Efficient Combined Estimation under Structural Breaks*

Tae-Hwy Lee† Shahnaz Parsaeian‡ Aman Ullah§

December 30, 2020

Abstract

Hashem Pesaran has made many seminal contributions, among others, in the time series econometrics estimation and forecasting under structural break, see Pesaran and Timmermann (2005, 2007), Pesaran et al. (2006), and Pesaran et al. (2013). In our paper here we focus on the estimation of regression parameters under multiple structural breaks with heteroskedasticity across regimes. We propose a combined 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 operational optimal combination weight is between zero and one. The analytical finite sample risk is derived, and it is shown that the risk of the proposed combined estimator is lower than that of the unrestricted estimator under any break size and break points. Further, we show that the combined estimator outperforms over the unrestricted estimator in terms of the mean squared forecast errors. Properties of the estimator is also demonstrated in simulations. Finally, empirical illustrations for parameter estimators and forecasts are presented through macroeconomic and financial data sets.

Keywords: Structural breaks, Combined estimator

*The authors are thankful to the editors and an anonymous referee for helpful and constructive comments on the paper.

†Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu

‡Department of Economics, University of Kansas, Lawrence, KS 66045. E-mail: shahnaz.parsaeian@ku.edu

§Department of Economics, University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu

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 for 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) and Bai (1997a) 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 and Corradi (2003), Qu and Perron (2007), and 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 work on structural breaks and estimation, see Pesaran and Timmermann (2005, 2007), Pesaran et al. (2006), Pesaran et al. (2013), and a comprehensive survey by Casini and Perron (2018) inter alia.

The current paper 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 paper 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, see Bai and Perron (1998) and Bai and Perron (2003) which is a consistent global minimizers of the sum of squared residuals.

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 in each regime. 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, in this paper 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, in which the coefficients across different regimes are restricted to be the same as if there is no structural break, uses all the observations in the sample, $t = \{1, \dots, T\}$, to estimate the coefficients. So, it is under the restriction that there is no break in the model. The advantage of imposing this restriction is that sometimes the break size is small, so precisely detecting the break point is difficult or not possible. Even under detectable break points, ignoring that break point and estimating the coefficients by using all observations, gives a better estimate.

In this paper, we focus on the estimation of regression parameters under multiple structural breaks, when errors across regimes are heteroskedastic. We propose a minimal mean square error estimator of regression parameters based on combining the restricted estimator which ignores that there is a break in the parameters, with the unrestricted estimator which acknowledges the break. An operational optimal combination weight is introduced. We derive the condition under which the proposed combined estimator outperforms the unrestricted estimator, in the sense of minimizing the risk. The combination is convex with the weight between zero and one. We derive the finite sample properties for the combined estimator, and show that it has a lower risk than the unrestricted estimator which is the common practice in estimating the parameters under structural breaks. The analytical finite sample results of the combined estimator are derived based on the large-sample expansion proposed by Nagar (1959). We also show that the proposed combined estimator outperforms the unrestricted estimator in terms of the mean squared forecast errors.

Monte Carlo experiments to evaluate the performance of the proposed combined estimator are carried out. The results confirm the theoretically expected improvements in the combined estimator compared to the unrestricted estimator under any break size and break points. As an empirical example, with a large macroeconomic and financial time series, we forecast the US output growth for $1, 6$, and 12 month forecast horizons and show the outperformance of using our proposed combined estimator relative to the unrestricted estimator.

The outline of the paper is as follows. Section 2 sets up the model under multiple structural breaks model, and introduces a minimal mean squared error combined estimator. Section 3 derives its finite sample properties analytically. Monte Carlo experiments are presented in Section 4 while Section 5 presents an empirical study. Finally Section 6 concludes. All proofs are relegated to Appendix.

2 The Structural Breaks Model and Combined Estimator

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

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

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

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

where $Y = \big(Y_{(1)}', \ldots, Y_{(m+1)}'\big)'$ is a $T \times 1$ vector of dependent variable in which $Y_{(i)} = \big(y_{T_{i-1}+1}, \ldots, y_{T_i}\big)'$ is an $l_i \times 1$ vector, $X = \text{diag}\big(X_{(1)}, \ldots, X_{(m+1)}\big)$ is a $T \times (m+1)k$ block diagonal matrix of regressors in which $X_{(i)} = \big(x_{T_{i-1}+1}, \ldots, x_{T_i}\big)'$ is an $l_i \times k$ matrix, $i = \{1, \ldots, m+1\}$, $l_i = T_i - T_{i-1}$ such that $\sum l_i = T$, and we use the convention that $T_0 = 0$, and $T_{m+1} = T$. Also, $\beta = \big(\beta_{(1)}', \ldots, \beta_{(m+1)}'\big)'$ is an $(m+1)k \times 1$ vector of coefficients, $\epsilon = \big(\epsilon_{(1)}', \ldots, \epsilon_{(m+1)}'\big)'$ is a $T \times 1$ vector of error terms, with $\epsilon_{(i)} = \sigma_{(i)}\big(u_{T_{i-1}+1}, \ldots, u_{T_i}\big)'$, such that

$$\epsilon = \begin{cases} \sigma_{(1)}(u_1, \ldots, u_{T_1})' & \text{for} \quad 1\ < t \leq T_1 \\ \sigma_{(2)}(u_{T_{1+1}}, \ldots, u_{T_2})' & \text{for} \quad T_1 < t \leq T_2 \\ \vdots & \qquad\qquad \vdots \\ \sigma_{(m+1)}(u_{T_{m+1}}, \ldots, u_T)' & \text{for} \quad T_m < t < T. \end{cases} \tag{3}$$

We make the following assumptions:

Assumption 1. The $T \times 1$ vector of errors, ϵ, has a zero conditional mean

$$\mathbb{E}(\epsilon | X) = 0, \tag{4}$$

and $\mathbb{E}\,\epsilon_{(i)}\epsilon_{(j)}' = \sigma_{(i)}^2 I_{l_i}$ for $i = j$ and $\mathbf{0}$ for $i \neq j$ such that the conditional variance-covariance is

$$V(\epsilon|X) = \mathbb{E}(\epsilon\epsilon'|X) = \Omega = \text{diag}\big(\sigma_{(1)}^2 I_{l_1}, \ldots, \sigma_{(m+1)}^2 I_{l_{m+1}}\big), \tag{5}$$

where I_{l_i} is an $l_i \times l_i$ identity matrix and $\mathbf{0}$ is an $l_i \times l_i$ matrix of zeros.

Assumption 2. The errors are normally distributed with mean zero and variance-covariance matrix Ω.

We note that Assumption 1 implies uncorrelated errors within and across regimes, but they are heteroskedastic across regimes.

In this paper, we introduce a combined estimator which has a lower risk than the restricted estimator and unrestricted estimator. In this section, we introduce the structural break model with restrictions 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 $p \times (m+1)k$ matrix with rank p, which shows the number of restrictions, as

$$R = \begin{bmatrix} -I_k & I_k & 0 & & 0 & 0 & 0 \\ 0 & -I_k & I_k & & 0 & 0 & 0 \\ & & & \vdots \; \vdots & & & \\ 0 & 0 & 0 & & -I_k & I_k & 0 \\ 0 & 0 & 0 & & 0 & -I_k & I_k \end{bmatrix}. \tag{6}$$

The matrix R considers the difference between coefficients, $R\beta = \left(\beta'_{(2)} - \beta'_{(1)}, \beta'_{(3)} - \beta'_{(2)}, \ldots, \beta'_{(m+1)} - \beta'_{(m)}\right)'$. Under the alternative hypothesis, $R\beta \neq \mathbf{0}$.

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

$$\widehat{\beta}_\gamma = (1-\gamma)\widehat{\beta}_{ur} + \gamma\widehat{\beta}_r, \tag{7}$$

where $\widehat{\beta}_{ur}$ and $\widehat{\beta}_r$ are the infeasible unrestricted estimator and the infeasible restricted estimator, respectively. Using the generalized least squares (GLS) in (2) we have

$$\begin{aligned} \widehat{\beta}_{ur} &= \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \\ &= \beta + \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\epsilon. \end{aligned} \tag{8}$$

Further applying GLS in (2) under the restriction $R\beta = \mathbf{0}$, we have

$$\begin{aligned} \widehat{\beta}_r &= \widehat{\beta}_{ur} - \underbrace{\left(X'\Omega^{-1}X\right)^{-1} R' \left[R\left(X'\Omega^{-1}X\right)^{-1} R'\right]^{-1} R}\widehat{\beta}_{ur} \\ &= \widehat{\beta}_{ur} - L\widehat{\beta}_{ur}, \end{aligned} \tag{9}$$

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 an $(m+1)k \times (m+1)k$ matrix.

The idea behind the combined estimator in (7) is that when the difference between the restricted and unrestricted estimator is small, the combined estimator gives more weight to the restricted estimator which is an efficient estimator under $R\beta = \mathbf{0}$. However, when the difference is large, the restricted estimator is biased under $R\beta \neq \mathbf{0}$, and the combined estimator gives more weight to the unrestricted estimator. Thus, the proposed combined estimator in (7) is a Stein-like shrinkage estimator that incorporates the trade-off between the bias and variance of the two estimators. All these are also reflected in the optimal combination weight in (12), that we derive next.

As γ in (7) is unknown, the first step is to find its optimal value. We derive the exact risk for the combined estimator, and minimize the risk to find the optimal value of the weight. The risk of the combined estimator with a positive definite weight matrix W is

$$\begin{aligned} Risk(\widehat{\beta}_{\gamma}, W) &= \mathbb{E}\left[(\widehat{\beta}_{\gamma} - \beta)' W (\widehat{\beta}_{\gamma} - \beta)\right] \\ &= Risk(\widehat{\beta}_{ur}, W) + \gamma^2 \Big[\beta' L' W L \beta + \operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W L\big)\Big] - 2\gamma \operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big). \end{aligned} \tag{10}$$

By minimizing the risk with respect to γ in (10), the optimal value of the weight denoted by γ^* is

$$\gamma^* = \frac{\operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}{\beta' L' W L \beta + \operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}, \tag{11}$$

which by plugging the unbiased estimator of its denominator we have

$$\begin{aligned} \gamma^* &= \frac{\operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}{\widehat{\beta}'_{ur} L' W L \widehat{\beta}_{ur}} \\ &= \frac{\operatorname{tr}\big((X'\Omega^{-1}X)^{-1} L' W\big)}{(\widehat{\beta}_{ur} - \widehat{\beta}_{r})' W (\widehat{\beta}_{ur} - \widehat{\beta}_{r})}. \end{aligned} \tag{12}$$

See Appendix A.1 for the proof of (12).[1] Note that the optimal weight depends on the unknown value Ω which we will replace with its estimate.

Define notation $\bar{\bar{\beta}}$ as a feasible estimator of β. The feasible unrestricted estimator from (8) is

$$\bar{\bar{\beta}}_{ur} = \beta + \big(X'\widehat{\Omega}^{-1}X\big)^{-1} X'\widehat{\Omega}^{-1}\epsilon, \tag{13}$$

where $\beta = \big(\beta'_{(1)}, \ldots, \beta'_{(m+1)}\big)'$ is $(m+1)k \times 1$, and $\widehat{\Omega} = \operatorname{diag}\big(\widehat{\sigma}^2_{(1)}\ I_{l_1}, \ldots, \widehat{\sigma}^2_{(m+1)}\ I_{l_{m+1}}\big) = \operatorname{diag}\big(S_{(1)}\ I_{l_1}, \ldots, S_{(m+1)}\ I_{l_{m+1}}\big)$ where $S_{(i)}$ is a consistent estimates of the $\sigma^2_{(i)}$ in which $S_{(i)} = \frac{\epsilon'_{(i)} M_{(i)} \epsilon_{(i)}}{l_i - k}$ and $M_{(i)} = I_{l_i} - X_{(i)}\big(X'_{(i)}X_{(i)}\big)^{-1}X'_{(i)}$ with $i = \{1, \ldots, m+1\}$. See Appendix A.2 for details. We note

[1] A different combined estimator may be obtained from a Bayesian approach, as considered by Pesaran et al. (2006) and Maheu and Gordon (2008). It is well known that the posterior mean (a Bayesian shrinkage estimator) can be shown to combine a prior mean with the likelihood function based sample estimator for the parameters, resulting in our context a weighted average of the prior mean (the unrestricted estimator using regime specific observations) and the restricted estimator using pooled data of all regimes, with the weights given by the respective normalized precision matrices. Our combination weight γ^* in equation (12) might be related to and compared with those Bayesian shrinkage estimators. We leave this possibility for future work, and thank Allan Timmermann for bringing this to our attention.

that since the Ω matrix is diagonal, we can rewrite the unrestricted estimator as an ordinary least square estimator.

The feasible restricted estimator from (9) is

$$\begin{aligned}\bar{\bar{\beta}}_r &= \bar{\bar{\beta}}_{ur} - \left(X'\widehat{\Omega}^{-1}X\right)^{-1}R'\left[R\left(X'\widehat{\Omega}^{-1}X\right)^{-1}R'\right]^{-1}R\bar{\bar{\beta}}_{ur} \\ &= \bar{\bar{\beta}}_{ur} - \widehat{L}\bar{\bar{\beta}}_{ur},\end{aligned} \tag{14}$$

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, the feasible combination weight, $\widehat{\gamma}^*$, can be considered as

$$\widehat{\gamma}^* = \frac{\operatorname{tr}\left((X'\widehat{\Omega}^{-1}X)^{-1}\widehat{L}'W\right)}{\bar{\bar{\beta}}'_{ur}\widehat{L}'W\widehat{L}\bar{\bar{\beta}}_{ur}}. \tag{15}$$

Further, the feasible combined estimator from (7) is

$$\bar{\bar{\beta}}_\gamma = (1-\widehat{\gamma}^*)\bar{\bar{\beta}}_{ur} + \widehat{\gamma}^*\bar{\bar{\beta}}_r. \tag{16}$$

We note that for the restricted estimator, $R\beta = \mathbf{0}$. For example, we can impose a restriction that all coefficients across regimes are equal, a restriction that the coefficients in some specific regimes are equal to each other, or any other restrictions. Restricting some of the coefficients to be identical across some regimes converts the model to the partial structural change model which is useful since it allows for a broad range of practical interest.

3 Finite Sample Properties: Approximate Bias, MSE, and Risk

In this section, we use the large sample approximation method proposed by Nagar (1959) to analyze the bias, the mean squared error (MSE) matrix, and the risk for the proposed feasible combined estimator in (16).

Theorem 1: *Under Assumptions 1-2, the bias of the combined estimator, up to order $O(T^{-1})$, is given by*

$$Bias\left(\bar{\bar{\beta}}_\gamma\right) = \mathbb{E}\left(\bar{\bar{\beta}}_\gamma - \beta\right) = -\frac{\operatorname{tr}(Q)}{\phi}L\beta, \tag{17}$$

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, and the second order moment matrix of the combined estimator, up to order $O(T^{-2})$, is

$$
\begin{aligned}
MSE\big(\bar{\bar{\beta}}_\gamma\big) &= \mathbb{E}\left[\big(\bar{\bar{\beta}}_\gamma-\beta\big)\big(\bar{\bar{\beta}}_\gamma-\beta\big)'\right] \\
&= MSE\big(\bar{\bar{\beta}}_{ur}\big)+\frac{1}{\phi^2}\ L\beta\beta'L'\Big(\mathrm{tr}(Q)\Big)^2-\frac{2\ \mathrm{tr}(Q)}{\phi}L(X'\Omega^{-1}X)^{-1} \\
&\quad+\frac{2\ \mathrm{tr}(Q)}{\phi^2}\ L\beta\beta'L'WL(X'\Omega^{-1}X)^{-1}+\frac{2\ \mathrm{tr}(Q)}{\phi^2}\ (X'\Omega^{-1}X)^{-1}L'WL\beta\beta'L',
\end{aligned}
\tag{18}
$$

where $MSE\big(\bar{\bar{\beta}}_{ur}\big)=(X'\Omega^{-1}X)^{-1}$. Further, for a weight matrix W of order $O(T)$, the risk of the combined estimator, up to order $O(T^{-1})$, is given by

$$
\begin{aligned}
Risk\big(\bar{\bar{\beta}}_\gamma,W\big) &= \mathbb{E}\left[\big(\bar{\bar{\beta}}_\gamma-\beta\big)'W\big(\bar{\bar{\beta}}_\gamma-\beta\big)\right] \\
&= Risk\big(\bar{\bar{\beta}}_{ur},W\big)-\frac{\Big(\mathrm{tr}(Q)\Big)^2}{\phi^2}\left\{\phi-\frac{4\big(\beta'L'WL(X'\Omega^{-1}X)^{-1}WL\beta\big)}{\mathrm{tr}(Q)}\right\},
\end{aligned}
\tag{19}
$$

where $Risk\big(\bar{\bar{\beta}}_{ur},W\big)=\mathrm{tr}\big((X'\Omega^{-1}X)^{-1}W\big)$. ■

See Appendix A.3 for the proof of Theorem 1. We note that $Q=W^{1/2}L(X'\Omega^{-1}X)^{-1}W^{1/2}=W^{1/2}L(X'\Omega^{-1}X)^{-1}L'W^{1/2}$ is a symmetric matrix.

Corollary 1.1: The risk of the combined estimator, up to order $O(T^{-1})$, is less than the risk of the unrestricted estimator as long as

$$
d\equiv\frac{\mathrm{tr}(Q)}{\lambda_{\max}(Q)}>4,
\tag{20}
$$

where $\lambda_{\max}(Q)$ represents the maximum eigenvalues of Q. □

The proof of Corollary 1.1 is given in the Appendix A.4.

Corollary 1.2: The bias, up to order $O(T^{-1})$, and finite sample risk, up to order $O(T^{-1})$, for the combined estimator with $W=X'\Omega^{-1}X$ are

$$
Bias\big(\bar{\bar{\beta}}_\gamma\big)=-\frac{p}{\phi}L\beta,
\tag{21}
$$

$$
Risk\big(\bar{\bar{\beta}}_\gamma,W\big)=Risk\big(\bar{\bar{\beta}}_{ur},W\big)-\frac{p^2}{\phi}\left\{1-\frac{4}{p}\right\},
\tag{22}
$$

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. □

The proof of Corollary 1.2 follows immediately after substituting $W = X'\Omega^{-1}X$ and noting $\mathrm{tr}(Q) = p$. In a special case, when there is only one structural break point so that $m = 1$, and $p = 2k$, the dominance condition in Corollary 1.2 becomes $k > 2$ which is the Stein's well-known dominance condition of a shrinkage estimator. Further, for $m \geq 4$ points of structural breaks, our shrinkage combined estimator dominates the unrestricted estimator for any number of regressors.

We note that when we consider $V(\epsilon|X) = \sigma^2\Omega$ where Ω is as given in (5), then we get the following Corollary.

Corollary 1.3: The large sample approximation bias, up to $O(T^{-1})$, in (17) and risk, up to $O(T^{-1})$, in (19) for the combined estimator, are written as

$$Bias\big(\bar{\bar{\beta}}_{\gamma}\big) = -\frac{\sigma^2 \mathrm{tr}(Q)}{\phi} L\beta \tag{23}$$

and

$$Risk\big(\bar{\bar{\beta}}_{\gamma}, W\big) = Risk\big(\bar{\bar{\beta}}_{ur}, W\big) - \frac{\sigma^4\ \big[\mathrm{tr}(Q)\big]^2}{\phi^2}\left\{\phi - \frac{4\beta' L' W L (X'\Omega^{-1}X)^{-1} W L\beta}{\mathrm{tr}(Q)}\right\}, \tag{24}$$

where $Risk\big(\bar{\bar{\beta}}_{ur}, W\big) = \sigma^2 \mathrm{tr}\big((X'\Omega^{-1}X)^{-1}W\big)$. Therefore, the risk of the combined estimator is less than the unrestricted estimator as long as $d > 4$ in (20). $\square$

Remark 1: The result in (23) and (24) immediately follow after substituting Ω as $\sigma^2\Omega$ in (17) and (19), respectively, and they are under the assumption that $W = O(T)$. In a special case, when we substitute $W = X'\Omega^{-1}X/\sigma^2$ in (23) and (24), respectively, we get the bias, up to order $O(T^{-1})$ and risk, up to order $O(T^{-1})$ as

$$Bias\big(\bar{\bar{\beta}}_{\gamma}\big) = -\frac{\sigma^2 p}{\phi} L\beta, \tag{25}$$

$$Risk\big(\bar{\bar{\beta}}_{\gamma}, W\big) = Risk\big(\bar{\bar{\beta}}_{ur}, W\big) - \frac{\sigma^2 p^2}{\phi}\left\{1 - \frac{4}{p}\right\}, \tag{26}$$

where $\phi = \beta' L' X' \Omega^{-1} X L \beta$. Thus the risk of the combined estimator is less than the risk of the unrestricted estimator if $p > 4$. $\square$

The above results in Corollary 1.3 and in its remark are for large-T and fixed σ. However, if T is fixed and σ is small in Kadane (1971) small sigma sense, $W = \frac{X'X}{\sigma^2} = O(\frac{1}{\sigma^2})$, and $V(\epsilon|X) = \sigma^2 I$, then we get the following Corollary.

Corollary 1.4: The small sigma bias and finite sample risk for the combined estimator are

$$Bias\big(\bar{\bar{\beta}}_\gamma\big) = -\frac{\sigma^2 p}{\phi_s} L\beta, \tag{27}$$

$$Risk\big(\bar{\bar{\beta}}_\gamma, W\big) = Risk\big(\bar{\bar{\beta}}_{ur}, W\big) - \frac{\sigma^2 p^2}{\phi_s}\left\{\frac{n-2}{n} - \frac{4}{p}\right\} \tag{28}$$

where $\phi_s = \beta' L' X' X L \beta, n = T - (m+1)k$, and the risk of the combined estimator is less than the risk of the unrestricted estimator if $p > \frac{4n}{n-2}$. □

The proof of Corollary 1.4 is given in Appendix A.4.

Remark 2: We note from the result of Corollary 1.4 that when T is moderately large so that $n/(n-2) \simeq 1$, then the combined estimator has smaller risk compared to the unrestricted estimator if $p > 4$. That is, the small sigma results on bias, risk, and improvement condition in Corollary 1.4 are the same as those obtained by the large sample results in Remark 1. Further, the small sigma condition $p > \frac{4n}{n-2} = 4 + \frac{8}{n-2}$ implies the condition $p > 4$ so long as $n > 10$ since number of restrictions p is an integer. □

3.1 Forecasting under Structural Break

Besides estimating the parameters of the model, we can use our combined estimator for deriving forecast under the structural breaks model, and compute the Mean Squared Forecast Error (MSFE) for the estimators. One solution for forecasting under breaks is to use the observations of the latest regime and estimate the coefficient based on that. But one can improve the performance of the forecast in the sense that has a lower MSFE by using observations out of the latest regime, see Pesaran and Timmermann (2005, 2007), and Pesaran et al. (2013).

For simplicity, assume that we have only one break, $m = 1$. In order to use our introduced combined estimator for forecasting purpose, we define a selection matrix $G = [\mathbf{0} \quad I_k]$. By multiplying it to the combined estimator, we have $G\widehat{\beta}_\gamma = (1-\gamma)G\widehat{\beta}_{ur} + \gamma G\widehat{\beta}_r$. Basically, we are focusing on the second elements of the $\beta' s$. By having the combined estimator, we can derive the out of sample forecast. Define $MSFE\big(\widehat{\beta}_\gamma\big) \equiv Risk\big(\widehat{\beta}_\gamma, x_{T+1}x'_{T+1}\big)$, where $W = x_{T+1}x'_{T+1}$ makes the risk the one step ahead MSFE. Corollary 1.5 shows the finite sample MSFE for our proposed combined estimator.

Corollary 1.5: The finite sample MSFE of the combined estimator is

$$\begin{aligned} MSFE(\bar{\bar{\beta}}_\gamma) &= \mathbb{E}\left[(\bar{\bar{\beta}}_\gamma - \beta)' G' \; x_{T+1} x'_{T+1} \; G \; (\bar{\bar{\beta}}_\gamma - \beta)\right] \\ &= MSFE(\bar{\bar{\beta}}_{ur}) - \frac{p^2}{\varpi}\left\{1 - \frac{4}{p}\right\}, \end{aligned} \tag{29}$$

where $\varpi \equiv \beta' L' G' x_{T+1} x'_{T+1} G L \beta$. Thus the MSFE of the combined estimator is smaller than that of unrestricted estimator when $p > 4$. □

The proof of Corollary 1.5 is in Appendix A.4.

4 Monte Carlo Simulation

This section provides Monte Carlo study of the proposed combined estimator. The goal is to compare the risk of the unrestricted estimator with the proposed combined estimator. It is shown that the risk of the proposed combined estimator is lower than the risk of the unrestricted estimator, regardless of the breakpoints and break sizes. We consider $T \in \{100, 200\}$, $m = 1$, and $k \in \{5, 8\}$. We try different values for a true breakpoint, $b_1 = \frac{T_1}{T} \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$. We also set $W = X'\Omega^{-1}X$.[2] We generate x_t and u_t such that $x_t \sim N(0, 1)$, and $u_t \sim N(0, 1)$. The data generating process uses the following model:

$$y_t = \begin{cases} x'_t \beta_{(1)} + \sigma_{(1)} u_t & \text{for} \quad 1 \ \leq t \leq T_1 \\ x'_t \beta_{(2)} + \sigma_{(2)} u_t & \text{for} \quad T_1 < t \leq T. \end{cases} \tag{30}$$

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 true break size in the coefficients which covers both a weak break ($\lambda = 0$) and a strong break ($\lambda = 1$). 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 Monte Carlo replications is 1,000.

To incorporate the uncertainty regarding the unknown parameters (b_1, λ, q), we estimate the break point, break size in the coefficient and the ratio of break in error variance. Tables 1-6 show the results for this experiment. The tables show the relative MSE with respect to the unrestricted estimator, i.e. $RMSE_\gamma = \frac{MSE(\bar{\widehat{\beta}}_\gamma)}{MSE(\bar{\widehat{\beta}}_{ur})}$. As it is clear from the results, the proposed

[2] The results with $W = I_{2k}$ are similar and are not reported to save space.

combined estimator with weight γ in all situations, regardless of the choice of q or b_1, is better than the unrestricted estimator. This confirms our theoretical results. Besides, as we increase the number of regressors, k, the ratio of Mean Squared Error ($RMSE$) gets smaller, especially for small break size (approximately less than 0.5).

We note that for a small break (small λ), the combination weight assign more weight to the restricted estimator which uses more observations. Therefore, we expect to see a larger value for $\widehat{\gamma}$. But for a large break (large λ), more weight is assigned to the unrestricted estimator. So we expect to see a smaller value for the $\widehat{\gamma}$. Figure 1 shows the distribution of $\widehat{\gamma}$ for different break size in the slope coefficient. The horizontal axis shows the value of $\widehat{\gamma}$ and the vertical axis shows the frequency for 1000 Monte Carlo. In this figure, $T = 100, T_1 = 80, k = 8$, and $\lambda \in \{0, 0.25, 0.5, 0.75, 1\}$. To incorporate the uncertainty associated with the break point and break size, we have estimated the breakpoint (T_1), the ratio of the break in the error variance (q) and the break in the slope coefficient (λ). As it is clear from this figure, when the break size increases, the value of $\widehat{\gamma}$ becomes smaller and the opposite is true for the small break size. Similar pattern can be seen for other specifications.

Remark 3: When there is no break in the true model, $\lambda = 0$ and $q = 1$, it may happen that the model mistakenly detects a break. As the unrestricted estimator uses observations within each regime separately, fewer observations are used to estimate the slope coefficients within each regime. Therefore, the combined estimator outperforms the unrestricted estimator and causes the RMSE not be equal to one. $\square$

5 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 sample we use for the analysis is for 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 eight factors can be interpreted as real activity/employment, inflation, term spreads, housing, interest rate variables, stock market variables which is seen in two of the factors, and output and inventories factors.

We evaluate the usefulness of the estimated factors by forecasting the U.S industrial production at the $1, 6$ and 12 month horizons.[3] The model that we use for forecasting take the form of

$$y_{t+h}^h = \beta_h' \ \widehat{f}_{t+h-1} + \alpha_h \ y_{t+h-1} + \varepsilon_{t+h}^h, \tag{31}$$

where y_{t+h}^h denotes output growth over the next h months, expressed at an annual rate, that is, $y_{t+h}^h = (1200/h) \ \ln(\mathrm{IP}_{t+h}/\mathrm{IP}_t)$. 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. We also compare our results with MSFE based on Pesaran et al. (2013) estimator who propose a weighted least square method for forecasting under structural breaks. 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 out-of-sample period which we recursively make one step ahead forecast. Each time that we expand the window, we apply the Schwarz's Bayesian Information Criteria (BIC) to choose the predictors out of the nine predictors, and identify break points by the sequential procedure introduced by Bai and Perron (1998) and Bai and Perron (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

[3] We assume that no structural breaks occur in the forecast period. For forecasting with structural breaks over the forecast period see Pesaran et al. (2006) which is a Bayesian procedure that allows for such a possibility.

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. Breakpoints are stable in the recursive estimation procedure. Mainly, the program detects one break. Also, there are some cases that no break is detected. For example, for $h = 1$ and $n_1 = 370$, we have 362 out-of-sample periods. For the first 242 expanding windows, no break is detected. For the remaining expanding windows (243-362) only one break (around the financial crisis) is detected. Figure 2 shows the histogram of the combination weight, $\widehat{\gamma}$. We did the same analysis for $n_1 = 130$ and $n_1 = 250$. The results are similar to Figure 2.

We compare the forecast based on unrestricted estimator with our proposed combined estimator forecast. Table 7 reports the ratio of MSFE over the benchmark forecast. The results show that the proposed combined estimator delivers vastly improved forecasts (lower MSFE) compared to unrestricted estimator for all horizons. It also has outperformance relative to the Pesaran et al. (2013) estimator. Table 7 also reports the test results based on Diebold and Mariano (1995) statistic for testing the predictive ability of estimators compare to the benchmark forecast.

6 Conclusion

We introduce the combined estimator of the unrestricted estimator with the restricted estimator to estimate the coefficients under structural break. The proposed combined estimator turns out to be a Stein-type shrinkage estimator. We derive the analytical finite sample risk for this estimator and show that the risk of this estimator is lower than the unrestricted estimator. Monte Carlo experiments show the improvement in the risk over the unrestricted estimator. As we increase the number of regressors, we get an even lower risk by using the combined estimator. For some 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 US output growth. We find that the MSFE of the proposed estimator is smaller than the MSFE of the unrestricted estimator. Further, we show that the proposed combined estimator performs well for longer horizon forecasts.

References

Altissimo, F. and Corradi, V. (2003). "Strong rules for detecting the number of breaks in a time series." *Journal of Econometrics* 117, 207-244.

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

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

Andrews, D.W.K., Inpyo, Lee, and W. Ploberger, (1996). "Optimal change point tests for normal linear regression." *Journal of Econometrics* 70, 9–38.

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

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

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

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

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

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

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

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

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

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

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

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

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

Maheu, J. M., and Gordon, S. (2008). "Learning, Forecasting and Structural Breaks." *Journal of Applied Econometrics* 23, 553–583.

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

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.

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

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). "Forecasting time series subject to multiple structural breaks." *Review of Economic Studies* 73, 1057–1084.

Pesaran, M.H., A. Pick and M. Pranovich (2013). "Optimal forecasts in the presence of structural breaks." *Journal of Econometrics* 177, 134-152.

Pesaran, M.H. and A. Timmermann, (2005). "Small sample properties of forecasts from autoregressive models under structural breaks." *Journal of Econometrics* 129, 183-217.

Pesaran, M.H. and A. Timmermann, (2007)."Selection of estimation window in the presence of breaks." *Journal of Econometrics* 137, 134-161.

Qian, J. and Su, L. (2016). "Shrinkage estimation of regression models with multiple structural changes." *Econometric Theory* 32, 1376-1433.

Qu, Z. and P. Perron (2007). "Estimating and testing structural changes in multiple regressions." *Econometrica* 75, 459–502.

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.

Ullah, A. (2004). *Finite Sample Econometrics.* Oxford: Oxford University Press.

A Appendix:

A.1 Proof of equation (12)

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

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

By minimizing the risk, and noting that $\operatorname{tr}\big((X'\Omega^{-1}X)^{-1}L'WL\big) = \operatorname{tr}\big((X'\Omega^{-1}X)^{-1}L'W\big)$, 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

$$\mathbb{E}(\widehat{\beta}'_{ur} L' W L \widehat{\beta}_{ur}) = \beta' L' W L\beta + \operatorname{tr}\big((X'\Omega^{-1}X)^{-1}L'WL\big). \tag{a.3}$$

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

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

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

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

□

A.2 Proof of the estimated variance

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

Proof: Remember,

$$\widehat{\Omega} = \begin{bmatrix} S_{(1)}\ I_{l_1} & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & S_{(m+1)}\ I_{l_{m+1}} \end{bmatrix}, \tag{a.6}$$

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,

$$\mathbb{E}(S_{(i)}) = \sigma^2_{(i)}, \tag{a.7}$$

and

$$var(S_{(i)}) = \frac{2\sigma^4_{(i)}}{l_i - k} = O(\frac{1}{l_i}). \tag{a.8}$$

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. $\square$

A.3 Proof of Theorem 1

We derive the optimal value of the weight in (12) which depends 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. Notice that, knowing the order of Δ, by expanding $\widehat{\Omega}^{-1}$ we have

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

Thus,

$$\begin{aligned}
(X'\widehat{\Omega}^{-1}X)^{-1} &= (X'\Omega^{-1}X)^{-1}\Bigg(I_T - (X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1} + (X'\Omega^{-1}(\Delta\Omega^{-1})^2X)(X'\Omega^{-1}X)^{-1} \\
&\quad + O_p(T^{-3/2})\Bigg)^{-1} \\
&= A_{-1} + A_{-3/2} + A_{-2} + O_p(T^{-5/2}),
\end{aligned} \tag{a.10}$$

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, e.g., $A_{-2} = O_p(T^{-2})$.

Therefore,

$$\begin{aligned}
\Big[R(X'\widehat{\Omega}^{-1}X)^{-1}R'\Big]^{-1} &= \big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} - \big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}R(X'\Omega^{-1}X)^{-1} \\
&\quad (X'\Omega^{-1}\Delta\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}R'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1} + O_p(1).
\end{aligned} \tag{a.11}$$

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

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

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 (12) is

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

where $\Pi_{-1/2} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon$, and the feasible term for the numerator of (12) is

$$\begin{aligned}(X'\widehat{\Omega}^{-1}X)^{-1}\widehat{L}'W &= (X'\Omega^{-1}X)^{-1}L'W + (X'\Omega^{-1}X)^{-1}L'_{-1/2}W \\ &\quad - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta\Omega^{-1}X(X'\Omega^{-1}X)^{-1}L'W + O_p(T^{-1}) \\ &= O(1) + O_p(T^{-1/2}) + O_p(T^{-1/2}) + O_p(T^{-1}).\end{aligned} \tag{a.14}$$

Finally, using (a.13) and (a.14), we have

$$\begin{aligned}\bar{\bar{\beta}}_{\gamma} - \beta &= \bar{\bar{\beta}}_{ur} - \beta - \widehat{\gamma}^*\big(\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_{r}\big) \\ &= \bar{\bar{\beta}}_{ur} - \beta - \Big[\frac{1}{\phi} - \frac{2}{\phi^2}\ \beta'L'WL\Pi_{-1/2} - \frac{2}{\phi^2}\ \beta'L'WL_{-1/2}\ \beta\Big] \\ &\quad \mathrm{tr}\big((X'\Omega^{-1}X)^{-1}L'W\big)\big[L\beta + L\Pi_{-1/2}\big] + O_p(T^{-2}),\end{aligned} \tag{a.15}$$

where $\phi = \beta'L'WL\beta$. Thus, the MSE, to order $O_p(T^{-2})$, is

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

$$+\frac{2\ \mathrm{tr}(Q)}{\phi^2}\ L\beta\beta' L'WL\ (X'\Omega^{-1}X)^{-1}+\frac{2\ \mathrm{tr}(Q)}{\phi^2}\ (X'\Omega^{-1}X)^{-1}L'WL\beta\beta' L'. \tag{a.16}$$

Finally, writing $\mathbb{E}\left[\left(\bar{\bar{\beta}}_\gamma-\beta\right)\left(\bar{\bar{\beta}}_\gamma-\beta\right)'W\right]$, where $W=O(T)$, and taking its trace we get the risk of this estimator, to order $O_p(T^{-1})$, is

$$\begin{aligned} Risk(\bar{\bar{\beta}}_\gamma, W) &= Risk(\bar{\bar{\beta}}_{ur}, W)+\frac{1}{\phi}\ \Big(\mathrm{tr}(Q)\Big)^2-\frac{2}{\phi}\ \Big(\mathrm{tr}(Q)\Big)^2 \\ &+\frac{4\ \mathrm{tr}(Q)}{\phi^2}\ \mathrm{tr}\big(L\beta\beta' L'WL\ (X'\Omega^{-1}X)^{-1}\big). \end{aligned} \tag{a.17}$$

This completes the proof of Theorem 1. ■

In evaluating expectations in (a.16) we have used the following results. Let the $T\times 1$ random vector ϵ be such that $\epsilon\sim \mathrm{N}(0,\sigma^2 I_T)$, and M_1 and M_2 be arbitrary $T\times T$ matrices. Then,

$$\mathbb{E}\left[(\epsilon' M_1\epsilon)(\epsilon' M_2\epsilon)\right]=\sigma^4\Big[\mathrm{tr}(M_1)\mathrm{tr}(M_2)+\mathrm{tr}(M_1M_2)+\mathrm{tr}(M_1M_2')\Big], \tag{a.18}$$

$$\mathbb{E}\left[\epsilon\epsilon' M_1\epsilon\epsilon'\right]=\sigma^4\Big[\mathrm{tr}(M_1)I_T+M_1+M_1'\Big]. \tag{a.19}$$

See Ullah (2004).

A.4 Proof of Corollaries 1.1, 1.4, 1.5

Corollary 1.1:

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

$$\begin{aligned} \frac{4\ \beta' L'WL(X'\Omega^{-1}X)^{-1}WL\beta}{\phi} &< \mathrm{tr}(Q) \\ \sup_{W^{1/2}L\beta}\ \frac{4\ \beta' L'W^{1/2}W^{1/2}L(X'\Omega^{-1}X)^{-1}W^{1/2}W^{1/2}L\beta}{\beta' L'WL\beta} &< \mathrm{tr}(Q) \\ 4\ \lambda_{\max}(Q) &< \mathrm{tr}(Q). \end{aligned} \tag{a.20}$$

Thus, the risk of the combined estimator is less than that of the unrestricted estimator if $d\equiv \frac{\mathrm{tr}(Q)}{\lambda_{\max}(Q)}>4$. □

Corollary 1.4:

Using $W=X'X/\sigma^2=O(1/\sigma^2), \Omega=\sigma^2 I$, and $\sigma^2=\widehat{\sigma}^2$ in the combination weight (12), and expanding the terms up to order σ^3, we get

$$\begin{aligned} \widehat{\gamma}^* &= \frac{\widehat{\sigma^2}\ \mathrm{tr}(Q)}{\bar{\bar{\beta}}'_{ur}L'X'XL\bar{\bar{\beta}}_{ur}} \\ &= \nu_1+\nu_2+O_p(\sigma^4), \end{aligned} \tag{a.21}$$

where $L = (X'X)^{-1}R'\big(R(X'X)^{-1}R\big)^{-1}R = O(1)$, $Q = (X'X)^{1/2}L(X'X)^{-1}(X'X)^{1/2} = O(1)$, $\text{tr}(Q) = \text{tr}(L) = p$,

$\nu_1 \equiv \frac{\sigma^2 {\epsilon^*}' M\epsilon^* \ \text{tr}(Q)}{\phi_s \ \big(T-(m+1)k\big)} = O_p(\sigma^2)$,

$\nu_2 \equiv -\frac{2\sigma^3 {\epsilon^*}' M\epsilon^* \ \text{tr}(Q)}{\phi_s^2 \ \big(T-(m+1)k\big)}\Big[\epsilon^{*\prime} X(X'X)^{-1}L'WL\beta\Big] = O_p(\sigma^3)$,

$\phi_s = \beta' L' X' X L\beta = O(1)$, and $\epsilon \equiv \sigma\epsilon^*$ with $\epsilon^* \sim \text{N}(0, I)$. Now, rewrite the combined estimator as

$$\begin{aligned} \bar{\bar{\beta}}_\gamma - \beta &= \big(\bar{\bar{\beta}}_{ur} - \beta\big) - \widehat{\gamma}^*\big(\bar{\bar{\beta}}_{ur} - \bar{\bar{\beta}}_r\big) \\ &= \bar{B}_1 + \bar{B}_2 + \bar{B}_3 + O_p(\sigma^4), \end{aligned} \tag{a.22}$$

where

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

$\bar{B}_2 \equiv -\nu_1 L\beta = -\frac{\sigma^2 {\epsilon^*}' M\epsilon^* \ \text{tr}(Q)}{\phi_s \ \big(T-(m+1)k\big)} L\beta = O_p(\sigma^2)$,

$\bar{B}_3 \equiv -\frac{\sigma^3 {\epsilon^*}' M\epsilon^* \ \text{tr}(Q)}{\phi_s \ \big(T-(m+1)k\big)} L(X'X)^{-1}X'\epsilon^* + \frac{2\sigma^3 {\epsilon^*}' M\epsilon^* \ \text{tr}(Q)}{\phi_s^2 \ \big(T-(m+1)k\big)} L\beta \ \epsilon^{*\prime} X(X'X)^{-1}L'X'XL\beta = O_p(\sigma^3)$.

Thus, the risk of the combined estimator is

$$\begin{aligned} Risk\big(\bar{\bar{\beta}}_\gamma, W\big) &= \mathbb{E}\left[\big(\bar{\bar{\beta}}_\gamma - \beta\big)' W \big(\bar{\bar{\beta}}_\gamma - \beta\big)\right] \\ &= \mathbb{E}\left(\bar{B}_1' W \bar{B}_1\right) + \mathbb{E}\left(\bar{B}_2' W \bar{B}_2\right) + 2\,\mathbb{E}\left(\bar{B}_1' W \bar{B}_3\right), \end{aligned} \tag{a.23}$$

where, using (a.18) and (a.19)

$$\mathbb{E}(\bar{B}_1' \, \mathbb{W} \, \bar{B}_1) = (m+1)k, \tag{a.24}$$

$$\mathbb{E}(\bar{B}_2' W \bar{B}_2) = \frac{\sigma^2 p^2}{\phi_s \ \big(T-(m+1)k\big)}\Big[T-(m+1)k+2\Big], \tag{a.25}$$

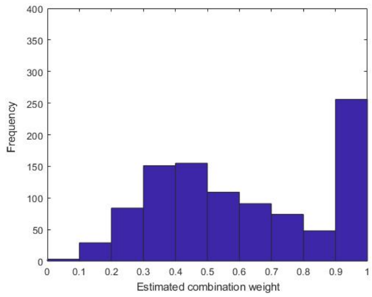
and

$$\mathbb{E}\left(\bar{B}_1' W \bar{B}_3\right) = -\frac{\sigma^2 p^2}{\phi_s} + \frac{2\sigma^2 p}{\phi_s}. \tag{a.26}$$

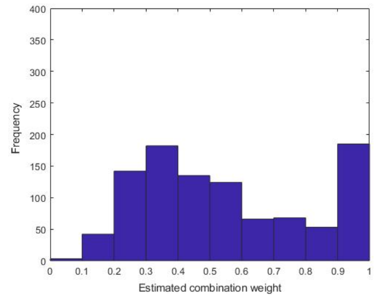
Substituting (a.24) to (a.26) in (a.23) we get the result in Corollary 1.4. □

Corollary 1.5:

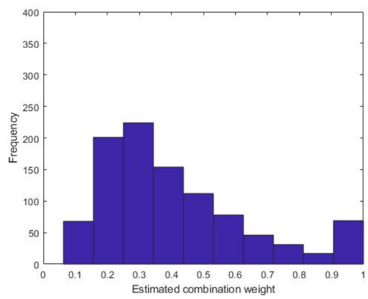
Getting the combination weight, $\widehat{\gamma}$, by using the in-sample prediction weight (see Corollary 1.2), and setting $W = x_{T+1}x_{T+1}'$, we can derive the one-step ahead MSFE, and have the result of this Corollary satisfied. □



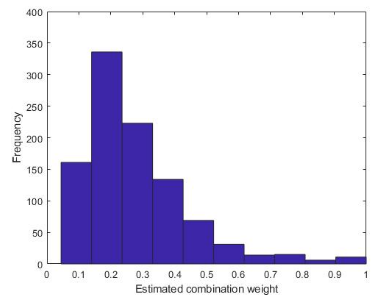
(a) $\lambda = 0$



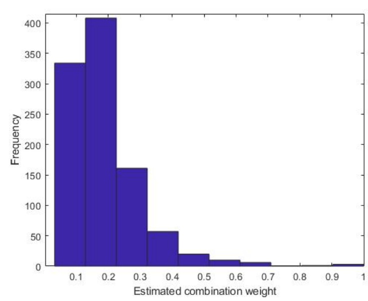
(b) $\lambda = 0.25$



(c) $\lambda = 0.5$



(d) $\lambda = 0.75$



(e) $\lambda = 1$

Figure 1: Frequency of $\widehat{\gamma}$ for $T = 100, T_1 = 80$, $k = 8$ and different λ

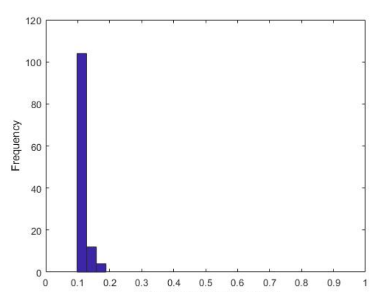


Figure 2: Frequency of $\widehat{\gamma}$ for $h = 1, n_1 = 370$

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

	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	0.591	0.563	0.535	0.506	0.470	0.437	0.410
	0.25	0.652	0.613	0.598	0.590	0.540	0.518	0.503
$T = 100$	0.50	0.762	0.727	0.736	0.717	0.682	0.659	0.658
	0.75	0.858	0.873	0.842	0.828	0.816	0.773	0.775
	1.00	0.936	0.944	0.934	0.918	0.901	0.856	0.853
	0.00	0.707	0.679	0.650	0.619	0.580	0.550	0.504
	0.25	0.808	0.762	0.744	0.721	0.681	0.675	0.667
$T = 200$	0.50	0.873	0.861	0.845	0.814	0.803	0.791	0.775
	0.75	0.951	0.953	0.939	0.919	0.908	0.890	0.858
	1.00	0.984	0.984	0.974	0.961	0.958	0.950	0.914

Note: This table represents the results of the relative MSE for which the benchmark model is the unrestricted estimator, $RMSE_\gamma \equiv MSE(\bar{\hat{\beta}}_\gamma)/MSE(\bar{\hat{\beta}}_{ur})$. The first column shows the sample size while the second column shows the true break size in the coefficient, λ.

Table 2: Simulation results, $k = 8$, $q = 0.5$, $W = X'\Omega^{-1}X$

	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	0.463	0.433	0.392	0.367	0.364	0.306	0.286
	0.25	0.536	0.498	0.474	0.451	0.439	0.394	0.366
$T = 100$	0.50	0.697	0.655	0.660	0.638	0.618	0.565	0.534
	0.75	0.831	0.842	0.830	0.821	0.776	0.728	0.689
	1.00	0.912	0.954	0.947	0.939	0.898	0.842	0.799
	0.00	0.595	0.555	0.521	0.484	0.440	0.401	0.355
	0.25	0.717	0.681	0.653	0.626	0.580	0.555	0.529
$T = 200$	0.50	0.854	0.852	0.846	0.825	0.781	0.759	0.706
	0.75	0.961	0.960	0.947	0.935	0.917	0.880	0.828
	1.00	0.977	0.981	0.974	0.965	0.962	0.935	0.899

Note: See the note of Table 1.

Table 3: Simulation results, $k = 5$, $q = 1$, $W = X'\Omega^{-1}X$

	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	0.578	0.580	0.581	0.588	0.579	0.590	0.599
	0.25	0.692	0.719	0.744	0.752	0.737	0.733	0.726
$T = 100$	0.50	0.827	0.851	0.895	0.913	0.904	0.875	0.875
	0.75	0.899	0.938	0.962	0.974	0.975	0.963	0.948
	1.00	0.938	0.965	0.981	0.987	0.991	0.984	0.977
	0.00	0.661	0.661	0.685	0.664	0.688	0.673	0.701
	0.25	0.801	0.816	0.835	0.837	0.834	0.830	0.831
$T = 200$	0.50	0.899	0.938	0.953	0.959	0.963	0.956	0.927
	0.75	0.949	0.972	0.982	0.981	0.986	0.984	0.968
	1.00	0.967	0.983	0.989	0.988	0.992	0.992	0.980

Note: See the note of Table 1.

Table 4: Simulation results, $k = 8$, $q = 1$, $W = X'\Omega^{-1}X$

	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	0.429	0.443	0.458	0.435	0.469	0.465	0.492
	0.25	0.595	0.637	0.688	0.675	0.687	0.651	0.636
$T = 100$	0.50	0.775	0.849	0.900	0.921	0.902	0.860	0.814
	0.75	0.882	0.939	0.961	0.973	0.966	0.948	0.910
	1.00	0.933	0.966	0.978	0.987	0.983	0.972	0.952
	0.00	0.567	0.576	0.581	0.583	0.576	0.578	0.592
	0.25	0.749	0.795	0.816	0.828	0.818	0.802	0.767
$T = 200$	0.50	0.898	0.937	0.957	0.960	0.966	0.944	0.912
	0.75	0.949	0.970	0.981	0.981	0.986	0.975	0.958
	1.00	0.969	0.982	0.988	0.988	0.992	0.987	0.975

Note: See the note of Table 1.

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

	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	0.383	0.425	0.447	0.472	0.501	0.564	0.620
	0.25	0.475	0.515	0.538	0.550	0.578	0.631	0.665
$T = 100$	0.50	0.618	0.647	0.683	0.685	0.706	0.738	0.771
	0.75	0.728	0.751	0.787	0.806	0.839	0.845	0.861
	1.00	0.796	0.826	0.873	0.912	0.929	0.935	0.936
	0.00	0.488	0.516	0.572	0.593	0.661	0.661	0.720
	0.25	0.623	0.643	0.659	0.705	0.734	0.750	0.802
$T = 200$	0.50	0.742	0.764	0.793	0.812	0.839	0.844	0.880
	0.75	0.819	0.861	0.894	0.923	0.937	0.952	0.954
	1.00	0.879	0.922	0.945	0.962	0.975	0.980	0.988

Note: See the note of Table 1.

Table 6: Simulation results, $k = 8$, $q = 2$, $W = X'\Omega^{-1}X$

	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	0.263	0.312	0.344	0.348	0.387	0.449	0.473
	0.25	0.346	0.392	0.430	0.425	0.461	0.520	0.552
$T = 100$	0.50	0.510	0.555	0.609	0.607	0.638	0.674	0.693
	0.75	0.655	0.728	0.786	0.821	0.840	0.840	0.824
	1.00	0.759	0.830	0.883	0.922	0.940	0.947	0.921
	0.00	0.342	0.400	0.456	0.487	0.518	0.555	0.585
	0.25	0.503	0.554	0.584	0.623	0.624	0.687	0.716
$T = 200$	0.50	0.689	0.746	0.775	0.817	0.829	0.847	0.846
	0.75	0.807	0.866	0.905	0.930	0.957	0.957	0.950
	1.00	0.883	0.922	0.951	0.964	0.978	0.977	0.980

Note: See the note of Table 1.

Table 7: Out-of-sample forecasting performance with monthly data

h	out-of-sample period	$MSFE_{\gamma}$	$MSFE_{ur}$	$MSFE_{PPP}$
1	1970:01-2020:03	0.6643**	0.6690	0.6680*
	1980:01-2020:03	0.5835	0.5862	0.5857
	1990:01-2020:03	0.5155**	0.5227	0.5226
6	1970:01-2020:03	0.7576***	0.7784	0.7674***
	1980:01-2020:03	0.5301*	0.5383	0.5370
	1990:01-2020:03	0.4667	0.4667	0.4667
12	1970:01-2020:03	0.2430***	0.2467	0.2449
	1980:01-2020:03	0.2182***	0.2211	0.2193
	1990:01-2020:03	0.2493***	0.2520	0.2496

Note: This table reports the ratio of MSFE for different estimators 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. $MSFE_{\gamma}$ represents the results for the γ weight combined estimator, and $MSFE_{PPP}$ represents the result based on the method proposed by Pesaran et al. (2013). ***, ** and * indicate significance at 1%, 5% and 10% based on Diebold and Mariano (1995).