrestricted estimator. In addition, we analytically compare its asymptotic risk with that of the Stein-like shrinkage estimator. The results show the out-performance of the Stein-like shrinkage estimator over the minimal mean squared error estimator under small break sizes. However, for large break sizes or large number of regressors, both estimators have almost equal performance. It is noteworthy that both of the proposed estimators uniformly outperform the unrestricted estimator, in the sense of having a smaller asymptotic risk, for any break size and break point. When choosing between the two proposed estimators, based on the analytical and numerical results, we recommend the Stein-like shrinkage estimator as it uniformly outperforms or performs as well as the minimal mean squared error estimator.³

We conduct a Monte Carlo simulation study to evaluate the performance of the proposed averaging estimators. The results confirm the theoretically expected improvements in the Stein-like shrinkage estimator and the minimal mean squared error estimator over the unrestricted estimator for any break size and break point. Further, we provide an empirical analysis of forecasting output growth rates of G7 countries using quarterly data from 1995:Q1 to 2016:Q4. Our empirical results show the benefits of using the proposed averaging estimators over the unrestricted estimator and other alternative estimators.

The paper is structured as follows. Section 2 presents the SUR model under the structural break. Section 3 introduces the Stein-like shrinkage estimator, as well as its asymptotic distribution and asymptotic risk. This section also introduces the minimal mean squared error estimator, and compares the asymptotic risk of this estimator with that of the Stein-like shrinkage estimator. Section 4 reports the Monte Carlo simulation. Section 5 presents the empirical analysis. Section 6 concludes the paper. Detailed proofs are provided in Appendix A.

³In panel data and multivariate equation models, shrinkage estimator and model averaging techniques have been considered on the issue of how to model potentially heterogeneous parameters across individual units, see [Maddala et al. \(2001\)](#), [Wang et al. \(2019\)](#), [Mehrabani and Ullah \(2020\)](#).

2 The model

Consider the following heterogeneous seemingly unrelated regressions model

$$y_{i,t} = x'_{i,t}\beta_i + u_{i,t}, \quad \text{for } i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where $x_{i,t}$ is a $k \times 1$ vector of regressors, and $u_{i,t}$ is the error term with zero mean that is allowed to have cross-sectional dependence as well as heteroskedasticity. The vector of coefficients, β_i , and the variance of the error term, $u_{i,t}$, are subject to a common break across individuals at time T_1 , where $b_1 \equiv T_1/T \in (0, 1)$, such that

$$\beta_i = \begin{cases} \beta_{i(1)} & \text{for } t = 1, \dots, T_1, \\ \beta_{i(2)} & \text{for } t = T_1 + 1, \dots, T. \end{cases} \quad (2)$$

Let Y_i, X_i, U_i denote the stacked data and errors for individuals $i = 1, \dots, N$ over the time period observed. Then,

$$Y_i = X_i\beta_i + U_i, \quad (3)$$

where $Y_i = (y'_{i(1)}, y'_{i(2)})'$ is a vector of $T \times 1$ dependent variable, with $y_{i(1)} = (y_{i,1}, \dots, y_{i,T_1})'$ and $y_{i(2)} = (y_{i,T_1+1}, \dots, y_{i,T})'$. Also, $X_i = \text{diag}(x_{i(1)}, x_{i(2)})$ is a $T \times 2k$ diagonal matrix in which $x_{i(1)} = (x_{i,1}, \dots, x_{i,T_1})'$, and $x_{i(2)} = (x_{i,T_1+1}, \dots, x_{i,T})'$. $\beta_i = (\beta'_{i(1)}, \beta'_{i(2)})'$ is a $2k \times 1$ vector of the slope coefficients, and $U_i = (u_{i(1)}, u_{i(2)})'$ is a $T \times 1$ vector of error terms with $u_{i(1)} = (u_{i,1}, \dots, u_{i,T_1})'$ and $u_{i(2)} = (u_{i,T_1+1}, \dots, u_{i,T})'$. It is convenient to stack the equations over N individuals as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 & \mathbf{0} & \dots & \mathbf{0} \\ 0 & X_2 & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}, \quad (4)$$

or compactly in matrix notation as

$$\underset{NT \times 1}{Y} = \underset{NT \times 2Nk}{X} \underset{2Nk \times 1}{b} + \underset{NT \times 1}{U}. \quad (5)$$

We make the following assumptions:

Assumption 1: $\{(\tilde{x}_t, \tilde{u}_t)\}$ is an i.i.d. sequence where $\tilde{x}_t = \text{diag}(x'_{1,t}, \dots, x'_{N,t})$, $\tilde{u}_t = (u_{1,t}, \dots, u_{N,t})'$, and $\mathbb{E}(\tilde{u}_t | \tilde{x}_t) = 0$.

Assumption 2: The disturbances are heteroscedastic across regimes, and are uncorrelated across time but correlated across individual equations,

$$\text{Var}(U) \equiv \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \dots & \Omega_{1N} \\ \Omega_{21} & \Omega_{22} & \dots & \Omega_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{N1} & \Omega_{N2} & \dots & \Omega_{NN} \end{bmatrix} \quad (6)$$

$$\text{where } \Omega_{ij} = \begin{bmatrix} \text{cov}(u_{i(1)}, u_{j(1)}) & \mathbf{0} \\ \mathbf{0} & \text{cov}(u_{i(2)}, u_{j(2)}) \end{bmatrix} = \begin{bmatrix} \sigma_{ij(1)} I_{T_1} & \mathbf{0} \\ \mathbf{0} & \sigma_{ij(2)} I_{(T-T_1)} \end{bmatrix}.$$

Assumption 3: The product moment matrix $(\frac{X'\Omega^{-1}X}{T})$ has full rank, and tends to a finite non-singular matrix as $T \rightarrow \infty$.

We note that Assumption 1 implies that the observations are not correlated across time so that conventional central limit theory applies. Assumption 2 implies the cross-sectional dependence across individuals. Assumption 3 guarantees that the Generalized least squares (GLS) estimation method used in Section 3 is uniquely defined.

We note that the method of break point estimation is based on the least-squares principle discussed in Bai and Perron (2003) for a univariate time-series model, and the extensions to the multivariate regression setting is described in Qu and Perron (2007). It has been shown that the estimated break fraction, \hat{b}_1 , converges to its true value, b_1 , at a rate that is fast enough not to affect the \sqrt{T} consistency of the estimated parameters asymptotically, see Qu and Perron (2007).

3 The proposed averaging estimators

This section introduces the Stein-like shrinkage estimator and the minimal mean squared error estimator. We also provide an analytical comparison of the two proposed estimators.

3.1 Stein-like shrinkage estimator

For the estimation of the slope parameters in model (5), we propose the Stein-like shrinkage estimator that can reduce the estimation error under structural breaks. Our proposed shrinkage estimator denoted by $\widehat{\mathbf{b}}_w$ is

$$\widehat{\mathbf{b}}_w = w_T \widehat{\mathbf{b}}_{ur} + (1 - w_T) \widehat{\mathbf{b}}_r, \quad (7)$$

where $\widehat{\mathbf{b}}_r$ is called the restricted estimator, which is under the restriction of no break in the coefficients. Thus, it estimates the parameters by ignoring the break. Therefore, the restricted estimator is biased when there is a break while it is efficient. The unrestricted estimator, $\widehat{\mathbf{b}}_{ur}$, estimates the coefficients by considering the common break point across all individuals; so this is the unbiased estimator but less efficient. As a result, the weighted average of the restricted and unrestricted estimators trades-off between the bias and variance efficiency. The shrinkage weight takes the form of

$$w_T = \left(1 - \frac{\tau}{D_T}\right)_+, \quad (8)$$

where weight takes the form of positive part function, $(x)_+ = x \mathbf{I}(x \geq 0)$. Also, τ is the shrinkage parameter that controls the degree of shrinkage, and D_T is a weighted quadratic loss equal to

$$D_T = T(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)' \mathbb{W}(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r). \quad (9)$$

The loss function in (9) measures the distance between the restricted and unrestricted estimators, with \mathbb{W} an arbitrary symmetric positive definite weight matrix. For example, when \mathbb{W} is equal to the inverse of the variance of the unrestricted estimator, D_T is a Wald-type statistic. This is also an appropriate weight choice as it simplifies the theoretical calculations and makes the loss function invariant to the rotations of the coefficient vector \mathbf{b} .

The idea behind the Stein-like shrinkage estimator in (7) is that for a large break size (a large value of D_T), the shrinkage estimator assigns a higher weight to the unrestricted estimator and a lower weight to the restricted estimator, since the restricted estimator adds large bias under the large break size. However, for a small break size (a small value of D_T), the shrinkage estimator gives more weight to the restricted estimator to gain from its efficiency. In other words, depending on the magnitude of the break, the shrinkage estimator assigns appropriate weight to each of the restricted

and unrestricted estimators to balance the trade-off between the bias and variance efficiency.

As mentioned earlier, the unrestricted estimator considers the common break and estimates the coefficients (pre- and post-break coefficients) using observations within each regime separately. This estimator is typically used for estimating the slope coefficients under structural break. Using the feasible GLS method for estimating the slope coefficients in (5), we have

$$\widehat{\mathbf{b}}_{ur} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}Y, \quad (10)$$

where $\widehat{\Omega}$ is the estimate of the unknown parameter Ω defined in (6). We obtain the estimates of the elements in Ω by using the OLS residuals for each equation. In practice, $\widehat{\sigma}_{ij(1)} = \frac{\widehat{u}'_{i(1)}\widehat{u}_{j(1)}}{T_1-k}$ is the estimate of the elements of Ω_{ij} , where $\widehat{u}_{i(1)} = y_{i(1)} - x_{i(1)}\widehat{\beta}_{i(1)}^{ols}$, and $\widehat{\beta}_{i(1)}^{ols} = (x'_{i(1)}x_{i(1)})^{-1}x'_{i(1)}y_{i(1)}$ is the pre-break estimator. Similarly, $\widehat{\sigma}_{ij(2)} = \frac{\widehat{u}'_{i(2)}\widehat{u}_{j(2)}}{T-T_1-k}$, where $\widehat{u}_{i(2)} = y_{i(2)} - x_{i(2)}\widehat{\beta}_{i(2)}^{ols}$, and $\widehat{\beta}_{i(2)}^{ols} = (x'_{i(2)}x_{i(2)})^{-1}x'_{i(2)}y_{i(2)}$ is the post-break estimator, for $i = 1, \dots, N$.

Alternatively, one can estimate the slope coefficients by imposing a restriction on the parameters, $R\mathbf{b} = \mathbf{r}_{p \times 1}$ in which R is a $p \times 2Nk$ restriction matrix with rank p . This is called the restricted estimator. By applying the feasible GLS method in (5), under the restriction $R\mathbf{b} = \mathbf{r}$, we have

$$\widehat{\mathbf{b}}_r = \widehat{\mathbf{b}}_{ur} - (X'\widehat{\Omega}^{-1}X)^{-1}R' \left[R (X'\widehat{\Omega}^{-1}X)^{-1}R' \right]^{-1} (R\widehat{\mathbf{b}}_{ur} - \mathbf{r}). \quad (11)$$

Under the assumption of no break in the slope coefficients, $R\mathbf{b} = \mathbf{0}$ where

$$R = \begin{bmatrix} I_k & -I_k & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_k & -I_k & \mathbf{0} & \dots & \mathbf{0} \\ & & & & \vdots & & \\ \mathbf{0} & \dots & \mathbf{0} & I_k & -I_k & & \end{bmatrix}, \quad (12)$$

and $p = Nk$.

3.1.1 Asymptotic results for the Stein-like shrinkage estimator

Our analysis is asymptotic as the time series dimension $T \rightarrow \infty$ while the number of individual units, N , is fixed. Under the local alternative assumption, consider the parameter sequences of the form

$$\mathbf{b} = \mathbf{b}_0 + \frac{\mathbf{h}}{\sqrt{T}}, \quad (13)$$

where \mathbf{b} is the true parameter value, \mathbf{b}_0 is the slope coefficients under the assumption of no break ($\beta_{i(1)} = \beta_{i(2)}$, for $i = 1, \dots, N$), and \mathbf{h} shows the magnitude of the break size in the coefficients. Thus, for any fixed \mathbf{h} , the break size \mathbf{h}/\sqrt{T} converges to zero as the sample size increases. We allow the size of the break to be different across individuals. That means that for each individual $i = 1, \dots, N$, we have $\beta_{i(1)} = \beta_{i(2)} + \frac{\delta_i}{\sqrt{T}}$, or generally

$$R\mathbf{b} = \begin{bmatrix} \beta_{1(1)} - \beta_{1(2)} \\ \beta_{2(1)} - \beta_{2(2)} \\ \vdots \\ \beta_{N(1)} - \beta_{N(2)} \end{bmatrix} = \frac{1}{\sqrt{T}} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{bmatrix} = \frac{\boldsymbol{\delta}}{\sqrt{T}}, \quad (14)$$

where $\boldsymbol{\delta} = (\delta'_1, \dots, \delta'_N)'$ is a vector of $Nk \times 1$, and the restriction matrix R is defined in (12). We note that $R\mathbf{h} = \boldsymbol{\delta}$, and $R\mathbf{b}_0 = \mathbf{0}$ under the assumption of no break. In the following theorem, we derive the asymptotic distributions of the unrestricted estimator, restricted estimator, and the Stein-like shrinkage estimator.

Theorem 1: *Under assumptions 1-3, along the sequences (13), the asymptotic distribution of the unrestricted estimator is*

$$\sqrt{T}(\widehat{\mathbf{b}}_{ur} - \mathbf{b}) \xrightarrow{d} Z \sim N(0, V_{ur}), \quad (15)$$

where $V_{ur} = \left(\mathbb{E} \left(\frac{X'\Omega^{-1}X}{T} \right) \right)^{-1}$ is the variance of the unrestricted estimator, and the asymptotic distribution of the restricted estimator is

$$\sqrt{T}(\widehat{\mathbf{b}}_r - \mathbf{b}) \xrightarrow{d} Z - V_{ur}R'(RV_{ur}R')^{-1}R(Z + \mathbf{h}). \quad (16)$$

In addition, the asymptotic distribution of the loss function, the weight and the Stein-like shrinkage estimator are

$$D_T = T(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)' \mathbb{W}(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r) \xrightarrow{d} (Z + \mathbf{h})' B (Z + \mathbf{h}), \quad (17)$$

$$w_T \xrightarrow{d} w(Z) = \left(1 - \frac{\tau}{(Z + \mathbf{h})' B (Z + \mathbf{h})}\right)_+, \quad (18)$$

$$\sqrt{T}(\widehat{\mathbf{b}}_w - \mathbf{b}) \xrightarrow{d} Z - w(Z) V_{ur} R' \left[R V_{ur} R' \right]^{-1} R (Z + \mathbf{h}), \quad (19)$$

where $B \equiv R' \left[R V_{ur} R' \right]^{-1} R V_{ur} \mathbb{W} V_{ur} R' \left[R V_{ur} R' \right]^{-1} R$ is a $2Nk \times 2Nk$ matrix. \blacksquare

See Appendix A.1 for the proof of this theorem. Theorem 1 shows that the asymptotic distribution of the Stein-like shrinkage estimator is a nonlinear function of the normal random vector Z , and the non-centrality parameter \mathbf{h} .

Remark 1: Lee et al. (2022a) consider a univariate time series model with a focus on forecasting. In this paper, we consider a SUR model and focus on improving the estimation of the slope coefficients within each regime, and also improving forecasts utilizing the improved estimator of the slope coefficients. The theoretical derivations of asymptotic distributions of the estimators provided in Theorem 1 (and subsequent theorems) are significantly different from those of Lee et al. (2022a) due to differences in the models. Besides, the restricted estimator considered in this paper is general which can be obtained using different linear restrictions on the model parameters. For example, under two breaks, it may be the case that slope coefficients in the first and third regimes are equal while different with the second regime. Another restriction would be to allow for partial changes in the slope coefficients, or if one is interested to shrink part of the slope coefficients. \square

When an estimator has an asymptotic distribution, $\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} \varpi$, we define its asymptotic risk as $\rho(\widehat{\beta}, \mathbb{W}) = \mathbb{E}(\varpi' \mathbb{W} \varpi)$. See Lehmann and Casella (1998). Using Theorem 1, we derive the asymptotic risk for the Stein-like shrinkage estimator. Theorem 2 provides the result.

Theorem 2: Under assumptions 1-3, for $0 < \tau \leq 2(Nk - 2)$, and for $\mathbb{W} = V_{ur}^{-1}$, the asymptotic risk for the Stein-like shrinkage estimator is

$$\rho(\widehat{\mathbf{b}}_w, \mathbb{W}) \leq \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) - \tau \left[\frac{2(Nk - 2) - \tau}{(c + 1)Nk} \right], \quad (20)$$

where $0 < c < \infty$. ■

See Appendix A.2 for the proof of this theorem.⁴ Theorem 2 shows that the asymptotic risk of the Stein-like shrinkage estimator is smaller than that of the unrestricted estimator. As the shrinkage parameter, τ , is unknown, we find it by minimizing the asymptotic risk. Theorem 3 shows the optimal value of the shrinkage parameter, denoted by τ_{opt}^* , and the associated asymptotic risk for the Stein-like shrinkage estimator.

Theorem 3: When $\mathbb{W} = V_{ur}^{-1}$ and $Nk > 2$, the optimal value of τ is

$$\tau_{opt}^* = Nk - 2, \quad (21)$$

where τ_{opt}^* is positive as long as $Nk > 2$. Also, the asymptotic risk for the Stein-like shrinkage estimator after plugging the τ_{opt}^* is

$$\rho(\widehat{\mathbf{b}}_w, \mathbb{W}) \leq \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) - \frac{(Nk - 2)^2}{(c + 1)Nk}. \quad (22)$$

Theorem 3 shows that the asymptotic risk of the Stein-like shrinkage estimator is smaller than that of the unrestricted estimator as long as $Nk > 2$. Besides, the out-performance of the Stein-like shrinkage estimator over the unrestricted estimator will increase for small break sizes, large number of individuals or large number of regressors. ■

3.2 Minimal mean squared error estimator

As an alternative to the shrinkage weight, we consider a weight between zero and one. Therefore, our second proposed estimator, the minimal mean squared error estimator, is

$$\widehat{\mathbf{b}}_\gamma = \gamma \widehat{\mathbf{b}}_r + (1 - \gamma) \widehat{\mathbf{b}}_{ur}, \quad \gamma \in [0, 1] \quad (23)$$

where $\widehat{\mathbf{b}}_{ur}$ and $\widehat{\mathbf{b}}_r$ are defined in (10) and (11), respectively. Using the asymptotic distribution of the estimators in (15) and (16), we derive the asymptotic risk for the minimal mean squared error

⁴The general result of Theorem 2 with any positive definite weight matrix \mathbb{W} is provided in Appendix A.2.

estimator as

$$\rho(\widehat{\mathbf{b}}_\gamma, \mathbb{W}) = \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + \gamma^2(\mathbf{h}'B\mathbf{h} + Nk) - 2\gamma Nk. \quad (24)$$

By minimizing the asymptotic risk with respect to γ in (24), the optimal value of the weight denoted by γ_{opt}^* is

$$\gamma_{opt}^* = \frac{Nk}{\mathbf{h}'B\mathbf{h} + Nk}, \quad (25)$$

which by plugging the unbiased estimator of its denominator we have

$$\widehat{\gamma}_{opt}^* = \frac{Nk}{\widehat{\mathbf{h}}'B\widehat{\mathbf{h}}} = \frac{Nk}{T(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)' \mathbb{W}(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)} = \frac{Nk}{D_T}, \quad (26)$$

where the second equality holds noting that $V_{ur}R'[RV_{ur}R']^{-1}R(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r) = (\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)$. Hence $T(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)'B(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r) = T(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)' \mathbb{W}(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)$.⁵

Remark 2: By comparing $\widehat{\gamma}_{opt}^*$ in (26) and τ_{opt}^*/D_T using (21), we see that the difference between the averaging weights of the Stein-like shrinkage estimator and the minimal mean squared error estimator is in their numerators. In other words, the optimal weight in (26) also depends on the distance between the restricted and unrestricted estimators. Basically, the idea behind the averaging estimator in (23) is similar to the Stein-like shrinkage estimator in (7). That is, when the difference between the restricted and unrestricted estimator is small, the minimal mean squared error estimator gives more weight to the restricted estimator, which is efficient under $R\mathbf{b} = \mathbf{0}$, and the opposite is true for the large distance between the two estimators. Therefore, the proposed averaging estimator in (23) is also a Stein-like shrinkage estimator that incorporates the trade-off between the bias and variance efficiency. \square

Given the optimal weight in (26), we derive the asymptotic risk for the minimal mean squared error estimator. Theorem 4 presents the results.

Theorem 4: *Under Assumptions 1-3, and given the optimal value of the weight in (26), for $\mathbb{W} = V_{ur}^{-1}$, the asymptotic risk of the minimal mean squared error estimator is*

$$\rho(\widehat{\mathbf{b}}_\gamma, \mathbb{W}) \leq \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) - \frac{Nk - 4}{(c + 1)}. \quad (27)$$

⁵Appendix A.3 obtains the optimal value of the weight, $\widehat{\gamma}_{opt}^*$, for any positive definite weight \mathbb{W} .

■

See Appendix A.4 for the proof of this theorem.⁶ Theorem 4 shows that the asymptotic risk of the minimal mean squared error estimator is smaller than that of the unrestricted estimator as long as $Nk > 4$.

Remark 3: Examining the results of Theorem 3 and Theorem 4, we find that the necessary condition for the outperformance of the Stein-like shrinkage estimator over the unrestricted estimator is $Nk > 2$, while the condition is $Nk > 4$ for the minimal mean squared error estimator, which is slightly a stronger condition. Furthermore, the difference between their risks, for $\mathbb{W} = V_{ur}^{-1}$, is

$$\rho(\hat{\mathbf{b}}_w, \mathbb{W}) - \rho(\hat{\mathbf{b}}_\gamma, \mathbb{W}) = \frac{-4}{(c+1)Nk} < 0. \quad (28)$$

Therefore, the Stein-like shrinkage estimator has a smaller asymptotic risk than the minimal mean squared error estimator. In addition, for small break sizes, we expect to see a better performance for the Stein-like shrinkage estimator relative to the minimal mean squared error estimator, in the sense of a smaller asymptotic risk. For a large break size, a large number of regressors, or a large number of individuals, we expect to see an equal performance between them.⁷ □

3.3 Forecasting under structural breaks

Generating accurate forecasts in the presence of structural breaks requires careful consideration of bias-variance trade-off. Our introduced Stein-like shrinkage estimator and the minimal mean squared error estimator consider the bias-variance trade-off in the estimation of parameters, and can thus be used to generate forecasts. As the true parameters that enter the forecasting period

⁶The general result of Theorem 4 with any positive definite weight matrix \mathbb{W} is provided in Appendix A.4.

⁷The difference between the asymptotic risks of the Stein-like shrinkage estimator and the minimal mean squared error estimator for any positive definite matrix \mathbb{W} is provided in Appendix A.4.

are the coefficients in the post-break sample, we define a $Nk \times 2Nk$ selection matrix G such that

$$G\mathbf{b} = \begin{bmatrix} \mathbf{0} & I_k & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I_k & \dots & \mathbf{0} \\ & & & & \vdots & \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & & I_k \end{bmatrix} \begin{bmatrix} \beta_{1(1)} \\ \beta_{1(2)} \\ \vdots \\ \beta_{N(1)} \\ \beta_{N(2)} \end{bmatrix} = \begin{bmatrix} \beta_{1(2)} \\ \beta_{2(2)} \\ \beta_{3(2)} \\ \vdots \\ \beta_{N(2)} \end{bmatrix}. \quad (29)$$

By multiplying G to the Stein-like shrinkage estimator, we have

$$G\widehat{\mathbf{b}}_w = w_T G\widehat{\mathbf{b}}_{ur} + (1 - w_T)G\widehat{\mathbf{b}}_r, \quad (30)$$

where $G\widehat{\mathbf{b}}_{ur}$ estimates the coefficients only by using the observations after the break point (also known as the post-break estimator), and $G\widehat{\mathbf{b}}_r$ is the restricted estimator under the assumption of no break in the model. The assumption of no break lines up with the fact that, with a small break, ignoring the break and estimating the coefficients using full-sample observations would result in a better forecast (lower MSFE), see [Boot and Pick \(2020\)](#).

Define the mean squared forecast error of the Stein-like shrinkage estimator as

$$\begin{aligned} \text{MSFE}(G\widehat{\mathbf{b}}_w) &= \mathbb{E} (y_{T+1} - x_{T+1}^* G\widehat{\mathbf{b}}_w)' (y_{T+1} - x_{T+1}^* G\widehat{\mathbf{b}}_w) \\ &= \mathbb{E} \left[(G\widehat{\mathbf{b}}_w - \beta_{(2)})' x_{T+1}^* x_{T+1}^* (G\widehat{\mathbf{b}}_w - \beta_{(2)}) \right] + \mathbb{E} (u_{T+1}' u_{T+1}) \\ &= \rho(G\widehat{\mathbf{b}}_w, x_{T+1}^* x_{T+1}^*) + \mathbb{E} (u_{T+1}' u_{T+1}), \end{aligned} \quad (31)$$

where $y_{T+1} = (y_{1,T+1}, \dots, y_{N,T+1})'$ is an $N \times 1$ vector of dependent variables at time $T + 1$, $x_{T+1}^* = \text{diag}(x_{1,T+1}, \dots, x_{N,T+1})$ is an $Nk \times N$ matrix of regressors, $\beta_{(2)} = (\beta_{1(2)}, \dots, \beta_{N(2)})'$, and $u_{T+1} = (u_{1,T+1}, \dots, u_{N,T+1})'$. Thus, by choosing \mathbb{W} accordingly, we use the asymptotic risk, $\rho(G\widehat{\mathbf{b}}_w, \mathbb{W})$, to approximate the first term on the right hand side of the MSFE. This along with $\mathbb{E} (u_{T+1}' u_{T+1})$ corresponds to the one-step-ahead MSFE. As the second term in (31), $\mathbb{E} (u_{T+1}' u_{T+1})$, does not depend on τ , minimizing the MSFE is equivalent to minimizing the asymptotic risk. Similarly, we define the MSFE for the minimal mean squared error estimator. Theorem 5 summarizes the results of MSFE for the estimators.

Theorem 5: Under Assumptions 1-3, the mean squared forecast errors of the Stein-like shrinkage estimator is

$$MSFE(\widehat{G\mathbf{b}}_w) \leq MSFE(\widehat{G\mathbf{b}}_{ur}) - \frac{[\text{tr}(\phi) - 2\lambda_{\max}(\phi)]^2}{(c+1)\text{tr}(\phi)}, \quad (32)$$

and the mean squared forecast error of the minimal mean squared errors estimator is

$$MSFE(\widehat{G\mathbf{b}}_\gamma) \leq MSFE(\widehat{G\mathbf{b}}_{ur}) - \frac{\text{tr}(\phi) - 4\lambda_{\max}(\phi)}{(c+1)}, \quad (33)$$

where $\phi \equiv \mathbb{W}^{1/2} G V_{ur} R' [R V_{ur} R']^{-1} R V_{ur} G' \mathbb{W}^{1/2}$, and $\mathbb{W} = x_{T+1}^* x_{T+1}^{*'}.$ ■

Theorem 5 shows that the MSFE of the Stein-like shrinkage estimator is smaller than that of the unrestricted estimator when $d = \frac{\text{tr}(\phi)}{\max(\phi)} > 2$, whereas the similar condition for the minimal mean squared error estimator is $d > 4$.

Remark 4: The proposed Stein-like shrinkage estimator can be generalized to account for the possibility of multiple common breaks. Similar to the single break case, the unrestricted estimator uses the observations within each regime separately, and the restricted estimator uses the full-sample of observations. In case of forecasting, the shrinkage estimator is the weighted average of the restricted estimator and the unrestricted estimator using the observations after the most recent break point.⁸ □

4 Monte Carlo simulations

This section employs Monte Carlo simulations to examine the performance of the theoretical results obtained in the paper. To do this, we consider the following data generating process

$$y_{i,t} = \begin{cases} x'_{i,t} \beta_{i(1)} + u_{i,t} & \text{for } i = 1, \dots, N, \quad t = 1, \dots, T_1, \\ x'_{i,t} \beta_{i(2)} + u_{i,t} & \text{for } i = 1, \dots, N, \quad t = T_1 + 1, \dots, T, \end{cases} \quad (34)$$

⁸We note that other restricted estimators can be considered using different restriction matrices. For example, in a two breaks case, a subsample estimate using the second and third subsamples can be combined with the unrestricted estimator.

where $x_{i,t} \sim N(0, 1)$, and we set the first column of that to be a vector of ones in order to allow for the fixed effect. Let the time series dimension be $T = 100$, the number of individual units be $N = 5$, and $k \in \{1, 3\}$. We consider different values for true break points, which are proportional to the sample size, $b_1 \equiv \frac{T_1}{T} \in \{0.2, 0.8\}$.⁹

Let $\beta_{i(2)}$ be a vector of ones, and $\delta_i = \beta_{i(1)} - \beta_{i(2)} = \frac{i}{N-1} \times s$ shows the true break size in the coefficients, where s varies from 0 to 1 in increments of 0.1. Moreover, define $\xi_{1(1)} \sim N(0, \sigma_{(1)}^2)$ and $\xi_{1(2)} \sim N(0, \sigma_{(2)}^2)$, and let $q \equiv \sigma_{(1)}/\sigma_{(2)}$ with $q \in \{0.5, 1, 2\}$. To allow for the cross-sectional dependence, we consider

$$\begin{cases} u_{i(1)} = 0.5\xi_{1(1)} + v_{i(1)} & \text{for } i = 1, \dots, N, \quad t = 1, \dots, T_1 \\ u_{i(2)} = 0.5\xi_{1(2)} + v_{i(2)} & \text{for } i = 1, \dots, N, \quad t = T_1 + 1, \dots, T, \end{cases} \quad (35)$$

where $v_{i(1)} \sim N(0, \sqrt{i} \sigma_{(1)}^2)$, and $v_{i(2)} \sim N(0, \sqrt{i} \sigma_{(2)}^2)$. As the break point, break sizes in the slope coefficients and the error variances are unknown in practice, we estimate them to incorporate the uncertainty regarding the estimation of these parameters in our analysis. The break point is estimated using the method proposed by [Qu and Perron \(2007\)](#).

In this study, we compare the performance of the Stein-like shrinkage estimator, the minimal mean squared error estimator, the unrestricted estimator, and the pre-test estimator. We calculate the pre-test estimator, denoted by $\widehat{\mathbf{b}}_{PT}$, as

$$\widehat{\mathbf{b}}_{PT} = \widehat{\mathbf{b}}_r I(D_T < c_v) + \widehat{\mathbf{b}}_{ur} I(D_T \geq c_v), \quad (36)$$

where c_v is the 5% critical value from the chi-square distribution, and D_T is the Wald-type test statistic in (9). Thus, the pretest estimator selects a restricted estimator when the Wald statistic is insignificant and selects the unrestricted estimator when the Wald statistic is significant. We report the relative mean squared error (MSE), and set the unrestricted estimator as the benchmark estimator so that the relative MSE of the unrestricted estimator is equal to one, i.e., the relative MSEs are $\frac{\rho(\widehat{\mathbf{b}}_w, \mathbb{W})}{\rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W})}$, $\frac{\rho(\widehat{\mathbf{b}}_\gamma, \mathbb{W})}{\rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W})}$, $\frac{\rho(\widehat{\mathbf{b}}_{PT}, \mathbb{W})}{\rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W})}$, and $\frac{\rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W})}{\rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W})}$. The relative MSE is a good measure for evaluating the estimation accuracy of different methods. In addition, it measures how different methods are compared with each others and what are the gains of methods relative to each other.

⁹Additional Monte Carlo results with other specifications are available in the Supplementary Online Appendix.

Figure 1 shows the results over 1000 Monte Carlo replications when $N = 5$. Since we consider different break sizes for individuals, we show the proportion of the break size in the coefficients (s) in the horizontal axis. The vertical axis shows the relative MSE.

4.1 Simulation results

Based on the results of Figure 1, the Stein-like shrinkage estimator has better performance than the unrestricted estimator, in the sense of having a smaller MSE, for any break sizes and break points. Figure 1 shows the results with $k = 1$ and $k = 3$. Based on Figure 1, for the small to medium break sizes in the coefficients, the Stein-like shrinkage estimator performs much better than the unrestricted estimator. In this case, the shrinkage estimator assigns more weight to the restricted estimator to gain from its efficiency. As the break size in the coefficients increases, the Stein-like shrinkage estimator performs close to the unrestricted estimator, but still slightly better. This may be related to the large bias that the restricted estimator adds to the shrinkage estimator under large break sizes. Therefore, the shrinkage estimator assigns more weight to the unrestricted estimator and less weight to the restricted estimator. We note that even for the large break sizes, we still do not observe the under-performance of the Stein-like shrinkage estimator over the unrestricted estimator. When $q = 0.5$ the error variance of the pre-break data is less than that of the post-break data while the opposite holds for $q = 2$. Looking at Figures 1(a)–1(c), we see the performance of the Stein-like shrinkage estimator is improving. When $q = 0.5$ and $T_1 = 20$, Figure 1(a), the pre-break estimator performs poorly as it only has 20 observations. When we combine it with the restricted estimator, it cannot benefit significantly, as the post-break observations are more volatile ($q = 0.5$). However, for $q = 2$ and $T_1 = 20$, Figure 1(c), when we combine the pre-break estimator with the restricted estimator, the estimator performs better, since the post-break observations are less volatile ($q = 2$). For $T_1 = 80$ and $q = 0.5$, Figure 1(d), the post-break estimator performs poorly since it only has 20 observations. When we combine it with the restricted estimator, it benefits more compared to the previous case that $T_1 = 20$, because the pre-break observations are less volatile ($q = 0.5$). This pattern is observed for $k = 3$ as well. Besides, as we increase the number of regressors, the gain obtained from the Stein-like shrinkage estimator increases. In addition, the Stein-like shrinkage estimator performs better than the minimal mean squared error estimator for the small-to-medium break sizes in the coefficients. For large break sizes, these two

estimators perform almost equally.

All simulation results confirm the results of Theorems 3 and 4 that the performance of the Stein-like shrinkage estimator and the minimal mean squared error estimator are uniformly better than the unrestricted estimator in the sense of having lower MSE, for any break points, break sizes in the coefficients, and any q . Generally, for large break sizes, large N or large number of regressors, k , the shrinkage estimators have equal performance. This confirms our theoretical findings in Remark 3.

The pretest estimator either uses the restricted estimator or the unrestricted estimator depending on the Wald test results. Generally, the MSE of the pretest estimator is similar to the restricted estimator for a small break size and is similar to the unrestricted estimator for a large break size, because these two extreme cases are better caught by the test statistic. For a moderate break size, however, the pretest estimator performs worse than the unrestricted estimator.

In order to see how the Stein-like shrinkage estimator assigns weights to the restricted and unrestricted estimators, we plot the estimated weights, \hat{w}_T , for the simulation results of Figures 1(a) and 1(g). The estimated weights are shown in Figure 2, in which the horizontal axis shows the break size in the mean, and the vertical axis shows the estimated Stein-like shrinkage weights. The results show that for small break sizes, the Stein-like shrinkage estimator mainly assigns a higher weight to the restricted estimator (estimated weight is small), while as the break size increases it assigns a larger weight to the unrestricted estimator (estimated weight is large). As the number of regressors increases, the range of the estimated weight becomes smaller. Similar patterns are seen under other specifications which are omitted to save space.

5 Empirical analysis

This section provides some empirical analysis for forecasting the growth rate of real output using a quarterly data set in G7 countries from 1995:Q1 to 2016:Q4.¹⁰ The predictors for each country are the log real equity prices, the real short term interest rate, and the difference between the long and short-term interest rates. The data are taken from the Global VAR (GVAR) dataset (2016 vintage). The data are available with the GVAR Toolbox, [Mohaddes and Raissi \(2018\)](#).

We evaluate the out-of-sample forecasting performance of the proposed Stein-like shrinkage

¹⁰The seven countries are Canada, France, Germany, Italy, Japan, the UK, and the USA.

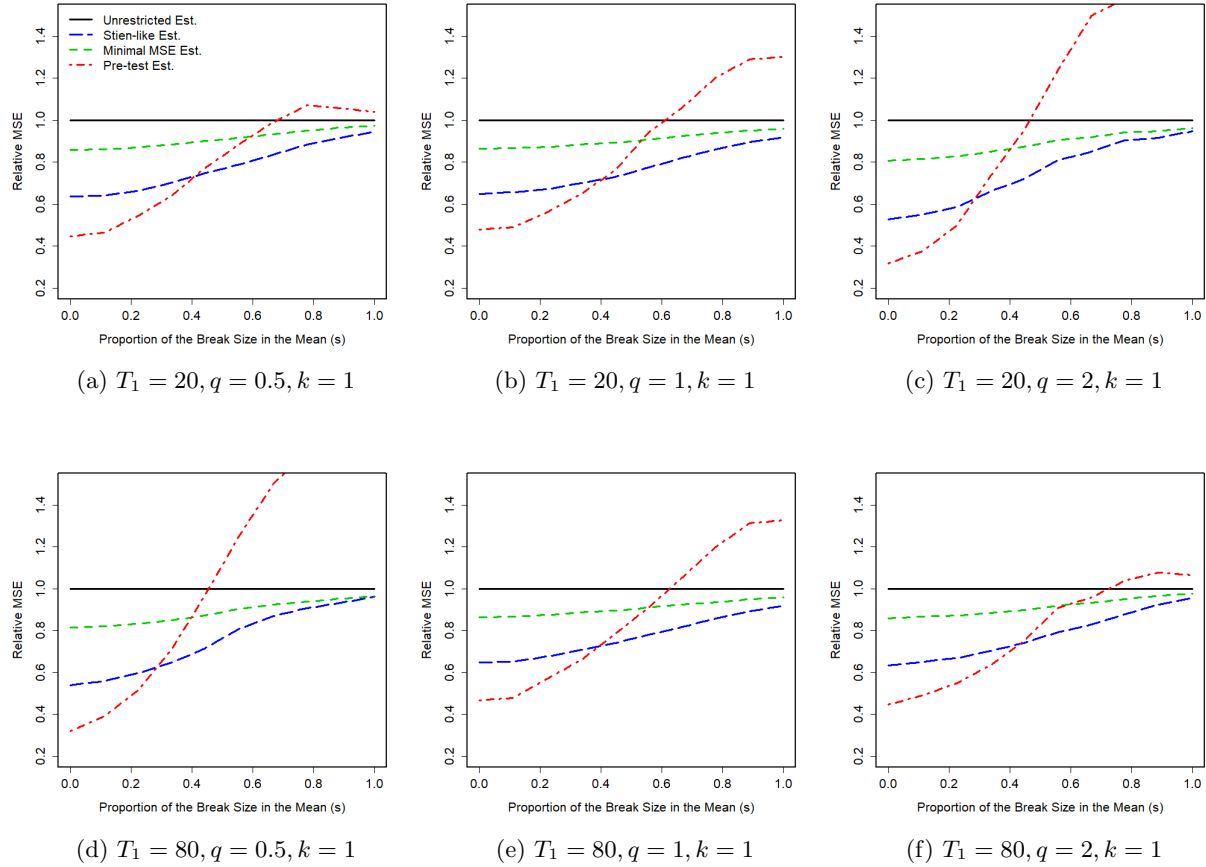


Figure 1: Monte Carlo results for $T = 100, N = 5$

estimator and the minimal mean squared error estimator with a range of alternative methods in terms of their MSFEs. The first two alternative methods are the SUR model that estimates the post-break slope coefficients across the entire cross-section (the unrestricted estimator), and the pre-test estimator. We also consider some of the existing univariate time-series forecasting approaches that provide forecasts independently in each cross-section series as alternative methods, that is, these time-series methods ignore the cross-sectional dependence. These are the method proposed by [Pesaran et al. \(2013\)](#) (labeled as “PPP” in tables), the five methods used in [Pesaran and Timmermann \(2007\)](#), namely, “Postbk”, “Troff”, “Pooled”, “WA”, “CV”, the full-sample forecast that ignores the break and uses the full-sample of observations (“Full”), the average window forecast proposed by [Pesaran and Pick \(2011\)](#) (“AveW”), the method proposed by [Lee et al. \(2022a\)](#) (“LPU”), and the forecast using the optimal window size proposed by [Inoue et al. \(2017\)](#) (“IJR”),

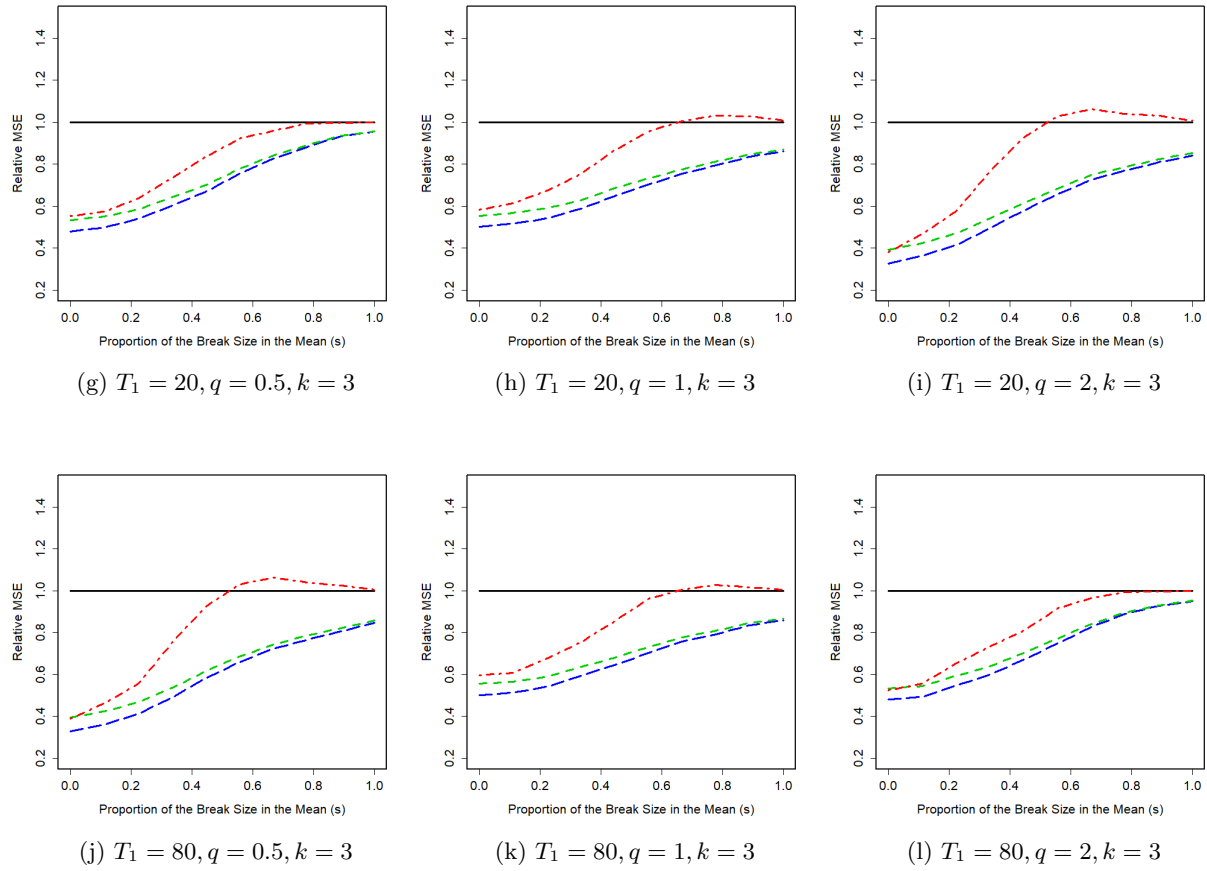


Figure 1: Monte Carlo results for $T = 100, N = 5$ (Cont.)

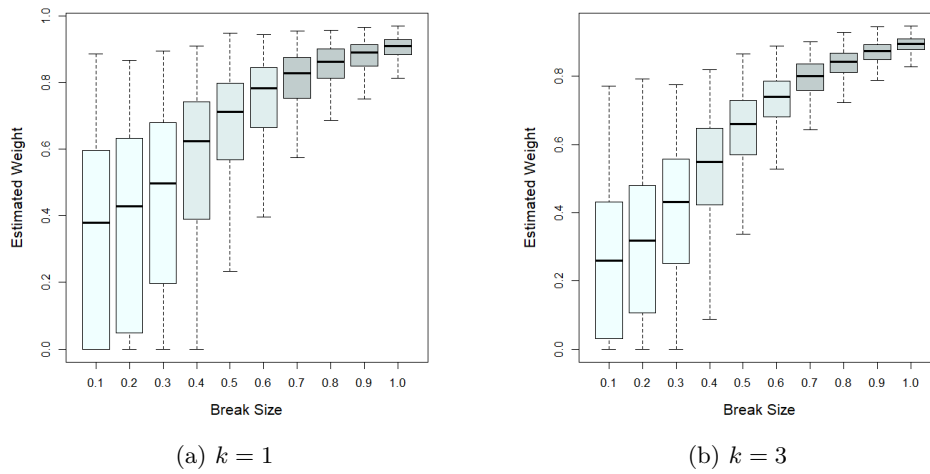


Figure 2: Boxplot for the estimated Stein-like shrinkage weight

which is designed for smoothly time varying parameters.

We compute h -step-ahead forecasts ($h = 1, 2, 3, 4$) for different forecasting methods described above, using both rolling and expanding windows. The rolling window forecasts is based on the most recent 10 years (40 quarters) of observations, the same window size used by [Stock and Watson \(2003\)](#). For the expanding window, we divide the sample of observations into two parts. The first T observations are used as the initial in-sample estimation period, and the remaining observations are the pseudo out-of-sample evaluation period. The initial estimation period is from 1995:Q1 to 2004:Q4 ($T = 40$), which leaves the out-of-sample period from 2005:Q1 to 2016:Q4. In both rolling and expanding window approaches, we initially estimate the break point using the [Qu and Perron \(2007\)](#) approach. Then, we generate h -quarters-ahead forecasts.¹¹ Furthermore, we present the forecasting results using the approach proposed by [Clark and McCracken \(2010\)](#) (labeled as “CM” in tables), which is the equally weighted average of the rolling and expanding window forecasts.

To evaluate the performance of our proposed estimators, we compute their MSFE and compare them with those of the alternative estimators. Tables 1–2 present the cross-country averages MSFE results for different methods across different forecast horizons. In the aggregation of the individual country MSFEs, we use both GDP with Purchasing Power Parity based weights (GDP-PPP) and the equal weights. The GDP weighted average uses weights $w_i = Y_i / (\sum_{j=1}^s Y_j)$, where Y_i is the 2008 GDP in purchasing power terms for country i available from the GVAR data set and $s = 7$ is the number of countries. The equal weights average uses $w_i = 1/s$. The results of cross country averages with GDP-PPP scheme are presented in Table 1, while the results for the equal weights are presented in Table 2. In each of these tables, Panels A–C show the MSFE results of different methods using the rolling window of most recent 40 observations, the expanding window, and the [Clark and McCracken \(2010\)](#) method, respectively.

Based on the results in Tables 1–2, the forecasts based on the Stein-like shrinkage estimator and the minimal mean squared error estimator provide improved forecasts over the unrestricted estimator for different forecast horizons. With the GDP-PPP weighted average scheme, for the Stein-like shrinkage estimator (and the minimal mean squared error estimator), the improvement ranges from 2.7 to 11.3 (1.8 to 7.5) percent for the rolling window, from 2 to 16 (1.1 to 7.5) for

¹¹For example, with the expanding window, when the in-sample period is from 1995:Q1-2007:Q7, the estimated break date is 1999:Q4 which is related to the dot-com bubble, while when the in-sample period includes the 2008 and beyond, the estimated break date is 2008:Q1 which is related to the financial crisis.

the expanding window, and from 3.1 to 13.9 (1.7 to 7.6) percent for the CM method. With the equally weighted average scheme, these improvement range from 5.5 to 12.9 (2.9 to 9.5) percent for the rolling window, from 4.6 to 17.5 (2.4 to 8.6) for the expanding window, and from 5.1 to 15.3 (2.7 to 9.5) percent for the CM method. The results suggest that improvement using the equally weighted average scheme are slightly higher than those of the GDP-PPP weighted average scheme. Besides, the MSFE of the Stein-like shrinkage estimator is less than the minimal mean squared error estimator. All these results are consistent with our theoretical findings in Theorems 3–4, and Remark 3.

Furthermore, the MSFEs of the Stein-like shrinkage estimator outperform the other alternative estimators. To determine whether the Stein-like shrinkage estimator significantly outperforms the alternative methods, we report the panel version of the Diebold and Mariano (1995) test (DM) introduced by Pesaran et al. (2009) in Tables 1–2, in which we take each of the alternative methods as a benchmark (i.e., DM_i is the benchmark method in which i represent the alternative methods such as MMSE, Unres., Pretest, Postbk, PPP, Troff, CV, WA, Pooled, Full, AveW, and IJR, respectively). In Tables 1–2, an asterisk represents statistically significance evidence in favor of the Stein-like shrinkage estimator according to the panel Diebold–Mariano test statistic. The 1%, 5%, and 10% significance levels are denoted by ***, ** and *, respectively. The results show that the Stein-like shrinkage estimator significantly outperforms all benchmark models in the majority of cases at the 1%, 5%, and 10% levels, for various forecast horizons, $h = 1, 2, 3, 4$. Overall, the results show the advantages of using the SUR model in forecasting relative to a univariate time-series model in terms of providing lower MSFE. As discussed in Bai et al. (1998), and Qu and Perron (2007), the break-point problem is one in which there are substantial payoffs for using multivariate rather than univariate techniques. The accuracy of break point estimation can be significantly improved by adding series that have common breaks, and the precision of the estimates increases with the number of equations in the system. The more accurate the break point is estimated, the better estimate of slope coefficients is obtained, which ultimately helps to have a better forecasts. Besides, the SUR model uses the cross-sectional information rather than ignoring them which helps to have a better estimate and hence forecast.

Table 1: Empirical results for forecasting output growth

h	Stein	MMSE	Unres.	Pretest	Postbk	LPU	PPP	Troff	CV	WA	Pooled	Full	AveW	IJR
Panel A: Rolling, GDP weighted average														
1	0.590	0.615**	0.665***	0.655**	0.691***	0.646**	0.690***	0.709***	0.667*	0.649*	0.652*	0.654*	0.663*	0.653**
2	1.974	1.993**	2.029**	2.065*	2.597***	2.302***	2.561***	2.687***	2.668***	2.295**	2.347**	2.348*	2.323**	2.328**
3	3.919	4.044**	4.244***	4.215***	5.162***	4.534**	5.154***	5.258***	5.236**	4.596**	4.749**	4.544*	4.675*	4.750**
4	6.173	6.293	6.441**	6.416**	8.800***	7.745***	8.771***	8.551***	8.655***	7.585**	7.752**	7.926**	7.302***	7.283***
Panel B: Expanding, GDP weighted average														
1	0.615	0.657***	0.732***	0.733***	0.767***	0.636*	0.754***	0.787***	0.637*	0.622*	0.626*	0.653*	0.656*	—
2	2.063	2.114**	2.175**	2.175**	2.610***	2.139**	2.590***	2.680***	2.293**	2.150*	2.169*	2.229*	2.362*	—
3	3.981	4.111**	4.254**	4.254**	5.147***	4.348*	5.137***	5.385**	4.728**	4.390**	4.410**	4.382*	4.775*	—
4	6.061	6.113	6.184**	6.184**	8.469***	7.380***	8.438***	8.383***	7.963**	7.245*	7.173*	7.621**	7.541**	—
Panel C: CM, GDP weighted average														
1	0.602	0.636***	0.699***	0.694***	0.729***	0.641**	0.722***	0.748***	0.652**	0.635*	0.639**	0.653*	0.659*	—
2	2.018	2.053**	2.102*	2.120**	2.603***	2.220**	2.575***	2.684***	2.480**	2.223**	2.258*	2.289**	2.343**	—
3	3.954	4.077**	4.249**	4.235**	5.154***	4.441**	5.146***	5.321**	4.982*	4.493**	4.576**	4.463*	4.725*	—
4	6.117	6.203	6.312**	6.300**	8.635***	7.562***	8.605***	8.467**	8.309**	7.415**	7.462**	7.774**	7.421**	—

Note: This table presents the $1000 \times \text{MSFE}$ for different estimators. h in the first column shows the forecast horizon. Panels A-C report the results based on the GDP weighted average with rolling window of most recent 40 observations, the expanding window with the initial in-sample estimation period of 10 years, and the [Clark and McCracken \(2010\)](#) method (CM), respectively. The GDP weighted average uses weights $w_i = Y_i / (\sum_{j=1}^s Y_j)$, where Y_i is the 2008 GDP in purchasing power terms for country i available from the GVAR data base and $s = 7$ is the number of countries. In the heading of the table, Stein shows the MSFE for the proposed Stein-like shrinkage estimator, MMSE is for the proposed minimal mean squared error estimator, Unres. is for the unrestricted estimator, Pretest is for the pre-test estimator, LPU is the one proposed by [Lee et al. \(2022a\)](#), PPP is the one proposed by [Pesaran et al. \(2013\)](#), Postbk, Troff, CV, WA, and Pooled are the five methods used in [Pesaran and Timmermann \(2007\)](#), Full uses the full-sample of observations by ignoring the break, AveW is the method proposed by [Pesaran and Pick \(2011\)](#) with $w_{\min} = 0.1$ and twenty windows, and IJR is the optimal window method proposed by [Inoue et al. \(2017\)](#) with $\underline{R} = \max(1.5T^{2/3}, 20)$ and $\overline{R} = \min(4T^{2/3}, T - h)$. This table also shows the panel Diebold and Mariano (DM) test statistic for the Stein-like shrinkage forecast relative to each of the alternative methods as the benchmark. An asterisk represents statistically significance evidence in favor of the Stein-like shrinkage estimator according to the panel Diebold–Mariano test statistic. The 1%, 5%, and 10% significance levels are denoted by ***, ** and *, respectively.

Table 2: Empirical results for forecasting output growth

h	Stein	MMSE	Unres.	Pretest	Postbk	LPU	PPP	Troff	CV	WA	Pooled	Full	AveW	IJR
Panel A: Rolling, Equally weighted average														
1	0.590	0.613**	0.677**	0.673**	0.690**	0.632**	0.691**	0.694**	0.673*	0.636*	0.642**	0.633	0.654*	0.650**
2	1.941	1.961	2.063*	2.111*	2.661***	2.257***	2.628***	2.691***	2.635**	2.241**	2.313***	2.258**	2.453**	2.308***
3	4.074	4.228**	4.478**	4.461**	5.413**	4.436**	5.371**	5.412**	5.376*	4.449*	4.643**	4.509*	5.145*	4.989*
4	6.223	6.394**	6.587**	6.555**	8.843***	7.287***	8.762***	8.524***	8.589***	7.125***	7.341***	7.383**	7.748*	7.590**
Panel B: Expanding, Equally weighted average														
1	0.593	0.637***	0.719**	0.725**	0.745***	0.608*	0.742***	0.757***	0.601	0.598	0.597	0.666**	0.646*	–
2	2.060	2.125**	2.222**	2.222**	2.740***	2.073**	2.719***	2.711**	2.254**	2.082*	2.087*	2.127*	2.471**	–
3	4.097	4.277**	4.490**	4.490**	5.451**	4.119*	5.438**	5.467**	4.810**	4.138*	4.205*	4.295*	5.091**	–
4	6.072	6.217*	6.368*	6.368*	8.583***	7.298***	8.548***	8.380***	7.895**	6.603*	6.606*	7.004*	7.732**	–
Panel C: CM, Equally weighted average														
1	0.591	0.625**	0.698**	0.699**	0.717***	0.620**	0.716***	0.726***	0.641**	0.617*	0.612*	0.649*	0.650*	–
2	2.005	2.047*	2.143*	2.166**	2.705***	2.165**	2.673***	2.701***	2.445**	2.161*	2.200**	2.192*	2.462**	–
3	4.085	4.252**	4.484**	4.475**	5.432**	4.277*	5.409**	5.439**	5.118*	4.293*	4.424*	4.402*	5.118*	–
4	6.147	6.306*	6.478*	6.462*	8.713***	7.292***	8.655***	8.452***	8.242**	6.864**	6.974**	7.194**	7.740**	–

Note: See the notes to Table 1. The equal weights average uses $w_i = 1/s$ with $s = 7$.

6 Conclusion

In this paper, we introduce two estimators for estimating the slope coefficients of seemingly unrelated regression models with cross-sectional dependence under structural breaks, which are then used for forecasting. The proposed Stein-like shrinkage estimator and the minimal mean squared error estimator are the weighted averages of the unrestricted estimator (i.e., using observations within each regime separately) and the restricted estimator. The averaging weight is proportional to the difference between the restricted and unrestricted estimators. Thus, for a large break size, a lower weight is assigned to the restricted estimator (which is biased) and more weight is assigned to the unrestricted estimator. The opposite is true for a small break size. We establish the asymptotic distribution and asymptotic risk for the proposed estimators and find the conditions under which they uniformly outperform the unrestricted estimator for any break size and break point. We also analytically and numerically compare the performance of the two proposed estimators. The results show that the Stein-like shrinkage estimator performs better than the minimal mean squared error estimator due to its smaller asymptotic risk. This out-performance is more clear under small break sizes. Monte Carlo simulations and the empirical application of forecasting output growth rates of G7 countries show the significant superiority of using the Stein-like shrinkage estimator over the alternative estimators.

References

- Bai, J. and Ng, S. (2002). “Determining the number of factors in approximate factor models.” *Econometrica* 70, 191-221.
- Barnett, W., Chauvet, M. and Leiva-Leon D. (2016). “Real-time nowcasting of nominal gdp with structural breaks.” *Journal of Econometrics* 191, 312–324.
- Bai, J. (2000). “Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices.” *Annals of Economics and Finance* 1, 303-339.
- Bai, J., R. Lumsdaine, and J. Stock (1998). “Testing for and dating common breaks in multivariate time series.” *Review of Economic Studies* 65, 395-432.

- Bai, J. and Perron, P. (2003). “Computation and analysis of multiple structural change models.” *Journal of Applied Econometrics* 18, 1-22.
- Boot, T. and Pick, A. (2020). “Does modeling a structural break improve forecast accuracy?” *Journal of Econometrics* 215, 35-59.
- Bernstein, D. S. (2005). *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*. Princeton University Press.
- Coakley, J., A. Fuertes, and Smith, R. P. (2002). “A principal components approach to cross-section dependence in panels.” Unpublished Manuscript, Birkbeck College, University of London.
- Chudick, A., and Pesaran M. H. (2015). “Large panel data models with cross-sectional dependence a survey.” *The Oxford Handbook of Panel Data*, Edited By Badi H. Baltagi New York: Oxford University Press.
- Clark, T. E. and McCracken, M. W. (2010). “Averaging forecasts from VARs with uncertain instabilities.” *Journal of Applied Econometrics* 25, 5-29.
- Clements, M.P. and Hendry, D.F. (2006). “Forecasting with breaks.” In Elliott, G., C.W.J. Granger and A. Timmermann (Eds.), *Handbook of Economic Forecasting* vol. 1, 605-658. North-Holland.
- Clements, M. P. and Hendry, D.F. (2011). *The Oxford Handbook of Economic Forecasting*. Oxford University Press.
- Diebold, F. X. and Mariano, R. S. (1995). “Comparing predictive accuracy.” *Journal of Business & Economic Statistics* 13, 253-263.
- Giacomini, R. and Rossi, B. (2009). “Detecting and predicting forecast breakdowns.” *Review of Economic Studies* 76, 669–705.
- Hansen, B. (2016). “Efficient shrinkage in parametric models.” *Journal of Econometrics* 190, 115-132.
- Hansen, B. (2017). “Stein-like 2SLS estimator.” *Econometrics Reviews* 36, 840-852.
- Inoue, A., Jin, L. and Rossi, B. (2017). “Rolling window selection for out-of-sample forecasting with time-varying parameters.” *Journal of Econometrics* 196, 55-67.

- Inoue, A. and Rossi, B. (2011). “Identifying the sources of instabilities in macroeconomic fluctuations.” *Review of Economics and Statistics* 164, 158–172.
- James, W. and Stein, C. M. (1961). “Estimation with quadratic loss.” *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 361-380.
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. 2nd ed., New York, Springer.
- Lee, T., Parsaeian, S. and Ullah, A. (2022). “Efficient combined estimation under structural breaks.” *Advances in Econometrics*, vol. 43, ed. by Chudik, A., Hsiao C., and A. Timmermann.
- Lee, T., Parsaeian, S. and Ullah, A. (2022a). “Optimal forecast under structural breaks.” *Journal of Applied Econometrics* 37, 965-987.
- Liu, L. (2022). “Density Forecasts in Panel Data Models: A Semiparametric Bayesian Perspective.” *Journal of Business & Economic Statistics*.
- Maddala, G. S., Li H., and Srivastava, V. K. (2001). “A comparative study of different shrinkage estimators for panel data models.” *Annals of Economics and Finance* 2, 1-30.
- Maasoumi, E. (1978). “A modified Stein-like estimator for the reduced form coefficients of simultaneous equations.” *Econometrica* 46, 695–703.
- Mehrabani, A. and Ullah, A. (2020). “Improved average estimation in seemingly unrelated regressions.” *Econometrics* 8, 1-22.
- Mohaddes, K., and Raissi M. (2018). “Global VAR (GVAR) database, 1979Q2-2016Q4”. <https://doi.org/10.17863/CAM.27337>.
- Phillips, P. C. B., and Sul, D. (2003). “Dynamic panel estimation and homogeneity testing under cross section dependence.” *The Econometrics Journal* 6, 217-259.
- Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). “Forecasting time series subject to multiple structural breaks.” *Review of Economic Studies* 73, 1057–1084.
- Pesaran, M.H. and Pick, A. (2011). “Forecast combination across estimation windows.” *Journal of Business & Economic Statistics* 29, 307–318.

- Pesaran, M.H., Pick, A. and Pranovich, M. (2013). “Optimal forecasts in the presence of structural breaks.” *Journal of Econometrics* 177, 134-152.
- Pesaran, M.H., Schuermann, T., Smith, L.Vanessa, (2009). “Forecasting economic and financial variables with global VARs.” *International Journal of Forecasting* 25, 642–675.
- Pesaran, M.H. and Timmermann, A. (2002). “Market timing and return prediction under model instability.” *Journal of Empirical Finance* 9, 495–510.
- Pesaran, M.H. and Timmermann, A. (2007). “Selection of estimation window in the presence of breaks.” *Journal of Econometrics* 137, 134-161.
- Qu, Z. and Perron, P. (2007). “Estimating and testing structural changes in multiple regressions.” *Econometrica* 75, 459–502.
- Rossi, B. (2013). “Advances in forecasting under instability.” In G. Elliott, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Vol. 2, Part B, 1203 - 1324. Elsevier.
- Smith, S. and Timmermann, A. (2018). “Detecting breaks in real time: a panel forecasting approach.” *Econometrics: Econometric & Statistical Methods - General eJournal*.
- Smith, S. (2018). “Forecasting panel data with structural breaks and regime-specific grouped heterogeneity.” Available at SSRN: <https://ssrn.com/abstract=3268759>.
- Stein, Charles M., (1956). “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution.” *Proceedings of the Third Berkeley Symposium on Math. Statist. and Probability*, Vol. 1, 197–206.
- Stock, J. H. and Watson, M. W. (2003). “Forecasting output and inflation: the role of asset prices.” *Journal of Economic Literature*, 788-829.
- Wang, W., X. Zhang, R. Paap, (2019). “To pool or not to pool: what is a good strategy?” *Journal of Applied Econometrics* 34, 724-745.
- Zellner, A. (1962). “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias.” *Journal of the American Statistical Association* 57, 348–368.

A Appendix: Mathematical details

Lemma 1: Suppose C and D are two matrices, where $C_{n \times n}$ and $D_{n \times m}$, then

$$D'CD \leq (D'D)\lambda_{\max}(C)$$

in which $\lambda_{\max}(C)$ is the maximum eigenvalue of C . See [Bernstein \(2005\)](#), page 271 for the proof.

□

A.1 Proof of Theorem 1

Define $\tilde{y}_t = (y_{1,t}, \dots, y_{N,t})'$, $\tilde{x}_t = \text{diag}(x'_{1,t}, \dots, x'_{N,t})$ and $\tilde{u}_t = (u_{1,t}, \dots, u_{N,t})'$. By stacking the model over individuals, $i = 1, \dots, N$, for each time, we have

$$\tilde{y}_t = \tilde{x}_t \beta_{(t)} + \tilde{u}_t, \quad \text{for } t = 1, \dots, T, \quad (\text{A.1})$$

where

$$\beta_{(t)} \equiv \begin{cases} \beta_{(1)} & \text{for } t = 1, \dots, T_1, \\ \beta_{(2)} & \text{for } t = T_1 + 1, \dots, T, \end{cases} \quad (\text{A.2})$$

in which $\beta_{(1)} = (\beta'_{1(1)}, \dots, \beta'_{N(1)})'$ and $\beta_{(2)} = (\beta'_{1(2)}, \dots, \beta'_{N(2)})'$ are $Nk \times 1$ vectors of the slope coefficients. Define $2Nk \times Nk$ selection matrices

$$G_1 = \begin{bmatrix} I_k & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & I_k & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ & & & \vdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}, \quad (\text{A.3})$$

and

$$G_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ I_k & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ & & & \vdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & I_k \end{bmatrix}, \quad (\text{A.4})$$

such that $\mathbf{b} = G_1\beta_{(1)} + G_2\beta_{(2)}$.

Rewrite Ω in (6) as $\Omega = \Sigma_{(1)} \otimes J_{(1)} + \Sigma_{(2)} \otimes J_{(2)}$, where

$$\Sigma_{(1)} = \begin{bmatrix} \sigma_{11(1)} & \dots & \sigma_{1N(1)} \\ \vdots & \vdots & \vdots \\ \sigma_{N1(1)} & \dots & \sigma_{NN(1)} \end{bmatrix}, \quad (\text{A.5})$$

$$\Sigma_{(2)} = \begin{bmatrix} \sigma_{11(2)} & \dots & \sigma_{1N(2)} \\ \vdots & \vdots & \vdots \\ \sigma_{N1(2)} & \dots & \sigma_{NN(2)} \end{bmatrix}, \quad (\text{A.6})$$

are $N \times N$ matrices, $J_{(1)} = \begin{bmatrix} I_{T_1} & \mathbf{0}_{T_1 \times T - T_1} \\ \mathbf{0}_{T - T_1 \times T_1} & \mathbf{0}_{T - T_1 \times T - T_1} \end{bmatrix}$, and $J_{(2)} = \begin{bmatrix} \mathbf{0}_{T_1 \times T_1} & \mathbf{0}_{T_1 \times T - T_1} \\ \mathbf{0}_{T - T_1 \times T_1} & I_{T - T_1} \end{bmatrix}$.

In the rest of the proof, we show the asymptotic normality of the feasible unrestricted estimator, $\hat{\mathbf{b}}_{ur} = G_1\hat{\beta}_{(1)}^{fgls} + G_2\hat{\beta}_{(2)}^{fgls}$. Initially, we derive the asymptotic distribution of the infeasible pre-break and post-break estimators, denoted by $\hat{\beta}_{(1)}^{gls}$ and $\hat{\beta}_{(2)}^{gls}$, respectively.

$$\begin{aligned} \sqrt{T}(\hat{\beta}_{(1)}^{gls} - \beta_{(1)}) &= \frac{1}{\sqrt{b_1}} \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{x}_t}{T_1} \right)^{-1} \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) \\ &= \frac{1}{\sqrt{b_1}} Q_1^{-1} \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) + \frac{1}{\sqrt{b_1}} \left(\left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{x}_t}{T_1} \right)^{-1} - Q_1^{-1} \right) \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) \\ &= \frac{1}{\sqrt{b_1}} Q_1^{-1} \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) + o_p(1), \end{aligned} \quad (\text{A.7})$$

where the last equality holds because $\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{x}_t}{T_1} \xrightarrow{p} \mathbb{E}(\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{x}_t) \equiv Q_1$ which is implied by

Assumptions 1 and 3. Also, we note that

$$\text{var}\left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}}\right) = (1/T_1) \mathbb{E}\left(\sum_{t=1}^{T_1} \tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{x}_t\right) = Q_1 = O(1) < \infty, \quad (\text{A.8})$$

which implies that $\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} = O_p(1)$. Therefore, by central limit theorem,

$$\sqrt{T}(\hat{\beta}_{(1)}^{gls} - \beta_{(1)}) \xrightarrow{d} N(0, Q_1^{-1}/b_1). \quad (\text{A.9})$$

Similarly, the asymptotic distribution of the infeasible post-break estimator is

$$\begin{aligned} \sqrt{T}(\hat{\beta}_{(2)}^{gls} - \beta_{(2)}) &= \frac{1}{\sqrt{1-b_1}} \left(\sum_{t=T_1+1}^T \frac{\tilde{x}'_t \Sigma_{(2)}^{-1} \tilde{x}_t}{T-T_1} \right)^{-1} \left(\sum_{t=T_1+1}^T \frac{\tilde{x}'_t \Sigma_{(2)}^{-1} \tilde{u}_t}{\sqrt{T-T_1}} \right) \\ &\xrightarrow{d} N(0, Q_2^{-1}/(1-b_1)), \end{aligned} \quad (\text{A.10})$$

where $\sum_{t=T_1+1}^T \frac{\tilde{x}'_t \Sigma_{(2)}^{-1} \tilde{x}_t}{T-T_1} \xrightarrow{p} \mathbb{E}(\tilde{x}'_t \Sigma_{(2)}^{-1} \tilde{x}_t) \equiv Q_2$.

Now, we provide the asymptotic distribution of the feasible pre-break estimator, $\hat{\beta}_{(1)}^{fgls}$.

$$\begin{aligned} \sqrt{T}(\hat{\beta}_{(1)}^{fgls} - \beta_{(1)}) &= \frac{1}{\sqrt{b_1}} \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \hat{\Sigma}_{(1)}^{-1} \tilde{x}_t}{T_1} \right)^{-1} \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \hat{\Sigma}_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) \\ &= \sqrt{T}(\hat{\beta}_{(1)}^{gls} - \beta_{(1)}) + \frac{1}{\sqrt{b_1}} [\Pi_1 + \Pi_2 + \Pi_3], \end{aligned} \quad (\text{A.11})$$

where

$$\Pi_1 \equiv \left[\left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \hat{\Sigma}_{(1)}^{-1} \tilde{x}_t}{T_1} \right)^{-1} - \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{x}_t}{T_1} \right)^{-1} \right] \left[\left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \hat{\Sigma}_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) - \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) \right], \quad (\text{A.12})$$

$$\Pi_2 \equiv \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{x}_t}{T_1} \right)^{-1} \left[\left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \hat{\Sigma}_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) - \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right) \right], \quad (\text{A.13})$$

and

$$\Pi_3 \equiv \left[\left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \hat{\Sigma}_{(1)}^{-1} \tilde{x}_t}{T_1} \right)^{-1} - \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{x}_t}{T_1} \right)^{-1} \right] \left(\sum_{t=1}^{T_1} \frac{\tilde{x}'_t \Sigma_{(1)}^{-1} \tilde{u}_t}{\sqrt{T_1}} \right). \quad (\text{A.14})$$

We note that

$$\begin{aligned}
\widehat{\mathbf{u}}_t &= \widetilde{\mathbf{y}}_t - \widetilde{\mathbf{u}}_t \widehat{\beta}_{(1)}^{ols} \\
&= \widetilde{\mathbf{u}}_t - \widetilde{\mathbf{x}}_t (\widehat{\beta}_{(1)}^{ols} - \beta_{(1)}) \\
&= \widetilde{\mathbf{u}}_t + o_p(1),
\end{aligned} \tag{A.15}$$

where $\widehat{\beta}_{(1)}^{ols} = (\widehat{\beta}_{1(1)}^{ols'}, \dots, \widehat{\beta}_{N(1)}^{ols'})'$. Therefore, $\widehat{\Sigma}_{(1)} = \frac{1}{T_1} \sum_{t=1}^{T_1} \widehat{\mathbf{u}}_t \widehat{\mathbf{u}}_t' = \Sigma_{(1)} + o_p(1)$. It follows that $(\sum_{t=1}^{T_1} \frac{\widetilde{\mathbf{x}}_t' \widehat{\Sigma}_{(1)}^{-1} \widetilde{\mathbf{x}}_t}{T_1}) - (\sum_{t=1}^{T_1} \frac{\widetilde{\mathbf{x}}_t' \Sigma_{(1)}^{-1} \widetilde{\mathbf{x}}_t}{T_1}) = o_p(1)$, $(\sum_{t=1}^{T_1} \frac{\widetilde{\mathbf{x}}_t' \widehat{\Sigma}_{(1)}^{-1} \widetilde{\mathbf{x}}_t}{T_1})^{-1} - (\sum_{t=1}^{T_1} \frac{\widetilde{\mathbf{x}}_t' \Sigma_{(1)}^{-1} \widetilde{\mathbf{x}}_t}{T_1})^{-1} = o_p(1)$, and $(\sum_{t=1}^{T_1} \frac{\widetilde{\mathbf{x}}_t' \widehat{\Sigma}_{(1)}^{-1} \widetilde{\mathbf{u}}_t}{\sqrt{T_1}}) - (\sum_{t=1}^{T_1} \frac{\widetilde{\mathbf{x}}_t' \Sigma_{(1)}^{-1} \widetilde{\mathbf{u}}_t}{\sqrt{T_1}}) = o_p(1)$. Thus, $\Pi_1 = o_p(1)$, $\Pi_2 = o_p(1)$ and $\Pi_3 = o_p(1)$. As a result, using central limit theorem, the asymptotic distribution of the feasible pre-break estimator in (A.11) is

$$\begin{aligned}
\sqrt{T}(\widehat{\beta}_{(1)}^{fgls} - \beta_{(1)}) &= \sqrt{T}(\widehat{\beta}_{(1)}^{gls} - \beta_{(1)}) + o_p(1), \\
&\xrightarrow{d} N(0, Q_1^{-1} / \sqrt{b_1}).
\end{aligned} \tag{A.16}$$

Similarly, the asymptotic distribution of the feasible post-break estimator is

$$\sqrt{T}(\widehat{\beta}_{(2)}^{fgls} - \beta_{(2)}) \xrightarrow{d} N(0, Q_2^{-1} / (1 - b_1)). \tag{A.17}$$

Therefore, it immediately follows that

$$\sqrt{T}(\widehat{\mathbf{b}}_{ur} - \mathbf{b}) \xrightarrow{d} Z \sim N(0, V_{ur}), \tag{A.18}$$

where $V_{ur} = (\mathbb{E}(\frac{X' \Omega^{-1} X}{T}))^{-1}$.

Furthermore, the asymptotic distribution of the restricted estimator is

$$\begin{aligned}
\sqrt{T}(\widehat{\mathbf{b}}_r - \mathbf{b}) &= \sqrt{T}(\widehat{\mathbf{b}}_{ur} - \mathbf{b}) - (X' \widehat{\Omega}^{-1} X)^{-1} R' \left[R (X' \widehat{\Omega}^{-1} X)^{-1} R' \right]^{-1} \sqrt{T}(R \widehat{\mathbf{b}}_{ur} - \mathbf{r} - R \mathbf{b} + R \mathbf{b}) \\
&\xrightarrow{d} Z - V_{ur} R' \left[R V_{ur} R' \right]^{-1} R (Z + \mathbf{h}),
\end{aligned} \tag{A.19}$$

where $R \mathbf{b} = \mathbf{r} + R \mathbf{h} / \sqrt{T}$. Thus, the difference between the restricted and unrestricted estimators is

$$\sqrt{T}(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r) \xrightarrow{d} V_{ur} R' \left[R V_{ur} R' \right]^{-1} R (Z + \mathbf{h}). \tag{A.20}$$

Using (A.20), distribution of D_T in (17) is straightforward. Also, using (A.18) and (A.20), the

asymptotic distribution of the Stein-like shrinkage estimator is

$$\begin{aligned}\sqrt{T}(\widehat{\mathbf{b}}_w - \mathbf{b}) &= \sqrt{T}(\widehat{\mathbf{b}}_{ur} - \mathbf{b}) - w_T \sqrt{T}(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r) \\ &\xrightarrow{d} Z - w(Z) V_{ur} R' [R V_{ur} R']^{-1} R(Z + \mathbf{h}).\end{aligned}\tag{A.21}$$

This completes the proof of Theorem 1. ■

A.2 Proof of Theorem 2

By having the results of Theorem 1, we follow along the lines of Hansen (2016) to prove Theorem 2. We derive the asymptotic risk of the Stein-like shrinkage estimator for any positive definite weight matrix \mathbb{W} . Ultimately, we set $\mathbb{W} = V_{ur}^{-1}$ which gives the results presented in Theorem 2. The asymptotic risk of the Stein-like shrinkage estimator can be calculated as

$$\begin{aligned}\rho(\widehat{\mathbf{b}}_w, \mathbb{W}) &= T \mathbb{E} \left[(\widehat{\mathbf{b}}_w - \mathbf{b})' \mathbb{W} (\widehat{\mathbf{b}}_w - \mathbf{b}) \right] \\ &= T \mathbb{E} \left[(\widehat{\mathbf{b}}_{ur} - \mathbf{b}) - w_T (\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r) \right]' \mathbb{W} \left[(\widehat{\mathbf{b}}_{ur} - \mathbf{b}) - w_T (\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r) \right] \\ &= \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + \tau^2 \mathbb{E} \left[\frac{1}{(Z + \mathbf{h})' B(Z + \mathbf{h})} \right] - 2\tau \mathbb{E} \left[\frac{(Z + \mathbf{h})' R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W} Z}{(Z + \mathbf{h})' B(Z + \mathbf{h})} \right] \\ &= \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + \tau^2 \mathbb{E} \left[\frac{1}{(Z + \mathbf{h})' B(Z + \mathbf{h})} \right] - 2\tau \mathbb{E} \left[\eta(Z + \mathbf{h})' R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W} Z \right],\end{aligned}\tag{A.22}$$

where $\eta(x) = (\frac{1}{x' B x})x$. Using the Stein's lemma (see lemma 2 in the appendix of Hansen (2016)), we simplify the last term in (A.22) as

$$\begin{aligned}\mathbb{E} \left[\eta(Z + \mathbf{h})' R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W} Z \right] &= \mathbb{E} \operatorname{tr} \left[\frac{\partial}{\partial (Z + \mathbf{h})} \eta(Z + \mathbf{h})' R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W} V_{ur} \right] \\ &= \mathbb{E} \operatorname{tr} \left[\frac{R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W} V_{ur}}{(Z + \mathbf{h})' B(Z + \mathbf{h})} \right] \\ &\quad - 2 \mathbb{E} \operatorname{tr} \left[\frac{B(Z + \mathbf{h})(Z + \mathbf{h})' R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W} V_{ur}}{\left((Z + \mathbf{h})' B(Z + \mathbf{h}) \right)^2} \right] \\ &= \mathbb{E} \operatorname{tr} \left[\frac{\mathbb{W}^{1/2} V_{ur} R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W}^{1/2}}{(Z + \mathbf{h})' B(Z + \mathbf{h})} \right] \\ &\quad - 2 \mathbb{E} \operatorname{tr} \left[\frac{(Z + \mathbf{h})' R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W} V_{ur} B(Z + \mathbf{h})}{\left((Z + \mathbf{h})' B(Z + \mathbf{h}) \right)^2} \right]\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[\frac{\text{tr}(A)}{(Z + \mathbf{h})'B(Z + \mathbf{h})} \right] - 2 \mathbb{E} \text{tr} \left[\frac{(Z + \mathbf{h})'B_1'AB_1(Z + \mathbf{h})}{\left((Z + \mathbf{h})'B(Z + \mathbf{h}) \right)^2} \right] \\
&\geq \mathbb{E} \left[\frac{\text{tr}(A)}{(Z + \mathbf{h})'B(Z + \mathbf{h})} \right] - 2 \mathbb{E} \left[\frac{\lambda_{\max}(A)}{(Z + \mathbf{h})'B(Z + \mathbf{h})} \right],
\end{aligned} \tag{A.23}$$

where $A \equiv \mathbb{W}^{1/2} V_{ur} R' [R V_{ur} R']^{-1} R V_{ur} \mathbb{W}^{1/2}$, $B \equiv R' [R V_{ur} R']^{-1} R V_{ur} \mathbb{W} V_{ur} R' [R V_{ur} R']^{-1} R$, $B_1 \equiv \mathbb{W}^{1/2} V_{ur} R' (R V_{ur} R')^{-1} R$, $R' (R V_{ur} R')^{-1} R V_{ur} \mathbb{W} V_{ur} B = B_1' A B_1$, and $B_1' B_1 = B$. Note that $\frac{\partial}{\partial x} \eta(x)' = (\frac{1}{x' B x}) I - \frac{2}{(x' B x)^2} B x x'$. Also, we use Lemma 1 to get the last inequality. Plugging (A.23) into (A.22) produces

$$\begin{aligned}
\rho(\widehat{\mathbf{b}}_w, \mathbb{W}) &< \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + \tau^2 \mathbb{E} \left[\frac{1}{(Z + \mathbf{h})'B(Z + \mathbf{h})} \right] - 2\tau \left[\frac{\text{tr}(A) - 2\lambda_{\max}(A)}{(Z + \mathbf{h})'B(Z + \mathbf{h})} \right] \\
&= \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) - \tau \mathbb{E} \left[\frac{2(\text{tr}(A) - 2\lambda_{\max}(A)) - \tau}{(Z + \mathbf{h})'B(Z + \mathbf{h})} \right] \\
&< \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) - \tau \left[\frac{2(\text{tr}(A) - 2\lambda_{\max}(A)) - \tau}{\mathbb{E}(Z + \mathbf{h})'B(Z + \mathbf{h})} \right] \\
&= \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) - \tau \left[\frac{2(\text{tr}(A) - 2\lambda_{\max}(A)) - \tau}{(c+1)\text{tr}(A)} \right],
\end{aligned} \tag{A.24}$$

where the last inequality is based on the Jensen's inequality. Notice that

$$\begin{aligned}
\mathbb{E}(Z + \mathbf{h})'B(Z + \mathbf{h}) &= \mathbf{h}'B\mathbf{h} + \mathbb{E}(Z'BZ) \\
&= \mathbf{h}'B\mathbf{h} + \text{tr}(B\mathbb{E}(ZZ')) \\
&= \mathbf{h}'B\mathbf{h} + \text{tr}(B V_{ur}) \\
&= \mathbf{h}'B\mathbf{h} + \text{tr}(A) \\
&\leq (c+1)\text{tr}(A),
\end{aligned} \tag{A.25}$$

in which for any $0 < c < \infty$, we define a ball such that $H(c) = \{\mathbf{h} : \mathbf{h}'B\mathbf{h} \leq \text{tr}(A) c\}$. When $\mathbb{W} = V_{ur}^{-1}$, $\text{tr}(A) = Nk$. This completes the proof of Theorem 2. \blacksquare

A.3 Proof of Eq. (26)

Here, we obtain the optimal weight, $\widehat{\gamma}_{opt}^*$, for the minimal mean squared error estimator for any positive definite weight matrix \mathbb{W} . Using the asymptotic distribution of the estimators in (15) and (16), the asymptotic risk for the minimal mean squared error estimator is

$$\begin{aligned}
\rho(\widehat{\mathbf{b}}_\gamma, \mathbb{W}) &= \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + \gamma^2 \mathbb{E} [(Z + \mathbf{h})' R' (RV_{ur} R')^{-1} RV_{ur} \mathbb{W} V_{ur} R' (RV_{ur} R')^{-1} R(Z + \mathbf{h})] \\
&\quad - 2\gamma \mathbb{E} [(Z + \mathbf{h})' R' (RV_{ur} R')^{-1} RV_{ur} \mathbb{W} Z] \\
&= \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + \gamma^2 \mathbf{h}' B \mathbf{h} + \gamma^2 \text{tr} \mathbb{E} (ZZ' B) - 2\gamma \text{tr} \mathbb{E} (ZZ' R' (RV_{ur} R')^{-1} RV_{ur} \mathbb{W}) \\
&= \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + \gamma^2 (\mathbf{h}' B \mathbf{h} + \text{tr}(A)) - 2\gamma \text{tr}(A),
\end{aligned} \tag{A.26}$$

where $\text{tr}(V_{ur} B) = \text{tr}(A)$. By minimizing the asymptotic risk with respect to γ in (A.26), the optimal value of the weight denoted by γ_{opt}^* is

$$\gamma_{opt}^* = \frac{\text{tr}(A)}{\mathbf{h}' B \mathbf{h} + \text{tr}(A)}. \tag{A.27}$$

To find an unbiased estimator for the denominator of the weight in (A.27), we note that

$$\begin{aligned}
\mathbb{E}(\widehat{\mathbf{h}}' B \widehat{\mathbf{h}}) &= T \mathbb{E} [(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)' B (\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)] \\
&= \mathbf{h}' B \mathbf{h} + \text{tr}(A),
\end{aligned} \tag{A.28}$$

where the unbiased estimator for $\mathbf{h}' B \mathbf{h}$ is $\mathbb{E}(\widehat{\mathbf{h}}' B \widehat{\mathbf{h}}) - \text{tr}(A)$. By plugging this into the denominator of (A.27) we have

$$\widehat{\gamma}_{opt}^* = \frac{\text{tr}(A)}{\widehat{\mathbf{h}}' B \widehat{\mathbf{h}}} = \frac{\text{tr}(A)}{T(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)' \mathbb{W} (\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r)} = \frac{\text{tr}(A)}{D_T}, \tag{A.29}$$

where the second equality holds noting that $V_{ur} R' [RV_{ur} R']^{-1} R(\widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r) = \widehat{\mathbf{b}}_{ur} - \widehat{\mathbf{b}}_r$. ■

A.4 Proof of Theorem 4

The asymptotic risk of the minimal mean squared error estimator after plugging the optimal value of the weight in (A.29) is

$$\rho(\widehat{\mathbf{b}}_\gamma, \mathbb{W}) = \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + (\text{tr}(A))^2 \mathbb{E} \left[\frac{1}{(Z + \mathbf{h})' B (Z + \mathbf{h})} \right]$$

$$\begin{aligned}
& -2 \operatorname{tr}(A) \mathbb{E} \left[\eta(Z + \mathbf{h})' R' (RV_{ur} R')^{-1} R (X' \Omega^{-1} X)^{-1} \mathbb{W} Z \right] \\
& \leq \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) + (\operatorname{tr}(A))^2 \mathbb{E} \left[\frac{1}{(Z + \mathbf{h})' B (Z + \mathbf{h})} \right] \\
& - 2 \operatorname{tr}(A) [\operatorname{tr}(A) - 2\lambda_{\max}(A)] \mathbb{E} \left[\frac{1}{(Z + \mathbf{h})' B (Z + \mathbf{h})} \right] \\
& \leq \rho(\widehat{\mathbf{b}}_{ur}, \mathbb{W}) - \frac{\operatorname{tr}(A) - 4\lambda_{\max}(A)}{(c+1)}, \tag{A.30}
\end{aligned}$$

where in the last two inequalities we use (A.23) and the Jensen's inequality, respectively. Plugging $\mathbb{W} = V_{ur}^{-1}$ completes the proof of Theorem 4.

Furthermore, the difference between the asymptotic risk of the Stein-like shrinkage estimator and that of the minimal mean squared error estimator for any positive definite matrix \mathbb{W} is

$$\rho(\widehat{\mathbf{b}}_w, \mathbb{W}) - \rho(\widehat{\mathbf{b}}_\gamma, \mathbb{W}) = \frac{-4 \lambda_{\max}(A)}{(c+1) d} < 0, \tag{A.31}$$

where $d \equiv \operatorname{tr}(A)/\lambda_{\max}(A)$. Therefore, the Stein-like shrinkage estimator has a smaller asymptotic risk than the minimal mean squared error estimator. ■