Title
Essays on Estimation and Forecasting Under Structural Break Models

Author
Parsaeian, Shahnaz

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Essays on Estimation and Forecasting Under Structural Break Models

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Shahnaz Parsaeian

June 2020

Dissertation Committee:

Professor Aman Ullah, Co-Chairperson
Professor Tae-Hwy Lee, Co-Chairperson
Professor Gloria Gonzalez-Rivera

The Dissertation of Shahnaz Parsaeian is approved:

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside

Acknowledgments

I would like to express my most sincere gratitude to all those who provided me the support and encouragements during my Ph.D. study. Especially, I would like to give my sincere appreciation to my advisors, Professor Aman Ullah and Professor Tae-Hwy Lee for their guidance, encouragement and support throughout the whole career of my Ph.D study. I would like to thank them for allowing me to grow as a research scientist. I would also like to thank Professor Gloria Gonzalez-Rivera for serving as my committee member and for her constructive comments, and encouragement.

I would like to give my special thanks to Professor Marcelle Chauvet for her support and kindness during my PhD program. Many thanks to Professor Jean Helwege for her time and feedbacks. Thanks must also go to Mr. Gary Kuzas for his constant assistance during the last five years.

I would also like to express my thank to my family. For my parents who raised me with a love of science and supported me in all my pursuits. Words cannot express how grateful I am to my mother for all of the sacrifices that she has made on my behalf. Her prayer for me was what sustained me thus far. A special appreciation to my husband. There are no words to convey how much I love him. He has been a true and great supporter during my good and bad times. These past several years have not been an easy ride, both academically and personally. I truly thank him for sticking by my side.

Thanks must also go to all my instructors, professors, and supervisors who throughout my educational career have supported and encouraged me to do my best. Specially, I would like to thank Mr. Ali Bagherian, Professor GholamReza Keshavarz Haddad, Professor Farshad Fatemi, Professor Mahyar Hosseiny, and Professor Shiva Zamani for their unfailing support and kindness.

Thank you all from the bottom of my heart for never letting me down.

I dedicate this thesis to my mother and my husband for their constant support and unconditional love.

This thesis is particularly dedicated to the memory of my beloved father, whose love and support are still with me every single day!

I love you all dearly.

ABSTRACT OF THE DISSERTATION

Essays on Estimation and Forecasting Under Structural Break Models

by

Shahnaz Parsaeian

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2020
Professor Aman Ullah, Co-Chairperson
Professor Tae-Hwy Lee, Co-Chairperson

This dissertation covers topics in estimation and forecasting under structural breaks, in time-series and panel data models.

Chapter two considers the linear structural breaks model with m breaks ($m + 1$ regimes), and aim in improving the estimation of the parameters within each regime. We form an optimal combined estimator of regression parameters based on combining restricted estimator under the restriction of no breaks in the parameters, with unrestricted estimator which considers the observations within each regime separately. We derive the analytical finite sample risk and asymptotic risk and show that the risk of the combined estimator is less than the unrestricted estimator. The simulation study and the empirical example of forecasting the U.S. output growth confirm our theoretical findings.

Chapter three develops an optimal combined estimator to forecast out-of-sample under structural breaks. We propose the combined estimator of the post-break estimator with the full-sample estimator which uses all observations in the sample. Using a local asymptotic framework, we obtain the asymptotic risk for the combined estimator and show

that it is strictly less than the risk of the post-break estimator, which is a common solution for forecasting under structural breaks. We also introduce a semi-parametric estimator. Using a discrete kernel, this estimator assigns full weight of one to the post-break observations and down-weights the pre-break sample observations. The kernel is found by cross validation. Simulation study and the empirical example of forecasting equity premium confirm our analytical findings.

Chapter four proposes an efficient Stein-like shrinkage estimator for estimating the slope parameters in the heterogeneous panel data models with cross-sectional dependence. We combine the unrestricted estimator with the restricted one. The unrestricted estimator estimates the parameters by considering the break points and only uses the observations within each regime, while the restricted estimator estimates the parameters under the restriction of no breaks in the coefficients. We show analytically that the asymptotic risk of the combined estimator is less than the unrestricted estimator. We also show the superiority of the combined estimator over the unrestricted estimator in terms of the mean square forecast error. Simulation study verifies the main results of this chapter.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Time series structural breaks (structural changes) models have been intensively investigated over the last fifty years. Different kinds of estimation methods and testing procedures have been proposed in the econometric and statistical literature. There are many structural change tests in the literature including but not limited to the CUSUM test by Brown, Durbin and Evans (1975), sup-type test by Andrews (1993), exponential-type and average-type test by Andrews and Ploberger (1994), tests by Bai (1995) and Bai (1998) which consider the median estimation of a regression model with a single break and multiple breaks, extension of the sup-type test to models with multiple changes by Bai and Perron (1998, 2003), sup Wald test for a single change in a multivariate system by Bai et al. (1998), tests of structural changes in semi-parametric models by Su and White (2010), and group fused Lasso method by Qian and Su (2016). For other studies on structural change in a finite dimensional setup, see the comprehensive survey by Perron (2006), and Casini and Perron (2018).

In order to make well informed policy decisions, It is very important for government, central banks and other economic agents to consider the breaks in their analysis to determine whether the response of economic variables to any changes are immediate or gradual. This knowledge is also helpful for predicting the economy in the future. If we ignore the parameter changes, standard estimators will be inconsistent and statistical inference will be misleading. However, finding the true model and estimates carry uncertainty, and a forecaster needs to take this uncertainty into account to make an accurate forecast.

In this dissertation, we introduce different estimators that can improve the performance of estimation and forecasting in the sense of having smaller risk or smaller mean square forecast error. To evaluate the forecast accuracy, we use the quadratic loss function which by far is the most prevalent loss function. Based on this symmetric quadratic function, negative and positive forecast errors of the same magnitude have the same loss. Since this loss function is everywhere differentiable, it is easily optimized either analytically or by means of numerical techniques. To mention a few use of this type of loss function, see Hansen (2007, 2008, 2016, 2017), and Pesaran et al. (2013). As the mean square error is the sum of squared bias and variance, it is very crucial to find an estimator that ends up giving the lower possible square errors. Under the correctly specified model, whether we have breaks in the model or not, one can use the unbiased estimator and therefore the mean square error depends only on the variance. However, misspecification or insufficient data within each regime can potentially increase the variance of the estimator, and as a result increase the mean square error. So, there should be a tradeoff between the bias and variance efficiency that results in better estimates with lower error. Here, we try to find estimates and forecasts that optimally tradeoff the bias and variance efficiency to make an improvement in the performance of the estimators.

Chapter two focuses on the estimation of regression parameters under multiple structural breaks. The common method for estimating the coefficients under structural breaks is to use the observations within each regimes separately, and estimate the coefficients. But this estimator by itself may not necessarily minimize the risk, specially for the cases that there are not enough observations within each regimes. If the breakpoints are detected correctly, this results in an unbiased estimator but with high variance, due to insufficient observations. Therefore the mean square error increases which is not appealing. To solve this problem, in this chapter we propose a minimal mean square error estimator of regression parameters based on combining restricted estimator under the situation that there is no breaks in the parameters, with unrestricted estimator which estimates parameters using observations within each regime. The combination weight is between zero and one. The analytical finite sample risk is derived to obtain the optimal combined weight by minimizing the risk. We discuss the finite sample results based on the two well known approaches of the large-sample expansion proposed by Nagar (1959) and the small-disturbance method proposed by Kadane (1971). We show that for any break sizes and/or any break points, the finite sample risk of the combined estimator is strictly less than the unrestricted estimator, when the number of restrictions exceeds four. We also derive the asymptotic risk of the combined estimator and show the superiority of this estimator over the unrestricted estimator. We conduct a Monte Carlo simulation study for different break points and different break sizes in the coefficient and the error terms, and the results confirm our theoretical findings. Finally we use the big macroeconomic and financial time series database, described by McCracken and Ng (2016), to forecast the U.S. industrial production, and show that the proposed combined estimator results in lower mean square forecast error than the unrestricted estimator.

Chapter three develops an optimal combined estimator to forecast out-of-sample under structural breaks. The common solution for forecasting under structural breaks is to use the observations after the most recent breakpoints, and estimate the coefficient using those observations. As discussed in Pesaran and Timmermann (2007), Pesaran and Pick (2011) and Pesaran et al. (2013), this solution may not necessarily be optimal in the sense of mean square forecast error, especially when there are not enough observations in the post-break sample. Assuming that the break dates are accurately estimated, the post-break estimator is unbiased but its variance may be large due to the relatively small post-break observations. One solution is to include some or all of the pre-break observations which may bias the forecast, but decreases the variance. For this purpose, in this chapter we propose two different estimation methods that exploit the pre-break observations. The first proposed estimator is called the Stein-like combined estimator in which the combination weight takes the form of the James and Stein (1961). Massoumi (1978) introduced the Stein-like estimator for the simultaneous equation, and Hansen (2016, 2017) has used the Stein-type weight in different contexts.

In the proposed Stein-like combined estimator, we show how to combine the estimator using the full-sample (i.e., both the pre-break and post-break data) and the estimator using only the post-break sample. The full-sample estimator is inconsistent when there is a break in the coefficients while it is efficient. The post-break estimator is consistent but less efficient. Hence, depending on the severity of the breaks, the full-sample estimator and the post-break estimator can be combined to balance the consistency and efficiency. The combination weight depends on the break severity, which we measure by the Hausman statistic, see Hausman (1978). Small value of Hausman statistic can be interpreted as the small break size in the coefficients. This is the case that the bias of the full-sample estimator is small, and we can gain a lot from the efficiency of the full-sample estimator. Therefore, the combined estimator assigns more weight (or even full weight) to the full estimator. On the other hand, very large Hausman statistic is interpreted as a big break size in the coefficients which results in a large bias. Thus, the combined estimator assigns more weight (or even full weight) to the unrestricted estimator. Basically having the Hausman statistic in the combination weight helps to balance the bias-variance trade-off. Besides, using the local asymptotic alternative, we derive the asymptotic risk for the proposed Stein-like combined estimator, and show that its risk is less than the risk of the post-break estimator.

The second proposed method in chapter three that exploits the pre-break observations is a semi-parametric estimator. Motivation for the semi-parametric estimator comes from the paper by Li, Ouyang and Racine (2013), in which they focus on the categorical data. Inspired by that, we develop the semi-parametric estimator for the time series analysis, and specifically apply that for the structural break models. Using a discrete kernel, this estimator assigns the full weight of one to the post-break observations and down-weights the pre-break observations by a weight between zero and one. The kernel is based on the work of Aitchison and Aitken (1976) who offers the kernel smoothing of discrete covariates by borrowing information from neighboring subsets. This approach is semi-parametric since it uses kernel smoothing for covariates while the relationship between dependent variables and independent variables is parametrically specified. We find the kernel numerically by cross-validation (CV), and prove theoretically that the weight estimated by CV is optimal in the sense of optimality introduced by K.-C. Li (1987). That means, the weight estimated by CV is asymptotically equivalent to the infeasible optimal weight. Beside, we show that given the kernel, the mean square forecast error of the semi-parametric estimator is less than the post-break estimator. The great point about this estimator is that it facilitates our theoretical point without being involved with estimation error of some unknown parameters, like break size in the coefficients or the error terms. We examine the properties of the proposed methods, numerically by simulation, and also empirically in predicting the equity premium. The results confirm the out-performance of the both proposed estimators relative to the post-break estimators.

While most of the existing literature consider the time series models, as more data become available, it is necessary to test for the constancy of parameters in panel data models and improve the estimation performance of panel regression models under structural breaks. For this purpose, chapter four extends the idea of exploiting the pre-break observations to the panel data model in order to make an improvement in the estimation of the parameters and also forecasting under structural breaks. For detection of break points under panel data model see Bai (2010), Chen (2013), Kim (2011, 2014), Baltagi et al. (2016), Baltagi et al. (2017) among others. Given the break points, this chapter proposes an efficient Stein-like shrinkage estimator for estimating the slope parameters in the heterogeneous panel data regression models with cross-sectional dependence, which then can be used for forecasting by allowing for common structural breaks. The proposed method is the weighted average of the two estimators. One is the restricted estimator which estimates the parameters under the restriction of no breaks in the coefficients. This estimator is bias if we have breaks in the model, but efficient as we use more observations. The other one is called the unrestricted estimator in which it estimates the parameters by considering the break points, so it only uses the observations within each regime across all individuals. This is the unbiased estimator but less efficient. Thus, combining these two estimators trade-off the bias and variance efficiency. The combination weight is proportional to the weighted quadratic loss function that based on the severity of the breaks, assigns appropriate weight to each of the estimators. We derive the asymptotic distribution and asymptotic risk for the shrinkage estimator and find the optimal averaging weight by minimizing the risk. We show that the proposed Stein-like shrinkage estimator performs better than the unrestricted estimator in the sense of having smaller risk. We also show the superiority of the shrinkage estimator over the unrestricted estimator in terms of the mean square forecast error. Additionally, Monte Carlo simulations and empirical study in forecasting U.S. industry level inflation rates are used to verify the main results of the proposed estimator.

Chapter 2

Efficient Combined Estimation under Structural Breaks

2.1 Introduction

Many macroeconomic and financial time series are subject to structural breaks. Structural break in linear regressions was considered early on by Chow (1960), and Quandt (1960). Seminal works were mostly designed for the specific case of testing a single break. See Andrews (1993) who proposes a supremum-type test, Andrews and Ploberger (1994) consider the exponential-type and average-type tests, Bai (1995), Bai (1997a), Bai (1998) inter alia. Later on, these methods were extended to detect the multiple structural breaks. Sequential tests for the null of m versus $m+1$ breaks are provided in Bai and Perron (1998) and in Bai (1997b). Besides, Bai (1999) proposes a sequential likelihood ratio test for the null of m versus $m+1$ breaks, where all break points are jointly estimated. See also Bai et al. (1998) for multivariate time series. There are many other statistical procedures that can be used for detection of break points, such as Andrews et al. (1996); Bai and Perron (2003); Altissimo et al. (2003); Qu and Perron (2007); Qian and Su (2016). The literature on detecting the structural break is massive and there are some cost efficient programs to detect the breaks. For a comprehensive survey on structural changes, see Perron (2006), and Casini and Perron (2018).

This chapter does not focus on methods for identifying the break points, as this issue has been paid enough attention in the literature. Instead, the goal of this chapter is to propose a combined estimator with a minimum risk under the assumption that structural break has in fact occurred. For estimation of the break points, we follow Bai and Perron (1998, 2003) which is a consistent global minimizer method of the error term.

The common method for estimating the coefficients under structural breaks (after detecting the break points) is to use the information within each regimes separately, and estimate the coefficients. But this estimator by itself may not necessarily minimize the risk in the case that break points are close to each other or there are not enough data to accurately estimate the coefficients within each regime. If the distance to break is short, then the parameters are likely to be poorly estimated relative to those obtained using more data. To overcome this problem, we propose the combined estimator of the unrestricted estimator, in which we estimate the coefficients within each regime separately only by using the observations on that specific regime, and the restricted estimator. The restricted estimator uses all the observations in the sample, $t = \{1, \ldots, T\}$, to estimate the coefficients. So, it is under the restriction that there is no break in the model or it can be replaced by any user specific restriction on the parameters of the model. The advantage of imposing this restriction is that sometimes the break size are small, so precisely detecting the break points is difficult or not possible. Even under detectable break points, ignoring the break point and estimating the coefficients by using all observations, gives us the better estimate.

In this chapter, we focus on the estimation of regression parameters under multiple structural breaks. We propose a minimal mean square error estimator of regression parameters based on combining restricted estimator under the situation that there is no break in the parameters, with unrestricted estimator under the break. The combination weight is between zero and one, and we derive the condition under which this estimator outperforms the unrestricted estimator, in the sense of minimizing the risk. We derive the finite sample properties for the proposed combined estimator, and show that the combined estimator has a lower risk than unrestricted estimator which is the common solution for estimating the parameters under structural break. Doing that, we discuss the finite sample results based on the two well known approaches of the large-sample expansion proposed by Nagar (1959) and the small-disturbance method proposed by Kadane (1971). Besides, we derive the asymptotic results for our estimator and show the dominance condition of that over the unrestricted estimator.

We perform Monte Carlo experiments to evaluate the performance of the proposed combined estimator. The results confirm the theoretically expected improvements in the combined estimator in compare to the unrestricted estimator. As an empirical example, we work with a large macroeconomic and financial time series. We forecast the output growth rate for $1, 6$, and 12 forecast horizons and show the outperformance of our proposed combined estimator relative to the unrestricted estimator.

The outline of the chapter is as follows. Section 2.2 sets up the model under multiple structural break models. Section 2.3 introduces a minimal mean square error combined estimator with weight between zero and one, and derives its finite sample risk based on the large-sample expansion method and the small-disturbance method. Section 2.4 derives the asymptotic risk for the proposed combined estimator. Monte Carlo experiments are presented in Section 2.5 while Section 2.6 presents the results of an empirical study. Finally Section 2.7 concludes the chapter. All proofs are relegated to the Appendix.

2.2 The structural breaks model

Consider the linear structural break model with m breaks or $m+1$ regimes. There are T observations and m is assumed known. The break dates occur at $\{T_1, T_2, \dots, T_m\}$. Suppose the structural breaks model has the following form

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}u_t & \text{for} \quad 1 < t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)}u_t & \text{for} \quad T_1 < t \leq T_2 \\ \vdots & \\ x_t'\beta_{(m+1)} + \sigma_{(m+1)}u_t & \text{for} \quad T_m < t < T, \end{cases} \tag{2.1}$$

where x_t is $k \times 1$ exogenous regressors, and $u_t \sim$ i.i.d.$(0,1)$. In matrix notation,

$$Y = X\beta + \epsilon, \tag{2.2}$$

where $Y = (y_1, \dots, y_T)'$ is a $T \times 1$ vector of dependent variables, $X = (X_1', \dots, X_{m+1}')'$ is a $T \times (m+1)k$ block diagonal matrix of regressors, where $X_i = (x_{T_{i-1}+1}, \dots, x_{T_i})'$, with $i = \{1, \dots, m+1\}$, and the convention that $T_0 = 0$, and $T_{m+1} = T$. Also, $\beta = (\beta_{(1)}', \dots, \beta_{(m+1)}')'$

is a $(m+1)k \times 1$ vector of coefficients, $\epsilon = (\epsilon_1, \ldots, \epsilon_{m+1})'$ is a $T \times 1$ vector of error terms, with $\epsilon_i = \sigma_{(i)}(u_{T_{i-1}+1}, \ldots, u_{T_i})'$, such that

$$\epsilon = \begin{cases} \sigma_{(1)}(u_1, \ldots, u_{T_1})' & \text{for} \quad 1 \; < t \leq T_1 \\ \sigma_{(2)}(u_{T_{1+1}}, \ldots, u_{T_2})' & \text{for} \quad T_1 < t \leq T_2 \\ \vdots & \qquad\qquad \vdots \\ \sigma_{(m+1)}(u_{T_{m+1}}, \ldots, u_T)' & \text{for} \quad T_m < t < T, \end{cases} \tag{2.3}$$

in which $\epsilon \sim \mathrm{N}(0, \Omega)$ with $\Omega = \mathrm{diag}\big(\sigma^2_{(1)}\iota_{l_1}, \ldots, \sigma^2_{(m+1)}\iota_{l_{m+1}}\big)$, where ι is a vector of ones and $l_i \equiv T_i - T_{i-1}$. Later on, in the asymptotic part in Section 2.4, we will relax the normality assumption on the error term.

As we mention earlier, the focus of this chapter is not on estimating the break points. So, given the break points, we want to introduce a combined estimator which has a lower risk in compare to both restricted estimator and unrestricted estimator. In the following Section, we introduce the combined estimator which consists of the unrestricted estimator and the restricted estimator with restriction on coefficients. Under the null hypothesis, we define $R\beta = r = \mathbf{0}$, in which r is a $p \times 1$ vector of zero and R is a $p \times (m+1)k$ convention matrix with rank p, which shows the number of restrictions, as

$$R = \begin{bmatrix} -I_k & I_k & 0 & & 0 & 0 & 0 \\ 0 & -I_k & I_k & & 0 & 0 & 0 \\ & & & \vdots \; \vdots & & & \\ 0 & 0 & 0 & & -I_k & I_k & 0 \\ 0 & 0 & 0 & & 0 & -I_k & I_k \end{bmatrix}. \tag{2.4}$$

Basically, the convention matrix R is in a way that shows the difference between coefficients, $R\beta = \big(\beta_{(2)} - \beta_{(1)}, \beta_{(3)} - \beta_{(2)}, \ldots, \beta_{(m+1)} - \beta_{(m)}\big)'$. Under the alternative hypothesis, $R\beta \neq r$.

2.3 Combined estimator and its finite sample risk

We propose a minimal mean square error combined estimator of β as the combination of the restricted estimator and the unrestricted estimator with a combination weight $\gamma \in [0\ 1]$ as

$$\widehat{\beta}_\gamma = (1-\gamma)\widehat{\beta}_{ur} + \gamma\widehat{\beta}_r, \tag{2.5}$$

where $\widehat{\beta}_{ur}$ and $\widehat{\beta}_r$ are the infeasible unrestricted estimator and the infeasible restricted estimator, respectively, which can be estimated using the generalized least square method (GLS) as

$$\widehat{\beta}_{ur} = \beta + \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\epsilon, \tag{2.6}$$

and

$$\begin{aligned}\widehat{\beta}_r &= \widehat{\beta}_{ur} - \underbrace{\left(X'\Omega^{-1}X\right)^{-1}R'\left[R\left(X'\Omega^{-1}X\right)^{-1}R'\right]^{-1}R}\widehat{\beta}_{ur} \\ &= \widehat{\beta}_{ur} - L\widehat{\beta}_{ur},\end{aligned} \tag{2.7}$$

where $L \equiv \left(X'\Omega^{-1}X\right)^{-1}R'\left[R\left(X'\Omega^{-1}X\right)^{-1}R'\right]^{-1}R$ is a $(m+1)k \times (m+1)k$ matrix.

In the following subsections, we derive the finite sample results based on the large-T expansion and the small-disturbance methods.

2.3.1 Large-sample expansion method

Here we want to analyze the finite sample properties of the proposed combined estimator using its approximate moments. There are two well known and frequently used approach in this area. One is the large-sample expansion method proposed by Nagar (1959) which we discuss it in this Section, and the other is small-disturbance method proposed by Kadane (1971) which we relegate for Subsection 2.3.2.

As γ in (2.5) is unknown, the first step is to find its optimal value. We derive the exact risk for the combined estimator given a known Ω, and minimize the risk to find the optimal value of the weight. The risk of the combined estimator is

$$\begin{aligned} Risk(\widehat{\beta}_\gamma, W) &= \mathbb{E}\left[(\widehat{\beta}_\gamma - \beta)' W (\widehat{\beta}_\gamma - \beta)\right] \\ &= Risk(\widehat{\beta}_{ur}, W) + \gamma^2 \Big[\beta' L' W L \beta + \mathrm{tr}\big((X'\Omega^{-1}X)^{-1} L' W L\big)\Big] \\ &- 2\gamma \ \mathrm{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big). \end{aligned} \tag{2.8}$$

By minimizing the risk with respect to γ in (2.8), the optimal value of the weight denoted by γ^* is

$$\gamma^* = \frac{\mathrm{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}{\beta' L' W L \beta + \mathrm{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}, \tag{2.9}$$

which by plugging the unbiased estimator of its denominator we have

$$\gamma^* = \frac{\mathrm{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}{\widehat{\beta}'_{ur} L' W L \widehat{\beta}_{ur}}. \tag{2.10}$$

See Appendix A for the proof of (2.10). Note that the optimal weight depend on the unknown value Ω which later we will replace it with its estimate.

Define notation $\bar{\bar{\widehat{\beta}}}$ as a feasible estimator of β. As we defined the unrestricted estimator earlier, it is for the structural break case that we estimate the coefficients within each regime separately, only by using the observations within that regime. The feasible unrestricted estimator is

$$\bar{\bar{\widehat{\beta}}}_{ur} = \beta + \left(X'\widehat{\Omega}^{-1}X\right)^{-1} X'\widehat{\Omega}^{-1}\epsilon, \tag{2.11}$$

where $\beta = \left(\beta'_{(1)}, \ldots, \beta'_{(m+1)}\right)'$ is $(m+1)k \times 1$, and

$\widehat{\Omega} = \mathrm{diag}\big(\widehat{\sigma^2_{(1)}} \ \iota_{l_1}, \ldots, \widehat{\sigma^2_{(m+1)}} \ \iota_{l_{m+1}}\big) = \mathrm{diag}\big(S_1 \ \iota_{l_1}, \ldots, S_{m+1} \ \iota_{l_{m+1}}\big)$ where ι is a vector of ones and S_i is a consistent estimates of the $\sigma^2_{(i)}$ in which $S_i = \frac{\epsilon_i' M_i \epsilon_i}{l_i - k}$, $l_i \equiv T_i - T_{i-1}$ and

$M_i = I_{l_i} - X_i(X_i'X_i)^{-1}X_i'$ with $i = \{1, \ldots, m+1\}$. See Appendix A for the proof and details.

Note that since the Ω matrix is diagonal, we can rewrite the unrestricted estimator as a ordinary least square method. That means that, we do not need to plug in the estimates of Ω into equation (2.11) and expanding the terms. Therefore, the feasible unrestricted estimator is

$$\begin{aligned}\bar{\bar{\beta}}_{ur} - \beta &= \left(X'\widehat{\Omega}^{-1}X\right)^{-1}X'\widehat{\Omega}^{-1}\epsilon \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon.\end{aligned} \tag{2.12}$$

For the restricted estimator, $R\beta = r = \mathbf{0}$. For example, we can impose the restriction that all coefficients across regimes are equall or the coefficients in some specific regimes are equal to each other or any other restrictions. Under the alternative hypothesis $R\beta \neq \mathbf{0}$. Restricting some of the coefficients to be identical across some regimes converts the model to the partial structural change model and it is useful since it allows for a broader range of practical interest, like partial structural change model. Perron and Qu (2006) show that the estimate of the break points have the same asymptotic properties with or without the restriction. Also, in finite sample the improvements for the estimation of breakpoints can be obtained by imposing the restriction. The feasible restricted estimator is

$$\begin{aligned}\bar{\bar{\beta}}_{r} &= \bar{\bar{\beta}}_{ur} - \left(X'\widehat{\Omega}^{-1}X\right)^{-1}R'\left[R\left(X'\widehat{\Omega}^{-1}X\right)^{-1}R'\right]^{-1}R\bar{\bar{\beta}}_{ur} \\ &= \bar{\bar{\beta}}_{ur} - \widehat{L}\bar{\bar{\beta}}_{ur},\end{aligned} \tag{2.13}$$

where $\widehat{L} = \left(X'\widehat{\Omega}^{-1}X\right)^{-1}R'\left[R\left(X'\widehat{\Omega}^{-1}X\right)^{-1}R'\right]^{-1}R$.

Having the feasible restricted and unrestricted estimators, and the feasible combination weight, $\widehat{\gamma}^{*}$, we can calculate the bias, MSE and the risk of the feasible combined estimator. Theorem 1 show the results. Note that the feasible combination weight is

$$\widehat{\gamma}^* = \frac{\mathrm{tr}\big((X'\widehat{\Omega}^{-1}X)^{-1}\widehat{L}'W\big)}{\bar{\bar{\beta}}'_{ur}\widehat{L}'W\widehat{L}\bar{\bar{\beta}}_{ur}}. \tag{2.14}$$

Theorem 1 *The bias, to order $O_p(T^{-1})$, of $\bar{\bar{\beta}}_\gamma$ is given by*

$$bias(\bar{\bar{\beta}}_\gamma) = -\frac{tr(Q)}{\phi} L\beta, \tag{2.15}$$

where $Q \equiv W^{1/2}L(X'\Omega^{-1}X)^{-1}W^{1/2}$, $\phi \equiv \beta'L'WL\beta$, and $W > 0$ is any user specific choice of weight which is assumed to be O(T), and the second order moment matrix, to order $O(T^{-2})$, of $\bar{\bar{\beta}}_\gamma$ around β is

$$\begin{aligned} MSE(\bar{\bar{\beta}}_\gamma) &= \mathbb{E}\left[(\bar{\bar{\beta}}_\gamma - \beta)(\bar{\bar{\beta}}_\gamma - \beta)'\right] \\ &= MSE(\bar{\bar{\beta}}_{ur}) + \frac{1}{\phi^2}\ L\beta\beta'L'\Big(tr(Q)\Big)^2 - \frac{2\ tr(Q)}{\phi} L(X'\Omega^{-1}X)^{-1} \\ &+ \frac{2\ tr(Q)}{\phi^2}\ L\beta\beta'L'WL(X'\Omega^{-1}X)^{-1} + \frac{2\ tr(Q)}{\phi^2}\ (X'\Omega^{-1}X)^{-1}L'WL\beta\beta'L', \end{aligned} \tag{2.16}$$

where $MSE(\bar{\bar{\beta}}_{ur}) = (X'\Omega^{-1}X)^{-1}$, and the risk, to order $O(T^{-1})$, associated with $\bar{\bar{\beta}}_\gamma$ is given by

$$Risk(\bar{\bar{\beta}}_\gamma, W) = Risk(\bar{\bar{\beta}}_{ur}, W) - \frac{\Big(tr(Q)\Big)^2}{\phi^2}\left\{\phi - \frac{4\big(\beta'L'WL(X'\Omega^{-1}X)^{-1}WL\beta\big)}{tr(Q)}\right\}, \tag{2.17}$$

where $Risk(\bar{\bar{\beta}}_{ur}, W) = tr\big((X'\Omega^{-1}X)^{-1}W\big)$, and the risk of the combined estimator is less than the unrestricted estimator as long as the term inside the curly bracket be positive, $tr(Q) > 4\ \lambda_{max}(Q)$, where $\lambda_{max}(Q)$ represents the maximum eigenvalues of Q. ■

See Appendix A for the proof of Theorem 1.

Corollary 2 *The finite sample risk for the combined estimator with $W = X'\Omega^{-1}X$ is*

$$Risk(\bar{\bar{\beta}}_{\gamma}, W) = Risk(\bar{\bar{\beta}}_{ur}, W) - \frac{p^2}{\phi}\left\{1 - \frac{4}{p}\right\}, \tag{2.18}$$

where the risk of the combined estimator is less than the unrestricted estimator if $p > 4$ where p is the number of restrictions for the restricted estimator. □

2.3.2 Small-disturbance method

Here, we want to find the finite sample MSE for the combined estimator using the Small-disturbance method developed by Kadane (1971) who applied this method to a linear, normal simultaneous equation system. For this section, we assume that $\sigma_{(1)} = \cdots = \sigma_{(m+1)} = \sigma$, so the model (2.1) in matrix notation can be written as $Y = X\beta + \sigma\epsilon$, where $\epsilon \sim \mathrm{N}(0, I_T)$ [1]. Therefore, the feasible unrestricted estimator is

$$\bar{\bar{\beta}}_{ur} - \beta = \sigma\big(X'X\big)^{-1}X'\epsilon, \tag{2.19}$$

and the feasible restricted estimator is

$$\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_{r} = L\bar{\bar{\beta}}_{ur}, \tag{2.20}$$

where $L = \big(X'X\big)^{-1}R'\left[R\ \big(X'X\big)^{-1}R'\right]^{-1}R$. Similar to Subsection 2.3.1, by having the combination weight, $\gamma^* = \frac{\sigma^2 \mathrm{tr}(Q_s)}{\bar{\bar{\beta}}'_{ur}L'WL\bar{\bar{\beta}}_{ur}}$ where $Q_s \equiv W^{1/2}L(X'X)^{-1}W^{1/2}$, we can derive the moments for the combined estimator using the small-disturbance method.

Replacing σ^2 with $\widehat{\sigma^2}$, we can get the feasible combination weight, $\widehat{\gamma}^*$, and the feasible combined estimator accordingly. Having that, we calculate the bias up to order σ^2, and the finite sample risk up to order σ^4 for the feasible combined estimator. Theorem 3 shows the results.

[1] Because of this transformation, we substitute Ω by $\sigma^2 I_T$ in this subsection.

Theorem 3 *The bias up to order σ^2 for the combined estimator is*

$$bias(\bar{\bar{\hat{\beta}}}_\gamma) = -\frac{\sigma^2 tr(Q_s)}{\phi} L\beta. \tag{2.21}$$

Also, the MSE, up to order σ^4, is

$$\begin{aligned} MSE(\bar{\bar{\hat{\beta}}}_\gamma) = MSE(\bar{\bar{\hat{\beta}}}_{ur}) &+ \frac{\sigma^4 \left(T-(m+1)k+2\right) \left[tr(Q_s)\right]^2}{\phi^2 \left(T-(m+1)k\right)} L\beta\beta' L' \\ &- \frac{2\sigma^4 \ tr(Q_s)}{\phi} L(X'X)^{-1} + \frac{4\sigma^4 \ tr(Q_s)}{\phi^2} L\beta\beta' L' W L (X'X)^{-1}, \end{aligned} \tag{2.22}$$

where $MSE(\bar{\bar{\hat{\beta}}}_{ur}) = \sigma^2 (X'X)^{-1}$, and the finite sample risk, up to order σ^4, and for any user specific choice of W is

$$\begin{aligned} Risk(\bar{\bar{\hat{\beta}}}_\gamma, W) = Risk(\bar{\bar{\hat{\beta}}}_{ur}, W) &- \frac{\sigma^4 \left[tr(Q_s)\right]^2}{\phi^2} \left\{ \left(1 - \frac{2}{T-(m+1)k}\right) \phi \right. \\ &\left. - \frac{4\beta' L' W L (X'X)^{-1} W L \beta}{tr(Q_s)} \right\}, \end{aligned} \tag{2.23}$$

where $Risk(\bar{\bar{\hat{\beta}}}_{ur}, W) = \sigma^2 tr\big((X'X)^{-1} W\big)$. Therefore, the risk of the combined estimator is less than the unrestricted estimator as long as the term inside the curly bracket be positive which requires $tr(Q_s)\Big(1 - \frac{2}{T-(m+1)k}\Big) > 4 \ \lambda_{max}(Q_s)$. ■

See Appendix A for the proof and details of Theorem 3. To have a better sense of Theorem 3, we can derive the risk for a specific choice of $W = X'X/\sigma^2$ which simplifies calculations a bit. Note that if $W = X'X/\sigma^2$, then $\text{tr}(Q_s) = \frac{1}{\sigma^2}\text{tr}(L) = p/\sigma^2$. Corollary 4 shows the risk for this specific choice of the weight.

Corollary 4 *The finite sample risk for the combined estimator with $W = X'X/\sigma^2$ is*

$$Risk(\bar{\bar{\hat{\beta}}}_\gamma, W) = Risk(\bar{\bar{\hat{\beta}}}_{ur}, W) - \frac{\sigma^2 \ p^2}{\beta' L' X' X L \beta} \left\{ 1 - \frac{2}{T-(m+1)k} - \frac{4}{p} \right\}, \tag{2.24}$$

where the risk of the combined estimator is less than the unrestricted estimator if the term inside the curly bracket be positive, $p > 4 + \frac{8}{T-(m+1)k-2}$. □

2.4 Asymptotic risk for the combined estimator

In this section we relax the normality assumption on the error term and derive the asymptotic risk for our proposed combined estimator in equation (2.5). Also, $\Omega = \mathrm{diag}\big(\sigma^2_{(1)}\iota_{l_1}, \ldots, \sigma^2_{(m+1)}\iota_{l_{m+1}}\big)$, which shows the breaks in the error term. Let $\widehat{\Omega}$ denote any consistent estimate of the asymptotic variance of the error term, Ω. Theorem 5 shows the asymptotic distribution for the unrestricted estimator and the restricted estimator in which we use the local asymptotic framework. So, under the null hypothesis there is no break in the model, and under the alternative, $\beta = \beta_0 + \frac{h}{\sqrt{T}}$ where β_0 is the true parameter value and h shows the size of the break in the coefficients.

Theorem 5 *The asymptotic distribution of the unrestricted estimator is*

$$\sqrt{T}\big(\bar{\widehat{\beta}}_{ur} - \beta\big) \xrightarrow{d} Z \sim N(0, V_{ur}), \tag{2.25}$$

where $V_{ur} = \underset{T\to\infty}{plim}\Big(\frac{X'\ \Omega^{-1}\ X}{T}\Big)^{-1}$. *Also, the asymptotic distribution of the restricted estimator is*

$$\sqrt{T}\big(\bar{\widehat{\beta}}_{ur} - \bar{\widehat{\beta}}_{r}\big) \xrightarrow{d} V_{ur}\ PR(Z + h), \tag{2.26}$$

where $P \equiv R'\ \Big[R\ \Big(\frac{X'\ \Omega^{-1}X}{T}\Big)^{-1} R'\Big]^{-1}$ *which is* $(m+1)k \times p$. ■

See Appendix A for the proof of Theorem 5. Using the results of Theorem 5, we can find the unknown value of the combination weight by minimizing the risk. The risk of the combined estimator is

$$\begin{aligned} Risk(\widehat{\beta}_\gamma, W) &= T\,\mathbb{E}\left[\big(\widehat{\beta}_\gamma - \beta\big)'W\ \big(\widehat{\beta}_\gamma - \beta\big)\right] \\ &= \mathrm{tr}(V_{ur}W) - \gamma\Big[2\ \mathrm{tr}(Q) - \gamma\big(h'Bh + \mathrm{tr}(Q)\big)\Big], \end{aligned} \tag{2.27}$$

where $Q = W^{1/2} V_{ur} R' P' V_{ur} W^{1/2}$, $B \equiv R' P' V_{ur} W V_{ur} P R$. Minimizing the risk with respect to γ, we find the optimal value of weight as

$$\gamma^* = \frac{\text{tr}(Q)}{h' B h + \text{tr}(Q)}, \tag{2.28}$$

and by replacing the denominator of (2.28) with its unbiased estimator, we have

$$\gamma^* = \frac{\text{tr}(Q)}{\widehat{h}' B \widehat{h}}. \tag{2.29}$$

Having the combination weight, we can derive the asymptotic distribution and the asymptotic risk for the combined estimator. Theorem 6 shows the results.

Theorem 6 The asymptotic distribution of the proposed combined estimator is

$$\begin{aligned} \sqrt{T}\big(\bar{\bar{\beta}}_\gamma - \beta\big) &= \sqrt{T}\big(\bar{\bar{\beta}}_{ur} - \beta\big) - \widehat{\gamma}^* \sqrt{T}\big(\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r\big) \\ &\xrightarrow{d} Z - \gamma^* \ V_{ur} P R (Z + h), \end{aligned} \tag{2.30}$$

and its asymptotic risk is

$$Risk(\bar{\bar{\beta}}_\gamma, W) \leq Risk(\bar{\bar{\beta}}_{ur}, W) - \frac{tr(Q) - 4\lambda_{max}(Q)}{c+1}, \tag{2.31}$$

where $Risk(\bar{\bar{\beta}}_{ur}, W) = tr(V_{ur} W)$*, and* $0 < c < \infty$*.* Thus, the risk of the combined estimator is less than the unrestricted estimator if $tr(Q) > 4 \ \lambda_{max}(Q)$. ■

See Appendix A for the proof of Theorem 6.

Corollary 7 *The asymptotic risk for the combined estimator with weight* $\gamma \in [0 \ 1]$ *with* $W = V_{ur}^{-1}$ *is*

$$Risk(\bar{\bar{\beta}}_\gamma, W) \leq Risk(\bar{\bar{\beta}}_{ur}, W) - \frac{p-4}{c+1}, \tag{2.32}$$

where the risk of the combined estimator is less than the unrestricted estimator if $p > 4$*.* □

2.5 Monte Carlo simulation

This section provides some Monte Carlo results for the proposed combined estimators. The goal is to compare the risk of the unrestricted estimator with the proposed combined estimator, and show that the risk of the proposed combined estimators are lower than the unrestricted estimator if the number of restrictions exceeds four, regardless of the breakpoints and the break size. To do this, let $t = 1, ..., T$ with $T \in \{100, 200\}$, and $k \in \{5, 8\}$. We try different values for breakpoints which are proportional to the sample observations, $b_1 = \frac{T_1}{T} \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$. We also set $W \in \{I_{2k}, V_{ur}^{-1}\}$, where $W = I$ shows the MSE. We generate x_t and ε_t such that $x_t \sim N(0,1)$, and $u_t \sim N(0,1)$. The data generating process uses the following model

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)} u_t & \text{for} \quad 1 \leq t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)} u_t & \text{for} \quad T_1 < t \leq T. \end{cases} \tag{2.33}$$

Let $\beta_{(1)}$ be a vector of ones, and $\lambda \equiv \beta_{(2)} - \beta_{(1)} \in \{0, 0.25, 0.5, 0.75, 1\}$ shows the break size in the coefficients. Also, $q = \sigma_{(1)}/\sigma_{(2)}$ shows the ratio of the break in the error term where we set $q \in \{0.5, 1, 2\}$. The number of replications is 1,000.

Tables 2.1-2.12 show the results for this experiment. The Tables show the relative MSE with respect to unrestricted estimator, i.e.

$$RMSE_\gamma = \frac{MSE(\widehat{\beta}_\gamma)}{MSE(\widehat{\beta}_{ur})}, \tag{2.34}$$

for different break points and different number of regressors. In all tables, the benchmark model is the unrestricted estimator. The first column in tables shows the sample size while the second column shows different break size in the coefficient, λ. As it is clear from the results, the performance of the proposed combined estimator with weight γ in all situations, regardless of the choice of W or the breakpoints, is better than the unrestricted estimator in the sense of having a lower risk or MSE. This confirms our results in the theoretical part. Besides, as we increase the number of regressors, k, the ratio of Mean Square Error ($RMSE$) gets smaller, especially for the small break sizes (approximately less than 0.5).

2.6 Empirical analysis

We asses the performance of our proposed method by applying that to the 130 macroeconomic and financial time series from the St. Louis Federal Reserve (FRED-MD) database. We use the monthly data from Jan 1959 up to Mar 2020. The data are described by McCracken and Ng (2016), who suggest various transformations to render the series stationary and to deal with missing values. After losing two observations to data transformation, the panel we use for analysis is for the sample 1959 : 03 to 2020 : 03 with $T = 732$ observations. The data are split into 8 groups: output and income (17 series), labor market (32 series), consumption and orders (10 series), orders and inventories (11 series), money and credit (14 series), interest rates and exchange rates (21 series), prices (21 series) and stock market (4 series).

As suggested by McCracken and Ng (2016), in a large N and large T dimension, we can use diffusion index forecasting and estimate the factor augmented regression to reduce the dimension. We estimate the static factors by principal component analysis (PCA) adapted to allow for missing values. We then select the number of significant factors using the criteria developed in Bai and Ng (2002), which is a generalization of Mallow's C_p criteria for large dimensional panels. The criterion finds eight factors in this sample. The seven factors can be interpreted as real activity/employment, inflation, term spreads, housing, interest rate variables, stock market variables, output and inventories factors.

We can evaluate the usefulness of the estimated factors by forecasting the U.S. output growth at the $1, 6$ and 12 month horizons. The model that we use for forecasting takes the form of

$$y_{t+h}^h = \beta_h' \ \widehat{f}_{t+h-1} + \gamma_h \ y_{t+h-1} + \varepsilon_{t+h}^h, \tag{2.35}$$

where y_{t+h}^h denote output growth over the next h months, expressed at an annual rate, i.e., $y_{t+h}^h = (1200/h) \ \ln(\text{IP}_{t+h}/\text{IP}_t)$. Also $\widehat{f}_{t+h-1}$ is the estimated eight factors at time $t+h-1$. In order to evaluate the performance of our proposed estimator, we compute the out of sample MSFE and compare them with MSFE from the unrestricted estimator. For this purpose, we divide the sample of T observations into two parts. The first n_1 observations is used as an in-sample estimation period, and the remaining $n_2 = T - n_1$ observations is the out-of-sample period which we recursively do one step ahead forecast. Each time that we expand the window, initially we identify break points by the sequential procedure introduced by Bai and Perron (1998, 2003), where we search for up to eight breaks and set the trimming parameter to 0.1 and the significance level to 5%. Using an initial estimation period of $n_1 = 130$ months (around 11 years) forecasts are recursively generated at each point in the out-of-sample period using only the information available at the time the forecast is made. As the selection of the forecast evaluation period is always somewhat arbitrary, we also report the results with an alternative estimation window sizes, so the beginning of the various forecast evaluation periods runs from $1970:01$ ($n_1 = 130$) through $1990:01$ ($n_1 = 370$). The results are qualitatively similar when a larger number of estimation period is used. The baseline forecast uses the observations after the last break. We compare the forecast based on the unrestricted estimator with our proposed combined estimator forecast. Table 2.13 reports the ratio of MSFE over the benchmark model. h in the first column shows the forecast horizon. The second column shows the start date of the out-of sample period which all ends at 2020:03. In the heading of table, $MSFE_{ur}$ is for the case that we only use post-break observations, and $MSFE_{\gamma}$ represents the results for the γ weight combined estimator. ** and * indicate significance at 5% and 10% based on the Diebold and Mariano (1995). Based on the results, the proposed estimator delivers vastly improved forecasts (lower MSFE) compared to the unrestricted estimator for all horizons.

2.7 Conclusion

We introduce the combined estimator of the unrestricted estimator with the restricted estimator to deal with the estimation of the coefficients under structural breaks. The combination weight is between zero and one. We derive the finite sample risk and asymptotic risk for this estimator and prove that the risk of this estimator is lower than the unrestricted estimator. Monte Carlo experiments show the improvement in the risk of the combined estimator over the unrestricted estimator. Based on the Monte Carlo results, for the small break sizes (approximately less than 0.5), we can see a vast improvement relative to the unrestricted estimator. Also, as we increase the number of regressors, we get a lower risk by using the introduced combined estimator. For the large break sizes, we can still see improvement relative to the unrestricted estimator, but not as much as the small breaks. We also apply our estimator for generating the out of sample forecast and use the model for forecasting the U.S. output growth. We show that the MSFE of the proposed estimator is less than the unrestricted estimator, and has a power to forecast under longer horizons.

Table 2.1: Simulation results for $k = 5$, $q = 0.5$, $W = X'\Omega^{-1}X$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.845	0.801	0.783	0.788	0.783	0.776	0.785
	0.25	0.976	0.984	0.979	0.976	0.966	0.955	0.919
$T = 100$	0.50	0.995	0.996	0.996	0.997	0.993	0.992	0.993
	0.75	0.993	0.998	0.997	0.999	0.999	0.999	1.001
	1.00	0.995	0.997	0.999	0.999	0.999	0.998	0.994
	0.00	0.902	0.839	0.843	0.832	0.826	0.750	0.833
	0.25	0.988	0.995	0.996	0.993	0.990	0.984	0.970
$T = 200$	0.50	0.996	0.999	1.000	1.000	1.000	1.001	0.997
	0.75	0.998	1.000	1.000	0.998	0.999	1.001	0.997
	1.00	0.998	0.999	1.000	1.000	1.000	1.000	0.998

Table 2.2: Simulation results for $k = 8$, $q = 0.5$, $W = X'\Omega^{-1}X$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.665	0.630	0.612	0.554	0.537	0.518	0.541
	0.25	0.955	0.971	0.964	0.952	0.932	0.905	0.830
$T = 100$	0.50	0.989	0.991	0.991	0.989	0.983	0.976	0.958
	0.75	0.994	0.997	0.995	0.995	0.994	0.990	0.980
	1.00	0.994	0.996	0.997	0.996	0.995	0.993	0.983
	0.00	0.687	0.666	0.631	0.588	0.557	0.470	0.446
	0.25	0.983	0.988	0.986	0.982	0.976	0.965	0.941
$T = 200$	0.50	0.995	0.997	0.997	0.996	0.995	0.995	0.987
	0.75	0.997	0.998	0.999	0.998	0.999	0.999	0.996
	1.00	0.998	0.999	1.000	1.000	0.999	0.999	0.996

Table 2.3: Simulation results for $k = 5$, $q = 1$, $W = X'\Omega^{-1}X$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.749	0.791	0.791	0.790	0.891	0.816	0.843
	0.25	0.924	0.960	0.962	0.972	0.966	0.957	0.921
$T = 100$	0.50	0.972	0.980	0.985	0.992	0.991	0.991	0.992
	0.75	0.978	0.990	0.992	0.995	0.996	0.998	1.000
	1.00	0.985	0.992	0.996	0.997	0.997	0.998	0.994
	0.00	0.788	0.810	0.890	0.823	0.984	0.804	0.787
	0.25	0.952	0.977	0.982	0.986	0.988	0.985	0.973
$T = 200$	0.50	0.983	0.992	0.993	0.996	0.997	0.999	0.996
	0.75	0.991	0.995	0.996	0.996	0.998	1.000	0.997
	1.00	0.992	0.995	0.997	0.998	0.998	0.999	0.998

Table 2.4: Simulation results for $k = 8$, $q = 1$, $W = X'\Omega^{-1}X$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.553	0.591	0.654	0.626	0.652	0.622	0.566
	0.25	0.844	0.920	0.936	0.944	0.933	0.912	0.834
$T = 100$	0.50	0.949	0.967	0.977	0.983	0.981	0.976	0.957
	0.75	0.975	0.987	0.990	0.992	0.992	0.989	0.979
	1.00	0.981	0.989	0.992	0.993	0.994	0.993	0.983
	0.00	0.572	0.619	0.670	0.672	0.664	0.607	0.544
	0.25	0.936	0.966	0.974	0.978	0.977	0.968	0.945
$T = 200$	0.50	0.978	0.988	0.991	0.993	0.994	0.995	0.988
	0.75	0.989	0.994	0.996	0.996	0.998	0.999	0.996
	1.00	0.993	0.995	0.998	0.998	0.998	0.999	0.997

Table 2.5: Simulation results for $k = 5$, $q = 2$, $W = X'\Omega^{-1}X$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.705	0.739	0.751	0.735	0.735	0.798	0.802
	0.25	0.914	0.949	0.954	0.969	0.969	0.974	0.967
$T = 100$	0.50	0.969	0.975	0.981	0.988	0.990	0.993	0.996
	0.75	0.975	0.988	0.989	0.993	0.994	0.997	0.998
	1.00	0.984	0.990	0.994	0.995	0.995	0.997	0.996
	0.00	0.714	0.715	0.711	0.710	0.815	0.817	0.815
	0.25	0.943	0.970	0.975	0.982	0.987	0.989	0.990
$T = 200$	0.50	0.979	0.989	0.988	0.992	0.994	0.996	0.996
	0.75	0.989	0.993	0.994	0.994	0.996	0.998	0.998
	1.00	0.991	0.993	0.995	0.996	0.997	0.998	0.998

Table 2.6: Simulation results for $k = 8$, $q = 2$, $W = X'\Omega^{-1}X$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.480	0.476	0.525	0.527	0.599	0.645	0.604
	0.25	0.839	0.911	0.927	0.947	0.953	0.958	0.942
$T = 100$	0.50	0.948	0.963	0.973	0.981	0.985	0.987	0.986
	0.75	0.974	0.986	0.988	0.991	0.993	0.994	0.993
	1.00	0.981	0.988	0.991	0.993	0.994	0.996	0.993
	0.00	0.468	0.482	0.509	0.518	0.591	0.652	0.651
	0.25	0.929	0.960	0.970	0.979	0.984	0.987	0.985
$T = 200$	0.50	0.976	0.986	0.989	0.991	0.994	0.997	0.996
	0.75	0.988	0.993	0.994	0.995	0.997	0.999	0.999
	1.00	0.992	0.994	0.997	0.997	0.998	0.999	0.999

Table 2.7: Simulation results for $k = 5$, $q = 0.5$, $W = I$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.935	0.813	0.799	0.843	0.872	0.883	1.000
	0.25	0.983	0.986	0.982	0.982	0.976	0.978	0.984
$T = 100$	0.50	0.996	0.996	0.997	0.999	0.997	1.000	1.002
	0.75	0.993	0.998	0.998	0.999	1.000	1.002	1.000
	1.00	0.996	0.998	0.999	0.999	0.999	1.001	0.998
	0.00	0.902	0.845	0.858	0.872	0.857	0.792	0.885
	0.25	0.989	0.995	0.997	0.994	0.992	0.988	0.984
$T = 200$	0.50	0.997	1.000	1.001	1.001	1.001	1.002	1.000
	0.75	0.998	1.000	1.000	0.999	0.999	1.002	0.998
	1.00	0.998	0.999	1.000	1.000	1.000	1.000	0.998

Table 2.8: Simulation results for $k = 8$, $q = 0.5$, $W = I$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.793	0.654	0.632	0.597	0.630	0.657	0.690
	0.25	0.971	0.973	0.967	0.959	0.944	0.930	0.906
$T = 100$	0.50	0.993	0.992	0.992	0.991	0.986	0.982	0.979
	0.75	0.995	0.997	0.996	0.996	0.996	0.995	0.990
	1.00	0.995	0.997	0.997	0.997	0.997	0.997	0.988
	0.00	0.710	0.673	0.644	0.604	0.608	0.524	0.527
	0.25	0.984	0.989	0.987	0.983	0.979	0.970	0.956
$T = 200$	0.50	0.996	0.997	0.997	0.996	0.996	0.997	0.992
	0.75	0.997	0.998	0.999	0.998	0.999	1.000	0.997
	1.00	0.998	0.999	1.000	1.000	0.999	0.999	0.997

Table 2.9: Simulation results for $k = 5$, $q = 1$, $W = I$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	1.005	0.860	0.837	0.800	0.902	0.913	0.991
	0.25	0.967	0.969	0.968	0.975	0.970	0.973	0.983
$T = 100$	0.50	0.979	0.982	0.986	0.992	0.993	0.997	1.001
	0.75	0.979	0.990	0.992	0.995	0.997	1.000	1.001
	1.00	0.988	0.992	0.996	0.997	0.997	1.000	0.997
	0.00	0.832	0.843	0.864	0.826	0.952	0.847	0.852
	0.25	0.961	0.981	0.984	0.986	0.989	0.989	0.985
$T = 200$	0.50	0.984	0.992	0.993	0.996	0.998	1.000	0.999
	0.75	0.992	0.996	0.996	0.996	0.998	1.001	0.998
	1.00	0.992	0.995	0.997	0.998	0.998	0.999	0.998

Table 2.10: Simulation results for $k = 8$, $q = 1$, $W = I$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.849	0.693	0.675	0.649	0.701	0.714	0.724
	0.25	0.912	0.935	0.942	0.947	0.941	0.934	0.913
$T = 100$	0.50	0.964	0.970	0.977	0.983	0.982	0.981	0.979
	0.75	0.979	0.988	0.990	0.992	0.993	0.993	0.989
	1.00	0.982	0.989	0.992	0.994	0.995	0.996	0.987
	0.00	0.639	0.634	0.677	0.684	0.681	0.640	0.613
	0.25	0.943	0.969	0.975	0.979	0.979	0.972	0.959
$T = 200$	0.50	0.980	0.989	0.992	0.993	0.994	0.996	0.993
	0.75	0.990	0.994	0.996	0.996	0.998	0.999	0.997
	1.00	0.993	0.995	0.998	0.998	0.998	0.999	0.997

Table 2.11: Simulation results for $k = 5$, $q = 2$, $W = I$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.958	0.776	0.830	0.784	0.770	0.818	0.862
	0.25	0.961	0.960	0.961	0.972	0.972	0.975	0.979
$T = 100$	0.50	0.975	0.976	0.980	0.988	0.990	0.993	0.998
	0.75	0.976	0.986	0.989	0.992	0.994	0.997	0.999
	1.00	0.987	0.990	0.994	0.995	0.995	0.998	0.996
	0.00	0.757	0.758	0.723	0.730	0.837	0.819	0.828
	0.25	0.952	0.974	0.977	0.983	0.987	0.989	0.991
$T = 200$	0.50	0.980	0.988	0.989	0.991	0.994	0.996	0.997
	0.75	0.989	0.993	0.994	0.994	0.996	0.998	0.998
	1.00	0.990	0.993	0.995	0.996	0.997	0.998	0.998

Table 2.12: Simulation results for $k = 8$, $q = 2$, $W = I$

	λ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.806	0.607	0.580	0.559	0.629	0.668	0.696
	0.25	0.899	0.924	0.936	0.951	0.956	0.962	0.958
$T = 100$	0.50	0.959	0.965	0.973	0.981	0.985	0.988	0.990
	0.75	0.977	0.985	0.988	0.991	0.993	0.994	0.995
	1.00	0.981	0.987	0.991	0.993	0.994	0.996	0.993
	0.00	0.549	0.518	0.535	0.539	0.609	0.667	0.674
	0.25	0.935	0.963	0.972	0.981	0.985	0.987	0.987
$T = 200$	0.50	0.976	0.986	0.989	0.991	0.994	0.997	0.997
	0.75	0.988	0.992	0.994	0.996	0.998	0.999	0.999
	1.00	0.992	0.994	0.997	0.997	0.998	0.999	0.999

Table 2.13: Empirical results for forecasting output growth

h	out-of-sample period	$MSFE_{\gamma}$	$MSFE_{ur}$
1	1970:01-2020:03	0.6577**	0.6690
	1980:01-2020:03	0.5792*	0.5862
	1990:01-2020:03	0.5091**	0.5227
6	1970:01-2020:03	0.7262***	0.7784
	1980:01-2020:03	0.5167*	0.5383
	1990:01-2020:03	0.4667	0.4667
12	1970:01-2020:03	0.1878***	0.2467
	1980:01-2020:03	0.1487***	0.2211
	1990:01-2020:03	0.1565***	0.2520

Chapter 3

Optimal Forecast under Structural Breaks

3.1 Introduction

In a regression framework, structural breaks can be considered as a regression equation with a shift in some/all of the regression coefficients and/or error term. One of the biggest concerns of economists is how to have an accurate forecast in the case of the structural breaks. Since the seminal work by Bates and Granger (1969), forecast combination is an effective way to improve forecasting performance. Especially under the model uncertainty, the performance of the forecast can be boosted by forecast combinations method, see Diebold and Pauly (1987), Clements and Hendry (1998, 1999, 2006), Stock and Watson (2004), Pesaran and Timmermann (2002, 2005, 2007), Timmerman (2006), Hansen (2007), Pesaran and Pick (2011), Rossi (2013), and Pesaran et al. (2013) inter alia. Normally in the forecast combinations, we need to closely pay attention to two points: first the selected forecast models, and secondly the combination weights. The common method for forecasting under structural break is to use a post-break estimator. But this estimator by itself may not minimize the MSFE specifically in the case that the break point is close to the end of the sample, and thus there are only a few observations in the post-break sample. In this case, the post-break parameters are likely to be poorly estimated relative to those obtained using pre-break data. Therefore, as pointed out by Pesaran and Timmermann (2007), Hansen (2009) and Pesaran et al. (2013), it is not always optimal to base the forecast only on post-break observations. Actually, using pre-break data biases the forecast but also reduces the forecast error variance which is helpful to lower the MSFE.

A key question that comes to mind under the presence of structural break is how to include the pre-break data to estimate the parameters of the forecasting model such that minimizes the loss function like MSFE. In this chapter we propose the combined estimator of the post-break estimator and the full-sample estimator which uses all observations in the sample, $t = \{1, \ldots, T\}$. The combination weight takes the form of the James-Stein weight, see Stein (1956) and James and Stein (1961). Massoumi (1978) introduced the Stein-like estimator for the simultaneous equation. Also, Hansen (2016, 2017) has used this Stein-type weight in different context. See Saleh (2006) for a comprehensive explanation for the Stein-like estimators.

The focus of this chapter is not on the classical problem of identifying the break points. Instead, the goal is to propose an estimator with a minimum MSFE under the assumption that structural break has in fact occurred. For estimation of the break points, we follow Bai and Perron (1998) which is a consistent global minimizer method of the error term. Bai and Perron (2003) propose an efficient dynamic programming algorithm that even with a large sample size, has a small computing cost and is super fast. We also assume that the break points are bounded away from the beginning or end of the sample which is crucial for consistency of the estimated break points, Andrews (1993), Bai and Perron (1998).

This chapter has two main contributions. First, we introduce the Stein-like combined estimator of the full-sample estimator and the post-break estimator which is a trade-off between bias and forecast error variance. The full-sample estimator (which uses the pre-break and post-break data) is bias under breaks, but at the same time it decreases the forecast error variance because it uses more observations. Post-break estimator (which is the standard solution under the structural break models) is an unbiased estimator under breaks but less efficient. As discussed in papers by Pesaran and Timmermann (2007), Pesaran and Pick (2011), Pesaran et al. (2013), we are able to improve the performance of forecast measured by MSFE by the trade-off between bias and forecast error variance. Our proposed Stein-like combined estimator is the trade-off between bias and variance of the two estimators. The main difference between our paper and Hansen (2017) paper which uses the Stein-like weigh in different context is that, here we minimized the weighted risk under the general positive semi-definite form of weight and did not restrict our theoretical proof for a specific choice of weight. We show that if we set this weight to the inverse of the difference between variances of estimators, then we get similar result as Hansen (2017).

An alternative approach to the fully parametric regression model is semi-parametric kernel based regression model which has become a popular method in applied work, and this is the second contribution of this chapter. See Robinson (1988) and Ichimura (1993) for the semi-parametric kernel-based regression models. The proposed semi-parametric estimator facilitate implementation of our theoretical point that using pre-break data can improve the forecasting performance. This method is a kernel weight of the observations. Motivation for the semi-parametric method comes from the paper by Li, Ouyang and Racine (2013), in which they focus on the categorical variables. Examples of the categorical variables are race, sex, age group and educational level, see also Kiefer and Racine (2017). Inspired by the categorical data, we introduce the semi-parametric method for the time series analysis, and specifically apply that for the structural break models. The point of using this method is that we can find the combination weights numerically by Cross-Validation (CV) without being involved to the estimation errors of parameters. We prove theoretically that the weight estimated by CV is optimal in the sense of optimality introduced by K.-C. Li (1987), the weight derives by CV is asymptotically equivalent to the infeasible optimal weight.

We undertake an empirical analysis of forecasting equity premium that compares the forecasting performance of our proposed Stein-like combined estimator, and some of the alternative methods, including post-break estimator, by using monthly, quarterly and annual data frequencies starting from 1927 up to 2018. This analysis which allows for multiple breaks at unknown times confirms the out-performance of using pre-break data in forecasting relative to forecasting with only the post-break data.

The outline of this chapter is as follows. Section 3.2 sets up the model under the structural breaks and introduces the Stein-like combined estimator and its risk. For simplicity, we discuss the problem under a single break, but generalization of the method to the multiple breaks is straightforward. Section 3.3 introduces the semi-parametric combined estimator. Section 3.4 discusses an alternative combined estimator proposed by Pesaran et al. (2013). Section 3.5 extends the model with multiple structural break. Section 3.6 reports Monte Carlo simulation and empirical example is in Section 3.7. Finally Section 3.8 concludes.

3.2 The structural break model

Consider the linear structural break model as $y_t = x_t'\beta_t + \sigma_t\varepsilon_t$, in which the $k \times 1$ vector of coefficients, β_t, and the error variance, σ_t, are subject to breaks. Let m denotes the number of breaks, and for simplicity assume we have one break ($m = 1$) that happens at time T_1. So we can rewrite the model as

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{for} \quad 1 \ < t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{for} \quad T_1 < t < T, \end{cases} \tag{3.1}$$

where x_t is $k \times 1$, $\varepsilon_t \sim$ i.i.d.$(0, 1)$, $t \in \{1, \ldots, T\}$, T_1 is the break point with $1 < T_1 < T$,

$$\beta_t = \begin{cases} \beta_{(1)} & \text{for} \quad 1 \ < t \leq T_1 \\ \beta_{(2)} & \text{for} \quad T_1 < t < T \end{cases} \tag{3.2}$$

and

$$\sigma_t = \begin{cases} \sigma_{(1)} & \text{for} \quad 1 \ < t \leq T_1 \\ \sigma_{(2)} & \text{for} \quad T_1 < t < T. \end{cases} \tag{3.3}$$

In this set up we have only one break (two regimes).

3.2.1 Stein-like combined estimator

Our proposed combined estimator of β is

$$\widehat{\beta}_{\alpha} = \alpha\widehat{\beta}_{Full} + (1-\alpha)\widehat{\beta}_{(2)}, \tag{3.4}$$

where $\widehat{\beta}_{\alpha}$ is the Stein-like combined estimator which is $k \times 1$, $\widehat{\beta}_{Full}$ is the estimator under the assumption that there is no break in the model, so we estimate the coefficient using all observations in the sample, $t \in \{1, \ldots, T\}$. $\widehat{\beta}_{(2)}$ estimates the coefficient only by using the post-break observations, and it is called the post-break estimator. We define the combination weight as

$$\alpha = \begin{cases} \frac{\tau}{H_T} & \text{if } H_T \geq \tau \\ 1 & \text{if } H_T < \tau, \end{cases} \tag{3.5}$$

where τ controls the degree of shrinkage, H_T is the Hausman statistic that measures the break size in the coefficients and is equal to

$$H_T = T(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full})'(\widehat{V}_{(2)} - \widehat{V}_{Full})^{-1}(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}), \tag{3.6}$$

where $\widehat{V}_{Full}$ and $\widehat{V}_{(2)}$ are the consistent estimates of the asymptotic variances of the full-sample estimator and the post-break estimator, respectively. The degree of shrinkage depends on the ratio of τ/H_T . When $H_T < \tau$, then $\alpha = 1$ and $\widehat{\beta}_{\alpha} = \widehat{\beta}_{Full}$. Small H_T can be interpreted as the small break size in the coefficients. This is the case that the bias of the full-sample estimator is small, and we can gain a lot from the efficiency of the full-sample estimator. On the other hand, very large H_T is interpreted as a big break size in the coefficients which results in a large bias. So, for the extreme case of large H_T, the combination weight would be very close to zero and $\widehat{\beta}_{\alpha} = \widehat{\beta}_{(2)}$. Other than these extreme cases, for $H_T > \tau$, $\widehat{\beta}_\alpha$ is a weighted average of the full-sample estimator and the post-break estimator.

It is easy to see that $\text{cov}(\widehat{\beta}_{(2)}, \widehat{\beta}_{Full}) = V_{Full}$, where $V_{Full} < V_{(2)}$. This means that the covariance between the estimators is equal to the variance of the efficient estimator. The interesting idea behind the Hausman statistic is that the efficient estimator, $\widehat{\beta}_{Full}$, must have zero asymptotic covariance with $\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}$ under the null hypothesis of no break in the coefficients ($\beta_{(1)} = \beta_{(2)}$). If this condition does not hold, it means that we can find another linear combination which would have smaller asymptotic variance than $\widehat{\beta}_{Full}$ which is assumed to be asymptotically efficient. Holding this condition in the combined estimator, shows the great choice of the full-sample and post-break estimators. See Hausman (1978) for more discussion.

In the next subsection, we develop the asymptotic distribution for the estimators under a local asymptotic framework which means that the break size is local to zero. That is, $\beta_{(1)} = \beta_{(2)} + \frac{\delta_1}{\sqrt{T}}$ in which δ_1 shows the break size between the coefficients. Simply it means that, as the sample size goes to infinity, we will have a weak break. Actually, to obtain a meaningful asymptotic distribution for the full-sample estimator, we consider such a local to zero assumption otherwise the risk explodes.

Full-sample estimator: There is no break in the model

As we introduced the Stein-like combined estimator earlier, it includes the full-sample estimator. The full-sample estimator is constructed under the null hypothesis that there is no break in the coefficients, $\beta_{(1)} = \beta_{(2)}$, so it uses all of the observations to estimate β. This assumption lines up with the fact that for the small break sizes, usually ignoring the break and estimating the coefficient by all of the observations results in better forecast (lower MSFE), see Boot and Pick (2019). Under the alternative hypothesis the break size is $\beta_1 = \beta_2 + \frac{\delta_1}{\sqrt{T}}$ which is asymptotically local to zero. We denote the full-sample estimator by $\widehat{\beta}_{Full}$, and estimate the coefficient by the Generalized Least Square (GLS)

$$\widehat{\beta}_{Full} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y, \tag{3.7}$$

where $\Omega = \text{diag}(\sigma^2_{(1)}, \ldots, \sigma^2_{(1)}, \sigma^2_{(2)}, \ldots, \sigma^2_{(2)})$ is a $T \times T$ matrix and $X = (X_1' \; X_2')'$ is a $T \times k$ matrix of regressors. So X_1 is $T_1 \times k$ matrix of observations before the break point, and X_2 is $(T - T_1) \times k$ matrix of observations after the break point. Assume that $T - T_1 \geq k + 1$, so at least we have the minimum number of observations in the post-break sample to estimate the coefficient. The choice of shortest estimation window selected is arbitrary, but one would expect it to be set around two to three times the dimension of β to avoid the extreme variation in the post-break parameter estimates. Also, assume that $\frac{X_i'\Omega_i^{-1}X_i}{T_i - T_{i-1}}$, with $i = \{1, \ldots, m+1\}$ and $T_0 = 0, T_{m+1} = T$, converges in probability to some non-random positive definite matrix not necessarily the same for all i. So, $\left(\frac{X'\Omega^{-1}X}{T}\right) \xrightarrow{p} Q$ and $\left(\frac{X_i' \; \Omega_i^{-1} X_i}{\Delta T_i}\right) \xrightarrow{p} Q_i$ where Q and Q_i are some positive definite matrix, and $\Delta T_i = T_i - T_{i-1}$. Throughout this Section, $m = 1$ since we have only one break in the model. Therefore, the distribution of the full-sample estimator is

$$\sqrt{T}\left(\widehat{\beta}_{Full} - \beta_{(2)}\right) \xrightarrow{d} N\left(Q^{-1}Q_1 b_1 \delta_1, Q^{-1}\right), \tag{3.8}$$

where $V_{Full} = \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \xrightarrow{p} Q^{-1}$, $\left(\frac{X_1'\Omega_1^{-1}X_1}{T_1}\right) \xrightarrow{p} Q_1$, $b_1 \equiv \frac{T_1}{T}$ shows the proportion of pre-break observations and $\Omega_1 = \text{diag}(\sigma^2_{(1)}, \ldots, \sigma^2_{(1)})$ is a $T_1 \times T_1$ matrix. Note that in practice, we need to estimate the value of the unknown parameters, $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$. For solving this problem, we can use the two-step GLS estimator method. The two-step estimator is computed by first obtaining the estimates of $\widehat{\sigma^2_{(1)}}$ and $\widehat{\sigma^2_{(2)}}$ by using ordinary least square residuals for each regime, and then plugging $\widehat{\Omega}$, which is asymptotically the consistent estimate for Ω, back into equation (3.8) to find the estimation of the coefficient.

Remark 8 *The full-sample estimator is under the null hypothesis that there is no break in the coefficient $\beta_{(1)} = \beta_{(2)}$, but we allow a break in the variance, $\sigma_{(1)} \neq \sigma_{(2)}$. Because we have variance heteroskedasticity in the full-sample estimator, we use GLS method which is more efficient than ordinary least squares (OLS).* □

Post-break estimator

The post-break estimator is for the case that we only focus on the observations after the most recent break point and is equal to

$$\widehat{\beta}_{(2)} = \left(X_2' \ \Omega_2^{-1} \ X_2\right)^{-1} X_2' \ \Omega_2^{-1} \ Y_2, \tag{3.9}$$

where $\Omega_2 = \text{diag}(\sigma^2_{(2)}, \ldots, \sigma^2_{(2)})$ is a $(T - T_1) \times (T - T_1)$ matrix. Basically, post-break estimator is the simple OLS estimator, since Ω_2^{-1} will be cancel out in this equation. But we are following the GLS format to be consistent with the full-sample estimator. The distribution of the post-break estimator is

$$\sqrt{T}\left(\widehat{\beta}_{(2)} - \beta_{(2)}\right) \xrightarrow{d} N\left(0, \frac{1}{1-b_1} Q_2^{-1}\right), \tag{3.10}$$

where $V_{(2)} = \frac{1}{1-b_1}\left(\frac{X_2' \ \Omega_2^{-1} \ X_2}{T-T_1}\right)^{-1} \xrightarrow{p} \frac{1}{1-b_1} Q_2^{-1}$. This is an unbiased estimator, and as the break happens towards the end of the sample, as b_1 increases, the variance of the post-break estimator will increase.

Stein-like combined estimator

For writing the distribution of the Stein-like combined estimator, at first we need to write the joint asymptotic distribution of the full-sample estimator and the post-break estimator. Theorem 9 shows the joint distribution, distribution of the Hausman statistics and finally distribution of the Stein-like combined estimator.

Theorem 9 *The joint asymptotic distribution of the full-sample estimator and the post-break estimator is*

$$\sqrt{T}\begin{bmatrix}\widehat{\beta}_{Full}-\beta_{(2)}\\ \widehat{\beta}_{(2)}-\beta_{(2)}\end{bmatrix}\xrightarrow{d}V^{1/2}Z, \tag{3.11}$$

where $Z\sim N(\theta,I_{2k})$, $\theta=V^{-1/2}\begin{bmatrix}Q^{-1}Q_1b_1\delta_1\\ 0\end{bmatrix}$, $V=\begin{bmatrix}V_{Full} & V_{Full}\\ V_{Full} & V_{(2)}\end{bmatrix}$,

$V_{Full}=\underset{T\to\infty}{plim}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1}=Q^{-1}$ *and* $V_{(2)}=\frac{1}{1-b_1}\underset{T\to\infty}{plim}\left(\frac{X_2'\Omega_2^{-1}X_2}{T-T_1}\right)^{-1}=\frac{1}{1-b_1}Q_2^{-1}$.

Besides, the distribution of the Hausman statistic is

$$\begin{aligned}H_T&=T(\widehat{\beta}_{(2)}-\widehat{\beta}_{Full})'(\widehat{V}_{(2)}-\widehat{V}_{Full})^{-1}(\widehat{\beta}_{(2)}-\widehat{\beta}_{Full})\\ &\xrightarrow{d}Z'V^{1/2}\;G\;(V_{(2)}-V_{Full})^{-1}\;G'\;V^{1/2}Z\\ &\equiv Z'MZ,\end{aligned} \tag{3.12}$$

where $G=(-I_k\quad I_k)'$ *and* $M\equiv V^{1/2}\;G\;(V_{(2)}-V_{Full})^{-1}\;G'\;V^{1/2}$ *is an idempotent matrix with rank* k*. Finally, the distribution of the Stein-like combined estimator is*

$$\begin{aligned}\sqrt{T}\big(\widehat{\beta}_{\alpha}-\beta_{(2)}\big)&=\sqrt{T}\big(\widehat{\beta}_{(2)}-\beta_{(2)}\big)-\alpha\sqrt{T}\big(\widehat{\beta}_{(2)}-\widehat{\beta}_{Full}\big)\\ &\xrightarrow{d}G_2'\;V^{1/2}Z-\left(\frac{\tau}{Z'MZ}\right)_1\;G'\;V^{1/2}Z,\end{aligned} \tag{3.13}$$

where $G_2 = (0 \;\; I_k)'$ *and* $(a)_1 = min[1, a]$. ■

See Appendix B for the proof of this Theorem. Based on Theorem 9, the joint asymptotic distribution of the full-sample and post-break estimators is normal. The Hausman statistics has asymptotic noncentral chi-square distribution, which later we deal with this non-centrality when we want to calculate the asymptotic risk. Finally, the asymptotic distribution of the Stein-like combined estimator is a function of the normal random vector, Z, and a function of the non-centrality parameter which appears because of having the Hausman statistic in the combination weight.

3.2.2 Asymptotic risk for the Stein-like combined estimator

In this section we find the asymptotic risk for the Stein-like combined estimator. Since our focus is on forecasting, it seems reasonable to consider $\beta_{(2)}$ as the true parameter vector in the definition of the risk.

Lemma 10 *When an estimator has an asymptotic distribution,* $\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} \varpi$*, then the asymptotic risk of this estimator can be calculated as* $\rho(\widehat{\beta}, \mathbb{W}) = \mathbb{E}(\varpi' \, \mathbb{W} \, \varpi)$*. See Lehmann and Casella (1998).* □

Based on Lemma 10, we can write the asymptotic risk for the Stein-like combined estimator. Note that if we set up the risk with $\mathbb{W} = X'\Omega^{-1}X$, we will have the definition of the mean square forecast error (MSFE). Then, we can derive the optimal combination weight, α, in which it minimizes the MSFE. Having the combination weight, α, and consequently the Stein-like combined estimator, $\widehat{\beta}_\alpha$, we can derive the h step ahead out of sample forecast. Throughout the calculation of the risk, we did not plug any specific form for $\mathbb{W}$, and calculate the risk for any user specific choice of $\mathbb{W} > 0$. Theorem 11 shows the risk for the Stein-like combined estimator.

Theorem 11 *The risk of the Stein-like combined estimator is*

$$\rho(\widehat{\beta}_\alpha, \mathbb{W}) = \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \frac{2\tau\ \theta' V^{1/2} G\, \mathbb{W}\, G' V^{1/2}\theta}{k(k+2)\ \theta' M \theta}\left\{\tau - 2\left(\frac{tr\big(\mathbb{W}(V_{(2)} - V_{Full})\big)\ \theta' M\theta}{\theta' V^{1/2} G\, \mathbb{W}\, G' V^{1/2}\theta} - 2\right)\right\}$$
$$\times\, \mu\ e^{-\mu}{}_1F_1\Big(\frac{k}{2}; \frac{k}{2}+2; \mu\Big) + \frac{\tau\ tr\big(\mathbb{W}(V_{(2)} - V_{Full})\big)}{k(k-2)}\left\{\tau - 2(k-2)\right\}$$
$$\times\, e^{-\mu}{}_1F_1\Big(\frac{k}{2} - 1; \frac{k}{2}+1; \mu\Big). \tag{3.14}$$

■

Based on Theorem 11, the risk of the combined estimator is lower than the risk of the post-break estimator, if the terms inside the curly brackets be negative. These terms are negative if $\mathrm{tr}(\nu) > 2\ \lambda_{\max}(\nu)$, where $\nu \equiv (V_{(2)} - V_{Full})^{1/2}\, \mathbb{W}(V_{(2)} - V_{Full})^{1/2}$, and $k > 2$. For the special case that $\mathbb{W} = \big(V_{(2)} - V_{Full}\big)^{-1}$, both conditions simplify to $k > 2$ which means that as long as we have more than two regressors, the risk of the Stein-like combined estimator is lower than the risk of the post-break estimator for any break sizes and break points. See Appendix B for the complete proof of the Theorem 11.

Using Theorem 11, we can find the optimal τ, denoted by τ^*, in which it minimizes the risk. For $0 \leq \tau \leq 2\Big(\frac{\mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)\big(\theta' M\theta\big)}{\theta'\ V^{1/2}\ G\mathbb{W}\ \ G'\ V^{1/2}\theta} - 2\Big)$, the optimal τ^* which depends on $\mathbb{W}$ is

$$\tau^*(\mathbb{W}) = \frac{\mathrm{tr}\big(\mathbb{W}(V_{(2)} - V_{Full})\big)\ \big(\theta' M\theta\big)}{\theta'\ V^{1/2}\ G\mathbb{W}\ \ G'\ V^{1/2}\theta} - 2. \tag{3.15}$$

Notice that $\tau^*(\mathbb{W})$ is positive when $\mathrm{tr}(\nu) > 2\ \lambda_{\max}(\nu)$. This is the necessary condition for the efficiency of the Stein-like combined estimator, as shown in Theorem 11.

Remark 12 *As b_1 increases, $V_{(2)}$ increases, consequently τ^* increases. Besides, as b_1 increases, $V_{(2)}$ increases, and because H_T inversely depends on $V_{(2)}$, so H_T decreases which results in having bigger $\frac{\tau^*}{H_T}$. Thus the full-sample estimator gets a higher weight when b_1 is large. Remember that $\widehat{\beta}_\alpha = \alpha\widehat{\beta}_{Full} + (1-\alpha)\widehat{\beta}_{(2)}$.* □

By plugging back the optimal $\tau^*(\mathbb{W})$ into the risk function, we can derive the optimal risk. Theorem 13 summarizes the result for any $\mathbb{W} > 0$.

Theorem 13 *If $0 \leq \tau \leq 2\Big(\frac{tr\big(\mathbb{W}(V_{(2)}-V_{Full})\big)\ (\theta' M\theta)}{\theta'\ V^{1/2}\ G\mathbb{W}\ G'\ V^{1/2}\theta} - 2\Big)$ and $tr(\nu) > 2\lambda_{max}(\nu)$, then the risk of the Stein-like combined estimator for any user specific choice of $\mathbb{W} > 0$ is*

$$\rho(\widehat{\beta}_\alpha, \mathbb{W}) = \rho(\widehat{\beta}_{(2)}, \mathbb{W}) - \frac{1}{k-2}\left\{\frac{\Big[\ tr\big(\mathbb{W}(V_{(2)}-V_{Full})\big)(\theta' M\theta) - 2\ \big(\theta' V^{1/2}\ G\,\mathbb{W}\,G'\ V^{1/2}\theta\big)\ \Big]^2}{(\theta' M\theta)\ (\theta' V^{1/2}\ G\,\mathbb{W}\,G'\ V^{1/2}\theta)}\right\}$$
$$\times\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big)\right\} - \frac{1}{k-2}\left\{\frac{\Big[tr\big(\mathbb{W}(V_{(2)}-V_{Full})\big)\Big]^2\big(\theta' M\theta\big)^2}{\big(\theta' V^{1/2}\ G\,\mathbb{W}\,G'\ V^{1/2}\theta\big)^2} - 4\right\}$$
$$\times\left\{\frac{\big(\theta' V^{1/2}\ G\,\mathbb{W}\,G'\ V^{1/2}\theta\big)}{\theta' M\theta} - \frac{tr\big(\mathbb{W}(V_{(2)}-V_{Full})\big)}{k}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}, \tag{3.16}$$

where $\rho(\widehat{\beta}_{(2)}, \mathbb{W}) = tr(\mathbb{W}\,V_{(2)})$, and $\mu = \frac{\theta' M\theta}{2}$ is the non-centrality parameter. ■

Note that in Theorem 13, the terms inside the curly brackets are positive, so we could prove that the Stein-like combined estimator has a smaller risk than the post-break estimator.

Corollary 14 *For the special case that $\mathbb{W} = (V_{(2)} - V_{Full})^{-1}$, the risk of the Stein-like combined estimator simplifies to*

$$\rho(\widehat{\beta}_\alpha, \mathbb{W}) = \rho(\widehat{\beta}_{(2)}, \mathbb{W}) - (k-2)\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big)\right\}, \tag{3.17}$$

where the risk of the Stein-like combined estimator is less than the post-break estimator if we have more than two regressors, $k > 2$. □

Based on Corollary 14, the gain obtained by using the Stein-like combined estimator can be derived by calculating the percentage change between $\rho(\widehat{\beta}_{(2)}, \mathbb{W})$ and $\rho(\widehat{\beta}_{\alpha}, \mathbb{W})$. Figure 3.1 shows the relationship between the break size in the coefficients (the horizontal axis) and the percentage change in risks (vertical axis), $-\frac{\rho(\widehat{\beta}_{\alpha}, \mathbb{W}) - \rho(\widehat{\beta}_{(2)}, \mathbb{W})}{\rho(\widehat{\beta}_{(2)}, \mathbb{W})}$, which we draw it by R program. For example, when the vertical axis shows the percentage change equal to 50%, it means that, by using the Stein-like combined estimator instead of the post-break estimator, we can decrease the risk by 50%. We draw the graphs for different break ratios in the error term, q. Based on the figures, as the number of regressors, k, increases, the percentage change between risks increases in favor of the Stein-like combined estimator. Also, when the variance of the pre-break data is lower than variance of the post-break data, $q < 1$, then it is more gain to use pre-break observations. Besides, for the cases that the break point is near the end of the sample, $b_1 = 0.8$, the gain obtained from the Stein-like combined estimator is higher than the case that the break point is near the beginning of the sample, $b_1 = 0.2$. The reason is that when there is not enough observations in the post-break sample, post-break estimator performs poorly due to the lack of the observations. Furthermore, when the break size in the coefficient increases (as bias increases), the performance of the Stein-like estimator is close to the post-break estimator, as expected. But still we do not have any underperformance which confirms that there is no cost in using the proposed combined estimator.

Remark 15 *The difference between our proposed method and Hansen (2017) paper is that, here we calculate the risk for any user specific form of* $\mathbb{W}$ *, and derive the optimal* τ^* *which depends on* $\mathbb{W}$*. For the special case that* $\mathbb{W} = \left(V_{(2)} - V_{Full}\right)^{-1}$*, then* $\tau^* = k - 2$ *which is the* *well-known results of James and Stein (1961). In this case, τ^* is positive when the number of regressors is larger than two, $k > 2$. This choice of $\mathbb{W}$ is the special case of Hansen (2017) paper.* $\square$

3.3 Semi-parametric estimator

Li, Ouyang, and Racine (2013) introduce the semi-parametric method for estimating the parameters of the model under the categorical data. Inspired by their approach, we want to propose an estimator for the time series structural break models that breaks happen at both coefficients and the error terms. We can estimate the post-break coefficients using the following discrete kernel estimator

$$\widehat{\beta}_\gamma = \underset{\beta_{(2)}}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_t - x_t' \beta_{(2)} \right)^2 L(t, \gamma), \tag{3.18}$$

where

$$L(t, \gamma) = \gamma \, \mathbf{1}(t \leq T_1) + \mathbf{1}(t > T_1), \tag{3.19}$$

and t goes over all of the observations. $\widehat{\beta}_\gamma$ is the estimator for the post-break observations. The idea is that for estimating the post-break data, it is worth to use information from the first regime as well. As the more recent information is usually more relevant for forecasting, this kernel-weight estimator gives the constant weight 1 to the post-break observations and down-weights the pre-break observations by a weight $\gamma \in \mathcal{H} = [0\ 1]$. In other words, the post-break coefficient is estimated by weighting observations over the whole sample. For the case that $\gamma = 0$, this estimator is only using the post-break observations, and when $\gamma = 1$, all of the observations are weighted equally. For other cases than $\gamma \in (0\ 1)$, we have the combination of the pre-break and post-break observations. The first order condition of equation (3.18) is

$$\sum_{t=1}^{T} L(t,\gamma)\ x_t\Big(y_t - x_t'\beta_{(2)}\Big) = 0, \tag{3.20}$$

leading to

$$\begin{aligned}
\widehat{\beta}_\gamma &= \Big(\sum_{t=1}^{T} L(t,\gamma)\ x_t x_t'\Big)^{-1} \sum_{t=1}^{T} L(t,\gamma)\ x_t y_t \\
&= \Big(\gamma X_1'X_1 + X_2'X_2\Big)^{-1}\Big(\gamma X_1'Y_1 + X_2'Y_2\Big) \\
&= \Big(\gamma X_1'X_1 + X_2'X_2\Big)^{-1}\Big(\gamma X_1'X_1\ \widehat{\beta}_{(1)} + X_2'X_2\ \widehat{\beta}_{(2)}\Big) \\
&= \Delta\ \widehat{\beta}_{(1)} + (I-\Delta)\ \widehat{\beta}_{(2)},
\end{aligned} \tag{3.21}$$

where $\Delta = \Big(\gamma X_1'X_1 + X_2'X_2\Big)^{-1}(\gamma X_1'X_1)$, $\widehat{\beta}_{(1)} = (X_1'X_1)^{-1}X_1'Y_1$ and $\widehat{\beta}_{(2)} = (X_2'X_2)^{-1}X_2'Y_2$. We can find γ by minimizing the following leave-one-out cross-validation (CV) criterion function

$$CV(\gamma) = \frac{1}{T-T_1}\sum_{s=T_1+1}^{T}\Big(y_s - x_s'\widehat{\beta}_\gamma^{(-s)}\Big)^2, \tag{3.22}$$

where $\widehat{\beta}_\gamma^{(-s)}$ denotes the CV estimate of β in which s goes over the post-break observations, $s = \{T_1+1,\ldots,T\}$ and at each time deletes $s'th$ row of the observation. Note that CV for finding γ is dropping observations from the post-break observations only. The reason is that from $t = \{1,\ldots T_1\}$ the coefficient is $\beta_{(1)}$, and we are trying to find the estimator for the post-break observations ($t > T_1$) in order to use it for forecasting. So, we drop observations from the post-break data and estimate the weight such that the estimated weight fits best the post-break observations. Once we estimate γ, we can calculate $\widehat{\beta}_{\widehat{\gamma}}$ and MSFE accordingly. The good point about this semi-parametric estimator is that γ does not depend on the estimation of the break size in the coefficient or the error term. Actually the estimation error in this estimator is low. Since this method finds the weight by CV, it can fit to data better than other methods which involve the estimation of some other parameters as well.

We just need to prove that the $\widehat{\gamma}$ which we estimate by CV is optimal in the sense of optimality introduced by K.-C. Li (1987). The concept of this optimality is that the average squared error of the CV is asymptotically as small as the average squared error of the infeasible best possible estimator. Define $L(\gamma) = \left[\left(\widehat{\beta}_\gamma - \beta_{(2)}\right)' X_2' X_2 \left(\widehat{\beta}_\gamma - \beta_{(2)}\right)\right] = \left(\widehat{\mu}(\gamma) - \mu\right)'\left(\widehat{\mu}(\gamma) - \mu\right)$ to be the loss function, where $\mu = X_2\beta_{(2)}$ and $\widehat{\mu}(\gamma) = X_2\widehat{\beta}_\gamma$, and the expected loss is $R(\gamma) = \mathbb{E}\left[L(\gamma)\right]$. Theorem 16 is about the optimality of the γ weight.

Theorem 16 *As $T \to \infty$, then*

$$\frac{L(\hat{\gamma})}{\inf_{\gamma\in\mathcal{H}} L(\gamma)} \xrightarrow{p} 1, \tag{3.23}$$

$$\frac{R(\hat{\gamma})}{\inf_{\gamma\in\mathcal{H}} R(\gamma)} \xrightarrow{p} 1. \tag{3.24}$$

■

Theorem 16 shows that the optimal weight obtained by CV is asymptotically equivalent to the infeasible optimal weight. Appendix B has the complete proof and details of this Theorem. Also, see Zhang et al. (2013) and Ullah et al. (2017) for more reference.

Remark 17 *Note that, as we allow the break happens in the error term as well as the coefficients, one may use the generalized least square (GLS) method for estimation. Using model (3.1), define $\widetilde{Y}_1 = \Omega_1^{-1/2}Y_1, \widetilde{Y}_2 = \Omega_2^{-1/2}Y_2, \widetilde{X}_1 = \Omega_1^{-1/2}X_1,\ \widetilde{X}_2 = \Omega_2^{-1/2}X_2$ where*

$\Omega_1 = diag\big(\sigma^2_{(1)}, \ldots, \sigma^2_{(1)}\big)$ and $\Omega_2 = diag\big(\sigma^2_{(2)}, \ldots, \sigma^2_{(2)}\big)$. Then,

$$\begin{aligned}
\widehat{\beta}_\gamma &= \left(\sum_{t=1}^{T} L(t,\gamma)\ \widetilde{x}_t \widetilde{x}_t'\right)^{-1} \sum_{t=1}^{T} L(t,\gamma)\ \widetilde{x}_t \widetilde{y}_t \\
&= \left(\gamma \widetilde{X}_1' \widetilde{X}_1 + \widetilde{X}_2' \widetilde{X}_2\right)^{-1} \left(\gamma \widetilde{X}_1' \widetilde{Y}_1 + \widetilde{X}_2' \widetilde{Y}_2\right) \\
&= \left(\frac{\gamma}{\sigma^2_{(1)}} X_1' X_1 + \frac{1}{\sigma^2_{(2)}} X_2' X_2\right)^{-1} \left(\frac{\gamma}{\sigma^2_{(1)}} X_1' Y_1 + \frac{1}{\sigma^2_{(2)}} X_2' Y_2\right) \\
&= \left(\frac{\gamma}{q^2} X_1' X_1 + X_2' X_2\right)^{-1} \left(\frac{\gamma}{q^2} X_1' X_1\ \widehat{\beta}_{(1)} + X_2' X_2\ \widehat{\beta}_{(2)}\right) \\
&= \Delta^*\ \widehat{\beta}_{(1)} + (I - \Delta^*)\ \widehat{\beta}_{(2)}, \qquad (3.25)
\end{aligned}$$

where $\gamma^* \equiv \frac{\gamma}{q^2}$, $q = \frac{\sigma_{(1)}}{\sigma_{(2)}}$, $\Delta^* = \Big(\gamma^* X_1' X_1 + X_2' X_2\Big)^{-1} (\gamma^* X_1' X_1)$. Notice that, we only estimate γ^* based on this semi-parametric method. So, even though the combination weight is in the form of fraction, $\frac{\gamma}{q^2}$, we can consider it as a single unknown parameter that we estimate that by cross validation. Basically, GLS method is transformed to the OLS method in this semi-parametric estimator because there is no need to estimate q^2. This is the advantage of this method, as it decreases the estimation error. □

3.3.1 Efficiency of the semi-parametric estimator

We show the consistency of the estimated weight in Theorem 16, and now we want to show that the proposed semi-parametric estimator is more efficient than the post-break estimator. That is, the risk of the semi-parametric estimator is less than the post-break estimator. As shown in (3.21), the semi-parametric estimator is the combination of the pre-break estimator and the post-break estimator. We can estimate the pre-break coefficient using the observations in the first regime as

$$\widehat{\beta}_{(1)} = (X_1' X_1)^{-1} (X_1' Y_1). \qquad (3.26)$$

We apply the local asymptotic framework, $\beta_{(1)} = \beta_{(2)} + \frac{\delta}{\sqrt{T}}$ where δ measures the size of the break in the coefficient, to find the asymptotic distribution for this estimator. Since our focus is on forecasting, we derive the asymptotic distribution of the pre-break estimator around the true parameter $\beta_{(2)}$ as

$$\sqrt{T}\left(\widehat{\beta}_{(1)} - \beta_{(2)}\right) \xrightarrow{d} N\left(\delta, V_{(1)}\right), \tag{3.27}$$

where $V_{(1)} = \underset{T\to\infty}{\text{plim}} \frac{\sigma^2_{(1)}}{b_1}\left(\frac{X_1'X_1}{T_1}\right)^{-1} = \frac{\sigma^2_{(1)}}{b_1}Q_1^{-1}$, and $Q_1 \equiv \frac{X_1'X_1}{T_1}$ is a positive definite matrix. As expected, the bias of the pre-break estimator will increase as the break size in the coefficient increases. The post-break estimator is

$$\widehat{\beta}_{(2)} = (X_2'X_2)^{-1}(X_2'Y_2), \tag{3.28}$$

and its distribution is

$$\sqrt{T}\left(\widehat{\beta}_{(2)} - \beta_{(2)}\right) \xrightarrow{d} N\left(0, V_{(2)}\right), \tag{3.29}$$

where $V_{(2)} = \frac{\sigma^2_{(2)}}{1-b_1}\underset{T\to\infty}{\text{plim}}\left(\frac{X_2'X_2}{T-T_1}\right)^{-1} = \frac{\sigma^2_{(2)}}{1-b_1}Q_2^{-1}$, and $Q_2 \equiv \frac{X_2'X_2}{T-T_1}$ is a positive definite matrix. In order to find the risk of the semi-parametric estimator, initially we derive the joint asymptotic distribution of the pre-break estimator and the post-break estimator. Theorem 18 shows the results.

Theorem 18 *The joint asymptotic distribution of the pre-break and post-break estimators is*

$$\sqrt{T}\begin{bmatrix}\widehat{\beta}_{(1)} - \beta_{(2)} \\ \widehat{\beta}_{(2)} - \beta_{(2)}\end{bmatrix} \xrightarrow{d} V^{1/2}Z, \tag{3.30}$$

where $Z \sim N(\theta, I_{2k})$, $\theta = V^{-1/2}\begin{bmatrix}\delta_1 \\ 0\end{bmatrix}$, $V = \begin{bmatrix}V_{(1)} & \mathbf{0} \\ \mathbf{0} & V_{(2)}\end{bmatrix}$, $V_{(1)} = \underset{T\to\infty}{plim}\frac{\sigma^2_{(1)}}{b_1}\left(\frac{X_1'X_1}{T}\right)^{-1} \equiv$

$\frac{\sigma^2_{(1)}}{b_1} Q_1^{-1}$ *and* $V_{(2)} = \frac{\sigma^2_{(2)}}{1-b_1} \underset{T\to\infty}{plim} \left(\frac{X_2' X_2}{T-T_1}\right)^{-1} = \frac{\sigma^2_{(2)}}{1-b_1} Q_2^{-1}$. ■

See Appendix B for the proof of Theorem 18. Using the results of Theorem 18, we can derive the asymptotic distribution of the semi-parametric estimator in terms of the post-break estimator. Let $\rho(\widehat{\beta}_\gamma, \mathbb{W})$ denote the risk of the combined estimator for any $\mathbb{W} > 0$. Theorem 19 presents the result.

Theorem 19 *The asymptotic risk of the semi-parametric estimator is*

$$\begin{aligned} \rho(\widehat{\beta}_\gamma, \mathbb{W}) &= \mathbb{E}\left[T(\widehat{\beta}_\gamma - \beta_{(2)})' \, \mathbb{W} \, (\widehat{\beta}_\gamma - \beta_{(2)})\right] \\ &\leq \rho(\widehat{\beta}_{(2)}, \mathbb{W}) - tr(S_1)\big(1 - \theta'\theta\big), \end{aligned} \tag{3.31}$$

where $S_1 \equiv V^{1/2} G \Delta' \, \mathbb{W} \, \Delta G' V^{1/2}$, *and* $G = (-I_k \; I_k)'$. *Thus, the risk of the semi-parametric estimator is less than the post-break estimator if* $\theta'\theta < 1$. ■

See Appendix B for the proof of Theorem 19.

3.4 Alternative combined estimator

In a particularly insightful paper, Pesaran et al (2013), hereafter PPP, propose that we can decrease the MSFE under the structural break by using the whole observations in the sample instead of only post-break observations. Their proposed estimator is

$$\widehat{\beta}_{PPP} = (X'WX)^{-1}(X'WY). \tag{3.32}$$

They derive the optimal weight, W, such that MSFE of the one-step-ahead forecast is minimized, and found that the optimal weight takes a value for observations before the break point and a value for the post-break observations, $W = \text{diag}\Big(w_{(1)}, \ldots, w_{(1)}, w_{(2)}, \ldots, w_{(2)}\Big)$, where $w_{(1)}$ is a fixed weight for the pre-break observations and $w_{(2)}$ is a fixed weight for the post-break observations and these are defined as

$$\begin{cases} w_{(1)} = \frac{1}{T}\frac{1}{b_1+(1-b_1)(q^2+Tb_1\phi^2)}, \\ w_{(2)} = \frac{1}{T}\frac{q^2+Tb_1\phi^2}{b_1+(1-b_1)(q^2+Tb_1\phi^2)}, \end{cases} \tag{3.33}$$

where $b_1 = \frac{T_1}{T}$ is the proportion of observations before the break, $q = \frac{\sigma_{(1)}}{\sigma_{(2)}}$ is the ratio of break in the error term, $\phi = \frac{x'_{T+1}\lambda}{\sigma^2_{(2)}(x'_{T+1}Q^{-1}x_{T+1})^{1/2}}$, $Q = \mathbb{E}(x_t x_t')$ and $\lambda = \beta_{(1)} - \beta_{(2)}$ is the break size. By knowing this form of W, we can rewrite their estimator as

$$\begin{aligned} \widehat{\beta}_{PPP} &= \left(w_{(1)}\ X_1'X_1 + w_{(2)}\ X_2'X_2\right)^{-1}\left(w_{(1)}\ X_1'X_1\widehat{\beta}_1 + w_{(2)}\ X_2'X_2\widehat{\beta}_{(2)}\right) \\ &= \Lambda\widehat{\beta}_{(1)} + (I-\Lambda)\widehat{\beta}_{(2)}, \end{aligned} \tag{3.34}$$

which is the combined estimator of the pre-break and post-break estimators with combination weight $\Lambda = \left(\frac{w_{(1)}}{w_{(2)}}\ X_1'X_1 +\ X_2'X_2\right)^{-1}\left(\frac{w_{(1)}}{w_{(2)}}\ X_1'X_1\right)$. See Pesaran et al. (2013) for details.

Remark 20 *By looking carefully at this estimator and comparing that with the semi-parametric estimator, we can see that the combination weight in these two estimators are very close to each other. By comparing Δ in equation (3.21) and Λ in equation (3.34), we can see that PPP's estimator is the especial case of this semi-parametric estimator under the condition that*

$$\gamma = \frac{1}{q^2 + T_1\phi^2}. \tag{3.35}$$

□

Note that, in the semi-parametric method, we only estimate one unknown parameter, γ, which deals with the break in the coefficients and the break in error variance.

3.5 Extension to multiple breaks

So far, we have talked about the case of having a single break in the model, but in practice the time series model may be subject to multiple breaks. The case of multiple breaks is the straightforward extension of the aforementioned sections for the Stein-like combined estimator. The combined estimator always can be defined as the combination of $\widehat{\beta}_{Full}$ and the estimator after the most recent break point. Here we write the model with two breaks in details, and then we extend that for the case of having more than two breaks. Suppose we have two breaks (three regime) in the model. The break dates are $\{T_1, T_2\}$, and we have breaks in both the coefficients and the error terms, such that

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{for} \quad 1 < t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{for} \quad T_1 < t \leq T_2 \\ x_t'\beta_{(3)} + \sigma_{(3)}\varepsilon_t & \text{for} \quad T_2 < t < T. \end{cases} \tag{3.36}$$

Thus the combined estimator in the case of having two breaks is

$$\hat{\beta}_\alpha = \alpha\widehat{\beta}_{Full} + (1-\alpha)\widehat{\beta}_{(3)}, \tag{3.37}$$

where α is defined as in equation (3.5), in which $H_T = T(\widehat{\beta}_{(3)} - \widehat{\beta}_{Full})'(\widehat{V}_{(3)} - \widehat{V}_{Full})^{-1}(\widehat{\beta}_{(3)} - \widehat{\beta}_{Full})$ and $V_{(3)}$ is the variance of the post-break estimator. It is clear that $\widehat{\beta}_\alpha - \beta_{(3)} = \big(\widehat{\beta}_{(3)} - \beta_{(3)}\big) - \alpha\big(\widehat{\beta}_{(3)} - \widehat{\beta}_{Full}\big)$. So, $\rho(\widehat{\beta}_\alpha, \mathbb{W}) = \mathbb{E}\left[T(\widehat{\beta}_\alpha - \beta_{(3)})'\, \mathbb{W}\, (\widehat{\beta}_\alpha - \beta_{(3)})\right]$. In order to derive the risk, at first we need to find the distribution of $\big(\widehat{\beta}_{(3)} - \beta_{(3)}\big)$ and $\big(\widehat{\beta}_{(3)} - \widehat{\beta}_{Full}\big)$. Let $b_1 = \frac{T_1}{T}, b_2 = \frac{T_2}{T}$, and $b_3 = \frac{T_3}{T}$ where $b_1 < b_2 < b_3$.

Suppose under the local alternative, the break sizes are: $(\beta_{(1)} - \beta_{(2)}, \beta_{(2)} - \beta_{(3)}) =$

$(\frac{\delta_1}{\sqrt{T}}, \frac{\delta_2}{\sqrt{T}})$. So, the full-sample estimator is

$$\begin{aligned}
\sqrt{T}\Big(\widehat{\beta}_{Full} - \beta_{(3)}\Big) &= \Big(\sum_{t=1}^{T} x_t x_t' \frac{1}{T\sigma_t^2}\Big)^{-1} \Bigg[\sum_{t=1}^{T_1} x_t x_t' \ \frac{\sqrt{T}\Big(\beta_{(1)} - \beta_{(3)}\Big)}{T\sigma_{(1)}^2} \\
&\quad + \sum_{t=T_1+1}^{T_2} x_t x_t' \ \frac{\sqrt{T}\Big(\beta_{(2)} - \beta_{(3)}\Big)}{T\sigma_{(2)}^2} + \sum_{t=1}^{T} \frac{x_t \sigma_t \varepsilon_t}{\sqrt{T}\sigma_t^2}\Bigg] \\
&= \Big(\frac{X'\Omega^{-1}X}{T}\Big)^{-1}\Big(\frac{X_1'\Omega_1^{-1}X_1}{Tb_1}\Big) b_1(\delta_1+\delta_2) + \Big(\frac{X'\Omega^{-1}X}{T}\Big)^{-1}\Big(\frac{X_2'\Omega_2^{-1}X_2}{T(b_2-b_1)}\Big) \\
&\quad (b_2-b_1)\delta_2 + \Big(\frac{X'\Omega^{-1}X}{T}\Big)^{-1} \ \sum_{t=1}^{T} \frac{x_t\sigma_t\varepsilon_t}{\sqrt{T}\sigma_t^2} \\
&\xrightarrow{d} \mathrm{N}\big(Q^{-1}Q_1 b_1(\delta_1+\delta_2) + Q^{-1}Q_2(b_2-b_1)\delta_2, V_F\big),
\end{aligned} \tag{3.38}$$

where $X = (X_1' \ X_2' \ X_3')'$ is $T \times k$, and $\Omega = \mathrm{diag}\big(\sigma_{(1)}^2, \ldots, \sigma_{(1)}^2, \sigma_{(2)}^2, \ldots, \sigma_{(2)}^2, \sigma_{(3)}^2, \ldots, \sigma_{(3)}^2\big)$ is a $T \times T$ matrix.

When we have two breaks in the model, the post-break estimator is

$$\begin{aligned}
\sqrt{T}\Big(\widehat{\beta}_{(3)} - \beta_{(3)}\Big) &= \Big(\frac{X_3' \ \Omega_3^{-1} \ X_3}{T-T_2}\Big)^{-1}\Big(\frac{X_3' \ \Omega_3^{-1} \ \sigma_{(3)}\varepsilon_3}{\sqrt{1-b_2}\sqrt{T-T_2}}\Big) \\
&\xrightarrow{d} \mathrm{N}\Big(0, V_{(3)}\Big),
\end{aligned} \tag{3.39}$$

where $V_{(3)} = \frac{1}{1-b_2} \ \underset{T\to\infty}{\mathrm{plim}}\Big(\frac{X_3'\Omega_3^{-1}X_3}{T-T_2}\Big)^{-1} = \frac{1}{1-b_2}Q_3^{-1}$. Like what we did in Theorem 9, we can write the joint asymptotic distribution of the estimators as

$$\sqrt{T}\begin{bmatrix} \widehat{\beta}_{Full} - \beta_{(3)} \\ \widehat{\beta}_{(3)} - \beta_{(3)} \end{bmatrix} \xrightarrow{d} V^{1/2}Z, \tag{3.40}$$

where $Z \sim \mathrm{N}(\theta, I_{2k})$, $\theta = V^{-1/2}\begin{bmatrix} Q^{-1}Q_1 b_1(\delta_1+\delta_2) + Q^{-1}Q_2(b_2-b_1)\delta_2 \\ 0 \end{bmatrix}$ and

$V = \begin{bmatrix} V_{Full} & V_{Full} \\ V_{Full} & V_{(3)} \end{bmatrix}$. Actually the main difference between multiple breaks in compare to the single break is the bias term, θ. We can extend it to the case of having m breaks which breaks happen at times $\{T_1, T_2, \ldots, T_m\}$

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{if} \quad 1 \ < t \leq T_1, \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{if} \quad T_1 < t \leq T_2, \\ \vdots & \\ x_t'\beta_{(m+1)} + \sigma_{(m+1)}\varepsilon_t & \text{if} \quad T_m < t < T. \end{cases} \tag{3.41}$$

In the case of having m breaks in model, the matrix θ which is $2k \times 1$ can be written as

$$\theta = V^{-1/2} \begin{bmatrix} Q^{-1}\Big(Q_1 b_1(\delta_1 + \cdots + \delta_m) + \cdots + Q_m(b_m - B_{m-1})\delta_m\Big) \\ 0 \end{bmatrix}. \tag{3.42}$$

Regarding the extension of the semi-parametric estimator, if we have m breaks in the model, we need to estimate m unknown weight parameter by CV, subject to the condition that weights are in the range of [0 1]. One interesting aspect of the kernel weights under multiple breaks is that, the weights do not necessarily need to be decreasing as the observations get farther from the end of the sample. This allows for the possibility that observations in different regimes which are not next to each other be similar and have similar weight, and might therefore be useful for forecasting out-of-sample.

3.6 Monte Carlo simulation

This section provides some Monto Carlo results based on the Stein-like combined estimator, semi-parametric estimator, PPP's estimator, and the post-break estimator. The goal is to compare the $MSFEs$ for these estimators. To do this, Let $t = 1, ..., T$ with $T = \{100, 200\}$, $q_1 = \frac{\sigma_{(1)}}{\sigma_{(2)}} \in \{0.5, 1, 2\}$ and $k = \{3, 5, 8\}$. We try different values for T_1

which are proportional to the pre-break sample observations, $b_1 = \frac{T_1}{T} \in \{0.2, 0.4, 0.6, 0.8\}$. Suppose $x_t \sim N(0,1)$ and $\varepsilon_t \sim i.i.d.\ N(0,1)$. The data generating process for the single break case is

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{if} \quad 1 < t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{if} \quad T_1 < t \leq T. \end{cases} \tag{3.43}$$

Let $\beta_{(2)}$ be a vector of ones, and $\beta_{(1)} = \beta_{(2)} + \frac{\delta_1}{\sqrt{T}}$ under local alternative. Note that the magnitude of distance between parameters is determined by the localizing parameter δ_1 and sample size T. Assume that $\lambda \equiv \frac{\boldsymbol{\delta_1}}{\sqrt{T}} \in \{0, 0.5, 1\}$ represents different break sizes. The number of replications is 1000.

We report the results based on the relative MSFE with respect to the post-break estimator which is the benchmark estimator, i.e.,

$$RMSFE_i = \frac{MSFE(\hat{y}_i)}{MSFE(\hat{y}_2)}, \tag{3.44}$$

where $MSFE(\hat{y}_i)$ is $MSFE(\hat{y}_\alpha)$, $MSFE(\hat{y}_\gamma)$ and $MSFE(\hat{y}_{PPP})$. Tables 3.1 through 3.6 show the results of the Monte Carlo for the single break case. We have also demonstrate some results in Figures 3.2 and 3.3.

As we may have more than one break in practice, we also consider multiple breaks in our simulation design. Accordingly, we extend our simulation experiments to allow for two breaks that happen at $T_1 = \frac{T}{3}$, and $T_2 = \frac{2T}{3}$ respectively. Define $q_1 = \frac{\sigma_{(1)}}{\sigma_{(3)}}$ and $q_2 = \frac{\sigma_{(2)}}{\sigma_{(3)}}$. The data generating process for the two breaks model follows

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{if} \quad 1 < t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{if} \quad T_1 < t \leq T_2 \\ x_t'\beta_{(3)} + \sigma_{(3)}\varepsilon_t & \text{if} \quad T_2 < t \leq T, \end{cases} \tag{3.45}$$

In order to consider different possibilities that may happen under multiple breaks, we do the Monte Carlo under different experiments. Table 3.7 shows break points specifications with the assigned experiment numbers. The first experiment is under no structural breaks. We allow for both moderate break (0.5) and large break (1) in coefficients in either direction (experiment numbers 2 to 5). Also, we consider the cases that the direction of break in coefficients change from decreasing to increasing or vice versa (experiment numbers 6 to 9). Experiment 10 represents the case that we have partial change in the coefficients, i.e $\beta_{(1)} = \beta_{(2)} \neq \beta_{(3)}$. Experiment 11 also shows the partial break case which $\beta_{(1)} \neq \beta_{(2)} = \beta_{(3)}$. Finally, experiments 12 and 13 represent the higher and lower post-break volatility respectively.

Simulation results

Tables 3.1 to 3.6 show the results of the Monte Carlo for the single break case for different sample size and different q. These tables represent the results of the relative MSFE, $RMSE_{\alpha} \equiv \frac{MSFE(\hat{y}_{\alpha})}{MSFE(\hat{y}_2)}$, $RMSE_{\gamma} \equiv \frac{MSFE(\hat{y}_{\gamma})}{MSFE(\hat{y}_2)}$, and $RMSE_{PPP} \equiv \frac{MSFE(\hat{y}_{PPP})}{MSFE(\hat{y}_2)}$. The benchmark forecast is post-break estimator. The first column in tables shows the ratio of pre-break observations while the second column shows different break sizes, λ. $q = 0.5$ means the variance of the pre-break data is less than the post-break data. In such case, considering the pre-break data in forecasting will improve the forecast accuracy. $q = 1$ happens when there is no break in the variance. Finally $q = 2$ means that the pre-break data is more volatile than the post-break data, so one cannot gain a lot by considering the pre-break data. Based on the results in Tables, the performance of the Stein-like combined estimator is better than (in some cases equivalent) the post-break estimator in the sense of having lower MSFE. As we increase the number of regressors, the MSFE of the Stein-like combined estimator decreases more. When the break size is large ($\lambda = 1$), all estimators perform almost the same as the post-break estimator. As we expected, when we increase the number of regressors, semi-parametric estimator performs well especially for the small and moderate break sizes ($\lambda \leq 0.5$) and this is because of having a lower estimation error. PPP estimator has smaller (almost equal) MSFE than post-break estimator when $k = 3$ and break size is of moderate size (large size), but as we increase the k, for the large break sizes it has under-performance relative to the post-break estimator. We have the similar pattern for the results when $q = 1$ and $q = 2$, but MSFE of estimators are lower with $q = 0.5$. The similar results hold as we increase the sample size from 100 to 200, but when the number of observations increase, the MSFE of estimators are closer to each other relative to $T = 100$.

Tables 3.8 and 3.9 show the results for the two breaks case under the specified set up in Table 3.7. Based on the results Stein-like combined estimator has a lower MSFE than post-break estimator. Also the combined estimator with weight γ performs very well under this set up and even has a lower MSFE than the Stein estimator.

3.7 Empirical analysis

This section presents an empirical illustration of our method. We consider the application of the forecasting procedure for equity premium. The data are from Welch and Goyal (2008) which we examine part of it (1927-2018) where data for all variables are available based on relevant frequency. We consider all data frequencies (monthly, quarterly and yearly) to assess the performance of our model, and report the results based on quarterly data, monthly data and annual data in this section. We refer to Welch and Goyal (2008) for detailed description of the data and sources.

The equity premium (or market premium) is the return on the stock market minus the return on a short-term risk-free treasury bill. Based on the economic variables used to predict the equity premium, we consider all 14 predictors from Welch and Goyal (2008) which data are available from 1927:$Q$1-2018:$Q$4, giving a time-series dimension $T = 368$. The 14 variables are: D/P(dividend price ratio), D/Y(dividend yield), E/P (earning price ratio), D/E (dividend payout ratio), $SVAR$ (stock variance), B/M(book-to-market ratio), $NTIS$(net equity expansion), TBL(treasury bill rate), LTY(long term yield), LTR(long term return), TMS(term spread), DFY(default yield spread), DFR(default return spread), and $INFL$ (inflation).

We recursively compute one-step-ahead forecasts using different forecasting methods for the models described in this chapter. Each time that we expand the window, initially we apply the Schwarz's Bayesian Information Criterio (BIC) to choose the predictors out of $k = 14$ available ones which are critical in predicting the equity premium. In other words, we select the forecasting model using this criterion from all 2^k possible specifications which $k = 14$. As we did not put any restriction on the number of predictors, this criterion may choose all 14 predictors or choose nothing as the extreme cases. Since BIC has a larger penalty term, it is a good choice to use it against in-sample over fitting, Pesaran and Timmermann (1995). Besides, we identify break points by the sequential procedure introduced by Bai and Perron (1998), hereafter BP, where we search for up to eight breaks and set the trimming parameter to 0.1 and the significance level to 5%. Using an initial estimation period of $n_1 = 92$ quarters (23 years) forecasts are recursively generated at each point in the out-of-sample period using only the information available at the time the forecast is made. As the selection of the forecast evaluation period is always somewhat arbitrary, we also report the results with larger estimation window sizes of $\{152, 228\}$. The results are qualitatively similar when a larger number of estimation period is used. The baseline forecast uses the observations after the last break identified by the sequential procedure of BP. We compare our proposed Stein-like combined estimator forecast with the BP post-break forecast, and forecast from the semi-parametric estimator which we derive the weight (γ) by CV. Also we compare our results with the one proposed by Pesaran et al. (2013).

Table 3.10 displays the summary statistics for the equity premium and its predictors. The min, max and high standard deviation shows the higher volatility which can be attributed to D/P, D/Y, E/P and D/E in compare to other predictors.

3.7.1 Forecast evaluation with quarterly data

In order to evaluate the performance of our proposed estimator, we compute its out of sample MSFE and compare it with MSFE from other competing estimators. For this purpose, we divide the sample of T observations into two parts. The first n_1 observation is used as an in-sample estimation period, and the remaining $n_2 = T - n_1$ observations is the out-of-sample period which we recursively do one step ahead forecast. We consider different estimation window sizes to check the sensitivity of the estimators to this choice. The MSFE for the predictive regression model (3.41) over the forecast evaluation period is given by

$$MSFE_i = \frac{1}{n_2} \sum_{t=1}^{n_2} (y_{n_1+t} - \widehat{y}_{i,n_1+t})^2, \tag{3.46}$$

where $MSFE_i$ stands for different forecasting method such as the BP post-break estimator ($MSFE_{BP}$), our proposed Stein-like combined estimator ($MSFE_\alpha$), semi-parametric estimator ($MSFE_\gamma$) and PPP estimator ($MSFE_{PPP}$).

Table 3.11 shows the recursively out-of-sample forecasting results under different in-sample estimation period (n_1), so the beginning of the various forecast evaluation periods runs from 1970:Q1 ($n_1 = 92$) through 2004:Q1 ($n_1 = 228$). We evaluate the forecasts for horizons $h = 1, 2, 3, 4$ quarters. h in the first column shows the forecast horizon. The second column (n_2) shows the start date of the out-of sample period which all ends at $2018 : Q4$. Using the initial in-sample estimation period n_1, forecasts are recursively generated at each point in the out-of-sample period using only the information available at the time the forecast is made. In the heading of Table, $MSFE_{BP}$ is for the case that we only use post-break observations. $MSFE_\alpha$ represents the results based on our proposed Stein-like combined estimator with weight $\alpha = 1 - \frac{\tau}{H_T}$. $MSFE_\gamma$ column represents the results for the semi-parametric estimator which we estimate the weight (γ) by cross validation. $MSFE_{PPP}$ represents the results based on Pesaran et al. (2013) estimator. Based on the results of the Table 3.11, the Stein-like combined estimator delivers vastly improved forecasts (lower MSFE) compared to the BP post-break estimator for all horizons. Besides, the semi-parametric estimator has a lower MSFE than post-break estimator but still combined estimator with the Stein-type weight outperforms these estimators. Our proposed combined estimator performs almost the same as PPP's estimator when $h = 1$, but for $h > 1$, the Stein-like combined estimator outperforms the PPP's estimator. The reason that Stein-type weight performs well can be attributed to the form of its weight that beautifully captures the break size and adjusts the weight between the full-sample estimator and the post-break estimator accordingly.

We also use the out-of-sample R^2 statistics suggested by Campbell and Thompson (2008) to compare the forecasts based on our proposed combined estimators and other existing methods relative to the post-break estimator. We compute the out-of-sample R^2 by

$$R_i^2 = 1 - \frac{MSFE_i}{MSFE_{BP}}, \tag{3.47}$$

where $MSFE_{BP}$ is the mean square forecast error based on the Bai and Perron (1998) post-break estimator (benchmark model), and $MSFE_i$ shows the MSFE for each of the introduced models including "α", "γ", and "PPP". Obviously positive R^2 shows the out-performance of the chosen model relative to the benchmark model, and negative R^2 value indicates the under-performance. Finally, we evaluate whether any of the out/under-performance of our model is statistically significant or not corresponding to testing H_0 : $R^2 = 0$ ($MSFE_{BP} = MSFE_i$) against the alternative hypothesis that $H_1 : R^2 > 0$ ($MSFE_{BP} > MSFE_i$). We report the well known Diebold and Mariano (1995) and West (1996) statistic for testing the null of equal predictive ability. The results are displayed in Table 3.12. Based on the results, the Stein-like combined estimator has a larger positive R^2 compare to the semi-parametric estimator and the post-break estimator, and the value of R^2 is significant across different estimation window size at 5% significance level for all h. Besides, the semi-parametric estimator is performing well in terms of having lower MSFE which actually this estimator facilitate implementing of our theoretical point that using pre-break observations improve forecast accuracy. Based on the results, the semi-parametric estimator performs better for the longer forecast horizons.

3.7.2 Empirical results based on monthly and annual data

Monthly data starts from January 1927 to December 2018 which gives time series dimension of $T = 1104$. Table 3.13 reports the results for MSFE which are recursively generated using different initial in-sample estimation window sizes $\{276, 396, 516\}$ such that the beginning of the various forecast evaluation periods runs from Jan-1950 to Jan-1970. Based on the results, MSFE of the Stein-like combined estimator is lower than all other estimators in the table.

Annual data starts from 1927 to the end of 2018 which gives time series dimension of $T = 92$. We report the results under different in-sample estimation period which correspond to early 1970's ($n_1 = 43$) through early 1990's ($n_1 = 63$). Like what we did for the quarterly and monthly data, Table 3.14 reports the MSFE under different forecast horizons, h. The results are similar to the one that we get with quarterly and monthly data. Stein-like combined estimator has a lower MSFE compare to other estimators for all h.

3.8 Conclusion

In this chapter we introduce the Stein-like combined estimator of the full-sample estimator (using all observations in the sample), and the post-break estimator which uses the observations after the most recent break point. The standard solution for forecasting under structural break is to use the post-break estimator, but it has been shown that using pre-break observations can decrease the MSFE. As our combined estimator uses the pre-break observations, we are able to reduce the variance of forecast error at the cost of adding some bias. We also introduce the semi-parametric estimator which is a discrete kernel weight of all observations in the sample. This estimator gives weight one to the post-break observations, and down-weight pre-break observations by weight $\gamma \in [0\ 1]$, which we can find it numerically by CV. Both the Stein-like combined estimator and the semi-parametric estimator have lower MSFE than the post-break estimator. Especially for the case of large number of regressors, the semi-parametric estimator performs well because it does not involve with the estimation error of many parameters, and we only need to estimate the weight γ. We also compare the performance of our proposed Stein-like combined estimator with the alternative estimator proposed by Pesaran et al. (2013). Based on the simulation results, all of the estimators perform well, and when the number of regressors is high and break size is large, we do not see any underperformance in our proposed estimators. Besides, the results from the empirical example with equity premium shows that the Stein-like combined estimator perform better than other disccussed competing estimators and is more robust to the frequency of data and selection of the initial in-sample estimation window.

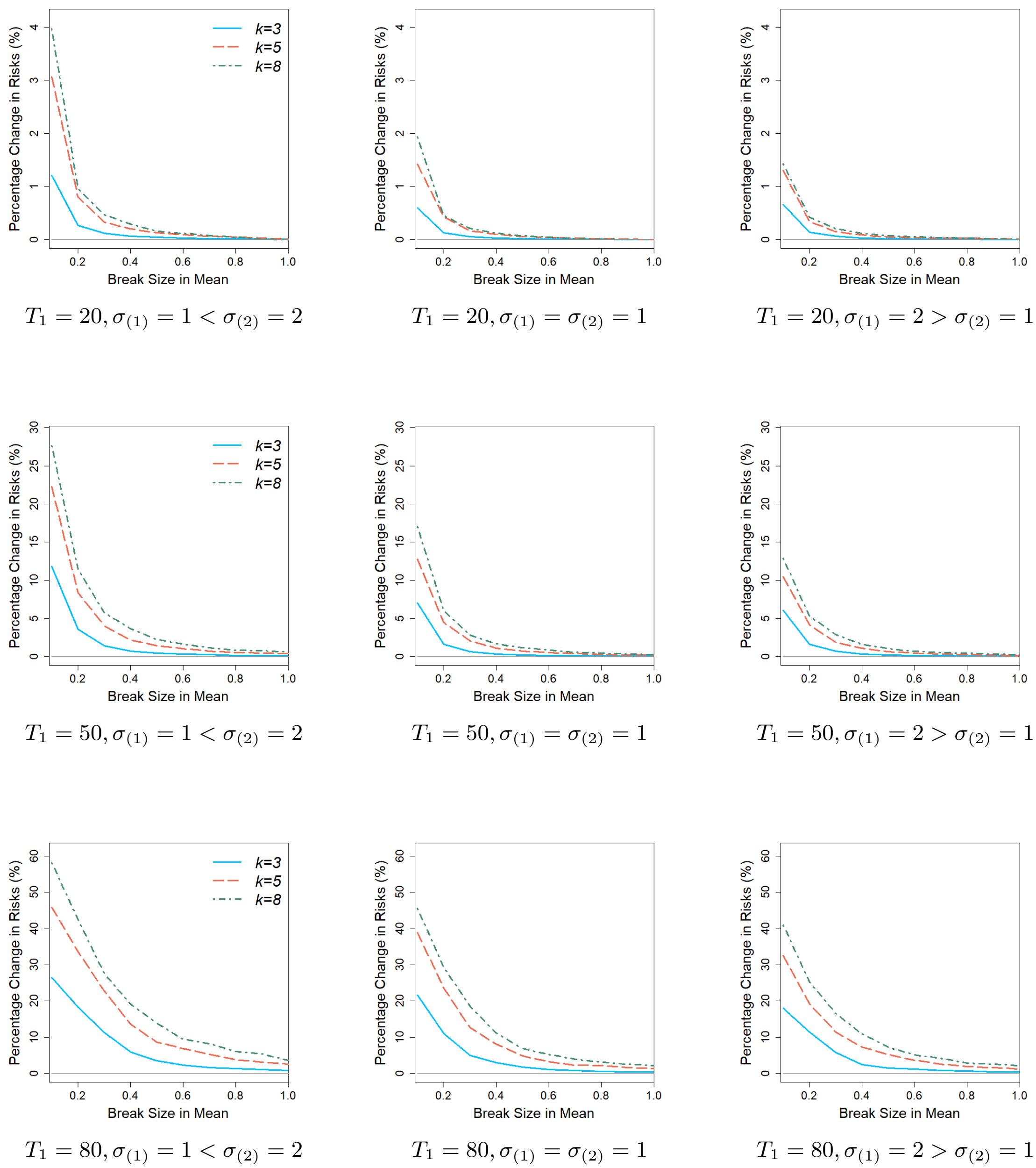
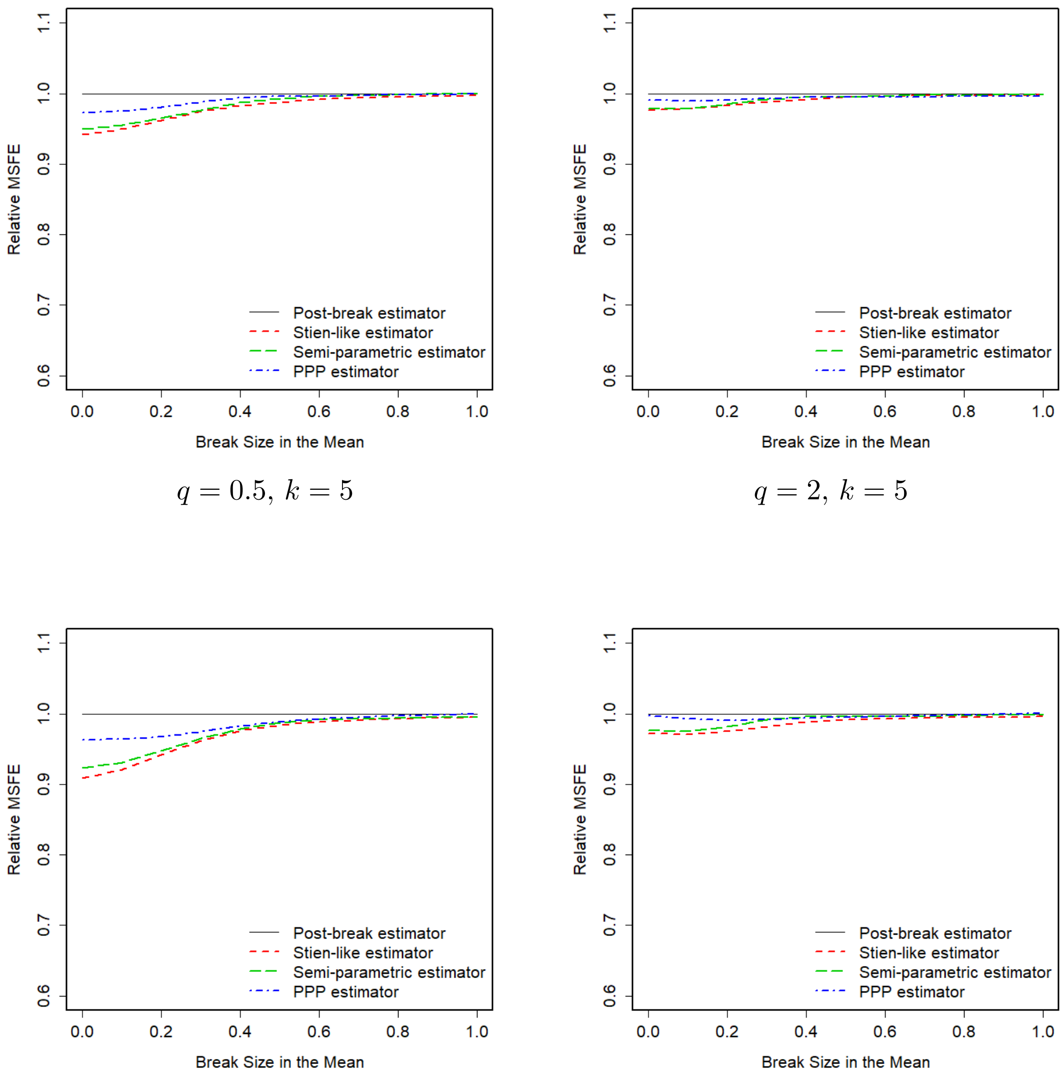


Figure 3.1: Risk-gain(%) between the Stein-like combined estimator and post break estimator, when $T = 100$



$q = 0.5,\ k = 5$ $\quad$ $q = 2,\ k = 5$

$q = 0.5,\ k = 8$ $\quad$ $q = 2,\ k = 8$

Figure 3.2: Simulation results with $T = 100$ and $b_1 = 0.4$

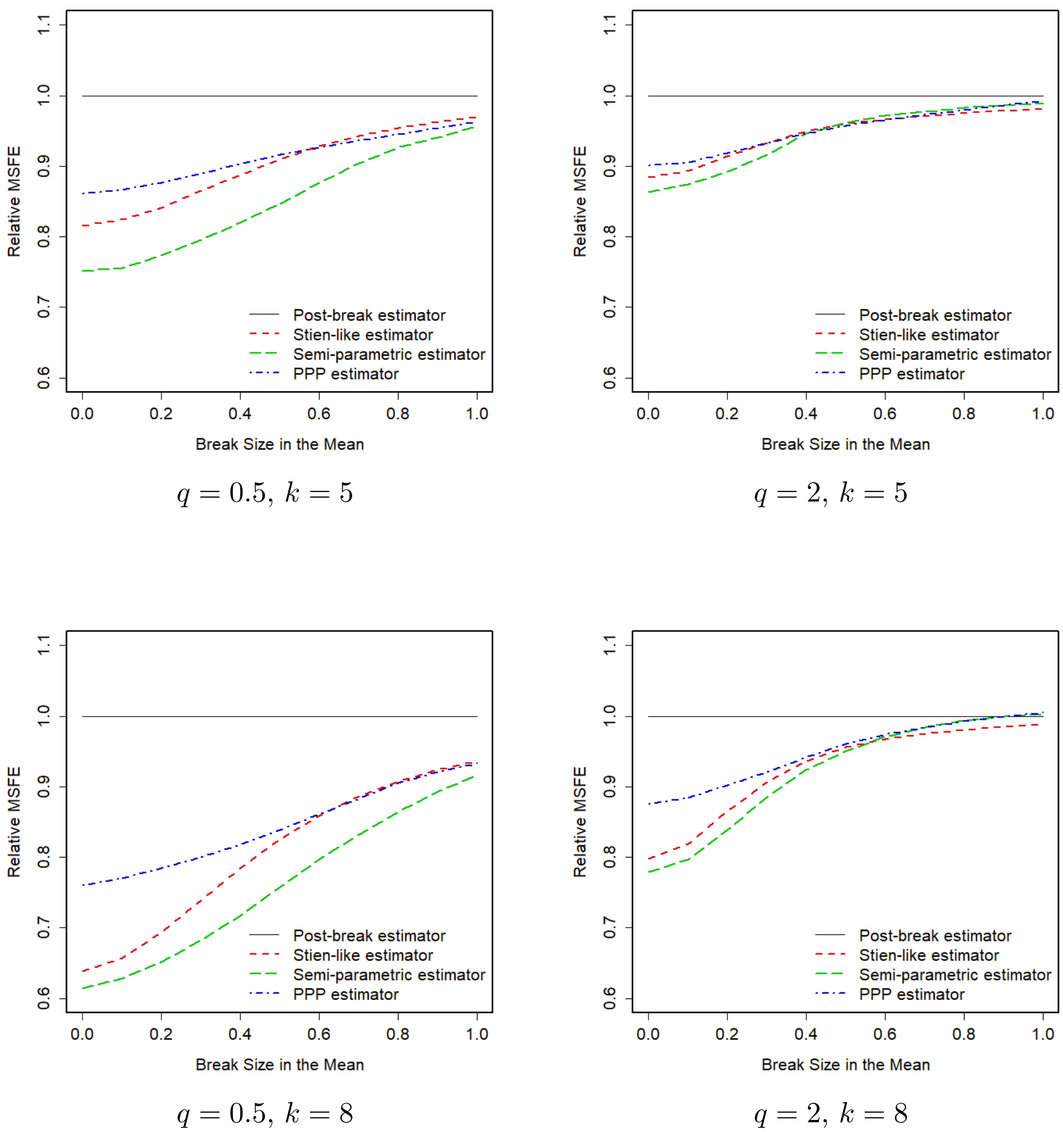


Figure 3.3: Simulation results with $T = 100$ and $b_1 = 0.8$

Table 3.1: Simulation results for a single break, $T = 100$, $q = 0.5$

		$k=3$			$k=5$			$k=8$		
	λ	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$
	0	0.988	0.986	0.996	0.968	0.975	0.992	0.952	0.967	0.989
$b = 0.2$	0.5	1.000	1.001	1.002	0.998	0.996	1.002	0.996	0.995	1.008
	1	1.000	1.002	1.006	1.000	0.999	1.005	0.999	0.999	1.011
	0	0.985	0.976	0.988	0.959	0.957	0.982	0.926	0.926	0.971
$b = 0.4$	0.5	0.998	1.000	1.005	0.996	0.996	1.003	0.993	0.995	1.000
	1	0.999	1.001	1.003	1.000	1.000	1.012	0.999	1.000	1.009
	0	0.975	0.951	0.977	0.925	0.917	0.959	0.871	0.857	0.922
$b = 0.6$	0.5	1.000	1.002	0.997	0.993	0.991	0.995	0.986	0.980	0.993
	1	0.999	1.001	0.999	0.999	0.996	1.002	0.997	0.992	1.012
	0	0.926	0.867	0.923	0.833	0.797	0.880	0.705	0.641	0.785
$b = 0.8$	0.5	0.998	1.002	0.992	0.977	0.976	0.974	0.953	0.921	0.929
	1	1.000	1.001	1.007	1.000	1.001	1.010	0.997	0.984	1.004

Table 3.2: Simulation results for a single break, $T = 100$, $q = 1$

		$k = 3$			$k = 5$			$k = 8$		
	λ	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$
	0	0.994	0.992	0.999	0.986	0.982	0.997	0.979	0.979	0.995
$b = 0.2$	0.5	1.000	1.000	1.001	0.999	0.996	1.004	0.996	0.995	1.003
	1	1.000	1.002	1.002	1.000	0.999	1.004	0.999	0.999	1.010
	0	0.992	0.987	0.993	0.977	0.974	0.990	0.957	0.952	0.983
$b = 0.4$	0.5	0.999	0.999	1.004	0.996	0.995	1.001	0.994	0.994	0.997
	1	0.999	1.001	1.002	1.000	1.000	1.004	0.998	1.000	1.001
	0	0.982	0.963	0.984	0.947	0.940	0.971	0.902	0.885	0.939
$b = 0.6$	0.5	1.000	1.002	0.996	0.991	0.989	0.994	0.986	0.979	0.988
	1	0.999	1.000	1.000	0.999	0.996	1.013	0.996	0.991	1.004
	0	0.940	0.882	0.930	0.860	0.822	0.893	0.737	0.672	0.803
$b = 0.8$	0.5	0.999	1.001	0.989	0.973	0.972	0.973	0.940	0.918	0.931
	1	1.000	1.001	1.005	0.999	1.001	1.001	0.994	0.982	1.003

Table 3.3: Simulation results for a single break, $T = 100$, $q = 2$

		$k = 3$			$k = 5$			$k = 8$		
	λ	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$	$RMSE_{\alpha}$	$RMSE_{\gamma}$	$RMSE_{PPP}$
	0	0.997	0.998	1.000	0.995	0.995	1.001	0.993	0.993	0.998
$b = 0.2$	0.5	1.000	1.000	1.000	0.998	0.996	1.001	0.996	0.994	0.998
	1	1.000	1.002	1.002	1.000	0.999	1.002	0.999	0.999	1.002
	0	0.996	0.997	0.997	0.991	0.990	0.996	0.983	0.986	0.993
$b = 0.4$	0.5	0.999	0.998	1.000	0.996	0.994	0.997	0.995	0.994	0.995
	1	0.999	1.001	1.000	1.000	1.000	1.010	0.999	1.000	1.001
	0	0.993	0.985	0.991	0.976	0.974	0.986	0.949	0.946	0.970
$b = 0.6$	0.5	0.999	1.000	0.995	0.990	0.986	0.988	0.985	0.978	0.984
	1	0.999	1.000	0.999	0.998	0.995	1.005	0.994	0.991	1.001
	0	0.964	0.911	0.951	0.910	0.890	0.925	0.807	0.772	0.843
$b = 0.8$	0.5	0.994	0.989	0.982	0.968	0.969	0.968	0.924	0.914	0.934
	1	1.000	1.001	1.002	0.998	1.000	1.003	0.986	0.979	0.997

Table 3.4: Simulation results for a single break, $T = 200$, $q = 0.5$

		$k = 3$			$k = 5$			$k = 8$		
	λ	$RMSE_\alpha$	$RMSE_\gamma$	$RMSE_{PPP}$	$RMSE_\alpha$	$RMSE_\gamma$	$RMSE_{PPP}$	$RMSE_\alpha$	$RMSE_\gamma$	$RMSE_{PPP}$
	0	0.998	0.997	0.999	0.991	0.992	0.996	0.979	0.984	0.995
$b = 0.2$	0.5	0.999	0.998	1.000	0.999	0.998	0.999	0.999	1.000	1.009
	1	1.000	1.000	1.003	1.000	1.000	1.005	1.000	1.000	1.007
	0	0.992	0.984	0.994	0.972	0.974	0.988	0.952	0.953	0.981
$b = 0.4$	0.5	0.998	0.996	0.996	0.998	0.997	0.997	0.999	0.999	1.010
	1	1.000	1.000	1.002	1.000	1.000	1.005	1.000	1.000	1.004
	0	0.983	0.970	0.984	0.955	0.951	0.976	0.914	0.916	0.960
$b = 0.6$	0.5	0.997	0.994	0.995	0.997	0.996	0.999	0.996	0.998	1.003
	1	1.000	1.000	1.003	0.999	1.000	1.006	0.999	0.999	1.005
	0	0.968	0.931	0.963	0.903	0.882	0.931	0.819	0.784	0.884
$b = 0.8$	0.5	0.996	0.994	0.996	0.994	0.995	0.997	0.987	0.984	1.002
	1	1.000	1.000	1.005	0.998	1.000	1.006	0.997	0.995	1.001

Table 3.5: Simulation results for a single break, $T = 200$, $q = 1$

		$k=3$			$k=5$			$k=8$		
	λ	$RMSE_\alpha$	$RMSE_\gamma$	$RMSE_{PPP}$	$RMSE_\alpha$	$RMSE_\gamma$	$RMSE_{PPP}$	$RMSE_\alpha$	$RMSE_\gamma$	$RMSE_{PPP}$
	0	0.999	1.001	1.000	0.997	0.999	0.999	0.994	0.995	0.998
$b=0.2$	0.5	0.999	0.997	0.999	0.999	0.998	0.997	1.000	1.000	1.002
	1	1.000	1.000	1.002	1.000	1.000	1.004	1.000	1.000	1.009
	0	0.998	0.994	0.997	0.986	0.986	0.994	0.974	0.974	0.988
$b=0.4$	0.5	0.998	0.996	0.996	0.999	0.997	0.998	0.999	1.000	1.005
	1	1.000	1.000	1.001	1.000	1.000	1.005	1.000	1.000	1.006
	0	0.992	0.982	0.991	0.972	0.967	0.984	0.944	0.942	0.975
$b=0.6$	0.5	0.997	0.994	0.996	0.998	0.996	0.999	0.997	0.999	1.004
	1	1.000	1.000	1.002	0.999	1.000	1.007	0.999	0.999	1.005
	0	0.978	0.946	0.973	0.922	0.900	0.942	0.848	0.811	0.900
$b=0.8$	0.5	0.996	0.993	0.995	0.994	0.995	0.996	0.987	0.985	1.005
	1	1.000	1.000	1.008	0.998	1.001	1.009	0.997	0.995	1.003

Table 3.6: Simulation results for a single break, $T = 200$, $q = 2$

	λ	$k=3$ $RMSE_{\alpha}$	$k=3$ $RMSE_{\gamma}$	$k=3$ $RMSE_{PPP}$	$k=5$ $RMSE_{\alpha}$	$k=5$ $RMSE_{\gamma}$	$k=5$ $RMSE_{PPP}$	$k=8$ $RMSE_{\alpha}$	$k=8$ $RMSE_{\gamma}$	$k=8$ $RMSE_{PPP}$
	0	1.000	1.001	1.000	1.001	1.002	1.000	1.000	1.001	0.999
$b = 0.2$	0.5	0.999	0.998	0.999	0.999	0.999	1.000	1.000	1.000	0.999
	1	1.000	1.000	1.000	1.000	1.000	1.002	1.000	1.000	0.999
	0	1.001	1.007	1.000	0.998	1.000	0.998	0.992	0.994	0.996
$b = 0.4$	0.5	0.998	0.996	0.996	0.999	0.998	0.998	1.000	1.000	1.000
	1	1.000	1.000	1.000	1.000	1.000	1.004	1.000	1.001	1.001
	0	0.999	1.002	0.997	0.990	0.990	0.994	0.978	0.982	0.991
$b = 0.6$	0.5	0.997	0.994	0.996	0.998	0.997	0.998	0.998	0.999	1.007
	1	1.000	1.000	1.000	0.999	1.000	1.004	0.999	0.999	1.003
	0	0.990	0.979	0.986	0.957	0.944	0.964	0.901	0.880	0.935
$b = 0.8$	0.5	0.996	0.993	0.993	0.995	0.996	0.997	0.988	0.986	1.005
	1	1.000	1.000	1.007	0.998	1.001	1.009	0.997	0.996	1.001

Table 3.7: Break point specifications by experiments

Experiment No.	$\beta_{(1)}$	$\beta_{(2)}$	$\beta_{(3)}$	$\sigma_{(1)}$	$\sigma_{(2)}$	$\sigma_{(3)}$
#1 : No break	1	1	1	1	1	1
#2 : Moderate break in coefficients (decline)	1.5	1	0.5	1	1	1
#3 : Large break in coefficients (decline)	2	1	0	1	1	1
#4 : Moderate break in coefficients (increase)	0.5	1	1.5	1	1	1
#5 : Large break in coefficients (increase)	0	1	2	1	1	1
#6 : Moderate decreasing and increasing break	1.5	1	1.5	1	1	1
#7 : Large decreasing and increasing break	2	1	2	1	1	1
#8 : Moderate increasing and decreasing break	0.5	1	0.5	1	1	1
#9 : Large increasing and decreasing break	0	1	0	1	1	1
#10 : No break, then increasing break	1	1	2	1	1	1
#11 : Decreasing, then no break	2	1	1	1	1	1
#12 : Higher post-break volatility	1	1	1	1	1	2
#13 : Lower post-break volatility	1	1	1	1	1	0.5

Table 3.8: Simulation results for multiple breaks, $T = 100$

	$k = 3$			$k = 5$			$k = 8$		
Experiment	$RMSFE_{\alpha}$	$RMSFE_{\gamma}$	$RMSFE_{PPP}$	$RMSFE_{\alpha}$	$RMSFE_{\gamma}$	$RMSFE_{PPP}$	$RMSFE_{\alpha}$	$RMSFE_{\gamma}$	$RMSFE_{PPP}$
#1	0.951	0.957	0.998	0.887	0.906	0.996	0.820	0.830	0.982
#2	0.999	0.994	1.019	0.995	0.983	1.018	0.991	0.979	1.009
#3	0.999	1.000	1.010	0.998	0.995	1.011	0.997	0.993	1.009
#4	0.999	0.997	1.008	0.996	0.992	1.012	0.995	0.978	1.003
#5	1.000	1.004	1.015	1.000	1.004	1.014	1.000	1.003	1.008
#6	0.983	0.964	1.004	0.967	0.921	0.992	0.911	0.860	0.990
#7	0.998	0.968	1.011	0.995	0.942	1.009	0.990	0.876	1.010
#8	0.991	0.975	1.001	0.960	0.929	0.996	0.955	0.868	0.995
#9	0.994	0.962	1.009	1.000	0.935	1.013	1.000	0.878	1.009
#10	0.999	1.002	1.009	0.999	0.997	1.009	0.995	0.994	1.015
#11	0.984	0.966	1.011	1.000	0.934	1.004	1.000	0.876	1.095
#12	0.934	0.949	1.011	0.859	0.892	1.002	0.782	0.816	1.001
#13	0.965	0.968	1.000	0.909	0.921	0.997	0.849	0.854	0.975

Table 3.9: Simulation results for multiple breaks, $T = 200$

	$k=3$			$k=5$			$k=8$		
Experiment	$RMSFE_{\alpha}$	$RMSFE_{\gamma}$	$RMSFE_{PPP}$	$RMSFE_{\alpha}$	$RMSFE_{\gamma}$	$RMSFE_{PPP}$	$RMSFE_{\alpha}$	$RMSFE_{\gamma}$	$RMSFE_{PPP}$
#1	0.982	0.981	0.999	0.953	0.959	0.999	0.933	0.938	0.995
#2	0.999	0.987	1.002	0.999	0.972	1.007	0.997	0.950	1.021
#3	0.999	0.983	1.007	0.998	0.974	1.016	0.996	0.939	1.011
#4	0.999	1.001	1.008	1.000	0.994	1.040	1.000	0.989	1.057
#5	0.999	1.001	1.038	0.999	1.000	1.040	0.999	0.999	1.066
#6	0.999	0.991	1.002	0.990	0.969	1.009	0.970	0.949	1.015
#7	0.999	0.986	1.001	1.000	0.973	1.015	1.000	0.950	1.039
#8	0.995	0.991	1.007	0.983	0.976	1.012	0.980	0.967	1.011
#9	0.998	0.988	1.007	0.993	0.976	1.023	0.991	0.956	1.055
#10	0.994	1.003	1.015	0.988	1.003	1.013	0.984	1.007	1.040
#11	1.000	0.991	1.007	0.993	0.965	1.010	0.989	0.939	1.029
#12	0.979	0.978	1.000	0.943	0.950	1.001	0.917	0.924	0.996
#13	0.983	0.985	0.998	0.959	0.969	1.000	0.941	0.950	0.998

Table 3.10: Summary statistics for quarterly data

Variables	Mean	St. dev.	Min.	Max.	Median
Equity Premium	0.0183	0.1058	-0.3654	0.6700	0.0280
D/P	-3.3785	0.4660	-4.4932	-1.9039	-3.3567
D/Y	-3.3643	0.4600	-4.4966	-2.0331	-3.3291
E/P	-2.7414	0.4223	-4.8074	-1.7750	-2.7926
D/E	-0.6372	0.3341	-1.2442	1.3795	-0.6304
$SVAR$	0.0086	0.0147	0.0004	0.1144	0.0039
BM	0.5695	0.2678	0.1252	2.0285	0.5462
$NTIS$	0.0167	0.0257	-0.0529	0.1635	0.0168
TBL	0.0339	0.0309	0.0001	0.1549	0.0293
LTY	0.0509	0.0279	0.0179	0.1482	0.0420
LTR	0.0144	0.0466	-0.1451	0.2437	0.0094
TMS	0.0170	0.0130	-0.0350	0.0453	0.0175
DFY	0.0113	0.0070	0.0034	0.0559	0.0090
DFR	0.0010	0.0219	-0.1184	0.1629	0.0018
$INFL$	0.0074	0.0130	-0.0411	0.0909	0.0072

Table 3.11: Out-of-sample forecasting performance with quarterly data

h	n_2	$MSFE_{\alpha}$	$MSFE_{\gamma}$	$MSFE_{BP}$	$MSFE_{PPP}$
1	1950:Q1-2018:Q4	0.0067	0.0076	0.0088	0.0068
	1960:Q1-2018:Q4	0.0073	0.0081	0.0097	0.0074
	1970:Q1-2018:Q4	0.0078	0.0089	0.0099	0.0080
2	1950:Q1-2018:Q4	0.0068	0.0083	0.0105	0.0087
	1960:Q1-2018:Q4	0.0073	0.0089	0.0116	0.0096
	1970:Q1-2018:Q4	0.0077	0.0097	0.0105	0.0086
3	1950:Q1-2018:Q4	0.0073	0.0086	0.0128	0.0239
	1960:Q1-2018:Q4	0.0079	0.0089	0.0138	0.0272
	1970:Q1-2018:Q4	0.0077	0.0093	0.0109	0.0082
4	1950:Q1-2018:Q4	0.0081	0.0079	0.0155	0.0078
	1960:Q1-2018:Q4	0.0087	0.0085	0.0173	0.0084
	1970:Q1-2018:Q4	0.0080	0.0089	0.0119	0.0087

Table 3.12: Statistical significance of predictive accuracy with quarterly data

h	n_1	R^2_α	R^2_γ	R^2_{PPP}
1	1950:Q1-2018:Q4	0.2414	0.1379	0.2299
		(0.0000)	(0.0001)	(0.0000)
	1960:Q1-2018:Q4	0.2474	0.1649	0.2371
		(0.0000)	(0.0001)	(0.0000)
	1970:Q1-2018:Q4	0.2121	0.1010	0.2020
		(0.0000)	(0.0026)	(0.0001)
2	1950:Q1-2018:Q4	0.3509	0.2632	0.3333
		(0.0000)	(0.0000)	(0.0000)
	1960:Q1-2018:Q4	0.3659	0.2846	0.3333
		(0.0000)	(0.0000)	(0.0000)
	1970:Q1-2018:Q4	0.2655	0.1593	0.3274
		(0.0000)	(0.0001)	(0.0000)
3	1950:Q1-2018:Q4	0.4527	0.3176	-0.6757
		(0.0000)	(0.0002)	(0.0001)
	1960:Q1-2018:Q4	0.4516	0.3484	-0.8194
		(0.0000)	(0.0003)	(0.0003)
	1970:Q1-2018:Q4	0.3231	0.1462	0.2077
		(0.0000)	(0.0007)	(0.0002)
4	1950:Q1-2018:Q4	0.5333	0.4444	0.4611
		(0.0000)	(0.0004)	(0.0001)
	1960:Q1-2018:Q4	0.5497	0.4712	0.4503
		(0.0000)	(0.0006)	(0.0003)
	1970:Q1-2018:Q4	0.3986	0.2238	0.2238
		(0.0000)	(0.0008)	(0.0015)

Table 3.13: Out-of-sample forecasting performance with monthly data

h	n_2	$MSFE_\alpha$	$MSFE_\gamma$	$MSFE_{BP}$	$MSFE_{PPP}$
1	1950:01-2018:12	0.0019	0.0021	0.0020	0.0329
	1960:01-2018:12	0.0020	0.0022	0.0021	0.0382
	1970:01-2018:12	0.0022	0.0023	0.0022	0.0458
2	1950:01-2018:12	0.0019	0.0022	0.0021	0.0027
	1960:01-2018:12	0.0020	0.0023	0.0021	0.0030
	1970:01-2018:12	0.0022	0.0024	0.0023	0.0033
3	1950:01-2018:12	0.0019	0.0022	0.0021	0.0021
	1960:01-2018:12	0.0020	0.0023	0.0022	0.0023
	1970:01-2018:12	0.0022	0.0025	0.0023	0.0025
4	1950:01-2018:12	0.0019	0.0022	0.0021	0.0029
	1960:01-2018:12	0.0021	0.0023	0.0023	0.0032
	1970:01-2018:12	0.0022	0.0024	0.0023	0.0036

Table 3.14: Out-of-sample forecasting performance with annual data

h	n_2	$MSFE_\alpha$	$MSFE_\gamma$	$MSFE_{BP}$	$MSFE_{PPP}$
1	1970-2018	0.0300	0.0323	0.0323	0.0337
	1980-2018	0.0265	0.0288	0.0288	0.0291
	1990-2018	0.0277	0.0279	0.0279	0.028
2	1970-2018	0.0360	0.0520	0.0520	0.0356
	1980-2018	0.0313	0.0508	0.0507	0.0259
	1990-2018	0.0282	0.0289	0.0288	0.0285
3	1970-2018	0.0288	0.0364	0.0363	0.0315
	1980-2018	0.0270	0.0360	0.0359	0.0260
	1990-2018	0.0268	0.0278	0.0278	0.0271
4	1970-2018	0.0320	0.0407	0.0407	0.3013
	1980-2018	0.0277	0.0356	0.0356	0.3378
	1990-2018	0.0275	0.0284	0.0284	0.0276

Chapter 4

Estimation and Forecasting of Heterogeneous Panel Data Models with Multiple Breaks

4.1 Introduction

By increasing the availability of longitudinal data sets, panel data models have become a popular tool and attracted much attention in statistics and econometrics. The extension of structural break models to the panel data setup is crucial because a structural break is regarded as an exogenous shock with permanent effects on the economic variables, and such a shock is likely to have impacts on many economic variables simultaneously. Recently there has been a growing literature on the estimation and tests of common breaks in panel data models in which there are N individual units and T time series observations for each individual. Much of the existing research concentrates on the detection of change points and asymptotic properties of their estimators.

Single change point estimation in the linear regression model is analyzed in Bai (1997a). Bai and Perron (1998, 2003) extends the results of Bai (1997a) to the multiple break points, and also propose tests for detecting the number of breaks. Bai et al. (1998) extends the work for multivariate time series and find that the number of series is positively related to the accuracy of the change point estimator. Joseph and Wolfson (1992, 1993) develop the idea of structural breaks to the panel data model. Bai (2010) studies the asymptotic properties of the change point estimator for the cross section independence model but allowing serially correlation within each panel. Hsu and Lin (2012) extends Bai (2010) theory to non-stationary panel data models where the error terms follow an $I(1)$ process. Results on testing for breaks have been extended to various forms of linear regression by Horváth and Hušková (2012), Pesaran and Yamagata (2008), Su and Chen (2013), Kim (2011, 2014), Baltagi et al. (2016), Baltagi et al. (2017) and Kao et al. (2012, 2018), among others. Furthermore, Chan et al. (2008) extend the testing procedure of Andrews (2003) from time series to heterogeneous panels where the breaks may occur at different time points across individuals. Smith (2018) develops a Bayesian approach to estimate noncommon breaks in panel regression models.

Despite the considerable attention to detection of break points with panel model, there are only a few papers that focusing on forecasting panel data model under the structural breaks. Liu (2018) constructs individual-specific density forecasts using a dynamic linear panel data model with common and heterogeneous coefficients and cross-sectional heteroskedasticity. Smith and Timmermann (2018) develop a Bayesian panel regression approach to estimating an unknown number of breaks and forecasting future outcomes. See also Smith (2018) who develops a Bayesian methodology by considering the regime-specific grouped heterogeneity, and Liu et al. (2020) who focus on point forecasts in dynamic panel data model.

This chapter provides a new averaging estimator, which is also well-known as the Stein-like shrinkage estimator, for the slope coefficients in the case of panel data models when the cross section dimension (N) is fixed while the time dimension (T) is allowed to increase without bounds. The importance of allowing for large T is crucial because for example technological changes or policy implementations are likely to happen over the long time horizons. We consider the common date of change over all cross sections, and we allow for the cross-sectional dependence to take advantage of more observations using panels, as opposed to time-series or cross-sectional data. The usefulness of the common breaks is obvious for the cases that global technological changes or financial shocks affect all markets or firms at the same time.

Our proposed estimator takes a weighted average of the unrestricted estimator and the restricted estimator using the combination weight constructed from a quadratic loss. See also Massoumi (1978) and Hansen (2016, 2017) for applying the stein-type weight in different context. The unrestricted estimator estimates the coefficient using the generalized least square method (GLS) by considering the common breaks across individuals. The restricted estimator is built under the restriction of no breaks in the coefficient, and therefore it will be bias when we have a break, but it is more efficient. We establish the asymptotic distribution for the Stein-like shrinkage estimator and show that the asymptotic risk of this estimator is strictly smaller than the unrestricted estimator under some conditions on the shrinkage parameter.

We undertake the Monte Carlo simulation study to evaluate the performance of the proposed estimator. Besides, we use the Stein-like shrinkage estimator to forecast industry level inflation rate, and the results confirms our theoretical findings.

This chapter is structured as follows. Section 4.2 sets up the heterogeneous panel data model under the structural breaks. Section 4.3 introduces the Stein-like shrinkage estimator and its asymptotic distributions and asymptotic risk. Section 4.4 extends the model with multiple structural breaks. Section 4.5 reports Monte Carlo simulation. Finally Section 4.7 concludes. Detailed proofs are provided in the Appendix.

4.2 The model and assumptions

Consider the following heterogeneous linear panel data model with a common break across individuals at time T_1

$$y_{i,t} = x_{i,t}'\beta_i + u_{i,t} \quad \text{for } i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{4.1}$$

where

$$\beta_i = \begin{cases} \beta_{i(1)} & \text{for} \quad t = 1, \ldots, T_1, \\ \beta_{i(2)} & \text{for} \quad t = T_1 + 1, \ldots, T, \end{cases} \tag{4.2}$$

$x_{i,t}$ is a $k \times 1$ vector of regressors, and $u_{i,t}$ is the error term with zero mean and allowed to have cross-sectional dependence as well as heteroskedasticity.

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

$$Y_i = X_i \boldsymbol{\beta}_i + U_i, \tag{4.3}$$

where $Y_i = \begin{bmatrix} y_{i(1)} \\ y_{i(2)} \end{bmatrix}$ is a vector of $T \times 1$ dependent variables, with $y_{i(1)} = \begin{bmatrix} y_{i,1} \\ \vdots \\ y_{i,T_1} \end{bmatrix}$ is $T_1 \times 1$, $y_{i(2)} = \begin{bmatrix} y_{i,T_1+1} \\ \vdots \\ y_{i,T} \end{bmatrix}$ is $(T - T_1) \times 1$. Also, $X_i = \begin{bmatrix} x_{i(1)} & \mathbf{0} \\ \mathbf{0} & x_{i(2)} \end{bmatrix}$ is a $T \times 2k$ matrix where $x_{i(1)} = \begin{bmatrix} x'_{i,1} \\ \vdots \\ x'_{i,T_1} \end{bmatrix}$ is $T_1 \times k$, and $x_{i(2)} = \begin{bmatrix} x'_{i,T_1+1} \\ \vdots \\ x'_{i,T} \end{bmatrix}$ is $(T - T_1) \times k$. $\boldsymbol{\beta}_i = \begin{bmatrix} \beta_{i(1)} \\ \beta_{i(2)} \end{bmatrix}$ is $2k \times 1$, and $U_i = \begin{bmatrix} u_{i(1)} \\ u_{i(2)} \end{bmatrix}$ is $T \times 1$ with $u_{i(1)} = \begin{bmatrix} u_{i,1} \\ \vdots \\ u_{i,T_1} \end{bmatrix}$ is $T_1 \times 1$, , and $u_{i(2)} = \begin{bmatrix} u_{i,T_1+1} \\ \vdots \\ u_{i,T} \end{bmatrix}$ is $(T - T_1) \times 1$.

We can stack the model over individuals and write it in a matrix form as

$$Y = X\boldsymbol{b} + U, \tag{4.4}$$

where $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$ is a vector of $NT \times 1$ dependent variables, $X = \begin{bmatrix} X_1 & \mathbf{0} & \ldots & \mathbf{0} \\ 0 & X_2 & \ldots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & X_N \end{bmatrix}$ is a matrix of $NT \times 2NK$ regressors, $\boldsymbol{b} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_N \end{bmatrix}$ is a vector of $2NK \times 1$, and $U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix}$.

4.2.1 Assumptions

A1. $\mathbb{E}(U_i | X_1, \ldots, X_N) = 0$.

A2. $\underset{T\to\infty}{\text{plim}} \left(\frac{X'\Omega^{-1}X}{NT}\right)^{-1} = Q$, where Q is a positive definite matrix. This guaranties that the GLS estimators are uniquely defined.

A3. We assume cross-sectional dependence such that the variance of the error term is

$$Var(U) \equiv \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \ldots & \Omega_{1N} \\ \Omega_{21} & \Omega_{22} & \ldots & \Omega_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{N1} & \Omega_{N2} & \ldots & \Omega_{NN} \end{bmatrix}$$ which is a matrix of $NT \times NT$,

where $\Omega_{ij} = \begin{bmatrix} cov\big(u_{i(1)}, u_{j(1)}\big) & \mathbf{0} \\ \mathbf{0} & cov\big(u_{i(2)}, u_{j(2)}\big) \end{bmatrix} = \begin{bmatrix} \sigma^2_{ij(1)} I_{T_1} & \mathbf{0} \\ \mathbf{0} & \sigma^2_{ij(2)} I_{(T-T_1)} \end{bmatrix}$. We can estimate the elements of Ω_{ij} as $\widehat{\sigma}^2_{ij(1)} = \frac{\widehat{u}'_{i(1)}\widehat{u}_{j(1)}}{T_1-k}$, where $\widehat{u}_{i(1)} = y_{i(1)} - x_{i(1)}\widehat{\beta}_{i(1)}$, and $\widehat{\beta}_{i(1)} = \big(x'_{i(1)}x_{i(1)}\big)^{-1} x'_{i(1)}y_{i(1)}$. Similarly, $\widehat{\sigma}^2_{ij(2)} = \frac{\widehat{u}'_{i(2)}\widehat{u}_{j(2)}}{T-T_1-k}$, where $\widehat{u}_{i(2)} = y_{i(2)} - x_{i(2)}\widehat{\beta}_{i(2)}$, and $\widehat{\beta}_{i(2)} = \big(x'_{i(2)}x_{i(2)}\big)^{-1} x'_{i(2)}y_{i(2)}$, for $i = 1, \ldots, N$.

4.3 Stein-like shrinkage estimator

For the estimation of the parameters in the introduced panel data model, we consider the Stein-like shrinkage estimator that can reduce the estimation error under structural breaks. Our proposed shrinkage estimator denoted by $\widehat{\boldsymbol{b}}_w$ is

$$\widehat{\boldsymbol{b}}_w = w\widehat{\boldsymbol{b}}_{ur} + (1-w)\widehat{\boldsymbol{b}}_r, \tag{4.5}$$

where $\widehat{\boldsymbol{b}}_r$ is called the restricted estimator and it is under the null hypothesis of no breaks in the coefficients, and it estimates the parameters by ignoring the breaks. So, this estimator is bias when we have a break but it is efficient. The unrestricted estimator, $\widehat{\boldsymbol{b}}_{ur}$, estimates the coefficients by considering the common break points across all individuals; This is the unbiased estimator but less efficient. Basically combining the restricted and unrestricted estimators trade-off the bias and variance efficiency. The combination weight takes the form of

$$w = \left(1 - \frac{\tau}{D_T}\right)_+, \tag{4.6}$$

which includes the positive part, τ is the shrinkage parameter and controls the degree of shrinkage, D_T is the weighted quadratic loss which is equal to

$$D_T = T(\widehat{\boldsymbol{b}}_{ur} - \widehat{\boldsymbol{b}}_r)' \, \mathbb{W}(\widehat{\boldsymbol{b}}_{ur} - \widehat{\boldsymbol{b}}_r), \tag{4.7}$$

and $\mathbb{W}$ is any positive definite weight matrix. For example, if $\mathbb{W} = \left(X'\Omega^{-1}X\right)^{-1}$, then D_T is equivalent to the Wald statistic. For large break sizes (large value of D_T), the shrinkage estimator assigns more weights to the unrestricted estimator. The reason for that is related to the large bias that restricted estimator adds to the model. On the other hand, for small break sizes (small value of D_T), the shrinkage estimator assigns more weight to the restricted estimator to gain from its efficiency. Basically having this form of weight helps to assign appropriate weight to each of the estimators based on the severity of the breaks.

4.3.1 Unrestricted estimator

As we mentioned earlier, the unrestricted estimator considers the common breaks across individuals, and estimate the coefficients using the available observations within each regime. Using GLS method, we estimate the coefficients as

$$\widehat{\boldsymbol{b}}_{ur} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y. \tag{4.8}$$

4.3.2 Restricted estimator

Alternatively, one can estimate the coefficients by ignoring the break points over all individuals. Basically, this estimator is constructed under the assumption of no breaks in the model. The restricted estimator is

$$\begin{aligned} \widehat{\boldsymbol{b}}_r &= \widehat{\boldsymbol{b}}_{ur} - \left(X'\Omega^{-1}X\right)^{-1} R' \left[R \left(X'\Omega^{-1}X\right)^{-1} R'\right]^{-1} R\widehat{\boldsymbol{b}}_{ur} \\ &= \widehat{\boldsymbol{b}}_{ur} - V_{ur} L R \widehat{\boldsymbol{b}}_{ur}, \end{aligned} \tag{4.9}$$

where we define $L \equiv R' \left[R \left(X' \; \Omega^{-1} X \; \right)^{-1} R'\right]^{-1}$ which is a $2Nk \times Nk$ matrix.

4.3.3 Asymptotic distribution

Our analysis is asymptotic as the time series dimension $T \to \infty$ while the number of individual units, N, is fixed. Under the local asymptotic normality approach, consider the parameter sequences of the form $\boldsymbol{b} = \boldsymbol{b}_0 + \frac{\mathbf{h}}{\sqrt{T}}$, where $\boldsymbol{b}_0$ is the true parameter space, and $\mathbf{h}$ shows magnitude of the break size in coefficients. So, for any fixed $\mathbf{h}$, the break size $\frac{\mathbf{h}}{\sqrt{T}}$ converges to zero as the sample size increases. We allow the size of the break be different across individuals. That means, for each individual $i = 1, \ldots, N$, we have $\beta_{i(1)} = \beta_{i(2)} + \frac{\delta_i}{\sqrt{T}}$

or generally

$$R\boldsymbol{b} = \begin{bmatrix} \beta_{1(1)} - \beta_{1(2)} \\ \beta_{2(1)} - \beta_{2(2)} \\ \vdots \\ \beta_{N(1)} - \beta_{N(2)} \end{bmatrix} = \frac{1}{\sqrt{T}} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{bmatrix} = \frac{\boldsymbol{\delta}}{\sqrt{T}}, \tag{4.10}$$

where $\boldsymbol{\delta} = (\delta_1', \ldots, \delta_N')'$ is a vector of $NK \times 1$, and the convention matrix R is defined as

$$R = \begin{bmatrix} I_k & -I_k & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_k & -I_k & \mathbf{0} & \ldots & \mathbf{0} \\ & & & & & \vdots & \\ \mathbf{0} & & \ldots & & \mathbf{0} & I_k & -I_k \end{bmatrix}, \tag{4.11}$$

with rank equal to p. Note that $R\boldsymbol{b}_0 = 0$ under the null hypothesis of no break, and therefore $R\mathbf{h} = \boldsymbol{\delta}$. By having the distribution of the unrestricted estimator and the restricted estimator, we can derive the asymptotic distribution for the Stein-like shrinkage estimator, and later on derive its asymptotic risk. Theorem 21 summarizes the results.

Theorem 21 *The asymptotic distribution of the unrestricted estimator is*

$$\sqrt{T}\big(\widehat{\boldsymbol{b}}_{ur} - \boldsymbol{b}\big) \xrightarrow{d} Z \sim N\big(0, V_{ur}\big), \tag{4.12}$$

and the asymptotic distribution of the restricted estimator is

$$\sqrt{T}\big(\widehat{\boldsymbol{b}}_{r} - \boldsymbol{b}\big) \xrightarrow{d} Z - V_{ur}R'\big(RV_{ur}R'\big)^{-1}R(Z + \mathbf{h}). \tag{4.13}$$

Besides, under the local alternative assumption and by having an smooth (locally quadratic) loss function in D_T, we have:

$$D_T = T(\widehat{\boldsymbol{b}}_{ur} - \widehat{\boldsymbol{b}}_r)'\,\mathbb{W}(\widehat{\boldsymbol{b}}_{ur} - \widehat{\boldsymbol{b}}_r) \xrightarrow{d} (Z + \mathbf{h})'B(Z + \mathbf{h}), \tag{4.14}$$

$$\alpha \equiv \alpha(Z) \xrightarrow{d} \left(1 - \frac{\tau}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\right)_+, \tag{4.15}$$

$$\sqrt{T}\big(\widehat{\boldsymbol{b}}_w - \boldsymbol{b}\big) \xrightarrow{d} Z - \alpha(Z)V_{ur}\ LR(Z+\mathbf{h}), \tag{4.16}$$

where $B \equiv R'L'V_{ur}\,\mathbb{W}\,V_{ur}LR$ *is a matrix of* $2Nk \times 2Nk$. ■

See Appendix C for the proof and details of Theorem 21. As we derive the asymptotic distribution for the Stein-like shrinkage estimator, we can derive its asymptotic risk. Theorem 22 shows the asymptotic risk for the Stein-like shrinkage estimator.

Theorem 22 *For* $0 < \tau \leq 2\big(tr(A) - 2\lambda_{max}(A)\big)$ *and for any* $\mathbb{W} > 0$, *the asymptotic risk for the Stein-like shrinkage estimator is equal to*

$$\rho(\widehat{\boldsymbol{b}}_w, \mathbb{W}) < \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) - \tau\Big[\frac{2\big(tr(A) - 2\lambda_{max}(A)\big) - \tau}{(c+1)tr(A)}\Big], \tag{4.17}$$

where $A \equiv \mathbb{W}^{1/2}\,V_{ur}R'L'V_{ur}\,\mathbb{W}^{1/2}$ *and* $0 < c < \infty$. *Therefore, the asymptotic risk of the Stein-like shrinkage estimator is less than the unrestricted estimator.* ■

Appendix C has the complete proof of Theorem 22. As the shrinkage parameter, τ, is unknown, we can find it by minimizing the asymptotic risk. Theorem 23 shows the optimal value of the shrinkage parameter denoted by τ^*_{opt}, and the associated asymptotic risk for the Stein-like shrinkage estimator.

Theorem 23 *The optimal value of* τ *is*

$$\tau^*_{opt} = tr(A) - 2\lambda_{max}(A), \tag{4.18}$$

and the asymptotic risk for the Stein-type shrinkage estimator after plugging the τ^*_{opt} *is*

$$\rho(\widehat{\boldsymbol{b}}_w, \mathbb{W}) < \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) - \frac{\Big[tr(A) - 2\lambda_{max}(A)\Big]^2}{(c+1)tr(A)}, \tag{4.19}$$

which shows that the asymptotic risk of the Stein-like shrinkage estimator is strictly smaller than the unrestricted estimator. ■

Up to now, we discuss about the risk of the Stein estimator for any $\mathbb{W} > 0$. If $\mathbb{W} = \left(X'\Omega^{-1}X\right)^{-1}$ which is a prediction weight, all calculations become simpler. Corollary 24 summarizes the results for this specific choice of $\mathbb{W}$.

Corollary 24 *For $\mathbb{W} = \left(X'\Omega^{-1}X\right)^{-1}$, the optimal value of τ is*

$$\tau^*_{opt} = p - 2, \tag{4.20}$$

where τ^*_{opt} is positive as long as $p > 2$, which p is the number of restrictions. Also, the associated asymptotic risk is

$$\rho(\widehat{\boldsymbol{b}}_w, \mathbb{W}) < \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) - \frac{\left(p-2\right)^2}{h'R'\left(R\ \left(X'\Omega^{-1}X\right)^{-1}R'\right)^{-1}Rh + p}. \tag{4.21}$$

□

4.3.4 Forecasting

Generating accurate forecasts in the presence of structural breaks requires careful consideration of bias-variance tradeoffs. As our introduced shrinkage estimator considers this tradeoff in the estimation of parameters, it is a perfect choice for generating forecast. Our approach is to use the proposed shrinkage estimator for forecasting by focusing on the post-break observations across all individuals, since the true parameters that enter to the forecasting period is the coefficients in the post-break sample. Define a $Nk \times 2Nk$ selection matrix G such that

$$G\boldsymbol{b} = \begin{bmatrix} \mathbf{0} & I_k & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I_k & \dots & \mathbf{0} & \mathbf{0} \\ & & & \vdots & & & \\ \mathbf{0} & \mathbf{0} & & \dots & \mathbf{0} & \mathbf{0} & I_k \end{bmatrix} \begin{bmatrix} \beta_{1(1)} \\ \beta_{1(2)} \\ \beta_{2(1)} \\ \beta_{2(2)} \\ \vdots \\ \beta_{N(1)} \\ \beta_{N(2)} \end{bmatrix} = \begin{bmatrix} \beta_{1(2)} \\ \beta_{2(2)} \\ \vdots \\ \beta_{N(2)} \end{bmatrix}. \tag{4.22}$$

Multiplying G to the shrinkage estimator, we have

$$G\widehat{\boldsymbol{b}}_w = wG\widehat{\boldsymbol{b}}_{ur} + (1-w)G\widehat{\boldsymbol{b}}_r, \tag{4.23}$$

where $G\widehat{\boldsymbol{b}}_{ur}$ estimates the coefficients only by using the observations after the break point over all individuals, also known as post-break estimator, and $G\widehat{\boldsymbol{b}}_r$ is the restricted estimator under the assumption of no breaks in the model. Define $MSFE(G\widehat{\boldsymbol{b}}_w) = \rho(G\widehat{\boldsymbol{b}}_w, x^*_{T+1}x^{*\prime}_{T+1})$ as the MSFE of the shrinkage estimator, where $x^*_{T+1} = \left(x'_{1,T+1}, \dots, x'_{N,T+1}\right)'$ is an $Nk \times 1$ column vector of regressors at time $T+1$, and $x^*_{T+1}x^{*\prime}_{T+1}$ shows the one step ahead out of sample forecast. Theorem 25 shows the MSFE for the shrinkage estimator.

Theorem 25 *The mean square forecast error of the Stein-like shrinkage estimator is*

$$MSFE(\widehat{G\boldsymbol{b}}_w) < MSFE(G\widehat{\boldsymbol{b}}_{ur}) - \frac{\left[tr(\phi) - 2\lambda_{max}(\phi)\right]^2}{(c+1)\,tr(\phi)}, \tag{4.24}$$

where $\phi \equiv \mathbb{W}^{1/2}\, GV_{ur}R'L'V_{ur}G'\,\mathbb{W}^{1/2}$*, and* $\mathbb{W} = x^*_{T+1}x^{*\prime}_{T+1}$. ■

4.4 Multiple breaks

For simplicity, so far we have focused on the case with a single structural break, but in practice a model may be subject to multiple breaks. The presence of multiple breaks complicates the relationship between the bias and variance efficiency, since more breakpoint scenarios become possible. But clearly, the bias-variance tradeoff does not rely on the absence of the multiple breaks. Therefore, the aforementioned procedure can readily be generalized to account for the possibility of multiple breaks. When multiple common break points occur at $\{T_1, \ldots, T_m\}$, there are $m+1$ regimes for each individuals as

$$y_{i,t} = \begin{cases} x'_{i,t}\beta_{i(1)} + u_{i,t} & \text{for} \quad 1 \ < t \leq T_1, \\ x'_{i,t}\beta_{i(2)} + u_{i,t} & \text{for} \quad T_1 < t \leq T_2, \\ \vdots & \\ x'_{i,t}\beta_{i(m+1)} + u_{i,t} & \text{for} \quad T_m < t < T, \end{cases} \tag{4.25}$$

for $i = 1, \ldots, N$. In this case, the parameter of interest, $\boldsymbol{b}$ is a vector of $(m+1)Nk \times 1$. Like the single break case, the unrestricted estimator uses the observations within each regimes separately and estimate the coefficients across individuals. As the restricted estimator is under the assumption of no breaks, $R\boldsymbol{b}_0 = 0$, when there are breaks in the model, the convention restriction matrix R should be in a way that

$$R\boldsymbol{b} = \begin{bmatrix} \beta_{1(1)} - \beta_{1(2)} \\ \vdots \\ \beta_{1(m)} - \beta_{1(m+1)} \\ \vdots \\ \beta_{N(1)} - \beta_{N(2)} \\ \vdots \\ \beta_{N(m)} - \beta_{N(m+1)} \end{bmatrix} = \frac{1}{\sqrt{T}} \begin{bmatrix} \delta_{11} \\ \vdots \\ \delta_{1m} \\ \vdots \\ \delta_{N1} \\ \vdots \\ \delta_{Nm} \end{bmatrix} = \frac{\boldsymbol{\delta}}{\sqrt{T}}. \tag{4.26}$$

Once the model is set up, the parameters, $\boldsymbol{b}$, can be estimated by generalized least squares as in (4.12) and (4.13). Besides, the asymptotic distributions and asymptotic risk can be obtained similarly as in Theorems 21 and 22.

4.5 Monte Carlo simulation

This section employs Monte Carlo simulations to examine the performance of the theoretical results developed in the chapter. In this study, we compare the risk of the Stein-like shrinkage estimator with the unrestricted estimator. To do this, we consider the following data generating process

$$y_{i,t} = \begin{cases} x'_{i,t}\beta_{i(1)} + u_{i,t} & \text{for } i = 1, \dots, N, \quad t = 1, \dots, T_1, \\ x'_{i,t}\beta_{i(2)} + u_{i,t} & \text{for } i = 1, \dots, N, \quad t = T_1 + 1, \dots, T, \end{cases} \tag{4.27}$$

where $x_{i,t} \sim \text{N}(0, 1)$, and we set the first column of that to be a vector of ones in order to allow for the fixed effect. let time series dimension be $T = 100$, the number of series $i = 1, \dots, N$ be $N \in \{5, 10\}$, and $k \in \{1, 3\}$. We consider different values for breakpoints which are proportional to sample observations, $b_1 \equiv \frac{T_1}{T} \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$, and set $\mathbb{W} = \left(X'\Omega^{-1}X\right)^{-1}$ which gives the results for the in sample prediction.

Let $\beta_{i(2)}$ be a vector of ones, and $\delta_i = \beta_{i(1)} - \beta_{i(2)} = \frac{i}{N-1} * s$, where s varies from 0 to 1 in increments of 0.1. Beside, we set $u_{1,t(1)} \sim$ i.i.d $\text{N}\big(0, \sigma^2_{(1)}\big)$ and $u_{1,t(2)} \sim$ i.i.d $\text{N}\big(0, \sigma^2_{(2)}\big)$, and define $q \equiv \sigma^2_{(1)}/\sigma^2_{(2)}$ where $q \in \{0.5, 1, 2\}$. To allow for the cross-sectional dependence, consider

$$\begin{cases} u_{i,t(1)} = 1.5u_{1,t(1)} + v_{i,t(1)} & \text{for} \quad i = 2, \dots, N, \quad t = 1, \dots, T_1 \\ u_{i,t(2)} = 1.5u_{1,t(2)} + v_{i,t(2)} & \text{for} \quad i = 2, \dots, N, \quad t = T_1 + 1, \dots, T, \end{cases} \tag{4.28}$$

where $v_{i,t(1)} \sim$ i.i.d $\text{N}(0, \frac{i}{N}\ \sigma^2_{(1)})$, and $v_{i,t(2)} \sim$ i.i.d $\text{N}(0, \frac{i}{N}\ \sigma^2_{(2)})$.

We report the ratio of the mean square error as $RMSE_w \equiv \frac{MSE(\widehat{\boldsymbol{b}}_w)}{MSE(\widehat{\boldsymbol{b}}_{ur})}$. Therefore, the value of $RMSE_w$ less than one shows the out-performance of the Stein-like shrinkage estimator relative to the unrestricted estimator. Tables 4.1 to 4.6 report the results. Number of Monte Carlo is 1,000.

Simulation results

Tables 4.1 to 4.6 represent the results of the relative MSE of the Stein-like shrinkage estimator relative to the unrestricted estimator, $RMSE_w \equiv \frac{\rho(\widehat{\boldsymbol{b}}_w,\mathbb{W})}{\rho(\widehat{\boldsymbol{b}}_{ur},\mathbb{W})}$, when $T = 100$. The benchmark model is the unrestricted estimator. The first column in the tables show the number of individuals while the second column show different break size in the coefficients, δ. Based on the results of Tables 4.1 to 4.6, the Stein-like shrinkage estimator has a better performance compare to the unrestricted estimator, in the sense of having the smaller mean square error. Tables 4.1-4.3 reporting the results for the case that we have only one regressors, $k = 1$, and therefore these are reporting the fixed effect. For the small break size in coefficients, the Stein-like shrinkage estimator performs much better than the unrestricted estimator compare to the larger break sizes. As the break size in the coefficients increases, the gain from using the Stein-like shrinkage estimator will drop and the relative mean square error will be close to one. The reason for that can be related to the large bias that the restricted estimator adds to the model under the large break size. Therefore the efficiency that we can get from the restricted estimator cannot offset the effect of the large bias. Note that even under this condition, still we do not see under-performance of the Stein-like shrinkage estimator over the unrestricted estimator which confirms that there is no cost in using the proposed estimator. Tables 4.4-4.6 report the results for $k = 3$. Based on the results, as we increase the number of regressors, the gain obtained from the shrinkage estimator will increase. Generally, for a fixed number of regressors, as the number of individual units, N, increases, the Stein-like shrinkage estimator performs better than the unrestricted estimator, for any break points, break size in the coefficients, and any q.

4.6 Empirical analysis

This section provides some empirical analysis for forecasting U.S. industry-level inflation rate. We work with monthly data set from January 2004 to March 2020 of seven industries, with the industry-level industrial production as the predictor. The seven industries are chemicals, computer and electronic products, electrical equipment, food, furniture and related products, industrial machinery, and manufacturing. All data are sourced from Federal Reserve Economic Data (FRED). As discussed in Stock and Watson (2003), there is evidence of instabilities in predictive performance of the growth of real output as a predictor to forecast inflation. Using the following panel model, we evaluate the out-of-sample forecasting performance. The panel model with common breaks is specified as

$$y_{i,t} = x'_{i,t-1}\beta_i + u_{i,t}, \tag{4.29}$$

where $y_{i,t}$ is the industry level inflation rate constructed as $y_{i,t} = 1200 \times \log\Big(PPI_{i,t}/PPI_{i,t-1}\Big)$ in which PPI is the Producer Price Index, and $x_{i,t}$ is monthly industrial production in real terms for each industry constructed as $x_{i,t} = 1200 \times \log\Big(IP_{i,t}/IP_{i,t-1}\Big)$, in which IP is the index of industrial production.

Table 4.7 presents summary statistics for the monthly inflation rates for each industry. The average monthly inflation rates range from -2.389 for furniture to 6.017 for computer. Also, three industries, chemicals, electrical and furniture, experiencing deflation over the sample. The high standard deviation along with the large range for minimum and maximum values reflect the large fluctuation in inflation, which can be attributed to furniture.

We evaluate the out-of-sample forecasting performance of the proposed shrinkage estimator with three alternative models in terms of their MSFEs. The first model is a panel model which estimates the post-break parameters across the entire cross-section (unrestricted estimator). The second model is a time series linear regression estimating post-break parameters independently to each series in the cross-section using ordinary least squares. The third model is the a panel with cross-sectional dependence but no breaks (restricted estimator). We estimate the model and the break point using initial estimation period of 5 years, from January 2004 up to the end of December 2008. Using the estimated parameters, we generate out-of-sample forecasts of inflation rates using a recursive (expanding) estimation window. Table 4.8 presents the results. In the heading of the table, $MSFE_w$ is the MSFE based on the Stein-like shrinkage estimator, $MSFE_{ur}$ is the MSFE of the unrestricted estimator (first alternative model), $MSFE_{ur_{ind}}$ is MSFE based on the time series model (second alternative model), and $MSFE_r$ is MSFE of the restricted estimator (third alternative model). * indicate significance at 1% based on Diebold and Mariano (1995) and West (1996) statistic.

Based on the results, the MSFE of the Stein-like shrinkage estimator is smaller than the unrestricted estimator, and this result is consistent with our theoretical findings. Furthermore, the MSFE of the Stein-like shrinkage estimator is lower than the other alternative estimators.

4.7 Conclusion

In this chapter, we introduce the Stein-like shrinkage estimator for estimating the coefficients of the heterogeneous panel data regression models with cross-sectional dependence under structural breaks. This shrinkage estimator is a combination of the unrestricted estimator and the restricted estimator. The restricted estimator is under the null hypothesis of no breaks in the coefficient. So, this is the efficient estimator, but it is bias when there is a break in the model. The unrestricted estimator estimate the unknown coefficient across all individuals by considering the break points. Therefore, the unrestricted estimator is the unbiased estimator but less efficient. Basically, the proposed Stein-like shrinkage estimator considers the tradeoff between bias and variance efficiency. The combination weight is proportion to the loss function, in which depending on the break size, it assigns weight to each of the estimators. We establish the asymptotic distribution and asymptotic risk for the proposed estimator and show that the Stein-like shrinkage estimator out-performs the unrestricted estimator, in the sense of having an smaller mean square error, for any break points and break size. Monte Carlo simulations and also empirical example of forecasting U.S. industry level inflation rates confirm the superiority of using the proposed Stein-like shrinkage estimator over the unrestricted estimator.

Table 4.1: Simulation results with $k = 1$, $q = 0.5$

	δ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
$N = 5$	0.00	0.679	0.634	0.580	0.522	0.477	0.459	0.444
	0.11	0.694	0.672	0.620	0.575	0.524	0.503	0.484
	0.22	0.775	0.754	0.722	0.679	0.634	0.583	0.539
	0.33	0.856	0.852	0.842	0.798	0.750	0.692	0.627
	0.44	0.928	0.937	0.923	0.891	0.866	0.801	0.717
	0.55	0.969	0.972	0.943	0.949	0.914	0.883	0.829
	0.66	0.989	0.985	0.986	0.973	0.986	0.945	0.867
	0.77	0.991	0.990	0.999	0.974	0.990	0.959	0.931
	0.88	0.995	0.993	0.996	0.998	0.991	0.988	0.954
	1.00	0.998	0.999	0.997	0.999	0.988	0.986	0.964
$N = 10$	0.00	0.635	0.580	0.524	0.458	0.414	0.406	0.441
	0.11	0.654	0.603	0.543	0.481	0.440	0.416	0.450
	0.22	0.688	0.639	0.594	0.535	0.486	0.451	0.472
	0.33	0.755	0.708	0.664	0.611	0.559	0.517	0.518
	0.44	0.793	0.764	0.740	0.689	0.639	0.587	0.577
	0.55	0.851	0.834	0.803	0.764	0.716	0.650	0.613
	0.66	0.896	0.890	0.850	0.833	0.763	0.709	0.667
	0.77	0.913	0.916	0.895	0.870	0.831	0.767	0.727
	0.88	0.933	0.928	0.917	0.890	0.868	0.816	0.753
	1.00	0.951	0.933	0.928	0.920	0.892	0.851	0.796

Table 4.2: Simulation results with $k = 1$, $q = 1$

	δ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
$N = 5$	0.00	0.578	0.620	0.644	0.661	0.640	0.631	0.574
	0.11	0.575	0.643	0.669	0.686	0.667	0.643	0.585
	0.22	0.645	0.686	0.711	0.735	0.711	0.700	0.633
	0.33	0.694	0.753	0.794	0.805	0.784	0.753	0.696
	0.44	0.756	0.818	0.854	0.867	0.859	0.832	0.769
	0.55	0.831	0.909	0.914	0.921	0.909	0.871	0.830
	0.66	0.880	0.932	0.952	0.961	0.945	0.927	0.873
	0.77	0.928	0.956	0.997	0.989	0.968	0.972	0.924
	0.88	0.961	0.985	0.984	0.995	0.983	0.971	0.955
	1.00	0.986	0.993	0.989	0.995	0.988	0.989	0.979
$N = 10$	0.00	0.550	0.577	0.602	0.609	0.606	0.555	0.549
	0.11	0.558	0.580	0.617	0.618	0.618	0.574	0.560
	0.22	0.581	0.605	0.633	0.650	0.632	0.611	0.582
	0.33	0.598	0.643	0.669	0.681	0.669	0.636	0.603
	0.44	0.639	0.686	0.722	0.731	0.723	0.678	0.639
	0.55	0.681	0.729	0.766	0.777	0.783	0.727	0.689
	0.66	0.719	0.775	0.809	0.829	0.814	0.775	0.723
	0.77	0.759	0.806	0.848	0.865	0.843	0.811	0.761
	0.88	0.796	0.844	0.881	0.888	0.869	0.852	0.790
	1.00	0.829	0.882	0.905	0.913	0.901	0.887	0.833

Table 4.3: Simulation results with $k = 1$, $q = 2$

	δ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
$N = 5$	0.00	0.463	0.450	0.487	0.536	0.590	0.640	0.677
	0.11	0.475	0.495	0.528	0.581	0.624	0.669	0.708
	0.22	0.554	0.588	0.629	0.666	0.718	0.754	0.780
	0.33	0.638	0.698	0.740	0.801	0.827	0.867	0.851
	0.44	0.721	0.795	0.844	0.896	0.935	0.920	0.933
	0.55	0.819	0.870	0.934	0.955	0.963	0.974	0.969
	0.66	0.878	0.928	0.963	0.969	0.980	0.985	0.989
	0.77	0.931	0.964	0.994	0.995	0.999	1.003	0.998
	0.88	0.945	0.970	0.992	0.993	0.994	0.993	0.998
	1.00	0.968	0.997	0.999	0.999	0.998	0.999	0.999
$N = 10$	0.00	0.456	0.390	0.419	0.459	0.523	0.583	0.637
	0.11	0.455	0.415	0.431	0.483	0.538	0.595	0.650
	0.22	0.492	0.444	0.480	0.534	0.586	0.653	0.696
	0.33	0.522	0.505	0.554	0.616	0.663	0.716	0.737
	0.44	0.562	0.575	0.633	0.692	0.741	0.795	0.796
	0.55	0.625	0.652	0.707	0.762	0.808	0.840	0.855
	0.66	0.669	0.714	0.765	0.825	0.859	0.862	0.884
	0.77	0.711	0.776	0.838	0.859	0.906	0.917	0.919
	0.88	0.760	0.812	0.853	0.902	0.908	0.934	0.931
	1.00	0.805	0.856	0.878	0.927	0.938	0.956	0.948

Table 4.4: Simulation results with $k = 3$, $q = 0.5$

	δ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
$N = 5$	0.00	0.574	0.534	0.470	0.406	0.350	0.301	0.297
	0.11	0.602	0.579	0.518	0.445	0.388	0.347	0.332
	0.22	0.682	0.690	0.644	0.579	0.509	0.440	0.392
	0.33	0.775	0.779	0.746	0.695	0.641	0.570	0.497
	0.44	0.842	0.841	0.825	0.789	0.734	0.666	0.585
	0.55	0.879	0.891	0.878	0.851	0.803	0.752	0.665
	0.66	0.909	0.930	0.904	0.891	0.846	0.801	0.716
	0.77	0.935	0.936	0.926	0.913	0.877	0.852	0.775
	0.88	0.944	0.960	0.949	0.927	0.911	0.875	0.815
	1.00	0.954	0.959	0.954	0.945	0.929	0.905	0.848
$N = 10$	0.00	0.597	0.544	0.479	0.409	0.361	0.354	0.441
	0.11	0.611	0.562	0.490	0.428	0.381	0.367	0.450
	0.22	0.642	0.600	0.544	0.475	0.428	0.409	0.463
	0.33	0.680	0.659	0.610	0.549	0.488	0.456	0.496
	0.44	0.732	0.721	0.672	0.622	0.562	0.512	0.535
	0.55	0.778	0.767	0.729	0.682	0.627	0.575	0.578
	0.66	0.811	0.814	0.778	0.738	0.683	0.636	0.622
	0.77	0.845	0.849	0.823	0.782	0.741	0.685	0.665
	0.88	0.869	0.879	0.850	0.824	0.784	0.729	0.700
	1.00	0.890	0.891	0.882	0.846	0.812	0.767	0.736

Table 4.5: Simulation results with $k = 3$, $q = 1$

	δ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
$N = 5$	0.00	0.437	0.489	0.531	0.553	0.536	0.489	0.430
	0.11	0.448	0.509	0.560	0.574	0.562	0.515	0.449
	0.22	0.496	0.570	0.629	0.639	0.626	0.571	0.495
	0.33	0.560	0.655	0.707	0.727	0.709	0.654	0.564
	0.44	0.639	0.726	0.772	0.796	0.772	0.715	0.635
	0.55	0.700	0.792	0.833	0.849	0.833	0.787	0.695
	0.66	0.751	0.830	0.870	0.877	0.860	0.830	0.746
	0.77	0.808	0.867	0.889	0.907	0.900	0.866	0.796
	0.88	0.834	0.895	0.913	0.932	0.925	0.893	0.830
	1.00	0.856	0.910	0.940	0.941	0.937	0.905	0.855
$N = 10$	0.00	0.523	0.514	0.549	0.554	0.544	0.516	0.519
	0.11	0.523	0.519	0.556	0.562	0.551	0.528	0.524
	0.22	0.540	0.547	0.578	0.601	0.576	0.541	0.536
	0.33	0.565	0.572	0.615	0.625	0.618	0.573	0.558
	0.44	0.596	0.613	0.657	0.676	0.654	0.617	0.590
	0.55	0.623	0.649	0.701	0.716	0.700	0.650	0.627
	0.66	0.657	0.697	0.736	0.757	0.744	0.697	0.656
	0.77	0.690	0.737	0.772	0.787	0.774	0.733	0.689
	0.88	0.724	0.769	0.804	0.823	0.808	0.764	0.719
	1.00	0.752	0.799	0.834	0.844	0.830	0.803	0.755

Table 4.6: Simulation results with $k = 3$, $q = 2$

	δ	$b_1 = 0.2$	$b_1 = 0.3$	$b_1 = 0.4$	$b_1 = 0.5$	$b_1 = 0.6$	$b_1 = 0.7$	$b_1 = 0.8$
	0.00	0.305	0.304	0.350	0.408	0.470	0.534	0.562
	0.11	0.321	0.343	0.396	0.459	0.524	0.583	0.601
	0.22	0.394	0.448	0.508	0.581	0.636	0.675	0.687
	0.33	0.483	0.555	0.644	0.701	0.744	0.780	0.771
$N = 5$	0.44	0.583	0.670	0.740	0.781	0.834	0.844	0.838
	0.55	0.658	0.745	0.807	0.853	0.871	0.895	0.878
	0.66	0.736	0.807	0.849	0.891	0.918	0.928	0.911
	0.77	0.787	0.844	0.882	0.911	0.932	0.942	0.928
	0.88	0.826	0.873	0.908	0.935	0.944	0.952	0.950
	1.00	0.849	0.895	0.928	0.947	0.956	0.959	0.950
	0.00	0.446	0.350	0.359	0.413	0.473	0.540	0.600
	0.11	0.447	0.364	0.379	0.427	0.493	0.557	0.606
	0.22	0.466	0.396	0.425	0.479	0.542	0.599	0.641
	0.33	0.498	0.454	0.495	0.547	0.609	0.658	0.685
$N = 10$	0.44	0.537	0.516	0.557	0.621	0.676	0.718	0.734
	0.55	0.576	0.574	0.630	0.689	0.739	0.768	0.772
	0.66	0.626	0.632	0.687	0.733	0.782	0.805	0.816
	0.77	0.666	0.680	0.733	0.785	0.822	0.840	0.845
	0.88	0.698	0.727	0.779	0.819	0.857	0.874	0.870
	1.00	0.735	0.761	0.807	0.852	0.881	0.892	0.892

Table 4.7: Summary statistics of industry level inflation rates

Industry	Mean	St. dev.	Min.	Max.
chemicals	-0.298	15.043	−83.982	65.366
computer	6.017	11.876	−50.616	36.793
electrical	-0.313	14.558	−48.341	50.393
food	0.756	9.779	−29.040	28.460
furniture	-2.389	18.238	−126.258	47.888
machinery	0.314	20.434	−68.609	43.802
manufacturing	0.271	10.700	−77.719	19.162

Table 4.8: Empirical results for forecasting inflation rates

out-of-sample period	$MSFE_w$	$MSFE_{ur}$	$MSFE_{ur_{ind}}$	$MSFE_r$
2009:01-2020:03	0.2358*	0.2488	0.2831	0.2526

Chapter 5

Conclusions

This dissertation proposes different combined estimators to make an improvement in the estimation and forecasting under structural break models. When it comes to estimation of parameters under structural breaks, the common solution is to use the observations within each regime separately, but this solution by itself may not necessarily minimize the mean square error, specially if there are only a few observations within each or some regimes. To improve the performance of the estimator, in chapter two we propose the combined estimator of the unrestricted estimator and the restricted estimator. Restricted estimator estimates the coefficients under the assumption of no breaks in the model while the unrestricted estimator only uses the observations within each regime. We set the combination weight between zero and one and derive the optimal value of that based on the finite sample and asymptotic theories. For the finite sample part, we use the two well-known approaches of the large-sample expansion method proposed by Nagar (1959) and the small-disturbance method proposed by Kadane (1971). For the asymptotic part, we use the local asymptotic framework, and derive the asymptotic risk for the combined estimator under the weak break size in the coefficient. We show analytically that the finite sample risk and also the asymptotic risk of the combined estimator is less than the unrestricted estimator. Monte Carlo simulation and also empirical study support the theoretical findings.

In chapter three, we mainly focus on finding the optimal forecast under structural breaks. One standard solution for forecasting under structural breaks is to use the observations after the most recent break-points, and estimate the coefficient using those observations. The reason behind this solution is the fact that the true parameters that enter to the forecasting period is the coefficient in the post-break sample. However, if the are only a few number of observations in the post-break sample, the estimates of the coefficients will be volatile, and this maps into the high mean square forecast error. To solve this issue, we propose two different combined estimators that exploit the pre-break observations. The first one is the Stein-like combined estimator that includes the combination of two estimators. One is the post-break estimator, and the other one is the full-sample estimator. The full-sample estimator constructs under the assumption of no breaks in the coefficients, and therefore uses all available observations in the sample. This is the efficient estimator but bias if there is a break in the model. Therefore, combining these two estimators tradeoffs the bias and variance efficiency. The combination weight is proportion to the Hausman statistic that based on the strength of the break, assigns weight to each of the two estimators. We show that the asymptotic risk of the Stein-like combined estimator is less than the unrestricted estimator.

The second proposed estimator is the semi-parametric one that uses the smooth discrete kernel. The idea behind this estimator is to estimate the post-break sample using all available observations in the full-sample and smooth them over time. Therefore, this estimator assigns the full weight of one to post-break sample observations and down-weight the pre-break sample by a weight between zero and one. The unknown weight can be found by cross-validation. We prove that the estimated weight by cross-validation is optimal in the sense that it is asymptotically equivalent to the infeasible possible weight. Given the weight, we also derive the asymptotic distribution of the semi-parametric estimator and show its efficiency compare to the post-break estimator. Simulation study and an empirical example evidenced the theoretical findings. We also compare the performance of our proposed estimator with the recent developed estimator proposed by Pesaran et. al (2013), and show the superiority of the proposed estimators in terms of the mean square forecast error.

In chapter four, we extend the idea of improving estimator from the time-series model to the panel data regression model. We consider the common breaks in a heterogenous panel data model with cross-sectional dependence. We propose the shrinkage estimator of the unrestricted estimator with the restricted one. The unrestricted estimator estimates the coefficients using the observations within each regime over all individual units. The restricted estimator is developed under the null hypothesis of no breaks in the model. So, this estimator is bias if we have a break but efficient. The combination weight is proportion to the Wald statistic that measure the severity of the break, and based on that gives weight to each of the estimators. We examine the properties of the proposed shrinkage estimator, analytically in asymptotic theory, and numerically by simulation and empirical analysis, and show that the shrinkage estimator has a lower risk than the unrestricted estimator.

Looking forward, there are several open avenues out of my dissertation that I leave them for the future works. One is considering non-common breaks in the panel data models. Also, forecasting heterogenous panel model by considering the individual specific random effects is another interesting topic. These are an important and practical topics, since there are only a few papers that focus on forecasting panel models under structural breaks, and mainly papers focus on testing for breaks or detecting the break points. Considering large panel models for structural breaks is crucial because of their application in the real world. For example, structural breaks due to technological change, or new policy implementation are more likely to occur over a longer time horizon.

Another open area of my research is the extension to the dynamic model. My current research are under the assumption that regressors are exogenous, and therefore, they exclude dynamic models. This is a challenging topic because derivation of mean square forecast error is complicated by the fact that the forecast errors are non-linear functions of past errors. In view of the widespread use of dynamic models in forecasting, this is clearly an important area to investigate, because the autoregressive models are frequently used in economics and business and they have been proved to perform well in practice.

Bibliography

[1] Aitchison J, Aitken CGG. (1976). "Multivariate binary discrimination by the kernel method." Biometrika 63, 413–420.

[2] Altissimo, Filippo, Corradi, Valentina, (2003). "Strong rules for detecting the number of breaks in a time series." Journal of Econometrics 117, 207-244.

[3] Andrews, D.W.K. (1993). "Tests for parameter instability and structural change with unknown change point." Econometrica 61, 821-856.

[4] Andrews, D.W.K. and W. Ploberger (1994). "Optimal tests when a nuisance parameter is present only under the alternative." Econometrica 62, 1383-1414.

[5] Andrews, Donald W.K., Inpyo, Lee, Werner, Ploberger, (1996). "Optimal change point tests for normal linear regression." Journal of Econometrics 70, 9–38.

[6] Andrews, Donald W.K., (2003). "End-of-sample instability tests." Econometrica 71, 1661–1694.

[7] Bai, J. (1995). "Least absolute deviation estimation of a shift." Econometric Theory 11, 403-436.

[8] Bai, J. (1998). "Estimation of multiple-regime regressions with least absolute deviation." Journal of Statistical Planning and Inference 74, 103-134.

[9] Bai, J. (1997a). "Estimation of a change point in multiple regression models." Review of Economics and Statistics 79, 551-563.

[10] Bai, J. (1997b). "Estimating multiple breaks one at a time." Econometric Theory 13, 315-352.

[11] Bai, J., (1999). "Likelihood ratio tests for multiple structural changes." Journal of Econometrics 91, 299-323.

[12] Bai, J. (2010). "Common breaks in means and variances for panel data." Journal of Econometrics 157, 78-92.

[13] Bai, J., Lumsdaine, R.L., Stock, J.H., (1998). "Testing for and dating breaks in multivariate time series." Review of Economic Studies 65, 395–432.

[14] Bai, J. and Perron, P. (1998). "Estimating and testing linear models with multiple structural changes." Econometrica 66, 47-78.

[15] Bai, J. (2010). "Common breaks in means and variances for panel data." Journal of Econometrics 157, 78-92.

[16] Bai, J., and Ng, S. (2002), "Determining the number of factors in approximate factor models." Econometrica 70, 191–221.

[17] Bai, J. and Perron, P. (2003). "Computation and analysis of multiple structural change models." Journal of Applied Econometrics 18, 1-22.

[18] Baltagi, B. H., Feng, Q., and Kao, C. (2016). "Estimation of heterogeneous panels with structural breaks." Journal of Econometrics 191, 176–195.

[19] Baltagi, B. H., Kao, C., and Liu, L. (2017). "Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term." Econometric Reviews 36, 85–102.

[20] Bates, J.M. and C.W.J. Granger (1969). "The Combination of Forecasts." Operations Research Quarterly 20, 451-468.

[21] Bernstein D. S. (2005)."Matrix mathematics: Theory, facts, and formulas with application to linear systems theory." Princeton University Press.

[22] Boot, T. and Andreas Pick (2019). "Does modeling a structural break improve forecast accuracy?" Journal of Econometrics.

[23] Brown, R. L., J. Durbin and J. M. Evans (1975). "Techniques for Testing the Constancy of Regression Relationships over Time." Journal of the Royal Statistical Society Series B 37, 149-192.

[24] Campbell, J. Y, and S. B. Thompson (2008). "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?" Review of Financial Studies 21, 1509-31.

[25] Casini, A. and Perron, P. (2018). "Structural breaks in time series." arXiv preprint arXiv:1805.03807.

[26] Chan, F., Mancini-Griffoli, T., Pauwels, L.L., (2008). "Stability Tests for Heterogenous Panel Data." Working paper, Curtin University of Technology.

[27] Chow, Gregory C. (1960). "Tests of equality between sets of coefficients in two linear regressions." Econometrica 28, 591-605.

[28] Clements, M.P. and D.F. Hendry (1998). "Forecasting Economic Time Series." Cambridge University Press.

[29] Clements, M.P. and D.F. Hendry (1999). "Forecasting Non-stationary Economic Time Series." The MIT Press.

[30] Clements, M.P. and D.F. Hendry (2006). "Forecasting with Breaks." In Elliott, G., C.W.J. Granger and A. Timmermann (Eds.), Handbook of Economic Forecasting vol. 1. North-Holland, 605-658.

[31] Diebold, F.X. and P. Pauly (1987). "Structural Change and the Combination of Forecasts." Journal of Forecasting 6, 21-40.

[32] Diebold, F. X. and R. S. Mariano (1995). "Comparing predictive accuracy."Journal of Business and Economic Statistics 13, 253-263.

[33] Hansen, Bruce (2007). "Least squares model averaging." Econometrica 75, 1175–1189.

[34] Hansen, Bruce (2009). "Averaging estimators for regressions with a possible structural break." Econometric Theory 25, 1498-1514.

[35] Hansen, Bruce (2016). "Efficient shrinkage in parametric models." Journal of Econometrics 190, 115-132.

[36] Hansen, Bruce (2017). "Stein-like 2SLS estimator." Econometrics Reviews 36, 840-852.

[37] Hausman, J. A. (1978). "Specification tests in econometrics." Econometrica 46, 1251-1271.

[38] Horváth, L., Hušková, M.(2012). "Change–point detection in panel data." Journal of Time Series Analysis 33, 631–648.

[39] Hsu, C.-C., Lin, C.-C., (2012). "Change-Point Estimation for Nonstationary Panel Data." Working Paper, National Central University.

[40] Ichimura H. (1993). "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models." Journal of Econometrics 58, 71–120.

[41] James, W., Stein, Charles M. (1961). "Estimation with quadratic loss." In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 361–380.

[42] Joseph, L., Wolfson, D. B. (1992). "Estimation in multi–path change–point problems." Communications in Statistics – Theory and Methods 21, 897-913.

[43] Joseph, L., Wolfson, D. B. (1993). "Maximum likelihood estimation in the multi-path changepoint problem." Annals of the Institute of Statistical Mathematics 45, 511-530.

[44] Judge, G.G. and M.E. Bock (1978). "The statistical implications of pretest and Stein-rule estimators in econometrics." North-Holland, Amsterdam.

[45] Kadane, J. (1971). "Comparison of k-class estimators when the variance is small." Econometrica 39, 723-737.

[46] Kao, C., Trapani, L., Urga, G. (2012). "Asymptotics for panel models with common shocks." Econometric Reviews 31, 390–439.

[47] Kao, C., Lorenzo Trapani, L., Urga, G. (2018). "Testing for instability in covariance structures." Bernoulli 24, 740–771.

[48] Nicholas M. Kiefer and Jeffrey S. Racine (2017). "The smooth colonel and the reverend find common ground." Econometric Reviews 36, 241-256.

[49] Kim, D. (2011). "Estimating a common deterministic time trend break in large panels with cross sectional dependence." Journal of Econometrics 164, 310–330.

[50] Kim, D., (2014). "Common breaks in time trends for large panel data with a factor structure." The Econometrics Journal 17, 301–337.

[51] Lebedev, N. N. (1972). "Special Functions and Their Applications." Dover Publications, Inc.; New York.

[52] Lehmann, E. L., Casella, G. (1998). "Theory of Point Estimation." 2nd ed. New York: Springer.

[53] Li, K.-C. (1987). "Asymptotic optimality for C_P , C_L , cross-validation and generalized cross-validation: discrete index set." The Annals of Statistics 15, 958-975.

[54] Liu, L., (2018). "Density Forecasts in Panel Data Models: A Semiparametric Bayesian Perspective." Preprint. Available at arXiv:1805.04178.

[55] Liu, L., H. R. Moon and F. Schorfheide (2020). "Forecasting with dynamic panel data models." Econometrica 88, 171–201.

[56] Li, Q., Ouyang, D., Racine, J. (2013). "Categorical semiparametric varying coefficient models." Journal of Applied Econometrics 28,551–579.

[57] Maasoumi, E. (1978). "A modified Stein-like estimator for the reduced form coefficients of simultaneous equations." Econometrica 46, 695–703.

[58] McCracken, M. W. and Ng, S. (2016). "FRED-MD: A monthly database for macroeconomic research." Journal of Business & Economic Statistics 34, 574-589.

[59] Nagar, A.L. (1959). "The bias and moment matrix of the general k-Class estimators of the parameters in simultaneous equations." Econometrica 27, 575-595.

[60] Perron, P. (2006). "Dealing with structural breaks." In T.C. Mills & K. Patterson (eds.), Palgrave Hand-book of Econometrics, Econometric Theory 1, pp. 278-352. Palgrave Macmillan.

[61] Perron, P. and Zhongjun Qu (2006). "Estimating restricted structural change models." Journal of Econometrics 134, 373-399.

[62] Pesaran, M.H., and A. Timmermann (1995). "Predictability of Stock Returns: Robustness and Economic Significance." Journal of Finance 50, 1201-28.

[63] Pesaran, M.H., and A. Timmermann (2002). “Market timing and return prediction under model instability.” Journal of Empirical Finance 9, 495–510.

[64] Pesaran, M.H. and A. Timmermann, (2005). “Small sample properties of forecasts from auto-regressive models under structural breaks.” Journal of Econometrics 129, 183-217.

[65] Pesaran, M.H., A. Timmermann, (2007).“Selection of estimation window in the presence of breaks.” Journal of Econometrics 137, 134-161.

[66] Pesaran, M.H., and Andreas Pick (2011). “Forecast Combination Across Estimation Windows.” Journal of Business and Economic Statistics 29, 307-318.

[67] Pesaran, M.H., Yamagata, T., (2008). “Testing slope homogeneity in large panels.” Journal of Econometrics 142, 50–93.

[68] Pesaran, M.H., A. Pick and M. Pranovich (2013). “Optimal forecasts in the presence of structural breaks.” Journal of Econometrics 177, 134-152.

[69] Qian, J., Su, L., (2016). “Shrinkage estimation of regression models with multiple structural changes.” Econometric Theory 32, 1376-1433.

[70] Qu, Z. and P. Perron (2007). “Estimating and testing structural changes in multiple regressions.” Econometrica 75, 459–502.

[71] Quandt, R. E. (1960). “Tests of the hypothesis that a linear regression obeys two separate regimes.” Journal of the American Statistical Association 55, 324-330.

[72] Robinson PM. (1988). “Root-n consistent semiparametric regression.” Econometrica 56, 931–954.

[73] Rao, C. R. (1973). “Linear Statistical Inference and Its Applications.” New York: John Wiley and Sons, Inc.

[74] Rapach, D. E., and Zhou, G. (2013). “Forecasting stock returns.” in Graham Elliott and Allan Timmermann, eds: Handbook of Economic Forecasting, Volume 2 (Elsevier, Amsterdam).

[75] Rossi, B. (2013). “Advances in forecasting under instability.” In G. Elliott, & A. Timmermann (Eds.), Handbook of Economic Forecasting, Chapter 21, 1203 - 1324. Elsevier volume 2, Part B.

[76] Saleh, A.K.Md.Ehsanses, 2006. Theory of Preliminary Test and Stein-Type Estimation with Applications. Wiley, Hoboken.

[77] Stein, Charles M., (1956). “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution.” In: Proc. Third Berkeley Symp. Math. Statist. Probab., vol. 1 pp. 197–206.

[78] Simon Smith (2018) “Noncommon breaks.” Unpublished Manuscript, University of Southern California.

[79] Smith, Simon and Timmermann, Allan (2018). "Detecting breaks in real time: a panel forecasting approach." Available at SSRN: https://ssrn.com/abstract=3238216.

[80] Smith, Simon (2018). "Forecasting panel data with structural breaks and regime specific grouped heterogeneity." Available at SSRN: https://ssrn.com/abstract=3268759.

[81] Stock, J. H., and Watson, M. W. (2004). "Combination Forecasts of Output Growth in a Seven-Country Data Set." Journal of Forecasting 23, 405–430.

[82] Su, L., Chen, Q., (2013). "Testing homogeneity in panel data models with interactive fixed effects." Econometric Theory 29, 1079–1135.

[83] Su, L. and H. White (2010) Testing structural change in partially linear models. Econometric Theory 26, 1761–1806

[84] Timmerman, Allan (2006). "Forecast Combinations." In G. Elliott, C. W. Granger, and A. Timmermann (Eds.), Handbook of economic forecasting (pp. 135-196). Elsevier.

[85] Ullah, A., (2004). "Finite Sample Econometrics". Oxford: Oxford University Press.

[86] Ullah, A., Wan, A. T. K., Wang, H., Zhang, X., Zou, G. (2017). "A semiparametric generalized ridge estimator and link with model averaging." Econometric Reviews 36, 370-384.

[87] Welch, I., and A. Goyal (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." Review of Financial Studies 21, 1455-508.

[88] West, K.D., (1996). "Asymptotic inference about predictive ability." Econometrica 64, 1067-1084.

[89] Zhang, X., Wan, A. T. K., Zou, G. (2013). "Model averaging by Jackknife criterion in models with dependent data." Journal of Econometrics 174(2), 82–94.

Appendix A

Appendix for Chapter 2

Proof of equation (2.10): To find the MSE and risk of the $\widehat{\beta}_\gamma$, we first find the optimal value for γ by minimizing the risk.

$$\begin{aligned}
Risk(\widehat{\beta}_\gamma, W) &= \mathbb{E}\left[(\widehat{\beta}_\gamma - \beta)' W (\widehat{\beta}_\gamma - \beta)\right] \\
&= \mathbb{E}\left[\big(\widehat{\beta}_{ur} - \beta\big) - \gamma\big(\widehat{\beta}_{ur} - \widetilde{\beta}_r\big)\right]' W \left[\big(\widehat{\beta}_{ur} - \beta\big) - \gamma\big(\widehat{\beta}_{ur} - \widetilde{\beta}_r\big)\right] \\
&= Risk(\widehat{\beta}_{ur}, W) + \gamma^2\, \mathbb{E}\left[\big(\widehat{\beta}_{ur} - \widetilde{\beta}_r\big)' W \big(\widehat{\beta}_{ur} - \widetilde{\beta}_r\big)\right] \\
&- 2\gamma\, \mathbb{E}\left[\big(\widehat{\beta}_{ur} - \widetilde{\beta}_r\big)' W \big(\widehat{\beta}_{ur} - \beta\big)\right] \\
&= Risk(\widehat{\beta}_{ur}, W) + \gamma^2\, \mathbb{E}\left[\widehat{\beta}_{ur}' L' W L \widehat{\beta}_{ur}\right] - 2\gamma\, \mathbb{E}\left[\widehat{\beta}_{ur}' L' W (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}\epsilon\right] \\
&= Risk(\widehat{\beta}_{ur}, W) + \gamma^2 \left[\beta' L' W L \beta + \operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W L\big)\right] \\
&- 2\gamma\ \operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big).
\end{aligned} \tag{A.1}$$

By minimizing the risk, we have

$$\gamma^* = \frac{\operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}{\beta' L' W L \beta + \operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}, \tag{A.2}$$

Note that, given a known Ω, the unbiased estimator for the denominator of the weight in (A.2) can be calculated as

$$\begin{aligned}\mathbb{E}(\widehat{\beta}'_{ur}L'WL\widehat{\beta}_{ur}) &= \mathbb{E}\left[\big(\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\big)'L'WL\big(\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\big)\right] \\ &= \beta'L'WL\beta + \mathrm{tr}\big((X'\Omega^{-1}X)^{-1}L'WL\big),\end{aligned} \tag{A.3}$$

So, the unbiased estimator for $\beta'L'WL\beta$ is

$$\widehat{\beta}'_{ur}L'WL\widehat{\beta}_{ur} - \mathrm{tr}\big((X'\Omega^{-1}X)^{-1}L'WL\big). \tag{A.4}$$

Therefore, by plugging the unbiased estimator of the denominator, (A.2) will be

$$\gamma^* = \frac{\mathrm{tr}\big((X'\Omega^{-1}X)^{-1}L'W\big)}{\widehat{\beta}'_{ur}L'WL\widehat{\beta}_{ur}}. \tag{A.5}$$

Proof of the estimates for the variance-covariance matrix:

Let us write $\Delta = \widehat{\Omega} - \Omega$ where Δ is a $T \times T$ matrix with elements of orders $O_p(T^{-1/2})$. Remember,

$$\widehat{\Omega} = \begin{bmatrix} S_1 \ \iota_{l_1} & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & S_{m+1} \ \iota_{l_{m+1}} \end{bmatrix}, \tag{A.6}$$

where $S_i = \frac{\epsilon_i' M_i \epsilon_i}{l_i - k}$, $l_i \equiv T_i - T_{i-1}$ with $i = \{1, \dots, m+1\}$. Thus,

$$\begin{aligned}\mathbb{E}(S_i) &= \frac{1}{l_i - k}\,\mathbb{E}(\epsilon_i' M_i \epsilon_i) \\ &= \frac{1}{l_i - k}\mathrm{tr}\big(\,\mathbb{E}(\epsilon_i \epsilon_i' M_i)\big) \\ &= \sigma^2_{(i)},\end{aligned} \tag{A.7}$$

and

$$
\begin{aligned}
var(S_i) &= \mathbb{E}(S_i - \sigma^2_{(i)})^2 \\
&= \mathbb{E}(S_i^2) + \sigma^4_{(i)} - 2\sigma^4_{(i)} \\
&= \frac{\sigma^4_{(i)}}{(l_i - k)^2}\Big(\big(\mathrm{tr}(M_i)\big)^2 + 2\mathrm{tr}(M_i)\Big) - \sigma^4_{(i)} \\
&= \frac{2\sigma^4_{(i)}}{l_i - k} = O(\frac{1}{l_i}).
\end{aligned} \tag{A.8}
$$

Define, $b_i = \frac{T_i}{T}$, such that $0 < b_1 < b_2 < \cdots < b_m < 1$ are constant. Thus, $S_i - \sigma^2_{(i)} = O_p(T^{-1/2})$, and this completes the proof. ■

Proof of Theorem 1: We derive the optimal value of the weight in (2.10) which depend on the unknown parameter, Ω. So, we need to plug in the estimate value for the Ω and find the feasible terms for its numerator and denominator. Note that knowing the order of Δ, by expanding $\widehat{\Omega}$ we have

$$
\begin{aligned}
\widehat{\Omega} &= (\Omega + \Delta)^{-1} \\
&= \Big((I_T + \Delta\Omega^{-1})\Omega\Big)^{-1} \\
&= \Omega^{-1}\big(I_T + \Delta\Omega^{-1}\big)^{-1} \\
&= \Omega^{-1}\Big(I_T - \Delta\Omega^{-1} + \Delta\Omega^{-1}\Delta\Omega^{-1} - \Delta\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1} + O_p(T^{-2})\Big) \\
&= \Omega^{-1} - \Omega^{-1}\Delta\Omega^{-1} + \Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1} - \Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1} + O_p(T^{-2}) \\
&= O_p(1) + O_p(T^{-1/2}) + O_p(T^{-1}) + O_p(T^{-3/2}) + O_p(T^{-2}).
\end{aligned} \tag{A.9}
$$

Thus,

$$
\begin{aligned}
(X'\widehat{\Omega}^{-1}X)^{-1} = \Big(& X'\Omega^{-1}X - X'\Omega^{-1}\Delta\Omega^{-1}X + X'\Omega^{-1}(\Delta\Omega^{-1})^2X - X'\Omega^{-1}(\Delta\Omega^{-1})^3X \\
&+ O_p(T^{-1})\Big)^{-1}
\end{aligned}
$$

$$= \left\{ \left(I_T - (X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1} + (X'\Omega^{-1}(\Delta\Omega^{-1})^2X)(X'\Omega^{-1}X)^{-1} \right.\right.$$

$$\left.\left. + \ldots \right)(X'\Omega^{-1}X) \right\}^{-1}$$

$$= (X'\Omega^{-1}X)^{-1} \left(I_T - (X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1} \right.$$

$$\left. + (X'\Omega^{-1}(\Delta\Omega^{-1})^2X)(X'\Omega^{-1}X)^{-1} + O_p(T^{-3/2}) \right)^{-1}$$

$$= (X'\Omega^{-1}X)^{-1} \left(I_T + (X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1} \right.$$

$$- (X'\Omega^{-1}(\Delta\Omega^{-1})^2X)(X'\Omega^{-1}X)^{-1}$$

$$\left. + (X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1} + O_p(T^{-3/2}) \right)$$

$$= (X'\Omega^{-1}X)^{-1} + (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}$$

$$- (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}(\Delta\Omega^{-1})^2X)(X'\Omega^{-1}X)^{-1}$$

$$+ (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}$$

$$+ O_p(T^{-5/2})$$

$$= A_{-1} + A_{-3/2} + A_{-2} + O_p(T^{-5/2}), \tag{A.10}$$

where

$A_{-1} = (X'\Omega^{-1}X)^{-1}$,

$A_{-3/2} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}$,

$A_{-2} = -(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}(\Delta\Omega^{-1})^2X)(X'\Omega^{-1}X)^{-1}$

$+ (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}$,

in which the suffixes of A indicate the order of magnitude in probability, i.e.,

$A_{-1} = O_p(T^{-1})$.

Using the above results, we have

$$
\begin{aligned}
\Big[R(X'\widehat{\Omega}^{-1}X)^{-1}R'\Big]^{-1} &= \Big[R(X'\Omega^{-1}X)^{-1}R' + R(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X) \\
&\quad (X'\Omega^{-1}X)^{-1}R' + O_p(T^{-2})\Big]^{-1} \\
&= \big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}\Big[I + R(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X) \\
&\quad (X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} + O_p(T^{-1})\Big]^{-1} \\
&= \big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}\Big[I - R(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X) \\
&\quad (X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} + O_p(T^{-1})\Big] \\
&= \big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} - \big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}R(X'\Omega^{-1}X)^{-1} \\
&\quad (X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} + O_p(1).
\end{aligned}
\tag{A.11}
$$

Using (A.11), we can calculate $\widehat{L}$ as

$$
\begin{aligned}
\widehat{L} &= (X'\widehat{\Omega}^{-1}X)^{-1}R'\Big[R(X'\widehat{\Omega}^{-1}X)^{-1}R'\Big]^{-1}R \\
&= \Big[(X'\Omega^{-1}X)^{-1} + (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1} + O_p(T^{-2})\Big]R' \\
&\quad \Big[\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} - \big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}R(X'\Omega^{-1}X)^{-1} \\
&\quad (X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} + O_p(1)\Big]R \\
&= (X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}R - (X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} \\
&\quad R(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}R \\
&\quad + (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}R + O_p(T^{-1}) \\
&= (X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}R - L(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)L \\
&\quad + (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)L + O_p(T^{-1}) \\
&= L_0 + L_{-1/2} + O_p(T^{-1}),
\end{aligned}
\tag{A.12}
$$

where

$$L_0 = L = (X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}R,$$

$$L_{-1/2} = \big(I - L\big)(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Delta\Omega^{-1}X)L.$$

Therefore, the feasible term for the denominator of (2.10) is

$$\begin{aligned}\bar{\bar{\beta}}_{ur}'\widehat{L}'W\widehat{L}\bar{\bar{\beta}}_{ur} &= \Big(\beta + \Pi_{-1/2}\Big)'\Big(L + L_{-1/2} + O_p(T^{-1})\Big)' \\ &\quad W\Big(L + L_{-1/2} + O_p(T^{-1})\Big)\Big(\beta + \Pi_{-1/2}\Big) \\ &= \beta'L'WL\beta + 2\ \beta'L'WL\Pi_{-1/2} + 2\ \beta'L'WL_{-1/2}\ \beta + O_p(1) \\ &= O_p(T) + O_p(T^{1/2}) + O_p(T^{1/2}) + O_p(1),\end{aligned} \tag{A.13}$$

where $\Pi_{-1/2} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon$, and the feasible term for the numerator of (2.10) is

$$\begin{aligned}(X'\widehat{\Omega}^{-1}X)^{-1}\widehat{L}'W &= \Big[(X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\ &\quad + O_p(T^{-2})\Big]\Big(L + L_{-1/2} + O_p(T^{-1})\Big)'W \\ &= (X'\Omega^{-1}X)^{-1}L'W + (X'\Omega^{-1}X)^{-1}L_{-1/2}'W \\ &\quad - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}L'W + O_p(T^{-1}) \\ &= O_p(1) + O_p(T^{-1/2}) + O_p(T^{-1/2}) + O_p(T^{-1}).\end{aligned} \tag{A.14}$$

Besides, the feasible term for $\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r$ is

$$\begin{aligned}\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r &= \widehat{L}\bar{\bar{\beta}}_{ur} \\ &= \Big(L + L_{-1/2} + O_p(T^{-1})\Big)\Big(\beta + \Pi_{-1/2} + \Pi_{-1} + O_p(T^{-3/2})\Big) \\ &= L\beta + L\Pi_{-1/2} + O_p(T^{-1}).\end{aligned} \tag{A.15}$$

Finally, using (A.13), (A.14) and (A.15), we have

$$\begin{aligned}
\bar{\bar{\beta}}_\gamma - \beta &= \bar{\bar{\beta}}_{ur} - \beta - \widehat{\gamma}^*\big(\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r\big) \\
&= \bar{\bar{\beta}}_{ur} - \beta - \mathrm{tr}\big((X'\widehat{\Omega}^{-1}X)^{-1}\widehat{L}'W\big)\big(\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r\big)\Big[\beta' L' W L \beta + 2\ \beta' L' W L \Pi_{-1/2} \\
&\quad + 2\ \beta' L' W L_{-1/2}\ \beta + O_p(1)\Big]^{-1} \\
&= \bar{\bar{\beta}}_{ur} - \beta - \ \frac{1}{\phi}\Big[1 - \frac{2}{\phi}\ \beta' L' W L \Pi_{-1/2} - \frac{2}{\phi}\ \beta' L' W L_{-1/2}\ \beta + O_p(T^{-1})\Big] \qquad \text{(A.16)} \\
&\quad \mathrm{tr}\big((X'\widehat{\Omega}^{-1}X)^{-1}\widehat{L}'W\big)\big(\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r\big) \\
&= \bar{\bar{\beta}}_{ur} - \beta - \Big[\frac{1}{\phi} - \frac{2}{\phi^2}\ \beta' L' W L \Pi_{-1/2} - \frac{2}{\phi^2}\ \beta' L' W L_{-1/2}\ \beta\Big] \\
&\quad \mathrm{tr}\big((X'\Omega^{-1}X)^{-1}L'W\big)\big[L\beta + L\Pi_{-1/2}\big] + O_p(T^{-2}),
\end{aligned}$$

where $\phi = \beta' L' W L \beta$. Thus, the bias for $\bar{\bar{\beta}}_\gamma$, to order $O_p(T^{-1})$, is

$$bias(\bar{\bar{\beta}}_\gamma) = -\frac{1}{\phi}\ \mathrm{tr}\big((X'\Omega^{-1}X)^{-1}L'W\big)L\beta,$$

and MSE, to order $O_p(T^{-2})$, is

$$\begin{aligned}
MSE\big(\bar{\bar{\beta}}_\gamma\big) &= \mathbb{E}\left[(\bar{\bar{\beta}}_\gamma - \beta)(\widehat{\beta}_\gamma - \beta)'\right] \\
&= \mathbb{E}\Big[\Big(\bar{\bar{\beta}}_{ur} - \beta - \widehat{\gamma}^*\big(\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r\big)\Big)\Big(\bar{\bar{\beta}}_{ur} - \beta - \widehat{\gamma}^*\big(\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r\big)\Big)'\Big] \\
&= MSE\big(\bar{\bar{\beta}}_{ur}\big) + \frac{1}{\phi^2}\ L\beta\beta' L'\ \Big(\mathrm{tr}(Q)\Big)^2 \\
&\quad - \frac{\mathrm{tr}(Q)}{\phi}\ \mathbb{E}\Big[L\Pi_{-1/2}\big(\bar{\bar{\beta}}_{ur} - \beta\big)'\Big] - \frac{\mathrm{tr}(Q)}{\phi}\ \mathbb{E}\Big[L\Pi_{-1/2}\big(\bar{\bar{\beta}}_{ur} - \beta\big)'\Big]' \\
&\quad + \mathrm{tr}(Q)\ \mathbb{E}\left(\Big[\frac{2}{\phi^2}\beta' L' W L \Pi_{-1/2} + \frac{2}{\phi^2}\beta' L' W L_{-1/2}\ \beta\Big] L\beta\big(\bar{\bar{\beta}}_{ur} - \beta\big)'\right) \\
&\quad + \mathrm{tr}(Q)\ \mathbb{E}\left(\Big[\frac{2}{\phi^2}\beta' L' W L \Pi_{-1/2} + \frac{2}{\phi^2}\beta' L' W L_{-1/2}\ \beta\Big] L\beta\big(\bar{\bar{\beta}}_{ur} - \beta\big)'\right)' \\
&= MSE\big(\bar{\bar{\beta}}_{ur}\big) + \frac{1}{\phi^2}\ L\beta\beta' L'\ \Big(\mathrm{tr}(Q)\Big)^2 \\
&\quad - \frac{\mathrm{tr}(Q)}{\phi}\ \mathbb{E}\Big[L\Pi_{-1/2}\Pi'_{-1/2}\Big] - \frac{\mathrm{tr}(Q)}{\phi}\ \mathbb{E}\Big[L\Pi_{-1/2}\Pi'_{-1/2}\Big]'
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{2\ \mathrm{tr}(Q)}{\phi^2}\ \mathbb{E}\left[\beta' L' W L \Pi_{-1/2} L \beta \Pi'_{-1/2}\right] + \frac{2\ \mathrm{tr}(Q)}{\phi^2}\ \mathbb{E}\left[\beta' L' W L_{-1/2} \beta L \beta \Pi'_{-1/2}\right] \\
&+ \frac{2\ \mathrm{tr}(Q)}{\phi^2}\ \mathbb{E}\left[\beta' L' W L \Pi_{-1/2} L \beta \Pi'_{-1/2}\right]' + \frac{2\ \mathrm{tr}(Q)}{\phi^2}\ \mathbb{E}\left[\beta' L' W L_{-1/2} \beta L \beta \Pi'_{-1/2}\right]' \\
&= MSE(\bar{\hat{\beta}}_{ur}) + \frac{1}{\phi^2}\ L\beta\beta' L'\ \Big(\mathrm{tr}(Q)\Big)^2 \\
&- \frac{\mathrm{tr}(Q)}{\phi}\,\mathbb{E}\left[L(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right] \\
&- \frac{\mathrm{tr}(Q)}{\phi}\ \mathbb{E}\left[L(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right]' \\
&+ \frac{2\ \mathrm{tr}(Q)}{\phi^2}\ L\beta\beta' L' W L\ \mathbb{E}\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right] + 0 \\
&+ \frac{2\ \mathrm{tr}(Q)}{\phi^2}\ \big(L\beta\beta' L' W L\big)'\ \mathbb{E}\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right]' + 0 \\
&= MSE(\bar{\hat{\beta}}_{ur}) + \frac{1}{\phi^2}\ L\beta\beta' L'\ \Big(\mathrm{tr}(Q)\Big)^2 - \frac{2\ \mathrm{tr}(Q)}{\phi}\ L\ (X'\Omega^{-1}X)^{-1} \\
&+ \frac{2\ \mathrm{tr}(Q)}{\phi^2}\ L\beta\beta' L' W L\ (X'\Omega^{-1}X)^{-1} + \frac{2\ \mathrm{tr}(Q)}{\phi^2}\ (X'\Omega^{-1}X)^{-1} L' W L \beta\beta' L',
\end{aligned}
\tag{A.17}
$$

and finally risk of this estimator, to order $O_p(T^{-1})$, is

$$
\begin{aligned}
Risk(\bar{\hat{\beta}}_{\gamma}, W) &= Risk(\bar{\hat{\beta}}_{ur}, W) + \frac{1}{\phi}\ \Big(\mathrm{tr}(Q)\Big)^2 - \frac{2}{\phi}\ \Big(\mathrm{tr}(Q)\Big)^2 \\
&+ \frac{4\ \mathrm{tr}(Q)}{\phi^2}\ \mathrm{tr}\big(L\beta\beta' L' W L\ (X'\Omega^{-1}X)^{-1}\big) \qquad \text{(A.18)} \\
&= Risk(\bar{\hat{\beta}}_{ur}, W) - \frac{\Big(\mathrm{tr}(Q)\Big)^2}{\phi^2}\ \left[\phi - \frac{4\ \beta' L' W L (X'\Omega^{-1}X)^{-1} W L \beta}{\mathrm{tr}(Q)}\right],
\end{aligned}
$$

where the risk of the combined estimator is less than the risk of the unrestricted estimator if

$$
\begin{aligned}
\frac{4\ \beta' L' W L (X'\Omega^{-1}X)^{-1} W L \beta}{\phi} &< \mathrm{tr}(Q) \\
\sup_{W^{1/2}L\beta} \frac{4\ \beta' L' W^{1/2} W^{1/2} L (X'\Omega^{-1}X)^{-1} W^{1/2} W^{1/2} L\beta}{\beta' L' W L \beta} &< \mathrm{tr}(Q) \\
4\ \lambda_{\max}\big(W^{1/2} L (X'\Omega^{-1}X)^{-1} W^{1/2}\big) &< \mathrm{tr}(Q) \\
4\ \lambda_{\max}(Q) &< \mathrm{tr}(Q).
\end{aligned}
\tag{A.19}
$$

Thus, risk of the combined estimator is less than the unrestricted estimator if $\mathrm{tr}(Q) > 4\lambda_{\max}(Q)$. This complete the proof of Theorem 1. ■

Proof of Theorem 3: In order to find the finite sample risk up to order σ^4 and bias up to order σ^2 for our proposed estimator, first we need to find $\widehat{\sigma^2}$. The estimate of σ^2 can be calculated as

$$
\begin{aligned}
\widehat{\sigma^2} &= \big(Y - X\widehat{\beta}_{ur}\big)'\big(Y - X\widehat{\beta}_{ur}\big)/\big(T-(m+1)k\big) \\
&= \Big(\sigma\epsilon - X\big(\widehat{\beta}_{ur} - \beta\big)\Big)'\Big(\sigma\epsilon - X\big(\widehat{\beta}_{ur} - \beta\big)\Big)/\big(T-(m+1)k\big) \\
&= \Big(\sigma\epsilon - \sigma X(X'X)^{-1}X'\epsilon\Big)'\Big(\sigma\epsilon - \sigma X(X'X)^{-1}X'\epsilon\Big)/\big(T-(m+1)k\big) \\
&= \Big(\sigma\big[I_T - X(X'X)^{-1}X'\big]\epsilon\Big)'\Big(\sigma\big[I_T - X(X'X)^{-1}X'\big]\epsilon\Big)/\big(T-(m+1)k\big) \\
&= \frac{\sigma^2\epsilon' M\epsilon}{T-(m+1)k},
\end{aligned}
\tag{A.20}
$$

where $M = I_T - X(X'X)^{-1}X'$.

Lemma 26 *Let the $T\times 1$ random vector ϵ be such that $\epsilon \sim N(0,\sigma^2 I_T)$, and M_1 and M_2 be arbitrary $T\times T$ matrices. Then,*

$$
\mathbb{E}\Big[(\epsilon' M_1\epsilon)(\epsilon' M_2\epsilon)\Big] = \sigma^4\Big[tr(M_1)tr(M_2) + tr(M_1M_2) + tr(M_1M_2')\Big], \tag{A.21}
$$

$$
\mathbb{E}\Big[\epsilon\epsilon' M_1\epsilon\epsilon'\Big] = \sigma^4\Big[tr(M_1)I_T + M_1 + M_1'\Big]. \tag{A.22}
$$

See Ullah (2004). □

Using the feasible combination weight and expanding the terms up to order σ^3 we have

$$\widehat{\gamma}^* = \frac{\widehat{\sigma^2}\ \mathrm{tr}(Q_s)}{\bar{\bar{\beta}}'_{ur} L'WL\hat{\bar{\beta}}_{ur}}$$

$$= \frac{\sigma^2 \epsilon' M\epsilon\ \mathrm{tr}(Q_s)/\big(T-(m+1)k\big)}{\big(\beta+\sigma(X'X)^{-1}X'\epsilon\big)'\ L'WL\ \big(\beta+\sigma(X'X)^{-1}X'\epsilon\big)}$$

$$= \frac{\sigma^2 \epsilon' M\epsilon\ \mathrm{tr}(Q_s)/(T-(m+1)k)}{\beta'L'WL\beta+\sigma^2\epsilon'X(X'X)^{-1}L'WL(X'X)^{-1}X'\epsilon+2\sigma\epsilon'X(X'X)^{-1}L'WL\beta}$$

$$= \frac{\sigma^2 \epsilon' M\epsilon\ \mathrm{tr}(Q_s)}{\phi\ \big(T-(m+1)k\big)}\left[1+\frac{\sigma^2}{\phi}\epsilon'X(X'X)^{-1}L'WL(X'X)^{-1}X'\epsilon+\frac{2\sigma}{\phi}\epsilon'X(X'X)^{-1}L'WL\beta\right]^{-1}$$

$$= \frac{\sigma^2 \epsilon' M\epsilon\ \mathrm{tr}(Q_s)}{\phi\ \big(T-(m+1)k\big)}\left[1-\frac{2\sigma}{\phi}\epsilon'X(X'X)^{-1}L'WL\beta+O_p(\sigma^4)\right]$$

$$= \nu_1+\nu_2+O_p(\sigma^4), \qquad \text{(A.23)}$$

where $Q_s = W^{1/2}L(X'X)^{-1}W^{1/2}$, $\phi=\beta'L'WL\beta$, $M=I_T-X\big(X'X\big)^{-1}X'$, and

$\nu_1 \equiv \frac{\sigma^2\epsilon'M\epsilon\ \mathrm{tr}(Q_s)}{\phi\ \big(T-(m+1)k\big)} = O_p(\sigma^2)$,

$\nu_2 \equiv -\frac{2\sigma^3\epsilon'M\epsilon\ \mathrm{tr}(Q_s)}{\phi^2\ \big(T-(m+1)k\big)}\Big[\epsilon'X(X'X)^{-1}L'WL\beta\Big] = O_p(\sigma^3)$.

Finally, by plugging the estimated optimal value of the weight into the combined estimator, we calculate the bias up to order σ^2 and the finite sample risk up to order σ^4 for the proposed combined estimator. To proof equation (2.23), rewrite the combined estimator as

$$\hat{\bar{\beta}}_\gamma - \beta = \big(\hat{\bar{\beta}}_{ur}-\beta\big) - \widehat{\gamma}^*\big(\hat{\bar{\beta}}_{ur}-\hat{\bar{\beta}}_r\big)$$

$$= \sigma(X'X)^{-1}X'\epsilon - \big[\nu_1+\nu_2+O_P(\sigma^4)\big]L\big[\beta+\sigma(X'X)^{-1}X'\epsilon\big] \qquad \text{(A.24)}$$

$$= \bar{B}_1+\bar{B}_2+\bar{B}_3+O_P(\sigma^4),$$

where

$\bar{B}_1 \equiv \sigma(X'X)^{-1}X'\epsilon$,

$$\bar{B}_2 \equiv -\nu_1 L\beta = -\frac{\sigma^2 \epsilon' M \epsilon \ \mathrm{tr}(Q_s)}{\phi \ \Big(T-(m+1)k\Big)} L\beta,$$

$$\bar{B}_3 \equiv -\sigma\nu_1 L(X'X)^{-1}X'\epsilon - \nu_2 L\beta$$

$$= -\frac{\sigma^3 \epsilon' M \epsilon \ \mathrm{tr}(Q_s)}{\phi \ \Big(T-(m+1)k\Big)} L(X'X)^{-1}X'\epsilon + \frac{2\sigma^3 \epsilon' M \epsilon \ \mathrm{tr}(Q_s)}{\phi^2 \ \Big(T-(m+1)k\Big)} L\beta \ \epsilon' X(X'X)^{-1}L' \, \mathbb{W} \, L\beta.$$

Thus, the bias for the combined estimator is

$$\begin{aligned} bias(\bar{\widehat{\beta}}_\gamma) &= \mathbb{E}(\bar{B}_1 + \bar{B}_2) \\ &= 0 + \mathbb{E}(\bar{B}_2) \\ &= -\frac{\sigma^2 \mathrm{tr}(Q_s)}{\phi} \ L\beta, \end{aligned} \tag{A.25}$$

and the risk of this combined estimator is

$$\begin{aligned} Risk(\bar{\widehat{\beta}}_\gamma, W) &= \mathbb{E}\left[(\bar{\widehat{\beta}}_\gamma - \beta)' W (\bar{\widehat{\beta}}_\gamma - \beta)\right] \\ &= \mathbb{E}\Big(\bar{B}_1 + \bar{B}_2 + \bar{B}_3\Big)' W \Big(\bar{B}_1 + \bar{B}_2 + \bar{B}_3\Big) \\ &= \mathbb{E}\left(\bar{B}_1' W \bar{B}_1\right) + \mathbb{E}\left(\bar{B}_2' W \bar{B}_2\right) + 2\,\mathbb{E}\left(\bar{B}_1' W (\bar{B}_2 + \bar{B}_3)\right) \\ &= \mathbb{E}\left(\bar{B}_1' W \bar{B}_1\right) + \mathbb{E}\left(\bar{B}_2' W \bar{B}_2\right) + 2\,\mathbb{E}\left(\bar{B}_1' W \bar{B}_3\right) \\ &= Risk(\bar{\widehat{\beta}}_{ur}, W) + \frac{\sigma^4}{\phi \ \Big(T-(m+1)k\Big)} \ \Big(\mathrm{tr}(Q_s)\Big)^2 \Big[T-(m+1)k+2\Big] \\ &- \frac{2\sigma^4 \ \Big(\mathrm{tr}(Q_s)\Big)^2}{\phi} + \frac{4\sigma^4 \ \mathrm{tr}(Q_s)}{\phi^2} \Big[\beta' L' W (X'X)^{-1} L' W L\beta\Big], \end{aligned} \tag{A.26}$$

where

$$\begin{aligned} \mathbb{E}(\bar{B}_1' \, \mathbb{W} \, \bar{B}_1) &= \sigma^2 \, \mathbb{E}\Big[\epsilon' X(X'X)^{-1} W (X'X)^{-1} X'\epsilon\Big] \\ &= \sigma^2 \mathrm{tr}\Big((X'X)^{-1}W\Big) = Risk(\widehat{\beta}_{ur}, W), \end{aligned} \tag{A.27}$$

$$\mathbb{E}(\bar{B}_2'W\bar{B}_2) = \frac{\sigma^4}{\phi^2}\ (\beta'LWL\beta)\big(\text{tr}(Q_s)\big)^2\,\mathbb{E}\Big[\epsilon'M\epsilon\epsilon'M\epsilon\Big]$$

$$= \frac{\sigma^4}{\phi^2\ \big(T-(m+1)k\big)^2}\ (\beta'LWL\beta)\big(\text{tr}(Q_s)\big)^2\Big[\big(T-(m+1)k\big)^2 \tag{A.28}$$

$$+ 2\ \big(T-(m+1)k\big)\Big],$$

and

$$\mathbb{E}\left(\bar{B}_1'W\bar{B}_3\right) = \mathbb{E}\Bigg[\sigma\epsilon'X(X'X)^{-1}W\Bigg(-\frac{\sigma^3\epsilon'M\epsilon\ \text{tr}(Q_s)}{\phi\ \big(T-(m+1)k\big)}L(X'X)^{-1}X'\epsilon$$

$$+\frac{2\sigma^3\epsilon'M\epsilon\ \text{tr}(Q_s)}{\phi^2\ \big(T-(m+1)k\big)}L\beta\ \epsilon'X(X'X)^{-1}L'WL\beta\Bigg)\Bigg]$$

$$= -\frac{\sigma^4\ \text{tr}(Q_s)}{\phi\ \big(T-(m+1)k\big)}\ \text{tr}\Bigg(\mathbb{E}\Big[(X'X)^{-1}WL(X'X)^{-1}X'\epsilon\epsilon'M\epsilon\epsilon'X\Big]\Bigg)$$

$$+\frac{2\sigma^4\ \text{tr}(Q_s)}{\phi^2\ \big(T-(m+1)k\big)}\ \mathbb{E}\Big[\beta'L'W(X'X)^{-1}X'\epsilon\epsilon'M\epsilon\epsilon'X(X'X)^{-1}L'WL\beta\Big]$$

$$= -\frac{\sigma^4\text{tr}(Q_s)}{\phi\ \big(T-(m+1)k\big)}\ \text{tr}\Bigg(\Big[(X'X)^{-1}WL(X'X)^{-1}X'$$

$$\Big(\big(T-(m+1)k\big)I_T+2M\Big)X\Big]\Bigg)+\frac{2\sigma^4\ \text{tr}(Q_s)}{\phi^2\ \big(T-(m+1)k\big)}\Big[\beta'L'W(X'X)^{-1}X'$$

$$\Big(\big(T-(m+1)k\big)I_T+2M)\Big)X(X'X)^{-1}L'WL\beta\Big]$$

$$= -\frac{\sigma^4\big(\text{tr}(Q_s)\big)^2}{\phi}-\frac{2\sigma^4\text{tr}(Q_s)}{\phi\ \big(T-(m+1)k\big)}\text{tr}\Big((X'X)^{-1}WL(X'X)^{-1}X'MX\Big)$$

$$+\frac{2\sigma^4\ \text{tr}(Q_s)}{\phi^2}\Big[\beta'L'W(X'X)^{-1}L'WL\beta\Big]$$

$$+\frac{4\sigma^4\ \text{tr}(Q_s)}{\phi^2\ \big(T-(m+1)k\big)}\Big[\beta'L'W(X'X)^{-1}X'MX(X'X)^{-1}L'WL\beta\Big]$$

$$= -\frac{\sigma^4\ \big(\text{tr}(Q_s)\big)^2}{\phi}+\frac{2\sigma^4\ \text{tr}(Q_s)}{\phi^2}\Big[\beta'L'W(X'X)^{-1}L'WL\beta\Big], \tag{A.29}$$

where $X'MX = 0$. Thus, the risk of the combined estimator is less than the unrestricted estimator if

$$
\begin{aligned}
\operatorname{tr}(Q_s)\Big(1-\frac{2}{T-(m+1)k}\Big) &> \sup_{W^{1/2}L\beta}\frac{4\ \beta' L' W (X'X)^{-1} L' W L\beta}{\beta' L' W L\beta} \\
\operatorname{tr}(Q_s)\Big(1-\frac{2}{T-(m+1)k}\Big) &> 4\ \lambda_{\max}(Q_s).
\end{aligned}
\tag{A.30}
$$

This completes the proof of Theorem 3. ■

Proof of Theorem 5:

The asymptotic distribution of the unrestricted estimator is

$$
\sqrt{T}(\bar{\bar{\beta}}_{ur}-\beta) = \Big(\frac{X'\ \widehat{\Omega}^{-1}\ X}{T}\Big)^{-1}\Big(\frac{X'\ \widehat{\Omega}^{-1}\ \epsilon}{\sqrt{T}}\Big) \xrightarrow{d} Z \sim N(0, V_{ur}), \tag{A.31}
$$

Also, the restricted estimator is

$$
\begin{aligned}
\bar{\bar{\beta}}_{r} &= \bar{\bar{\beta}}_{ur} - \Big(\frac{X'\ \widehat{\Omega}^{-1}X}{T}\Big)^{-1} R' \left[R\left(\frac{X'\ \widehat{\Omega}^{-1}X}{T}\right)^{-1} R'\right]^{-1} R\bar{\bar{\beta}}_{ur} \\
&= \bar{\bar{\beta}}_{ur} - \widehat{V}_{ur}\widehat{P}(R\bar{\bar{\beta}}_{ur}),
\end{aligned}
$$

where $\widehat{P} \equiv R'\left[R\left(\frac{X'\ \widehat{\Omega}^{-1}X}{T}\right)^{-1} R'\right]^{-1}$. So

$$
\begin{aligned}
\sqrt{T}(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}) &= \widehat{V}_{ur}\ \widehat{P}\sqrt{T}\Big(R\bar{\bar{\beta}}_{ur} - R\beta + R\beta\Big) \\
&= \widehat{V}_{ur}\ \widehat{P}R\sqrt{T}\Big(\bar{\bar{\beta}}_{ur} - \beta + \frac{h}{\sqrt{T}}\Big) \\
&\xrightarrow{d} V_{ur}\ PR(Z+h).
\end{aligned}
\tag{A.32}
$$

This completes the proof of Theorem 5. ■

Proof of Theorem 6:

Lemma 27 *Suppose A and B are two matrices, where $A_{n\times n}$ and $B_{n\times m}$, then*

$$
B'AB \leq (B'B)\lambda_{max}(A)
$$

in which $\lambda_{max}(A)$ is the maximum eigenvalue of A. See Bernstein D. S. (2005), page 271 for the proof. □

Having the feasible combination weight, we can derive the risk of the combined estimator as

$$
\begin{aligned}
\rho(\bar{\bar{\beta}}_{\gamma}, W) &= T\,\mathbb{E}\Big[\big(\bar{\bar{\beta}}_{ur}-\beta\big)-\widehat{\gamma}^{*}\big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)\Big]' W \Big[\big(\bar{\bar{\beta}}_{ur}-\beta\big)-\widehat{\gamma}^{*}\big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)\Big] \\
&= \rho(\bar{\bar{\beta}}_{ur}, W) + T\,\mathbb{E}\Big[(\widehat{\gamma}^{*})^2\big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)' W \big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)\Big] \\
&- 2T\,\mathbb{E}\Big[\widehat{\gamma}^{*}\big(\bar{\bar{\beta}}_{ur}-\widehat{\beta}_{r}\big)' W \big(\bar{\bar{\beta}}_{ur}-\beta\big)\Big] \\
&= \rho(\bar{\bar{\beta}}_{ur}, W) + \big(\mathrm{tr}(Q)\big)^2 \;\mathbb{E}\left[\frac{T\big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)' W \big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)}{\big(T\big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)' B \big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)\big)^2}\right] \\
&- 2\;\mathrm{tr}(Q)\;\mathbb{E}\left[\frac{T\big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)' W \big(\bar{\bar{\beta}}_{ur}-\beta\big)}{T\big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)' B \big(\bar{\bar{\beta}}_{ur}-\bar{\bar{\beta}}_{r}\big)}\right] \\
&= \rho(\bar{\bar{\beta}}_{ur}, W) + \big(\mathrm{tr}(Q)\big)^2 \;\mathbb{E}\left[\frac{(Z+h)'R'L'V_{ur}WV_{ur}LR(Z+h)}{\big((Z+h)'R'L'V_{ur}BV_{ur}LR(Z+h)\big)^2}\right] \\
&- 2\mathrm{tr}(Q)\;\mathbb{E}\left[\frac{(Z+h)'R'L'V_{ur}WZ}{(Z+h)'R'L'V_{ur}BV_{ur}LR(Z+h)}\right] \\
&= \rho(\bar{\bar{\beta}}_{ur}, W) + \big(\mathrm{tr}(Q)\big)^2 \;\mathbb{E}\left[\frac{1}{(Z+h)'B(Z+h)}\right] - 2\mathrm{tr}(Q)\;\mathbb{E}\left[\frac{(Z+h)'R'L'V_{ur}WZ}{(Z+h)'B(Z+h)}\right] \\
&= \rho(\bar{\bar{\beta}}_{ur}, W) + \big(\mathrm{tr}(Q)\big)^2 \mathbb{E}\left[\frac{1}{(Z+h)'B(Z+h)}\right] - 2\mathrm{tr}(Q)\,\mathbb{E}\Big[\eta(Z+h)'R'L'V_{ur}WZ\Big],
\end{aligned}
\tag{A.33}
$$

where $\eta(x) = \big(\frac{1}{x'Bx}\big)x$. Note that $\frac{\partial}{\partial x}\eta(x)' = (\frac{1}{x'Bx})I - \frac{2}{(x'Bx)^2}Bxx'$. We can simplify the second term in the above equation by using the Stein's lemma (see lemma 2 in the appendix of Hansen (2016)) as:

$$
\begin{aligned}
\mathbb{E}\Big[\eta(Z+h)'R'L'V_{ur}WZ\Big] &= \mathbb{E}\,\mathrm{tr}\left[\frac{\partial}{\partial(Z+h)}\eta(Z+h)'R'L'V_{ur}WV_{ur}\right] \\
&= \mathbb{E}\,\mathrm{tr}\left[\frac{R'L'V_{ur}WV_{ur}}{(Z+h)'B(Z+h)}\right]
\end{aligned}
$$

$$
\begin{aligned}
&- 2\,\mathbb{E}\,\mathrm{tr}\left[\frac{B(Z+h)(Z+h)'R'L'V_{ur}WV_{ur}}{\Big((Z+h)'B(Z+h)\Big)^2}\right] \\
&= \mathbb{E}\left[\frac{\mathrm{tr}(Q)}{(Z+h)'B(Z+h)}\right] \\
&- 2\,\mathbb{E}\,\mathrm{tr}\left[\frac{(Z+h)'R'L'V_{ur}WV_{ur}B\ (Z+h)}{\Big((Z+h)'B(Z+h)\Big)^2}\right] \\
&= \mathbb{E}\left[\frac{\mathrm{tr}(Q)}{(Z+h)'B(Z+h)}\right] - 2\,\mathbb{E}\,\mathrm{tr}\left[\frac{(Z+h)'B_1'QB_1\ (Z+h)}{\Big((Z+h)'B(Z+h)\Big)^2}\right] \\
&\geq \mathbb{E}\left[\frac{\mathrm{tr}(Q)}{(Z+h)'B(Z+h)}\right] - 2\,\mathbb{E}\left[\frac{\lambda_{\max}(Q)}{(Z+h)'B(Z+h)}\right], \qquad \text{(A.34)}
\end{aligned}
$$

where we use Lemma 27 to get the last inequality, and $B_1 = W^{1/2}V_{ur}LR$ with $B_1'B_1 = B$.

Now, by plugging equation (A.34) into (A.33) we have

$$
\begin{aligned}
\rho(\bar{\bar{\beta}}_{\gamma}, \mathbb{W}) &\leq \rho(\bar{\bar{\beta}}_{ur}, \mathbb{W}) + \big(\mathrm{tr}(Q)\big)^2\ \mathbb{E}\left[\frac{1}{(Z+h)'B(Z+h)}\right] - 2\mathrm{tr}(Q)\ \left[\frac{\mathrm{tr}(Q) - 2\ \lambda_{\max}(Q)}{(Z+h)'B(Z+h)}\right] \\
&= \rho(\bar{\bar{\beta}}_{ur}, \mathbb{W}) - \mathrm{tr}(Q)\left[2\big(\mathrm{tr}(Q) - 2\ \lambda_{\max}(Q)\big) - \mathrm{tr}(Q)\right]\mathbb{E}\left[\frac{1}{(Z+h)'B(Z+h)}\right] \\
&\leq \rho(\bar{\bar{\beta}}_{ur}, \mathbb{W}) - \mathrm{tr}(Q)\left[\mathrm{tr}(Q) - 4\ \lambda_{\max}(Q)\right]\left[\frac{1}{\mathbb{E}(Z+h)'B(Z+h)}\right] \\
&= \rho(\bar{\bar{\beta}}_{ur}, \mathbb{W}) - \left[\frac{\mathrm{tr}(Q) - 4\ \lambda_{\max}(Q)}{(c+1)}\right],
\end{aligned}
\qquad \text{(A.35)}
$$

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

$$
\begin{aligned}
\mathbb{E}(Z+h)'B(Z+h) &= h'Bh + \mathbb{E}(Z'BZ) \\
&= h'Bh + \mathrm{tr}\big(B\,\mathbb{E}(ZZ')\big) \\
&= h'Bh + \mathrm{tr}(BV_{ur}) \qquad \text{(A.36)} \\
&= h'Bh + \mathrm{tr}(Q) \\
&\leq (c+1)\mathrm{tr}(Q),
\end{aligned}
$$

where for any $0 < c < \infty$, we define a ball such that $H(c) = \{h : h'Bh \leq \text{tr}(Q)\ c\}$ and the inequality is for $h \in H(c)$. Therefore, the risk of the combined estimator is less than the unrestricted estimator if $\text{tr}(Q) > 4\ \lambda_{\text{max}}(Q)$. This completes the proof of Theorem 6. ■

Appendix B

Appendix for Chapter 3

Proof of Theorem 9: The proof of Theorem 9 is straightforward by having the distribution of the full-sample and post-break estimators. The full-sample estimator is

$$\begin{aligned}
\widehat{\beta}_{Full} &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \\
&= \left(\sum_{t=1}^{T} x_t x_t' \frac{1}{\sigma_t^2}\right)^{-1} \sum_{t=1}^{T} x_t y_t \frac{1}{\sigma_t^2} \\
&= \left(\sum_{t=1}^{T} x_t x_t' \frac{1}{\sigma_t^2}\right)^{-1} \left[\sum_{t=1}^{T_1} x_t x_t' \ \frac{\beta_{(1)}}{\sigma_{(1)}^2} + \sum_{t=T_1+1}^{T} x_t x_t' \ \frac{\beta_{(2)}}{\sigma_{(2)}^2} + \sum_{t=1}^{T} \frac{x_t \sigma_t \varepsilon_t}{\sigma_t^2}\right] \\
&= \left(\sum_{t=1}^{T} x_t x_t' \frac{1}{\sigma_t^2}\right)^{-1} \left[\sum_{t=1}^{T_1} x_t x_t' \ \frac{\beta_{(1)}}{\sigma_{(1)}^2} - \sum_{t=1}^{T_1} x_t x_t' \ \frac{\beta_{(2)}}{\sigma_{(1)}^2} + \sum_{t=1}^{T_1} x_t x_t' \ \frac{\beta_{(2)}}{\sigma_{(1)}^2}\right. \\
&\quad \left. + \sum_{t=T_1+1}^{T} x_t x_t' \ \frac{\beta_{(2)}}{\sigma_{(2)}^2} + \sum_{t=1}^{T} \frac{x_t \sigma_t \varepsilon_t}{\sigma_t^2}\right] \\
&= \left(\sum_{t=1}^{T} x_t x_t' \frac{1}{\sigma_t^2}\right)^{-1} \left[\sum_{t=1}^{T_1} x_t x_t' \ \frac{\beta_{(1)} - \beta_{(2)}}{\sigma_{(1)}^2} + \sum_{t=1}^{T} x_t x_t' \ \frac{\beta_{(2)}}{\sigma_t^2} + \sum_{t=1}^{T} \frac{x_t \sigma_t \varepsilon_t}{\sigma_t^2}\right],
\end{aligned} \tag{B.1}$$

and its distribution around the true parameter $\beta_{(2)}$ is

$$\sqrt{T}\Big(\widehat{\beta}_{Full} - \beta_{(2)}\Big) = \left(\sum_{t=1}^{T} x_t x_t' \frac{1}{T\sigma_t^2}\right)^{-1} \left[\sum_{t=1}^{T_1} x_t x_t' \; \frac{b_1\sqrt{T}\Big(\beta_{(1)} - \beta_{(2)}\Big)}{Tb_1 \; \sigma_{(1)}^2} + \sum_{t=1}^{T} \frac{x_t \sigma_t \varepsilon_t}{\sqrt{T}\sigma_t^2}\right]$$
$$= \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \left(\frac{X_1'\Omega_1^{-1}X_1}{Tb_1}\right) b_1\delta_1 + \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \left(\sum_{t=1}^{T} \frac{x_t\sigma_t\varepsilon_t}{\sqrt{T}\sigma_t^2}\right)$$
$$\xrightarrow{d} N\Big(Q^{-1}Q_1 b_1\delta_1, Q^{-1}\Big) \tag{B.2}$$

Besides, the distribution of the post-break estimator is

$$\sqrt{T}\Big(\widehat{\beta}_{(2)} - \beta_{(2)}\Big) = \left(\frac{X_2' \; \Omega_2^{-1} \; X_2}{T - Tb_1}\right)^{-1} \left(\frac{X_2' \; \Omega_2^{-1} \; \sigma_{(2)}\varepsilon_2}{\sqrt{1-b_1}\sqrt{T - Tb_1}}\right)$$
$$\xrightarrow{d} N\left(0, \frac{1}{1-b_1} Q_2^{-1}\right) \tag{B.3}$$

Having these distributions, we can write the joint distribution and the proof of Theorem 9 is compltete. ■

Proof for Theorem 11: The risk of the Stein-like combined estimator can be calculated as

$$\rho(\widehat{\beta}_\alpha, \mathbb{W}) = \mathbb{E}\left[T(\widehat{\beta}_\alpha - \beta_{(2)})' \, \mathbb{W} \, (\widehat{\beta}_\alpha - \beta_{(2)})\right]$$
$$= T\,\mathbb{E}\left[\big(\widehat{\beta}_{(2)} - \beta_{(2)}\big) - \alpha\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}\big)\right]' \mathbb{W} \left[\big(\widehat{\beta}_{(2)} - \beta_{(2)}\big) - \alpha\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}\big)\right]$$
$$= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \tau^2\,\mathbb{E}\left[\frac{Z' \; V^{1/2} \; G\,\mathbb{W}\,G' \; V^{1/2} \; Z}{(Z'MZ)^2}\right] - 2\tau\,\mathbb{E}\left[\frac{Z' \; V^{1/2} \; G\,\mathbb{W}\,G_2' \; V^{1/2} \; Z}{Z'MZ}\right]$$
$$= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \tau^2 \; \mathrm{tr}\Big[\mathbb{W}\,G' \; V^{1/2}\,\mathbb{E}\Big((Z'MZ)^{-2}ZZ'\Big)V^{1/2} \; G\Big]$$
$$- 2\tau \; \mathrm{tr}\Big[\mathbb{W}\,G_2' \; V^{1/2}\,\mathbb{E}\Big((Z'MZ)^{-1}ZZ'\Big)V^{1/2} \; G\Big]. \tag{B.4}$$

The challenging part for calculation of this risk function is to take expectation from the noncentral chi-square distribution, with noncentrality parameter equal to $\frac{\theta' M\theta}{2}$. For calculating these expectations, we need to use the following Lemmas.

Lemma 28 *Let $\chi^2_p(\mu)$ denote a noncentral chi-square random variable with noncentral parameter μ and p degree of freedom. Besides, let p denote a positive integer such that $p > 2r$. Then*

$$\mathbb{E}\left[\left(\chi^2_p(\mu)\right)^{-r}\right] = 2^{-r}e^{-\mu}\frac{\Gamma(\frac{p}{2}-r)}{\Gamma(\frac{p}{2})}\ {}_1F_1\Big(\frac{p}{2}-r;\frac{p}{2};\mu\Big),$$

where ${}_1F_1(.;.;.)$ is the confluent hypergeometric function which is defined as ${}_1F_1(a;b;\mu) = \sum_{n=0}^{\infty}\frac{(a)_n}{(b)_n}\frac{\mu^n}{n!}$, where $(a)_n = a(a+1)\dots(a+n-1)$, $(a)_0 = 1$. See Ullah (1974). □

Lemma 29 *The definition of the confluent hypergeometric function implies the following relations:*

1. ${}_1F_1(a;b;\mu) = {}_1F_1(a+1;b;\mu) - \frac{\mu}{b}\ {}_1F_1(a+1;b+1;\mu)$,
2. ${}_1F_1(a;b;\mu) = \frac{b-a}{b}\ {}_1F_1(a;b+1;\mu) + \frac{a}{b}\ {}_1F_1(a+1;b+1;\mu)$, *and*
3. $(b-a-1)\ {}_1F_1(a;b;\mu) = (b-1)\ {}_1F_1(a;b-1;\mu) - a\ {}_1F_1(a+1;b+1;\mu)$

See Lebedev (1972). □

Lemma 30 *Let the $T\times 1$ vector Z is distributed normally with mean vector θ and covariance matrix I_T, and M is any $T\times T$ idempotent matrix with rank r. Also assume $\phi(.)$ is a Borel measurable function. Then:*

$$\begin{aligned}\mathbb{E}\left[\phi(Z'MZ)ZZ'\right] &= \mathbb{E}\left[\phi\Big(\chi^2_{r+2}(\mu)\Big)\right]M + \mathbb{E}\left[\phi\Big(\chi^2_{r}(\mu)\Big)\right](I_T-M)\\ &+ \mathbb{E}\left[\phi\Big(\chi^2_{r+4}(\mu)\Big)\right]M\theta\theta'M + \mathbb{E}\left[\phi\Big(\chi^2_{r}(\mu)\Big)\right](I_T-M)\theta\theta'(I_T-M)\\ &+ \mathbb{E}\left[\phi\Big(\chi^2_{r+2}(\mu)\Big)\right]\Big(\theta\theta'M + M\theta\theta' - 2M\theta\theta'M\Big),\end{aligned}$$

where $\mu \equiv \frac{\theta' M\theta}{2}$ is the non-centrality parameter. See Judge and Bock (1978). □

Using Lemmas 28 - 30, we can calculate the expectations in (B.4). For simplicity, define $A \equiv G'V^{1/2}$, $\Psi \equiv G_2'V^{1/2}$, and let the non-centrality parameter of chi-square distribution based on Lemma 30 to be defined as $\mu \equiv \frac{\theta' M\theta}{2}$. For clarity, we focus on the second and third term in equation (B.4) one by one. The second term can be simplified as

$$
\begin{aligned}
\operatorname{tr}\Big[\mathbb{W}\, A\, \mathbb{E}\Big((Z'MZ)^{-2}ZZ'\Big)A'\Big] &= \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-2} \operatorname{tr}\Big(\mathbb{W}\, AMA'\Big) + \mathbb{E}\left[\chi^2_{k}(\mu)\right]^{-2} \\
&\quad \operatorname{tr}\Big(\mathbb{W}\, A(I_{2K}-M)A'\Big) \\
&\quad + \mathbb{E}\left[\chi^2_{k+4}(\mu)\right]^{-2} \operatorname{tr}\Big(\mathbb{W}\, AM\theta\theta' MA'\Big) \\
&\quad + \mathbb{E}\left[\chi^2_{k}(\mu)\right]^{-2} \operatorname{tr}\Big(\mathbb{W}\, A(I_{2K}-M)\theta\theta'(I_{2K}-M)A'\Big) \\
&\quad + \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-2} \operatorname{tr}\Big(\mathbb{W}\, A(\theta\theta' M + M\theta\theta' - 2M\theta\theta' M)A'\Big) \\
&= \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-2} \operatorname{tr}\Big(\mathbb{W}\, AA'\Big) + \mathbb{E}\left[\chi^2_{k+4}(\mu)\right]^{-2} \operatorname{tr}\Big(\mathbb{W}\, A\theta\theta' A'\Big) \\
&= \left[\frac{1}{4}\, e^{-\mu}\, \frac{\Gamma(\frac{k}{2}-1)}{\Gamma(\frac{k}{2}+1)}\ {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}+1; \mu\Big)\right] \operatorname{tr}\Big(\mathbb{W}\, AA'\Big) \\
&\quad + \left[\frac{1}{4}\, e^{-\mu}\, \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+2)}\ {}_1F_1\Big(\frac{k}{2}; \frac{k}{2}+2; \mu\Big)\right] \Big(\theta' A'\, \mathbb{W}\, A\theta\Big) \\
&= \left[\frac{\theta' A'\, \mathbb{W}\, A\theta}{(k-2)\theta' M\theta}\right] \left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}\Big)\right] \\
&\quad - \left[\frac{\theta' A'\, \mathbb{W}\, A\theta}{(k-2)\theta' M\theta} - \frac{\operatorname{tr}\big(\mathbb{W}\, AA'\big)}{k(k-2)}\right] \left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}+1\Big)\right],
\end{aligned} \tag{B.5}
$$

where the last equality derives by using Lemma 29 several times, $AM = A$, $MA' = A'$, $A\big(I_{2K}-M\big)A' = 0$, $A\big(I_{2K}-M\big)\theta\theta'\big(I_{2K}-M\big)A' = 0$, and $A\big(\theta\theta' M + M\theta\theta' - 2M\theta\theta' M\big)A' = 0$. Remember that M is an idempotent matrix. The third term in equation (B.4) can be simplified as

$$
\begin{aligned}
\operatorname{tr}\Big[\mathbb{W}\,\Psi\,\mathbb{E}\Big((Z'MZ)^{-1}ZZ'\Big)A'\Big] &= \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-1}\operatorname{tr}\Big(\mathbb{W}\,\Psi M A'\Big) + \mathbb{E}\left[\chi^2_{k}(\mu)\right]^{-1} \\
&\operatorname{tr}\Big(\mathbb{W}\,\Psi(I_{2K}-M)A'\Big) \\
&+\mathbb{E}\left[\chi^2_{k+4}(\mu)\right]^{-1}\operatorname{tr}\Big(\mathbb{W}\,\Psi M\theta\theta' M A'\Big) \\
&+\mathbb{E}\left[\chi^2_{k}(\mu)\right]^{-1}\operatorname{tr}\Big(\mathbb{W}\,\Psi(I_{2K}-M)\theta\theta'(I_{2K}-M)A'\Big) \\
&+\mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-1}\operatorname{tr}\Big(\mathbb{W}\,\Psi(\theta\theta' M + M\theta\theta' - 2M\theta\theta' M)A'\Big) \\
&= \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-1}\operatorname{tr}\Big(\mathbb{W}\,AA'\Big) + \mathbb{E}\left[\chi^2_{k+4}(\mu)\right]^{-1}\operatorname{tr}\Big(\mathbb{W}\,A\theta\theta' A'\Big) \\
&+0-\mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-1}\operatorname{tr}\Big(\mathbb{W}\,A\theta\theta' A'\Big) \\
&= \left[\frac{1}{2}e^{-\mu}\frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+1)}\,{}_1F_1(\frac{k}{2};\frac{k}{2}+1;\mu)\right]\operatorname{tr}\Big(\mathbb{W}\,AA' - \mathbb{W}\,A\theta\theta' A'\Big) \\
&+\left[\frac{1}{2}\;e^{-\mu}\;\frac{\Gamma(\frac{k}{2}+1)}{\Gamma(\frac{k}{2}+2)}\;{}_1F_1\Big(\frac{k}{2}+1;\frac{k}{2}+2;\mu\Big)\right]\operatorname{tr}\Big(\mathbb{W}\,A\theta\theta' A'\Big) \\
&= 2\;e^{-\mu}\;\left[\frac{2\big(\theta' A'\,\mathbb{W}\,A\theta\big)}{\big(\theta' M\theta\big)\big(k-2\big)} - \frac{\operatorname{tr}(\mathbb{W}\,AA')}{k-2}\right] \\
&\times\left[{}_1F_1(\frac{k}{2}-1;\frac{k}{2};\mu) - {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right], \qquad \text{(B.6)}
\end{aligned}
$$

where the last equality derives by using Lemma 29 several times, $\Psi M = A$, $\Psi\theta = 0$, $\Psi\big(I_{2K}-M\big)A' = 0$, $\Psi\big(I_{2K}-M\big)\theta\theta'\big(I_{2K}-M\big)A' = 0$, and $AA' = V_{(2)} - V_{Full}$. Finally, plugging (B.5) and (B.6) into (B.4), and defining $\mathcal{H} \equiv V^{1/2}G\,\mathbb{W}\,G'V^{1/2}$ for the ease of writing, produces

$$\rho(\widehat{\beta}_\alpha, \mathbb{W}) = \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \frac{\tau^2}{k-2}\left\{\frac{\theta'\mathcal{H}\theta}{\theta'M\theta}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big) - e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}$$

$$+ \frac{\tau^2\ \mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)}{k(k-2)}\ \left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}$$

$$- \frac{2\ \tau}{k-2}\left\{\mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big) - \frac{2\ \theta'\mathcal{H}\theta}{\theta'M\theta}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big)\right\}$$

$$+ \frac{2\ \tau}{k-2}\left\{\mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big) - \frac{2\ \theta'\mathcal{H}\theta}{\theta'M\theta}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}$$

$$- \frac{2\ \tau}{k-2}\left\{\mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big) - \frac{2\ \theta'\mathcal{H}\theta}{\theta'M\theta}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}-1;\mu\Big)\right\}$$

$$+ \frac{4\ \tau}{k-2}\left\{\frac{\mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)}{k} - \frac{\theta'\mathcal{H}\theta}{\theta'M\theta}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}$$

$$= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \frac{\tau^2}{k-2}\left\{\frac{\theta'\mathcal{H}\theta}{\theta'M\theta}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big) - e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}$$

$$+ \frac{\tau^2\ \mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)}{k(k-2)}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}$$

$$- \frac{2\ \tau}{k-2}\left\{\mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big) - \frac{2\ \theta'\mathcal{H}\theta}{\theta'M\theta}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big)\right.$$

$$\left. - e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}$$

$$- \frac{2\ \tau}{k-2}\left\{\frac{(k-2)\ \mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)}{k}\right\}\left\{e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}$$

$$= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \frac{2\ \tau\ \theta'\mathcal{H}\theta}{k(k+2)\ \theta'M\theta}\left\{\tau - 2\left(\frac{\mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)\ \theta'M\theta}{\theta'\mathcal{H}\theta} - 2\right)\right\}$$

$$\left\{\mu\ e^{-\mu}{}_1F_1\Big(\frac{k}{2};\frac{k}{2}+2;\mu\Big)\right\} + \frac{\tau\ \mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)}{k(k-2)}\left\{\tau - 2(k-2)\right\}$$

$$\left\{e^{-\mu}{}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right\}, \tag{B.7}$$

where ${}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big) - {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big) = \frac{2\ \mu\ (k-2)}{k(k+1)}\left\{{}_1F_1\Big(\frac{k}{2};\frac{k}{2}+2;\mu\Big)\right\}$ in the last equality. Thus, the risk of the Stein-like combined estimator is less than the risk of the post-break estimator if:

1. $0 \leq \tau < 2\left(\frac{\text{tr}\left(\mathbb{W}(V_{(2)}-V_{Full})\right)\ \theta' M\theta}{\theta'\mathcal{H}\theta} - 2\right)$,

2. $0 \leq \tau < 2(k-2)$,

where the upper bound in condition (I) is greater than zero if

$$\text{tr}\left(\mathbb{W}(V_{(2)} - V_{Full})\right) > \text{Sup}\ \frac{2\ \theta'\mathcal{H}\theta}{\theta' M\theta}$$

$$\text{tr}\left(\mathbb{W}(V_{(2)} - V_{Full})\right) > \text{Sup}\ \frac{2\ \theta' V^{1/2} G\,\mathbb{W}\,G' V^{1/2}\theta}{\theta' V^{1/2} G(V_{(2)} - V_{Full})^{-1/2}\ (V_{(2)} - V_{Full})^{-1/2} G' V^{1/2}\theta} \tag{B.8}$$

$$\text{tr}\left(\mathbb{W}(V_{(2)} - V_{Full})\right) > \lambda_{\max}\left((V_{(2)} - V_{Full})^{1/2}\,\mathbb{W}(V_{(2)} - V_{Full})^{1/2}\right).$$

Besides, the upper bound in condition (II) is positive if the number of regressors are greater than 2, $k > 2$. This completes the proof of Theorem 11. ∎

Proof of Theorem 16: This part, shows the step by step procedure that is needed for the proof of the optimality presented in Theorem 16. To do that, at first we need to derive $L(\gamma)$ and $R(\gamma)$. Rewrite equation (3.21) as

$$\widehat{\beta}_\gamma - \beta_{(2)} = \left(\gamma X_1' X_1 + X_2' X_2\right)^{-1}\left(\gamma\ X_1' X_1 \lambda + \gamma\ X_1' \sigma_{(1)}\epsilon_1 + X_2' \sigma_{(2)}\epsilon_2\right), \tag{B.9}$$

where $\lambda = \beta_{(1)} - \beta_{(2)}$, $\epsilon_1 = (\varepsilon_1, \ldots, \varepsilon_{T_1})'$, $\epsilon_2 = (\varepsilon_{T_1+1}, \ldots, \varepsilon_T)'$. Based on the loss function,

$$
\begin{aligned}
R(\gamma) &= \mathbb{E}\left[\left(\widehat{\beta}_\gamma - \beta_{(2)}\right)' \ \mathbb{W} \ \left(\widehat{\beta}_\gamma - \beta_{(2)}\right)\right] \\
&= \mathbb{E}\Bigg[\Big(\gamma \ \lambda' X_1' X_1 + \gamma \ \sigma_{(1)} \epsilon_1' X_1 + \sigma_{(2)} \epsilon_2' X_2\Big)\big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \\
&\quad \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1}\Big(\gamma \ X_1' X_1 \lambda + \gamma \ X_1' \sigma_{(1)} \epsilon_1 + X_2' \sigma_{(2)} \epsilon_2\Big)\Bigg] \\
&= \gamma^2 \lambda' X_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} X_1' X_1 \lambda \\
&+ \gamma^2 \, \mathbb{E}\left[\sigma_{(1)}^2 \ \epsilon_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} X_1' \epsilon_1\right] \\
&+ \mathbb{E}\left[\sigma_{(2)}^2 \ \epsilon_2' X_2 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} X_2' \epsilon_2\right] \\
&= \lambda' A_1 \lambda + \gamma^2 \sigma_{(1)}^2 \ \mathrm{tr}(A_2) + \sigma_{(2)}^2 \ \mathrm{tr}(A_3),
\end{aligned} \tag{B.10}
$$

where $A_1 = \gamma^2 X_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \ \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} X_1' X_1$,

$A_2 = X_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1}$ and

$A_3 = X_2' X_2 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1}$. Since λ is the break size which is not known, we can plug the unbiased estimator for this term instead. The unbiased estimator for $\lambda' A_1 \lambda$ is equal to

$$
\widehat{\lambda}' A_1 \widehat{\lambda} - \widehat{\sigma_{(1)}^2} \ \mathrm{tr}\Big(A_1 (X_1' X_1)^{-1}\Big) - \widehat{\sigma_{(2)}^2} \ \mathrm{tr}\Big(A_1 (X_2' X_2)^{-1}\Big) \tag{B.11}
$$

and by plugging it back into (B.10) we have:

$$
\begin{aligned}
\widehat{R}(\gamma) &= \widehat{\lambda}' A_1 \widehat{\lambda} + \widehat{\sigma_{(1)}^2} \ \mathrm{tr}\Big(\gamma^2 A_2 - A_1 (X_1' X_1)^{-1}\Big) + \widehat{\sigma_{(2)}^2} \ \mathrm{tr}\Big(A_3 - A_1 (X_2' X_2)^{-1}\Big) \\
&= \widehat{\lambda}' A_1 \widehat{\lambda} + 0 + 2\widehat{\sigma_{(2)}^2} \ \mathrm{tr}\Big(\mathbb{W} \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1}\Big) - \widehat{\sigma_{(2)}^2} \ \mathrm{tr}\Big(\big(X_2' X_2\big)^{-1} \mathbb{W}\Big),
\end{aligned} \tag{B.12}
$$

where $\gamma^2 A_2 - A_1 (X_1' X_1)^{-1} = 0$. Also, $A_3 - A_1 (X_2' X_2)^{-1} = 2 \ \mathrm{tr}\Big(\mathbb{W} \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1}\Big) - \mathrm{tr}\Big(\big(X_2' X_2\big)^{-1} \ \mathbb{W}\Big)$, and $\widehat{\sigma_{(2)}^2} = \frac{(Y_2 - X_2 \widehat{\beta}_{(2)})'(Y_2 - X_2 \widehat{\beta}_{(2)})}{T - T_1 - k}$ is an unbiased estimator for the variance of the post-break observations. Besides, we can rewrite $\widehat{\lambda}' A_1 \widehat{\lambda}$ as

$$
\begin{aligned}
\widehat{\lambda}' A_1 \widehat{\lambda} &= \big(\widehat{\beta}_{(1)} - \widehat{\beta}_{(2)}\big)' A_1 \big(\widehat{\beta}_{(1)} - \widehat{\beta}_{(2)}\big) \\
&= \widehat{\beta}'_{(1)} A_1 \widehat{\beta}_{(1)} + \widehat{\beta}'_{(2)} A_1 \widehat{\beta}_{(2)} - 2\widehat{\beta}'_{(2)} A_1 \widehat{\beta}_{(1)} \\
&= \gamma^2 Y_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \left(\gamma X_1' X_1 + X_2' X_2\right)^{-1} X_1' Y_1 + \widehat{\beta}'_{(2)} \mathbb{W} \widehat{\beta}_{(2)} - 2\widehat{\beta}'_{(2)} \mathbb{W} \widehat{\beta}_\gamma \\
&+ 2\widehat{\beta}'_{(2)} \mathbb{W} \left(\gamma X_1' X_1 + X_2' X_2\right)^{-1} \gamma X_1' Y_1 + \widehat{\beta}_\gamma \mathbb{W} \widehat{\beta}_\gamma - 2\widehat{\beta}'_\gamma \mathbb{W} \left(\gamma X_1' X_1 + X_2' X_2\right)^{-1} \gamma X_1' Y_1 \\
&+ \gamma^2 Y_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \left(\gamma X_1' X_1 + X_2' X_2\right)^{-1} X_1' Y_1 \\
&- 2\widehat{\beta}'_{(2)} \mathbb{W} \left(\gamma X_1' X_1 + X_2' X_2\right)^{-1} \gamma X_1' Y_1 + 2\widehat{\beta}'_\gamma \mathbb{W} \left(\gamma X_1' X_1 + X_2' X_2\right)^{-1} \gamma X_1' Y_1 \\
&- 2\gamma^2 \, Y_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} \mathbb{W} \left(\gamma X_1' X_1 + X_2' X_2\right)^{-1} X_1 Y_1 \\
&= \widehat{\beta}'_{(2)} \mathbb{W} \widehat{\beta}_{(2)} - 2\widehat{\beta}'_{(2)} \mathbb{W} \widehat{\beta}_\gamma + \widehat{\beta}'_\gamma \mathbb{W} \widehat{\beta}_\gamma \\
&= \widehat{\beta}'_{(2)} \; X_2' X_2 \; \widehat{\beta}_{(2)} - 2 \; Y_2' X_2 \widehat{\beta}_\gamma + \big(\widehat{\mu}(\gamma) - Y_2 + Y_2\big)' \big(\widehat{\mu}(\gamma) - Y_2 + Y_2\big) \\
&= \widehat{\beta}'_{(2)} \; X_2' X_2 \; \widehat{\beta}_{(2)} - 2 \; Y_2' \; \widehat{\mu}(\gamma) + \big(\widehat{\mu}(\gamma) - Y_2\big)' \big(\widehat{\mu}(\gamma) - Y_2\big) + Y_2' Y_2 + 2\big(\widehat{\mu}(\gamma) - Y_2\big)' Y_2 \\
&= \widehat{\beta}'_{(2)} \; X_2' X_2 \; \widehat{\beta}_{(2)} + \big(\widehat{\mu}(\gamma) - Y_2\big)' \big(\widehat{\mu}(\gamma) - Y_2\big) - Y_2' Y_2,
\end{aligned}
\tag{B.13}
$$

where we plug $\mathbb{W} = X_2' X_2$ to derive the MSFE.

Thus, based on equation (B.12), the MSFE for this semi-parametric estimator up to the relevant terms to γ is

$$
\widehat{R}(\gamma) = \big(\widehat{\mu}(\gamma) - Y_2\big)' \big(\widehat{\mu}(\gamma) - Y_2\big) + 2 \; \widehat{\sigma^2_{(2)}} \; \mathrm{tr}\Big(X_2' X_2 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1}\Big). \tag{B.14}
$$

Define $P(\gamma) \equiv X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}X_2'$. Thus, we can rewrite $\widehat{\mu}(\gamma)$ as

$$\begin{aligned}\widehat{\mu}(\gamma) &= X_2\ \widehat{\beta}_\gamma \\ &= X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}\big(\gamma X_1'Y_1 + X_2'Y_2\big) \\ &= X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}\big(\gamma X_1'Y_1\big) + X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}X_2'Y_2 \\ &= \Phi + P(\gamma)Y_2,\end{aligned} \tag{B.15}$$

where $\Phi \equiv X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}\big(\gamma X_1'Y_1\big)$. Based on this definition, we can rewrite the risk in (B.14) as

$$\begin{aligned}\widehat{R}(\gamma) &= \big(\widehat{\mu}(\gamma) - Y_2\big)'\big(\widehat{\mu}(\gamma) - Y_2\big) + 2\widehat{\sigma^2_{(2)}}\ \mathrm{tr}\Big(X_2'X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}\Big) \\ &= \Big(\widehat{\mu}(\gamma) - \mu - \sigma_{(2)}\epsilon_2\Big)'\Big(\widehat{\mu}(\gamma) - \mu - \sigma_{(2)}\epsilon_2\Big) + 2\widehat{\sigma^2_{(2)}}\ \mathrm{tr}\big(P(\gamma)\big) \\ &= L(\gamma) + \sigma^2_{(2)}\ \epsilon_2'\epsilon_2 - 2\sigma_{(2)}\ \epsilon_2'\Big(\widehat{\mu}(\gamma) - \mu\Big) + 2\widehat{\sigma^2_{(2)}}\ \mathrm{tr}\big(P(\gamma)\big) \\ &= L(\gamma) + \sigma^2_{(2)}\ \epsilon_2'\epsilon_2 - 2\sigma_{(2)}\ \epsilon_2'\ \widehat{\mu}(\gamma) + 2\sigma_{(2)}\ \epsilon_2'\mu + 2\widehat{\sigma^2_{(2)}}\ \mathrm{tr}\big(P(\gamma)\big) \\ &= L(\gamma) + \sigma^2_{(2)}\ \epsilon_2'\epsilon_2 - 2\sigma_{(2)}\ \epsilon_2'\ \big(\Phi + P(\gamma)Y_2\big) + 2\sigma_{(2)}\ \epsilon_2'\mu + 2\widehat{\sigma^2_{(2)}}\ \mathrm{tr}\big(P(\gamma)\big) \\ &= L(\gamma) + \sigma^2_{(2)}\ \epsilon_2'\epsilon_2 - 2\sigma_{(2)}\ \epsilon_2'\ \Phi - 2\sigma_{(2)}\ \epsilon_2'\ P(\gamma)\mu - 2\sigma_{(2)}\ \epsilon_2'P(\gamma)\ \sigma_{(2)}\epsilon_2 \\ &+ 2\sigma_{(2)}\ \epsilon_2'\ \mu + 2\widehat{\sigma^2_{(2)}}\ \mathrm{tr}\big(P(\gamma)\big) \\ &= L(\gamma) + \sigma^2_{(2)}\ \epsilon_2'\epsilon_2 - 2\sigma_{(2)}\ \epsilon_2'\ \Phi - 2\sigma_{(2)}\ \mu'\ P(\gamma)\ \epsilon_2 - 2\sigma^2_{(2)}\ \epsilon_2'P(\gamma)\ \epsilon_2 \\ &+ 2\sigma_{(2)}\ \epsilon_2'\ \mu + 2\widehat{\sigma^2_{(2)}}\ \mathrm{tr}\big(P(\gamma)\big).\end{aligned} \tag{B.16}$$

We want to show that the estimate of the weight, $\widehat{\gamma}$, that we get from CV is optimal in the sense that the average squared error of the CV is asymptotically as small as the average squared error of the infeasible best possible estimator. Also notice that

$$\begin{aligned}R(\gamma) &= \mathbb{E}\left[L(\gamma)\right] \\ &= \mathbb{E}\Big[\big(\widehat{\mu}(\gamma) - \mu\big)'\big(\widehat{\mu}(\gamma) - \mu\big)\Big]\end{aligned}$$

$$= \mathbb{E}\left[\big(\Phi + P(\gamma)Y_2 - \mu\big)'\big(\Phi + P(\gamma)Y_2 - \mu\big)\right]$$

$$= \mathbb{E}\left[\Big(\Phi + P(\gamma)\mu + P(\gamma)\ \sigma_{(2)}\epsilon_2 - \mu\Big)'\Big(\Phi + P(\gamma)\mu + P(\gamma)\ \sigma_{(2)}\epsilon_2 - \mu\Big)\right]$$

$$= \mathbb{E}\left[\Phi'\Phi + \sigma^2_{(2)}\ \epsilon_2' P^2(\gamma)\epsilon_2 + \big(P(\gamma)\ \mu - \mu\big)'\big(P(\gamma)\ \mu - \mu\big) + 2\Phi'\big(P(\gamma)\ \mu - \mu\big)\right]$$

$$= \mathbb{E}\left[\Phi'\Phi\right] + \sigma^2_{(2)}\ \mathrm{tr}\Big(P^2(\gamma)\Big) + \big(P(\gamma)\ \mu - \mu\big)'\big(P(\gamma)\ \mu - \mu\big) + 2\Phi'\big(P(\gamma)\ \mu - \mu\big)$$

$$= \Upsilon + \sigma^2_{(1)}\ \gamma\ \mathrm{tr}\big(P(\gamma) - P^2(\gamma)\big) + \sigma^2_{(2)}\ \mathrm{tr}\big(P^2(\gamma)\big) + \big(P(\gamma)\mu - \mu\big)'\big(P(\gamma)\mu - \mu\big)$$

$$+ 2\Phi'\big(P(\gamma)\ \mu - \mu\big)$$

$$= \Upsilon + \sigma^2_{(1)}\ \gamma\ \mathrm{tr}\big(P(\gamma)\big) + \big(\sigma^2_{(2)} - \gamma\sigma^2_{(1)}\big)\ \mathrm{tr}\big(P^2(\gamma)\big) + \Big(P(\gamma)\ \big(\mu + \sigma_{(2)}\epsilon_2 - \sigma_{(2)}\epsilon_2\big) - \mu\Big)'$$

$$\Big(P(\gamma)\ \big(\mu + \sigma_{(2)}\epsilon_2 - \sigma_{(2)}\epsilon_2\big) - \mu\Big) + 2\Phi'\big(P(\gamma)\ \mu - \mu\big)$$

$$= \Upsilon + \sigma^2_{(1)}\ \gamma\ \mathrm{tr}\big(P(\gamma)\big) + \big(\sigma^2_{(2)} - \gamma\sigma^2_{(1)}\big)\ \mathrm{tr}\big(P^2(\gamma)\big)$$

$$+ \Big(P(\gamma)Y_2 + \Phi - \mu - \Phi - \sigma_{(2)}\ P(\gamma)\epsilon_2\Big)'$$

$$\Big(P(\gamma)Y_2 + \Phi - \mu - \Phi - \sigma_{(2)}\ P(\gamma)\epsilon_2\Big) + 2\Phi'\big(P(\gamma)\ \mu - \mu\big)$$

$$= \Upsilon + \sigma^2_{(1)}\ \gamma\ \mathrm{tr}\big(P(\gamma)\big) + \big(\sigma^2_{(2)} - \gamma\sigma^2_{(1)}\big)\ \mathrm{tr}\big(P^2(\gamma)\big) + \Big(\widehat{\mu}(\gamma) - \mu - \Phi - \sigma_{(2)}\ P(\gamma)\epsilon_2\Big)'$$

$$\Big(\widehat{\mu}(\gamma) - \mu - \Phi - \sigma_{(2)}\ P(\gamma)\epsilon_2\Big) + 2\Phi'\big(P(\gamma)\ \mu - \mu\big)$$

$$= L(\gamma) + \Phi'\Phi + \sigma^2_{(2)}\ \epsilon_2' P^2(\gamma)\epsilon_2 - 2\big(\widehat{\mu}(\gamma) - \mu\big)'\big(\Phi + \sigma_{(2)}\ P(\gamma)\epsilon_2\big)$$

$$+ \Upsilon + \sigma^2_{(1)}\ \gamma\ \mathrm{tr}\big(P(\gamma)\big) + \big(\sigma^2_{(2)} - \gamma\sigma^2_{(1)}\big)\ \mathrm{tr}\big(P^2(\gamma)\big) + 2\Phi'\big(P(\gamma)\ \mu - \mu\big)$$

$$= L(\gamma) + \Phi'\Phi + \sigma^2_{(2)}\ \epsilon_2' P^2(\gamma)\epsilon_2 - 2\Big(\Phi + P(\gamma)\big(\mu + \sigma_{(2)}\epsilon_2\big) - \mu\Big)'\Big(\Phi + \sigma_{(2)}\ P(\gamma)\epsilon_2\Big)$$

$$+ \Upsilon + \sigma^2_{(1)}\ \gamma\ \mathrm{tr}\big(P(\gamma)\big) + \big(\sigma^2_{(2)} - \gamma\sigma^2_{(1)}\big)\ \mathrm{tr}\big(P^2(\gamma)\big) + 2\Phi'\big(P(\gamma)\ \mu - \mu\big)$$

$$= L(\gamma) + \Phi'\Phi + \sigma^2_{(2)}\ \epsilon_2' P^2(\gamma)\epsilon_2 - 2\Phi'\Phi - 2\sigma_{(2)}\ \Phi' P(\gamma)\epsilon_2$$

$$- 2\Big(P(\gamma)\mu - \mu\Big)'\Big(\Phi + \sigma_{(2)}\ P(\gamma)\epsilon_2\Big)$$

$$- 2\sigma^2_{(2)}\ \epsilon_2' P^2(\gamma)\epsilon_2 + \Upsilon + \sigma^2_{(1)}\ \gamma\ \mathrm{tr}\big(P(\gamma)\big) + \big(\sigma^2_{(2)} - \gamma\sigma^2_{(1)}\big)\ \mathrm{tr}\big(P^2(\gamma)\big) + 2\Phi'\big(P(\gamma)\ \mu - \mu\big)$$

$$= L(\gamma) - \Psi - \sigma_{(2)}^2\ \epsilon_2' P^2(\gamma)\epsilon_2 - 2\sigma_{(2)}\Phi' P(\gamma)\epsilon_2 - 2\ \big(P(\gamma)\mu - \mu\big)'\sigma_{(2)} P(\gamma)\epsilon_2$$

$$+ \sigma_{(1)}^2 \gamma\ \mathrm{tr}\big(P(\gamma)\big) + \big(\sigma_{(2)}^2 - \gamma\sigma_{(1)}^2\big)\ \mathrm{tr}\big(P^2(\gamma)\big), \tag{B.17}$$

where $\Phi'\Phi = \Upsilon + \Psi$, $\Upsilon \equiv \gamma^2 \beta_{(1)}' X_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} X_2' X_2\ \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} X_1' X_1 \beta_{(1)}$, $\Psi \equiv \gamma^2\ \sigma_{(1)}^2 \epsilon_1' X_1 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} X_2' X_2 \big(\gamma X_1' X_1 + X_2' X_2\big)^{-1} X_1' \epsilon_1$, and $\mathbb{E}\left[\Phi'\Phi\right] = \Upsilon + \sigma_{(1)}^2\ \gamma\ \mathrm{tr}\big(P(\gamma) - P^2(\gamma)\big)$.

Note that we can rewrite optimality condition presented in Theorem 16 as

$$\frac{\widehat{R}(\gamma)}{R(\gamma)} = \frac{\widehat{R}(\gamma) - R(\gamma) + R(\gamma)}{R(\gamma)} = 1 + \frac{\widehat{R}(\gamma) - R(\gamma)}{R(\gamma)}. \tag{B.18}$$

Thus, to prove the optimality condition in Theorem 16, it suffices to prove that the terms in $\frac{\widehat{R}(\gamma) - R(\gamma)}{R(\gamma)}$ are negligible. In other words, we need to prove that the following conditions hold.

$$\operatorname*{Sup}_{\gamma \in \mathcal{H}} \frac{|\sigma_{(2)} \epsilon_2' \Phi|}{R(\gamma)} = o_p(1), \tag{B.19}$$

$$\operatorname*{Sup}_{\gamma \in \mathcal{H}} \frac{|\sigma_{(2)}^2 \epsilon_2' P(\gamma)\epsilon_2|}{R(\gamma)} = o_p(1), \tag{B.20}$$

$$\operatorname*{Sup}_{\gamma \in \mathcal{H}} \frac{|\Psi|}{R(\gamma)} = o_p(1), \tag{B.21}$$

$$\operatorname*{Sup}_{\gamma \in \mathcal{H}} \frac{|\sigma_{(2)}^2\ \epsilon_2' P^2(\gamma)\epsilon_2|}{R(\gamma)} = o_p(1), \tag{B.22}$$

$$\operatorname*{Sup}_{\gamma \in \mathcal{H}} \frac{|\sigma_{(2)} \Phi' P(\gamma)\epsilon_2|}{R(\gamma)} = o_p(1), \tag{B.23}$$

$$\operatorname*{Sup}_{\gamma \in \mathcal{H}} \frac{|\sigma_{(2)}\ \mu' P^2(\gamma)\epsilon_2|}{R(\gamma)} = o_p(1), \tag{B.24}$$

$$\operatorname*{Sup}_{\gamma \in \mathcal{H}} \frac{|\mathrm{tr}(P(\gamma))|}{R(\gamma)} = o_p(1), \tag{B.25}$$

$$\operatorname*{Sup}_{\gamma \in \mathcal{H}} \frac{|\mathrm{tr}(P^2(\gamma))|}{R(\gamma)} = o_p(1). \tag{B.26}$$

In order to prove the conditions in equations (B.19)-(B.26), define $\lambda_{\max}(B)$ to be the largest eigenvalue of matrix B. Thus,

$$\operatorname*{Sup}_{\gamma \in H} \lambda_{\max}\big(P(\gamma)\big) = \operatorname*{Sup}_{\gamma \in H} \lambda_{\max}\Big(X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}X_2'\Big) \leq \lambda_{\max}\Big(X_2(X_2'X_2)^{-1}X_2'\Big) = 1. \tag{B.27}$$

Besides, it is easy to see that $R(\gamma) = O_p(T)$, $\mu'\mu = O(T)$, and $X_2'\sigma_{(2)}\epsilon_2 = O_p(\sqrt{T})$.

Proof of equation (B.19):

$$\begin{aligned}
\sigma_{(2)}\epsilon_2'\Phi &= \gamma\ \sigma_{(2)}\epsilon_2'X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}X_1'Y_1 \\
&\leq \gamma\ \sigma_{(2)}\epsilon_2'X_2\big(\gamma X_1'X_1\big)^{-1}X_1'Y_1 \\
&= \sigma_{(2)}\epsilon_2'X_2\big(X_1'X_1\big)^{-1}X_1'Y_1 \\
&= \sigma_{(2)}\epsilon_2'X_2\ \widehat{\beta}_{(1)} \\
&= O_p(\sqrt{T}).\ O_p(1) = O_p(\sqrt{T}).
\end{aligned} \tag{B.28}$$

Proof of equation (B.20):

$$\operatorname*{Sup}_{\gamma \in H} \sigma_{(2)}^2\ \epsilon_2'P(\gamma)\epsilon_2 \leq \epsilon_2'\sigma_{(2)}\ \big(X_2(X_2'X_2)^{-1}X_2\big)\ \sigma_{(2)}\epsilon_2 = O_p(1). \tag{B.29}$$

Proof of equation(B.21):

$$\begin{aligned}\Psi &= \gamma^2\ \sigma^2_{(1)}\epsilon_1' X_1\Big(\gamma X_1'X_1 + X_2'X_2\Big)^{-1} X_2'X_2\ \Big(\gamma X_1'X_1 + X_2'X_2\Big)^{-1} X_1'\epsilon_1 \\ &\leq \gamma^2\ \sigma^2_{(1)}\epsilon_1' X_1\Big(\gamma X_1'X_1\Big)^{-1} X_2'X_2\ \Big(\gamma X_1'X_1\Big)^{-1} X_1'\epsilon_1 \\ &= \epsilon_1'\sigma_{(1)}X_1(X_1'X_1)^{-1}X_2'X_2(X_1'X_1)^{-1}X_1'\sigma_1\epsilon_1 \\ &= \Big(\widehat{\beta}_{(1)} - \beta_{(1)}\Big)' X_2'X_2\ \Big(\widehat{\beta}_{(1)} - \beta_{(1)}\Big) = O_p(1).\end{aligned} \tag{B.30}$$

Proof of equation (B.22):

Lemma 31 *For any matrix Z and C, we have: $Z'CZ \leq Z'Z\ \lambda_{max}(C)$. Therefore, $Z'C^2Z \leq (Z'CZ)\ \lambda_{max}(C)$.* □

$$\operatorname*{Sup}_{\gamma\in H} \sigma^2_{(2)}\epsilon_2'P^2(\gamma)\epsilon_2 \leq \sigma^2_{(2)}\ \lambda_{\max}\big(P(\gamma)\big)\ \epsilon_2'P(\gamma)\epsilon_2 \leq \sigma_{(2)}\epsilon_2'X_2(X_2'X_2)^{-1}X_2'\sigma_{(2)}\epsilon_2 = O_p(1). \tag{B.31}$$

Proof of equation (B.23):

$$\begin{aligned}\sigma_{(2)}\Phi'P(\gamma)\epsilon_2 &= \gamma\ Y_1'X_1(\gamma X_1'X_1 + X_2'X_2)^{-1}\ X_2'X_2(\gamma X_1'X_1 + X_2'X_2)^{-1}X_2'\sigma_{(2)}\epsilon_2 \\ &\leq \gamma\ Y_1'X_1(X_2'X_2)^{-1}\ X_2'X_2\ \Big(\gamma X_1'X_1\Big)^{-1} X_2'\sigma_{(2)}\epsilon_2 \\ &= Y_1'X_1(X_1'X_1)^{-1}X_2'\sigma_{(2)}\epsilon_2 \\ &= Y_1'X_1(X_1'X_1)^{-1}X_2'\sigma_{(2)}\epsilon_2 \\ &= \widehat{\beta}_1\ X_2'\sigma_{(2)}\epsilon_2 \\ &= O_p(1)\ .\ O_p(\sqrt{T}) = O_p(\sqrt{T}).\end{aligned} \tag{B.32}$$

Proof of equation (B.24):

$$\begin{aligned}
\sigma_{(2)}\mu' P^2(\gamma)\epsilon_2 &= \mu' X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}X_2'X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}X_2'\sigma_{(2)}\epsilon_2 \\
&\leq \mu' X_2(X_2'X_2)^{-1}X_2'X_2(X_2'X_2)^{-1}X_2'\sigma_{(2)}\epsilon_2 \\
&= \beta_{(2)}'X_2'X_2(X_2'X_2)^{-1}X_2'\sigma_{(2)}\epsilon_2 \\
&= \beta_{(2)}'X_2'\sigma_{(2)}\epsilon_2 \\
&= O_p(1)\ .\ O_p(\sqrt{T}) = O_p(\sqrt{T}).
\end{aligned} \tag{B.33}$$

Proof of equation (B.25):

$$\operatorname*{Sup}_{\gamma\in H}\ \mathrm{tr}\big(P(\gamma)\big) \leq \mathrm{tr}\big(X_2'(X_2'X_2)^{-1}X_2\big) = k = O_p(1). \tag{B.34}$$

Proof of equation (B.26):

$$\begin{aligned}
\mathrm{tr}\Big(P^2(\gamma)\Big) &= \mathrm{tr}\Big[X_2\big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}X_2'X_2\ \big(\gamma X_1'X_1 + X_2'X_2\big)^{-1}X_2'\Big] \\
&\leq \mathrm{tr}\Big[X_2(X_2'X_2)^{-1}X_2'X_2\ (X_2'X_2)^{-1}X_2'\Big] = k = O_p(1).
\end{aligned} \tag{B.35}$$

Based on the above conditions, the proof of Theorem 16 thus follows. Therefore, the weight derives by CV is optimal. ■

Proof of Theorem 18: Define $b_1 \equiv \frac{T_1}{T}$ to be the proportion of the pre-break observations. The pre-break estimator can be written as

$$\begin{aligned}
\widehat{\beta}_{(1)} &= (X_1'X_1)^{-1}(X_1'Y_1) \\
&= (X_1'X_1)^{-1}(X_1'X_1\beta_{(1)} + X_1'X_1\beta_{(2)} - X_1'X_1\beta_{(2)} + X_1'\sigma_{(1)}\epsilon_1).
\end{aligned} \tag{B.36}$$

Thus, using the local alternative assumption, $\beta_{(1)} = \beta_{(2)} + \frac{\delta_1}{\sqrt{T}}$, we have

$$\begin{aligned}
\sqrt{T}\big(\widehat{\beta}_{(1)} - \beta_{(2)}\big) &= \Big(\frac{X_1'X_1}{T_1}\Big)^{-1}\Big(\frac{X_1'X_1}{T_1}\sqrt{T}(\beta_{(1)} - \beta_{(2)}) + \frac{X_1'\sigma_{(1)}\epsilon_1}{\sqrt{b_1}\ \sqrt{T_1}}\Big) \\
&= \delta_1 + \Big(\frac{X_1'X_1}{T_1}\Big)^{-1}\Big(\frac{X_1'\sigma_{(1)}\epsilon_1}{\sqrt{b_1}\ \sqrt{T_1}}\Big) \\
&\xrightarrow{d} \mathrm{N}\big(\delta_1, V_{(1)}\big).
\end{aligned} \tag{B.37}$$

Besides, the distribution of the post-break estimator is

$$\begin{aligned}\sqrt{T}\big(\widehat{\beta}_{(2)} - \beta_{(2)}\big) &= \Big(\frac{X_2'X_2}{T-T_1}\Big)^{-1}\Big(\frac{X_2'\sigma_{(2)}\epsilon_2}{\sqrt{1-b_1}\,\sqrt{T-T_1}}\Big) \\ &\xrightarrow{d} \mathrm{N}\big(0, V_{(2)}\big).\end{aligned} \tag{B.38}$$

Using (B.37) and (B.38) we can write the joint distribution of the pre-break estimator and the post-break estimator, and this completes the proof of Theorem 18. ■

Proof of Theorem 19: The risk of the semi-parametric estimator is

$$\begin{aligned}\rho(\widehat{\beta}_\gamma, \mathbb{W}) &= \mathbb{E}\Big[T(\widehat{\beta}_\gamma - \beta_{(2)})'\,\mathbb{W}\,(\widehat{\beta}_\gamma - \beta_{(2)})\Big] \\ &= T\,\mathbb{E}\Big[\big(\widehat{\beta}_{(2)} - \beta_{(2)}\big) - \Delta\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{(1)}\big)\Big]'\,\mathbb{W}\Big[\big(\widehat{\beta}_{(2)} - \beta_{(2)}\big) - \Delta\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{(1)}\big)\Big] \\ &= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + T\,\mathbb{E}\Big[\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{(1)}\big)'\Delta'\,\mathbb{W}\,\Delta\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{(1)}\big)\Big] \\ &- 2T\,\mathbb{E}\Big[\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{(1)}\big)'\Delta'\,\mathbb{W}\,\big(\widehat{\beta}_{(2)} - \beta_{(2)}\big)\Big] \\ &= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \mathbb{E}\Big[Z'V^{1/2}G\Delta'\,\mathbb{W}\,\Delta G'V^{1/2}Z\Big] - 2\,\mathbb{E}\Big[Z'V^{1/2}G\Delta'\,\mathbb{W}\,G_2'V^{1/2}Z\Big] \\ &= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \mathrm{tr}\Big[\mathbb{E}(ZZ')S_1\Big] - 2\mathrm{tr}\Big[\mathbb{E}(ZZ')S_2\Big] \\ &= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \mathrm{tr}\Big[\big(I_{2k} + \theta\theta'\big)S_1\Big] - 2\mathrm{tr}\Big[\big(I_{2k} + \theta\theta'\big)S_2\Big] \\ &= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \mathrm{tr}(S_1) + \theta' S_1 \theta - 2\mathrm{tr}(S_2)\end{aligned} \tag{B.39}$$

where $S_1 \equiv V^{1/2}G\Delta'\,\mathbb{W}\,\Delta G'V^{1/2}$, $S_2 \equiv V^{1/2}G\Delta'\,\mathbb{W}\,G_2'V^{1/2}$, $\theta' S_2 \theta = 0$. Note that, $I_k > \Delta$ since

$$\begin{aligned}I_k - \Delta &= \Big(X_1'X_1 + \frac{1}{\gamma}X_2'X_2\Big)^{-1}\Big(X_1'X_1 + \frac{1}{\gamma}X_2'X_2\Big) - \Big(X_1'X_1 + \frac{1}{\gamma}X_2'X_2\Big)^{-1}\Big(X_1'X_1\Big) \\ &= \Big(X_1'X_1 + \frac{1}{\gamma}X_2'X_2\Big)^{-1}\Big(\frac{1}{\gamma}X_2'X_2\Big) > 0.\end{aligned} \tag{B.40}$$

Also, using (B.40), we can see that $S_2 > S_1$ because

$$\begin{aligned} S_2 &= V^{1/2} G \Delta' \, \mathbb{W} \, I_k G_2' V^{1/2} \\ &> V^{1/2} G \Delta' \, \mathbb{W} \, \Delta G_2' V^{1/2} \\ &> V^{1/2} G \Delta' \, \mathbb{W} \, \Delta G' V^{1/2} = S_1. \end{aligned} \tag{B.41}$$

Lemma 32 *For any $T \times T$ matrix $A > 0$,*

$$A < tr(A) I_T.$$

□

Using Lemma 32 and substituting (B.40) and (B.41) in (B.39)

$$\begin{aligned} \rho(\widehat{\beta}_\gamma, \mathbb{W}) &< \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \mathrm{tr}(S_1) + \mathrm{tr}(S_1)\theta'\theta - 2\mathrm{tr}(S_1) \\ &< \rho(\widehat{\beta}_{(2)}, \mathbb{W}) - \mathrm{tr}(S_1)\big(1 - \theta'\theta\big). \end{aligned} \tag{B.42}$$

Since $\mathrm{tr}(S_1) > 0$, in order to $\rho(\widehat{\beta}_\gamma, \mathbb{W}) < \rho(\widehat{\beta}_{(2)}, \mathbb{W})$, we need $\theta'\theta < 1$ which is equivalent to $\delta_1' V_{(1)}^{-1} \delta_1 < 1$. This completes the proof of Theorem 19. ■

Appendix C

Appendix for Chapter 4

Proof of Theorem 21: The asymptotic distribution of the restricted estimator is

$$\begin{aligned}\sqrt{T}\big(\widehat{\boldsymbol{b}}_r-\boldsymbol{b}\big) &= \sqrt{T}\big(\widehat{\boldsymbol{b}}_{ur}-\boldsymbol{b}\big)-\big(X'\Omega^{-1}X\big)^{-1}R'\left[R\,\big(X'\Omega^{-1}X\big)^{-1}R'\right]^{-1}\sqrt{T}\big(R\widehat{\boldsymbol{b}}_{ur}-R\boldsymbol{b}+R\boldsymbol{b}\big)\\ &= \sqrt{T}\big(\widehat{\boldsymbol{b}}_{ur}-\boldsymbol{b}\big)-V_{ur}R'\left[R\,V_{ur}R'\right]^{-1}R\sqrt{T}\Big(\widehat{\boldsymbol{b}}_{ur}-\boldsymbol{b}+\frac{\mathbf{h}}{\sqrt{T}}\Big)\\ &\overset{d}{\to} Z-V_{ur}R'\Big[R\,V_{ur}R'\Big]^{-1}R(Z+\mathbf{h}),\end{aligned} \tag{C.1}$$

where $R\boldsymbol{b}_0=0$ under the null hypothesis, and therefore $R\boldsymbol{b}=R\mathbf{h}$. Besides, the difference between the restricted and unrestricted estimators is

$$\begin{aligned}\sqrt{T}\big(\widehat{\boldsymbol{b}}_{ur}-\widehat{\boldsymbol{b}}_r\big) &= V_{ur}\;LR\sqrt{T}\Big(\widehat{\boldsymbol{b}}_{ur}-\boldsymbol{b}+\frac{\mathbf{h}}{\sqrt{T}}\Big)\\ &\overset{d}{\to} V_{ur}\;LR(Z+\mathbf{h}).\end{aligned} \tag{C.2}$$

Using (C.2), distribution of D_T in (4.14) is straightforward. Also, note that we can rewrite the combined estimator as

$$\begin{aligned}\sqrt{T}\big(\widehat{\boldsymbol{b}}_w-\boldsymbol{b}\big) &= \sqrt{T}\big(\widehat{\boldsymbol{b}}_{ur}-\boldsymbol{b}\big)-w\sqrt{T}\big(\widehat{\boldsymbol{b}}_{ur}-\widehat{\boldsymbol{b}}_r\big)\\ &\overset{d}{\to} Z-\alpha(Z)V_{ur}\;LR(Z+\mathbf{h}).\end{aligned} \tag{C.3}$$

This completes the proof of Theorem 21. ■

Proof of Theorem 22:

Lemma 33 *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. □

The asymptotic risk of the Stein-type shrinkage estimator can be calculated as

$$\begin{aligned}
\rho(\widehat{\boldsymbol{b}}_w, \mathbb{W}) &= T\,\mathbb{E}\left[\left(\widehat{\boldsymbol{b}}_w - \boldsymbol{b}\right)' \mathbb{W}\ \left(\widehat{\boldsymbol{b}}_w - \boldsymbol{b}\right)\right] \\
&= T\,\mathbb{E}\left[\left(\widehat{\boldsymbol{b}}_{ur} - \boldsymbol{b}\right) - \alpha\left(\widehat{\boldsymbol{b}}_{ur} - \widehat{\boldsymbol{b}}_r\right)\right]' \mathbb{W}\left[\left(\widehat{\boldsymbol{b}}_{ur} - \boldsymbol{b}\right) - \alpha\left(\widehat{\boldsymbol{b}}_{ur} - \widehat{\boldsymbol{b}}_r\right)\right] \\
&= \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) + \tau^2\,\mathbb{E}\left[\frac{1}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\right] - 2\tau\,\mathbb{E}\left[\frac{(Z+\mathbf{h})'R'L'V_{ur}\,\mathbb{W}\,Z}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\right] \\
&= \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) + \tau^2\,\mathbb{E}\left[\frac{1}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\right] - 2\tau\,\mathbb{E}\left[\eta(Z+\mathbf{h})'R'L'V_{ur}\,\mathbb{W}\,Z\right],
\end{aligned} \tag{C.4}$$

where $\eta(x) = \left(\frac{1}{x'Bx}\right)x$. Using the Stein's lemma, we can simplify the last term in (C.4) as

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

$$
\begin{aligned}
&= \mathbb{E}\left[\frac{\operatorname{tr}(A)}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\right] - 2\,\mathbb{E}\operatorname{tr}\left[\frac{(Z+\mathbf{h})'B_1'AB_1\ (Z+\mathbf{h})}{\big((Z+\mathbf{h})'B(Z+\mathbf{h})\big)^2}\right] \\
&\geq \mathbb{E}\left[\frac{\operatorname{tr}(A)}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\right] - 2\,\mathbb{E}\left[\frac{\lambda_{\max}(A)}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\right],
\end{aligned}
\tag{C.5}
$$

where $A \equiv \mathbb{W}^{1/2}\, V_{ur}R'L'V_{ur}\, \mathbb{W}^{1/2}$, $B_1 \equiv \mathbb{W}^{1/2}\, V_{ur}LR$, $R'L'V_{ur}\, \mathbb{W}\, V_{ur}B = B_1'AB_1$, and $B_1'B_1 = B$. Note that $\frac{\partial}{\partial x}\eta(x)' = (\frac{1}{x'Bx})I - \frac{2}{(x'Bx)^2}Bxx'$. Also, we use Lemma 33 to get the last inequality. Plugging (C.5) into (C.4) produces

$$
\begin{aligned}
\rho(\widehat{\boldsymbol{b}}_w, \mathbb{W}) &< \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) + \tau^2\,\mathbb{E}\Big[\frac{1}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\Big] - 2\tau\Big[\frac{\operatorname{tr}(A) - 2\lambda_{\max}(A)}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\Big] \\
&= \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) - \tau\,\mathbb{E}\Big[\frac{2\big(\operatorname{tr}(A) - 2\lambda_{\max}(A)\big) - \tau}{(Z+\mathbf{h})'B(Z+\mathbf{h})}\Big] \\
&< \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) - \tau\Big[\frac{2\big(\operatorname{tr}(A) - 2\lambda_{\max}(A)\big) - \tau}{\mathbb{E}(Z+\mathbf{h})'B(Z+\mathbf{h})}\Big] \\
&= \rho(\widehat{\boldsymbol{b}}_{ur}, \mathbb{W}) - \tau\Big[\frac{2\big(\operatorname{tr}(A) - 2\lambda_{\max}(A)\big) - \tau}{(c+1)\operatorname{tr}(A)}\Big],
\end{aligned}
\tag{C.6}
$$

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

$$
\begin{aligned}
\mathbb{E}(Z+\mathbf{h})'B(Z+\mathbf{h}) &= \mathbf{h}'B\mathbf{h} + \mathbb{E}(Z'BZ) \\
&= \mathbf{h}'B\mathbf{h} + \operatorname{tr}\big(B\,\mathbb{E}(ZZ')\big) \\
&= \mathbf{h}'B\mathbf{h} + \operatorname{tr}(BV_{ur}) \\
&= \mathbf{h}'B\mathbf{h} + \operatorname{tr}(A) \\
&\leq (c+1)\operatorname{tr}(A),
\end{aligned}
\tag{C.7}
$$

in which for any $0 < c < \infty$, we define a ball such that $H(c) = \{\mathbf{h} : \mathbf{h}'B\mathbf{h} \leq \operatorname{tr}(A)\ c\}$. This completes the proof of Theorem 22. ■