Optimal Forecast under Structural Breaks[†]

Tae-Hwy Lee[‡] Shahnaz Parsaeian[§*] Aman Ullah[¶]

January 24, 2022

Abstract

This paper develops an optimal combined estimator to forecast out-of-sample under structural breaks. When it comes to forecasting, using only the post-break observations after the most recent break point may not be optimal. In this paper we propose a new estimation method that exploits the pre-break information. In particular, 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 while it is efficient. The post-break estimator is consistent but inefficient. 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. We derive the Stein-like combined estimator of the full-sample and the post-break estimators, to balance the bias-variance trade-off. The combination weight depends on the break severity, which we measure by the Wu-Hausman statistic. We examine the properties of the proposed method, analytically in theory, numerically in simulation, and also empirically in forecasting real output growth across nine industrial economies.

Keywords: Forecasting; Structural breaks; Stein-like combined estimator; Output growth

JEL Classification: C13, C32, C53

[†]We would like to thank the editor (Barbara Rossi) and anonymous referees for helpful and constructive comments on the paper. We are grateful to Tom Boot, Zongwu Cai, Ali Mehrabani, and Andreas Pick for helpful discussions.

[‡]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

[*]Correspondence to: Shahnaz Parsaeian, Department of Economics, University of Kansas, Lawrence, KS 66045. Email: shahnaz.parsaeian@ku.edu

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

1 Introduction

In a regression framework, structural breaks are shifts in the regression coefficients and/or error variance. An important problem is how to make an accurate forecast in the presence of possible structural breaks. Since the seminal work by Bates and Granger (1969), forecast combination has been a common practice to improve forecasting performance. Especially under the model uncertainty, the forecasting performance can be boosted by a forecast combination 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 (2009), Pesaran and Pick (2011), Rossi (2013), and Pesaran et al. (2013) inter alia.

An obvious method for forecasting under structural breaks is to use the post-break data to estimate the parameters of the model. But this post-break estimator by itself may not be optimal in the mean squared forecast errors (MSFE) either when a break occurs close to the end of the sample and thus there are only a few observations in the post-break sample or when a break is weak and therefore hard to detect. In such cases, the post-break parameters are likely to be inefficiently estimated relative to those obtained using the pre-break data as well. As pointed out by Pesaran and Timmermann (2007) and Pesaran et al. (2013), it is not always optimal to base the forecast only on post-break observations. However, using the pre-break data would make the forecast biased, while it reduces the forecast error variance.

A key question that arises under the presence of structural breaks therefore is how to use the pre-break data to estimate the forecasting model such that the loss function like MSFE is minimized. In this paper we propose a combined estimator of the post-break estimator and the full-sample estimator which uses all observations in the sample, $t = \{1, \ldots, T\}$, to improve the performance of a forecast model. The combination weight takes the form of the James-Stein weight, cf. Stein (1956) and James and Stein (1961). The proposed Stein-like combined estimator exploits the trade-off between the bias and variance of the forecast error. The full-sample estimator is biased if there are structural breaks, while it may have smaller forecast error variance because it uses more observations. The post-break estimator (which is a common solution under structural breaks) is unbiased but less efficient. Therefore, we can improve the performance of forecast measured by MSFE by exploiting the trade-off between the bias and forecast error variance. See Saleh (2006) for a comprehensive review for the Stein-like estimators.[1]

The goal of this paper is to develop an estimator with a minimum MSFE under a structural break, which we introduce in Section 2 and examine its asymptotic properties. In the simulation study and the empirical analysis presented in Sections 4 and 5, we estimate the break points following Bai and Perron (1998, 2003) which gives a consistent global minimizer of the sum of squared residuals.[2] We assume that the break points are away from the beginning and the end of the sample which is crucial for consistency of estimating break points, see Andrews (1993).

We undertake an empirical analysis for forecasting the real output growth across nine industrial economies by using quarterly data from 1979 to 2016 to compare the forecasting performance of our proposed Stein-like combined estimator with the post-break estimator and a range of alternative methods existing in the literature. Specifically, we compare the forecasting performance of our estimator with the five methods proposed by Pesaran and Timmermann (2007), the average window method proposed by Pesaran and Pick (2011), and the weighted least square estimator proposed by Pesaran et al. (2013). We also compare the methods using the optimal window proposed by Inoue et al. (2017), which is designed for the smoothly time-varying parameters. Our empirical results confirm the benefits of using the pre-break data in estimation of the forecasting model relative to using only the post-break data.

The outline of the paper is as follows. Section 2 sets up the structural break model with a single break, introduces the Stein-like combined estimator, and presents its asymptotic risk. For simplicity, we discuss the problem under a single break, which simplifies the essential idea without unduly complicating notation. However the generalization of the method for the multiple breaks is straightforward as considered in Section 3. Section 4 reports Monte Carlo simulation and Section 5 presents empirical analysis. Section 6 concludes.

[1] Hansen (2016, 2017) considers a Stein-like estimator in combining 2SLS and ordinary least square (OLS) estimators. A main difference between our paper and Hansen (2017) is that, here we minimize the asymptotic risk for any user specific positive definite weight matrix to find the optimal combination. Hansen (2017) sets this weight matrix to the inverse of the difference between variances of the two estimators.

[2] Alternatively, the break points can be detected by the Impulse Indicator Saturation (IIS) approach of Hendry et al. (2008). See also Castle and Hendry (2019). IIS creates a complete set of impluse indicator variables for every observation and a computer algorithm, such as the split-half approach by Hendry et al. (2008), can be employed to search the indicators that match with break points. As the IIS generally incurs a high-dimensionality problem, a machine learning algorithm such as L2-Boosting or Lasso can also be a natural consideration for selecting the impulse indicators to estimate the break points. We thank a referee for bringing the IIS method to our attention.

2 The structural break model

Consider the linear structural break model as $y_t = x_t'\beta_t + u_t$, where $u_t = \sigma_t\varepsilon_t$, $\varepsilon_t \sim \text{i.i.d.}(0,1)$, and x_t is a $k \times 1$ vector of regressors that may contain lagged values of y_t. In this model, the $k \times 1$ vector of coefficients, β_t, and the error variance, σ_t^2, are subject to breaks. Let m denote the number of breaks. For simplicity we assume one break ($m = 1$) that happens at time T_1 with $1 < T_1 < T$. So we can rewrite the model as

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

where

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

and

$$\sigma_t = \begin{cases} \sigma_{(1)} & \text{for} \quad 1 \ < t \leq T_1 \\ \sigma_{(2)} & \text{for} \quad T_1 < t \leq T. \end{cases} \tag{3}$$

In this set up we have only one break (two regimes). Let $U = \big(u_1, \ldots, u_T\big)'$, so the variance of the error term is $V(U|X) = \Omega = \text{diag}\big(\sigma_{(1)}^2 I_{T_1},\ \sigma_{(2)}^2 I_{T-T_1}\big)$, where I_{T_1} is a $T_1 \times T_1$ identity matrix and I_{T-T_1} is a $(T - T_1) \times (T - T_1)$ identity matrix. Therefore, we assume that the disturbances are uncorrelated within and across the regime, but heteroskedastic across the regime.

Assumption 1: $\mathbb{E}(u_t|x_t) = 0$, and $\sigma_t^2 = \mathbb{E}(u_t^2|x_t)$.

2.1 Stein-like combined estimator

As our interest is on forecasting, the parameter of interest is $\beta_{(2)}$. Our proposed combined estimator of $\beta_{(2)}$ is

$$\widehat{\beta}_\alpha = \alpha\widehat{\beta}_{Full} + (1 - \alpha)\widehat{\beta}_{(2)}, \tag{4}$$

where $\widehat{\beta}_\alpha$ is the Stein-like combined estimator which is $k \times 1$, $\widehat{\beta}_{Full}$ is the estimator using all observations in the sample, $t \in \{1, \ldots, T\}$, and $\widehat{\beta}_{(2)}$ estimates the coefficient only by using the post-break observations, $t > T_1$, and it is called the post-break estimator. We define the combination weight as

$$\alpha = \begin{cases} \frac{\tau}{H_T} & \text{if } H_T \geq \tau \\ 1 & \text{if } H_T < \tau, \end{cases} \tag{5}$$

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

$$H_T = T\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}\big)'\big(\widehat{V}_{(2)} - \widehat{V}_{Full}\big)^{-1}\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}\big), \tag{6}$$

where $\widehat{V}_{Full}$ and $\widehat{V}_{(2)}$ are the consistent estimators of the asymptotic variances of the full-sample estimator, $\widehat{\beta}_{Full}$, and the post-break estimator, $\widehat{\beta}_{(2)}$, respectively, as defined in Theorem 1 below. The degree of shrinkage depends on the ratio of τ/H_T . When $H_T < \tau$, then $\alpha = 1$ and $\widehat{\beta}_\alpha = \widehat{\beta}_{Full}$. A small H_T can be interpreted as a small break size in the coefficients. This is when 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, a large H_T is interpreted as a big break in the coefficients which results in a large bias in $\widehat{\beta}_{Full}$. So, for the extreme case of a large H_T, the combination weight α would be close to zero and $\widehat{\beta}_\alpha \approx \widehat{\beta}_{(2)}$. Other than these extreme cases, when $H_T > \tau$, $\widehat{\beta}_\alpha$ is a weighted average of the full-sample estimator and the post-break estimator.

It can be shown that $\mathrm{cov}\big(\widehat{\beta}_{(2)}, \widehat{\beta}_{Full}\big) = V_{Full}$, where $V_{Full} < V_{(2)}$, i.e., 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 ($\beta_{(1)} = \beta_{(2)}$), because otherwise we could find another linear combination which would have smaller asymptotic variance than $\widehat{\beta}_{Full}$ which contradicts as $\widehat{\beta}_{Full}$ is asymptotically efficient. Therefore we combine the full-sample and the post-break estimators. See Hausman (1978) for more discussion.

Remark 1: Model uncertainty arises due to lack of information about the timing and size of the break. If the break size is "small", then one may get a better forecast by ignoring the break and using the full-sample estimator which is efficient. If a break size is "large", the full-sample estimator suffers from a large bias which may dominate the benefit from its efficiency. Our combined estimator is to balance the trade-off between the bias and variance efficiency. As the Hausman statistic is the ratio of the bias over efficiency, it helps to assign appropriate weight to the full-sample and the post-break estimators.[3] □

In the next subsections, we develop the asymptotic distribution for the estimators under a local asymptotic framework where the break size is local to zero

$$\beta_{(1)} - \beta_{(2)} = \frac{\delta_1}{\sqrt{T}}. \tag{7}$$

2.1.1 The full-sample estimator using all observations

When there is no break in the coefficients, $\beta_{(1)} = \beta_{(2)}$, the full-sample estimator uses all of the observations to estimate β. This assumption lines up with the fact that, with a small break, ignoring the break and estimating the coefficient using all the observations would result in a better forecast (lower MSFE), see Boot and Pick (2020). Thus we consider the local alternative hypothesis that $\beta_{(1)} - \beta_{(2)} = \frac{\delta_1}{\sqrt{T}}$.

We denote the full-sample estimator by $\widehat{\beta}_{Full}$, and estimate the coefficient by the generalized least squares (GLS) estimator

$$\widehat{\beta}_{Full} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y, \tag{8}$$

where $X = \left(x_1, x_2, \ldots, x_T\right)' = \left(X'_{(1)}\ X'_{(2)}\right)'$ is a $T \times k$ matrix of regressors, $X_{(1)}$ is a $T_1 \times k$ matrix of pre-break observations, and $X_{(2)}$ is a $\left(T - T_1\right) \times k$ matrix of post-break observations. 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 we choose around two to three times the dimension of β to avoid the extreme variation in the post-break parameter estimates.

Assumption 2: We assume that $\frac{X'_{(i)}\Omega^{-1}_{(i)}X_{(i)}}{T_i - T_{i-1}}$, 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, i.e., $\left(\frac{X'\Omega^{-1}X}{T}\right) \xrightarrow{p} Q$ and $\left(\frac{X'_{(i)}\Omega^{-1}_{(i)}X_{(i)}}{\Delta T_i}\right) \xrightarrow{p} Q_i$ where Q and Q_i are positive definite matrices, and $\Delta T_i = T_i - T_{i-1}$.

Throughout this section, $m = 1$ since we have only one break. Therefore, the distribution of

[3] Other structural break tests that have power against the local alternatives may be used as long as they can be written as the weighted distance of the full-sample estimator and the post-break estimator. In addition, one needs to know the analytical form of the test to derive its asymptotic distribution and calculate the asymptotic distribution of the proposed Stein-like combined estimator and its asymptotic risk, as we have done in Theorems 1-3 using the Hausman statistic. This is beyond the scope of this paper and we leave it for future study.

the full-sample estimator is

$$\sqrt{T}\Big(\widehat{\beta}_{Full} - \beta_{(2)}\Big) \xrightarrow{d} N\Big(Q^{-1}Q_1 b_1 \delta_1, Q^{-1}\Big), \tag{9}$$

where $Q^{-1} \equiv \underset{T\to\infty}{\text{plim}}\Big(\frac{X'\Omega^{-1}X}{T}\Big)^{-1} = V_{Full}$, $Q_1 \equiv \underset{T\to\infty}{\text{plim}}\Big(\frac{X'_{(1)}\Omega^{-1}_{(1)}X_{(1)}}{T_1}\Big)$, $b_1 \equiv \lim_{T\to\infty}\big(\frac{T_1}{T}\big)$ denotes the proportion of pre-break observations, and $\Omega_{(1)} = \sigma^2_{(1)} I_{T_1}$ is a $T_1 \times T_1$ diagonal matrix. Note that in practice, we need to estimate the value of the unknown parameters, $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$. For this, we 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 the OLS residuals for each regime, and then plug $\widehat{\Omega}$ back into equation (8). Since $\widehat{\Omega}$ is a consistent estimator for Ω , we also have $\widehat{V}_{Full} = \Big(\frac{X'\widehat{\Omega}^{-1}X}{T}\Big)^{-1} \xrightarrow{p} Q^{-1}$.

Remark 2: The full-sample estimator is calculated 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 the GLS method which is more efficient than the OLS. □

2.1.2 The post-break estimator using post-break observations

The post-break estimator uses only the observations after the most recent break point and is equal to

$$\widehat{\beta}_{(2)} = \Big(X'_{(2)}\ \Omega^{-1}_{(2)}\ X_{(2)}\Big)^{-1} X'_{(2)}\ \Omega^{-1}_{(2)}\ Y_{(2)}, \tag{10}$$

where $\Omega_{(2)} = \sigma^2_{(2)} I_{T-T_1}$ is a $\big(T - T_1\big) \times \big(T - T_1\big)$ diagonal matrix, and $Y_{(2)}$ is the dependent variable in the second regime. The post-break estimator is the simple OLS estimator, since $\Omega^{-1}_{(2)}$ will be cancelled out in this equation. But we write it in the GLS format to be consistent with the full-sample estimator. The distribution of this estimator is

$$\sqrt{T}\Big(\widehat{\beta}_{(2)} - \beta_{(2)}\Big) \xrightarrow{d} N\Big(0, \frac{1}{1-b_1} Q_2^{-1}\Big), \tag{11}$$

where $\frac{1}{1-b_1} Q_2^{-1} \equiv \frac{1}{1-b_1} \underset{T\to\infty}{\text{plim}}\Big(\frac{X'_{(2)}\Omega^{-1}_{(2)}X_{(2)}}{T-T_1}\Big)^{-1} = V_{(2)}$. This is an unbiased estimator. As the break happens towards the end of the sample, b_1 increases, and the variance of the post-break estimator increases.

2.1.3 Stein-like combined estimator

To derive the distribution of the Stein-like combined estimator in (4), we first obtain the joint asymptotic distribution of the full-sample estimator and the post-break estimator. Theorem 1 below shows the joint distribution, the distribution of the Hausman statistic, and the distribution of the Stein-like combined estimator.

Theorem 1: *Under Assumptions 1-2, along the sequences (7), the joint asymptotic distribution of the full-sample estimator and the post-break estimator is*

$$\sqrt{T}\begin{bmatrix}\widehat{\beta}_{Full}-\beta_{(2)}\\ \widehat{\beta}_{(2)}-\beta_{(2)}\end{bmatrix}\xrightarrow{d} V^{1/2}Z, \tag{12}$$

where $Z\sim \mathrm{N}\big(\theta, I_{2k}\big)$, $\theta = V^{-1/2}\begin{bmatrix}Q^{-1}Q_1 b_1\delta_1\\ 0\end{bmatrix}$, and $V=\begin{bmatrix}V_{Full} & V_{Full}\\ V_{Full} & V_{(2)}\end{bmatrix}$.

Besides, the distribution of the Hausman statistic is

$$\begin{aligned} H_T &= T\big(\widehat{\beta}_{(2)}-\widehat{\beta}_{Full}\big)'\big(\widehat{V}_{(2)}-\widehat{V}_{Full}\big)^{-1}\big(\widehat{\beta}_{(2)}-\widehat{\beta}_{Full}\big)\\ &\xrightarrow{d} Z'V^{1/2}\ G\ \big(V_{(2)}-V_{Full}\big)^{-1}\ G'\ V^{1/2}Z \\ &\equiv Z'MZ, \end{aligned} \tag{13}$$

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

$$\begin{aligned} \sqrt{T}\big(\widehat{\beta}_{\alpha}-\beta_{(2)}\big) &= \sqrt{T}\big(\widehat{\beta}_{(2)}-\beta_{(2)}\big)-\alpha\sqrt{T}\big(\widehat{\beta}_{(2)}-\widehat{\beta}_{Full}\big)\\ &\xrightarrow{d} G_2'\ V^{1/2}Z-\left(\frac{\tau}{Z'MZ}\right)_1\ G'\ V^{1/2}Z, \end{aligned} \tag{14}$$

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

See Appendix A.1 for the proof of this theorem. The joint asymptotic distribution of the full-sample and post-break estimators is normal. The Hausman statistic has asymptotic noncentral chi-squared distribution. In the next subsection, we show that the asymptotic risk depends on this non-centrality. The asymptotic distribution of the Stein-like combined estimator is a function of the normal random vector, Z, with the non-centrality parameter θ which depends on the proportion of the pre-break observations b_1 and the break size δ_1.

2.2 Asymptotic risk for the Stein-like combined estimator

In this section we derive the asymptotic risk for the Stein-like combined estimator. The risk of an estimator $\widehat{\beta}$ may not be finite unless it has sufficient finite moments. Therefore, we use a trimmed loss to ensure its existence, and take limits as the sample size and the trimming parameter increase:

$$\rho\big(\widehat{\beta}, \mathbb{W}\big) = \lim_{\xi\to\infty} \liminf_{T\to\infty} \mathbb{E}\min\Big[T\Big(\widehat{\beta}-\beta\Big)'\,\mathbb{W}\,\Big(\widehat{\beta}-\beta\Big)', \xi\Big], \tag{15}$$

where the expected scaled loss is trimmed at an arbitrarily trimmed parameter ξ. We note that as $\xi \to \infty$, the trimming becomes negligible. This definition of the asymptotic risk is well-defined and easy to calculate whenever an estimator has an asymptotic distribution, $\sqrt{T}(\widehat{\beta}-\beta) \xrightarrow{d} \varpi$. For then, we define the asymptotic risk for this estimator as $\rho(\widehat{\beta}, \mathbb{W}) = \mathbb{E}(\varpi'\,\mathbb{W}\,\varpi)$, see Lemma 6.1.14 of Lehmann and Casella (1998).[4] In this paper, $\widehat{\beta}$ would be like the full-sample, the post-break, and the Stein-like combined estimators, and β would be $\beta_{(2)}$.

Given the asymptotic distribution of the Stein-like combined estimator in (14), we first write the asymptotic risk for this estimator. Since our focus is on forecasting, we consider $\beta_{(2)}$ as the true parameter vector in defining the asymptotic risk. Then, we derive the optimal combination weight, α, which minimizes the asymptotic risk, and consequently we obtain the optimal Stein-like combined estimator, $\widehat{\beta}_\alpha$. Throughout the calculation of the asymptotic risk, we do not specify any specific form for $\mathbb{W}$, and calculate the asymptotic risk for any positive definite choice of weight $\mathbb{W} > 0$.[5] Theorem 2 shows the asymptotic risk for the Stein-like combined estimator.

Theorem 2: *Under Assumptions 1-2, the asymptotic risk of the Stein-like combined estimator is*

$$\begin{aligned}\rho\big(\widehat{\beta}_\alpha, \mathbb{W}\big) = \rho\big(\widehat{\beta}_{(2)}, \mathbb{W}\big) &+ \frac{\tau\,\theta' A\theta}{k(k+2)}\left[\tau - 2\left(\frac{\mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)\ \theta' M\theta}{\theta' A\theta} - 2\right)\right] e^{-\mu}\,{}_1F_1\Big(\frac{k}{2}; \frac{k}{2}+2; \mu\Big) \\ &+ \frac{\tau\ \mathrm{tr}\big(\mathbb{W}(V_{(2)}-V_{Full})\big)}{k(k-2)}\Big[\tau - 2(k-2)\Big] e^{-\mu}\,{}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}+1; \mu\Big),\end{aligned} \tag{16}$$

provided $k > 2$, where $A \equiv V^{1/2} G\,\mathbb{W}\,G' V^{1/2}$, and ${}_1F_1(.;.;.)$ *is the confluent hypergeometric function*

[4] We use $\liminf_{T\to\infty}$ in (15) because we do not make any assumptions for the moment convergence, as in Hansen (2014, 2016, 2017) who adopted Lemma 6.1.14 of Lehmann and Casella (1998) in a related context.

[5] We note that the one-step-ahead mean squared forecast error for the Stein-like combined estimator is MSFE $= \mathbb{E}\left(y_{T+1} - x_T'\widehat{\beta}_\alpha\right)^2 = \sigma^2_{(2)} + \mathbb{E}\left(\big(\widehat{\beta}_\alpha - \beta_{(2)}\big)' x_T x_T' \big(\widehat{\beta}_\alpha - \beta_{(2)}\big)\right)$. By choosing $\mathbb{W}$ accordingly, we use the asymptotic risk in (16) to approximate the second term on the right hand side of the MSFE. This along with $\sigma^2_{(2)}$ corresponds to the one-step-ahead MSFE. Therefore, minimizing the MSFE is equivalent to minimizing the asymptotic risk.

which is defined as ${}_1F_1(a;b;\mu) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n} \frac{\mu^n}{n!}$, *where* $(a)_n = a(a+1)\ldots(a+n-1)$, $(a)_0 = 1$, *and* $\mu = \theta' M\theta/2$ *is the non-centrality parameter.* ■

Proof: See Appendix A.2.

We note that $k > 2$ is the condition for the existence of the asymptotic risk in Theorem 2 in the sense that the exact moment of a ratio of quadratic forms involved in the derivation exists. Further, from Theorem 2 the asymptotic risk of the combined estimator is lower than that of the post-break estimator, if the terms inside the square brackets be negative. These terms are negative if $\text{tr}(\nu) > 2\ \lambda_{\max}(\nu)$, where $\nu \equiv \left(V_{(2)} - V_{Full}\right)^{1/2} \mathbb{W} \left(V_{(2)} - V_{Full}\right)^{1/2}$, and $k > 2$. For the special case that $\mathbb{W} = \left(V_{(2)} - V_{Full}\right)^{-1}$, both conditions are simplified 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 size, δ_1 in Eq. (7), and any break point, b_1.

Using Theorem 2, we can find the optimal τ, denoted by τ_{opt}, which minimizes the asymptotic risk. For $0 \leq \tau \leq 2\Big(\frac{\text{tr}\left(\mathbb{W}(V_{(2)}-V_{Full})\right)\left(\theta' M\theta\right)}{\theta' A\theta} - 2\Big)$, the τ_{opt} which depends on $\mathbb{W}$ is

$$\tau_{opt}(\mathbb{W}) = \frac{\text{tr}\Big(\mathbb{W}(V_{(2)} - V_{Full})\Big)\ \left(\theta' M\theta\right)}{\theta' A\theta} - 2. \tag{17}$$

Notice that $\tau_{opt}(\mathbb{W})$ is positive when $\text{tr}(\nu) > 2\ \lambda_{\max}(\nu)$. This is a necessary condition for the efficiency of the Stein-like combined estimator.

Remark 3: We note that the optimal shrinkage parameter in (17) depends on the unknown variances, $V_{(2)}, V_{Full}$, and θ. As shown in Appendix A.2, the optimal value in (17) can be replaced with $\tau^*(\mathbb{W}) \equiv \frac{\text{tr}(\nu)}{\lambda_{\max}(\nu)} - 2$ which is positive. Thus, by replacing a consistent estimation of the variances, the results of Theorem 2 holds so long as $\widehat{\tau}^*(\mathbb{W}) \xrightarrow{p} \tau^*(\mathbb{W})$ as $T \to \infty$. □

Remark 4: As b_1 increases, $V_{(2)}$ increases, consequently τ_{opt} increases. Besides, because H_T inversely depends on $V_{(2)}$, H_T decreases. Hence, a larger b_1 results in a bigger $\alpha = \frac{\tau_{opt}}{H_T}$. Thus the full-sample estimator, $\widehat{\beta}_{Full}$, gets a bigger weight when b_1 is larger. □

By plugging back the optimal $\tau_{opt}(\mathbb{W})$ in (17) into the asymptotic risk function in (16), we can derive the optimal risk. Theorem 3 summarizes the result for any $\mathbb{W} > 0$.

Theorem 3: *Under Assumptions 1-2, if* $0 \leq \tau \leq 2\Big(\frac{\text{tr}\left(\mathbb{W}(V_{(2)}-V_{Full})\right)\ \left(\theta' M\theta\right)}{\theta' A\theta} - 2\Big)$ *and* $\text{tr}(\nu) > 2\lambda_{\max}(\nu)$, *then the risk of the Stein-like combined estimator for any user specific choice of* $\mathbb{W} > 0$ *is*

$$\rho\big(\widehat{\beta}_{\alpha}, \mathbb{W}\big) = \rho\big(\widehat{\beta}_{(2)}, \mathbb{W}\big) - \frac{1}{k-2}\left[\frac{\Big[\operatorname{tr}\big(\mathbb{W}(V_{(2)} - V_{Full})\big)\big(\theta' M\theta\big) - 2\big(\theta' A\theta\big)\Big]^2}{(\theta' M\theta)\ (\theta' A\theta)}\right]\left[e^{-\mu}{}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}; \mu\Big)\right]$$

$$- \frac{1}{k-2}\left[\frac{\Big[\operatorname{tr}\big(\mathbb{W}(V_{(2)} - V_{Full})\big)\Big]^2 \big(\theta' M\theta\big)^2}{\big(\theta' A\theta\big)^2} - 4\right]\left[\frac{\theta' A\theta}{\theta' M\theta} - \frac{\operatorname{tr}\big(\mathbb{W}(V_{(2)} - V_{Full})\big)}{k}\right]$$

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

where $\rho\big(\widehat{\beta}_{(2)}, \mathbb{W}\big) = \operatorname{tr}\big(\mathbb{W}\, V_{(2)}\big)$. ■

Note that in Theorem 3, all the terms inside the square brackets are positive, and so the Stein-like combined estimator has a smaller asymptotic risk than the post-break estimator.

Corollary 3.1: *For the special case that* $\mathbb{W} = \big(V_{(2)} - V_{Full}\big)^{-1}$, *the asymptotic risk of the Stein-like combined estimator simplifies to*

$$\rho\big(\widehat{\beta}_{\alpha}, \mathbb{W}\big) = \rho\big(\widehat{\beta}_{(2)}, \mathbb{W}\big) - \big(k-2\big)\left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}; \mu\Big)\right], \tag{18}$$

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

Remark 5: For the special case when $\mathbb{W} = \big(V_{(2)} - V_{Full}\big)^{-1}$, then from (17) we have $\tau_{opt} = k - 2$ which is independent of any unknown parameters. This is the well-known results of James and Stein (1961). In this case, τ_{opt} is positive when the number of regressors is larger than two, $k > 2$. This choice of $\mathbb{W}$ is considered in Hansen (2017). □

Remark 6: Our results in Theorems 2 and 3 show the asymptotic risk of the Stein estimator, while Hansen (2017) provides its upper bound. □

Based on Corollary 3.1, the gain obtained by using the Stein-like combined estimator can be derived by calculating the percentage difference between $\rho\big(\widehat{\beta}_{(2)}, \mathbb{W}\big)$ and $\rho\big(\widehat{\beta}_{\alpha}, \mathbb{W}\big)$, as numerically demonstrated in Figure 1. Figure 1 shows the relationship between the break size in the coefficients (the horizontal axis) and the percentage change in asymptotic risks, $\frac{\rho(\widehat{\beta}_{(2)}, \mathbb{W}) - \rho(\widehat{\beta}_{\alpha}, \mathbb{W})}{\rho(\widehat{\beta}_{(2)}, \mathbb{W})}$, (vertical axis). For example, when the vertical axis shows the percentage difference equal to 50%, it means that, by using the Stein-like combined estimator instead of the post-break estimator, we can reduce the asymptotic risk by 50%. We draw the graphs for different break ratios in the error variance, q. Comparing the three lines in each panel of Figure 1, we note that, as the number of regressors, k, increases, the percentage difference between risks increases in favor of the Stein-like combined estimator. Also, when the error variance of the pre-break period is lower than error variance of the post-break period, $q < 1$, there is more gain from using the pre-break observations. Besides, for the case that the break point is near the end of the sample, $b_1 = 0.8$, the gain from the Stein-like combined estimator is higher than for 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, when b_1 is large, the post-break estimator performs poorly due to the lack of the observations after the break. Furthermore, when the break size in the coefficient increases (as bias increases), the risk of the Stein-like estimator gets closer to that of the post-break estimator. But still the Stein estimator always outperforms the post-break estimator.

2.3 Comparing the Stein with an alternative estimator by Pesaran et al. (2013)

In a similar context, Pesaran et al. (2013) (hereafter "PPP") propose a method that can reduce MSFE under the structural breaks by weighting the full-sample observations. Their weighted least squares estimator takes the form

$$\widehat{\beta}_{PPP} = (X'WX)^{-1}(X'WY), \tag{19}$$

where W is diagonal taking a value $w_{(1)}$ for the pre-break observations and another value $w_{(2)}$ for the post-break observations, i.e., $W = \operatorname{diag}\big(w_{(1)}, \dots, w_{(1)}, w_{(2)}, \dots, w_{(2)}\big)$. They derive the following optimal weights with $k \geq 1$ and stationary regressors:

$$\begin{cases} w_{(1)} = \frac{1}{T}\frac{1}{b_1+(1-b_1)(q^2+Tb_1\phi^2)}, \\ w_{(2)} = \frac{1}{T}\frac{q^2+Tb_1\phi^2}{b_1+(1-b_1)(q^2+Tb_1\phi^2)}, \end{cases} \tag{20}$$

where $\lambda = \beta_{(1)} - \beta_{(2)}$ is the break size in the regression coefficient, $\phi = \frac{x'_{T+1}\lambda}{\sigma_{(2)}(x'_{T+1}\Omega_{xx}^{-1}x_{T+1})^{1/2}}$, and $\Omega_{xx} = \mathbb{E}(x_t x_t')$ is a positive definite matrix. See Pesaran et al. (2013) for more details.

Given the diagonal form of W, we rewrite the PPP estimator in equation (19) as

$$\widehat{\beta}_{PPP} = \Lambda\widehat{\beta}_{(1)} + (I - \Lambda)\widehat{\beta}_{(2)}, \tag{21}$$

where $\Lambda = \left(\frac{w_{(1)}}{w_{(2)}}\ X'_{(1)}X_{(1)} +\ X'_{(2)}X_{(2)}\right)^{-1}\left(\frac{w_{(1)}}{w_{(2)}}\ X'_{(1)}X_{(1)}\right)$. Basically, the PPP estimator can be viewed as a combined estimator of the pre-break and the post-break estimators, with the combination matrix weight Λ.

To compare the Stein estimator and the PPP estimator, we first rewrite the PPP estimator in terms of the full-sample estimator and the post-break estimator, noting that $\widehat{\beta}_{Full} = \Gamma\widehat{\beta}_{(1)} + \big(I_k - \Gamma\big)\widehat{\beta}_{(2)}$, where $\Gamma = \big(\frac{1}{q^2}X'_{(1)}X_{(1)} + X'_{(2)}X_{(2)}\big)^{-1}\big(\frac{1}{q^2}X'_{(1)}X_{(1)}\big)$. Under the stationarity assumption of the regressors, the PPP estimator is

$$\widehat{\beta}_{PPP} = \frac{b_1 + (1-b_1)q^2}{b_1 + (1-b_1)(q^2 + T_1\phi^2)}\ \widehat{\beta}_{Full} + \left(1 - \frac{b_1 + (1-b_1)q^2}{b_1 + (1-b_1)(q^2 + T_1\phi^2)}\right)\ \widehat{\beta}_{(2)}, \tag{22}$$

where the combination weight is between zero and one.

Next, we consider a general combined estimator, denoted by $\widehat{\beta}_c$, as

$$\widehat{\beta}_c = w_c\ \widehat{\beta}_{Full} + (1-w_c)\ \widehat{\beta}_{(2)}, \qquad w_c \in [0\ 1]. \tag{23}$$

Denote the asymptotic risk for this estimator by $\rho(\widehat{\beta}_c, \mathbb{W})$, where $\mathbb{W} > 0$. We find the optimal value of w_c, denoted by w_c^*, by minimizing the asymptotic risk or equivalently minimizing the weighted mean squared error, which takes the following form

$$w_c^* = \frac{\mathrm{tr}\big(\mathbb{W}(V_{(2)} - V_{Full})\big)}{\theta' A\theta + \mathrm{tr}\big(\mathbb{W}(V_{(2)} - V_{Full})\big)}. \tag{24}$$

Consider a statistic $H_T = T\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}\big)'\,\mathbb{W}\,\big(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}\big)$, where H_T is the Hausaman statistic if $\mathbb{W} = \big(\widehat{V}_{(2)} - \widehat{V}_{Full}\big)^{-1}$. We note that $\mathbb{E}\left(H_T\right) = \theta' A\theta + \mathrm{tr}\big(\mathbb{W}(V_{(2)} - V_{Full})\big)$. Therefore, an unbiased estimator for the denominator of the weight in (24) is H_T. Thus, the feasible optimal combination weight, denoted by $\widehat{w}_c^*$, is

$$\widehat{w}_c^* = \frac{\mathrm{tr}\big(\mathbb{W}(\widehat{V}_{(2)} - \widehat{V}_{Full})\big)}{H_T}. \tag{25}$$

Hence, when the difference between the full-sample estimator and the post-break estimator is small (a small H_T), the combined estimator in (23) assigns a higher weight to the full-sample estimator which is more efficient. The opposite is true for a large break size (a large H_T). Consequently, the combined estimator in (23) is in a class of the Stein-like estimator that balances the trade-off between the bias and variance efficiency of the full-sample estimator. Let us call the estimator in (23) as the "general combined estimator". Thus, the asymptotic risk of the general combined estimator for $\mathbb{W} = \left(V_{(2)} - V_{Full}\right)^{-1}$ is[6]

$$\rho\left(\widehat{\beta}_c, \mathbb{W}\right) = \rho\left(\widehat{\beta}_{(2)}, \mathbb{W}\right) - \frac{k(k-4)}{k-2}\left[e^{-\mu}\ {}_1F_1\left(\frac{k}{2}-1; \frac{k}{2}; \mu\right)\right], \tag{26}$$

where its asymptotic risk is less than that of the post-break estimator if $k > 4$.[7] The detailed proof can be found in the supplementary online appendix. By comparing the asymptotic risk of the general combined estimator and that of the Stein-like combined estimator presented in Corollary 3.1, we see that

$$\rho\left(\widehat{\beta}_\alpha, \mathbb{W}\right) - \rho\left(\widehat{\beta}_c, \mathbb{W}\right) = -\frac{2(k-1)}{k-2}\left[e^{-\mu}\ {}_1F_1\left(\frac{k}{2}-1; \frac{k}{2}; \mu\right)\right] \leq 0, \tag{27}$$

if $k > 2$. Therefore, the Stein-like combined estimator always out-performs or performs as good as the general combined estimator if $k > 2$. For large break sizes (large μ), we expect to see an almost equal performance between the two estimators.

Looking at the PPP estimator in (22), it is clear that the PPP estimator can be seen as the general combined estimator in (23) with weight $w_c = \left(b_1 + (1-b_1)q^2\right)/\left(b_1 + (1-b_1)(q^2 + T_1\phi^2)\right)$. Thus, as we show in (27), the Stein-like combined estimator always out-performs the general combined estimator and therefore the PPP estimator as long as $k > 2$. This is confirmed in the simulation and empirical study in Sections 4-5.

3 Extension to multiple breaks

So far, we have talked about the case of having a single break, but in practice a time series model may be subject to multiple breaks. The case of multiple breaks is a straightforward extension of the previous section. The combined estimator is defined as the combination of the full-sample estimator and the estimator using observations after the most recent break point. For example, consider a model with two breaks (three regimes)

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{for} \quad 1 < t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{for} \quad T_1 < t \leq T_2 \\ x_t'\beta_{(3)} + \sigma_{(3)}\varepsilon_t & \text{for} \quad T_2 < t < T. \end{cases} \tag{28}$$

[6]This choice of $\mathbb{W}$ simplifies the calculations and allows us to compare the asymptotic risks between the Stein-like combined estimator and the general combined estimator.

[7]The condition under which the Stein-like combined estimator outperforms the post-break estimator is $k > 2$. See Corollary 3.1 above.

Thus the combined estimator is

$$\widehat{\beta}_\alpha = \alpha\widehat{\beta}_{Full} + (1-\alpha)\widehat{\beta}_{(3)}, \tag{29}$$

where α is defined as in equation (5), in which $H_T = T\big(\widehat{\beta}_{(3)} - \widehat{\beta}_{Full}\big)'\big(\widehat{V}_{(3)} - \widehat{V}_{Full}\big)^{-1}\big(\widehat{\beta}_{(3)} - \widehat{\beta}_{Full}\big)$ and $V_{(3)}$ is the asymptotic variance of the post-break estimator. The asymptotic risk of the Stein-like combined estimator is $\rho\big(\widehat{\beta}_\alpha, \mathbb{W}\big) = \mathbb{E}\Big[T\big(\widehat{\beta}_\alpha - \beta_{(3)}\big)'\,\mathbb{W}\,\big(\widehat{\beta}_\alpha - \beta_{(3)}\big)\Big]$. In order to derive the asymptotic risk, first we need to find the asymptotic distributions of the two estimators. Let $b_1 \equiv \lim_{T\to\infty}\big(\frac{T_1}{T}\big)$ and $b_2 \equiv \lim_{T\to\infty}\big(\frac{T_2}{T}\big)$ where $b_1 < b_2$. Under the local asymptotic framework, $\big(\beta_{(1)} - \beta_{(2)}, \beta_{(2)} - \beta_{(3)}\big) = \big(\frac{\delta_1}{\sqrt{T}}, \frac{\delta_2}{\sqrt{T}}\big)$. Thus, the distribution of the full-sample estimator is

$$\sqrt{T}\Big(\widehat{\beta}_{Full} - \beta_{(3)}\Big) \xrightarrow{d} \mathrm{N}\Big(Q^{-1}Q_1 b_1(\delta_1+\delta_2) + Q^{-1}Q_2(b_2-b_1)\delta_2, V_{Full}\Big),$$

where $X = \big(X'_{(1)}\ X'_{(2)}\ X'_{(3)}\big)'$ is $T \times k$, with $X_{(1)}, X_{(2)}, X_{(3)}$ representing the regressors in the three regimes, and $\Omega = \mathrm{diag}\big(\sigma^2_{(1)}I_{T_1}, \sigma^2_{(2)}I_{T_2-T_1}, \sigma^2_{(3)}I_{T-T_2}\big)$ is a $T\times T$ matrix. With two breaks, the distribution of the post-break estimator is

$$\sqrt{T}\Big(\widehat{\beta}_{(3)} - \beta_{(3)}\Big) \xrightarrow{d} \mathrm{N}\Big(0, V_{(3)}\Big), \tag{30}$$

where $V_{(3)} = \frac{1}{1-b_2}\ \operatorname*{plim}_{T\to\infty}\Big(\frac{X'_{(3)}\Omega^{-1}_{(3)}X_{(3)}}{T-T_2}\Big)^{-1} = \frac{1}{1-b_2}Q_3^{-1}$. Also, the notations $Q_1, Q_2, Q_3, \Omega_{(1)}, \Omega_{(2)}, \Omega_{(3)}$ are defined similarly to those in Section 2. As in Theorem 1, we write the joint asymptotic distribution of the full and post-break estimators as:

$$\sqrt{T}\begin{bmatrix}\widehat{\beta}_{Full} - \beta_{(3)} \\ \widehat{\beta}_{(3)} - \beta_{(3)}\end{bmatrix} \xrightarrow{d} V^{1/2}Z, \tag{31}$$

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 $V = \begin{bmatrix}V_{Full} & V_{Full} \\ V_{Full} & V_{(3)}\end{bmatrix}$.

Having the joint asymptotic distribution, we calculate the asymptotic risk similarly to Theorem 2. The point is that we consider the multiple breaks model as if it were a single break. The main difference between the two cases with multiple breaks and a single break is the bias term, θ.

We can extend this to the case of m breaks occuring at $\{T_1, \ldots, T_m\}$

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{if} \quad 1 \ < t \leq T_1, \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{if} \quad T_1 < t \leq T_2, \\ \quad \vdots & \\ x_t'\beta_{(m+1)} + \sigma_{(m+1)}\varepsilon_t & \text{if} \quad T_m < t < T. \end{cases} \tag{32}$$

In the case of m breaks, the bias vector θ (which is $2k \times 1$) can be written as

$$\theta = V^{-1/2} \begin{bmatrix} Q^{-1}\Big(Q_1 b_1(\delta_1 + \cdots + \delta_m) + \cdots + Q_m(b_m - b_{m-1})\delta_m\Big) \\ 0 \end{bmatrix}. \tag{33}$$

Remark 7: Under two breaks, one may think of other combined estimators, e.g., the combination of the full-sample estimator and the subsample estimator based on the second and third subsamples. But, this subsample estimator is not consistent for $\beta_{(3)}$. Also, because the full-sample estimator is the most efficient one, the efficiency of the combined estimator can not be enhanced by combining with this inconsistent subsample estimator using the second and third subsamples. Therefore, this combined estimator does not balance the trade-off between the bias and variance efficiency.[8] $\square$

4 Monte Carlo evidence on forecasting performance

This section presents Monte Carlo results on the forecasting performance of the Stein-like combined estimator, in comparison with the following seven methods: (1) the post-break method (labeled as "Postbk" in figures and tables); (2) the trade-off method ("Troff"); (3) weighted average of forecasts ("WA"); (4) the pooled forecast combination ("Pooled"); (5) cross validation ("CV"), (6) the full-sample forecast ("Full"), and (7) the estimator proposed by Pesaran et al. (2013) ("PPP"). The first five methods are used in Pesaran and Timmermann (2007).

[8] Another example is to combine the full-sample estimator, a subsample estimator using the second and third subsamples, and another subsample estimator using the third subsample. Although combining three estimators does not fit into our current setup of the Stein-like combined estimator, similarly to the above discussion, this combined estimator does not balance between bias and variance efficiency. This can be extended to more than two break cases as well. We leave the possibility of using multiple estimators for future works.

4.1 Simulation designs

First, we consider $m = 1$. Let $t = 1, ..., T$ with $T = 100$, $b_1 \in \{0.2, 0.4, 0.6, 0.8\}$, $q_1 \equiv \frac{\sigma_{(1)}}{\sigma_{(2)}} \in \{0.5, 1, 2\}$ and $k = 5$. The results with $k \in \{3, 8\}$ are available in the supplementary online appendix. The data generating process is

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{if} \quad 1 \ < t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{if} \quad T_1 < t \leq T, \end{cases} \tag{34}$$

where $x_t \sim \mathrm{N}(0, 1)$, $\varepsilon_t \sim$ i.i.d. $\mathrm{N}(0, 1)$, $\beta_{(2)}$ is a vector of ones, and $\beta_{(1)} = \beta_{(2)} + \frac{\delta_1}{\sqrt{T}}$. Note that the break size is determined by δ_1 and T. We consider different break sizes in the coefficients, $\beta_{(1)} - \beta_{(2)}$, ranging from 0 to 2 in increments of 0.1. To incorporate the uncertainty associated with the estimation of the unknown parameters, we assume that T_1, q, and break size are unknown and have to be estimated.[9] This setup is similar to that in Pesaran et al. (2013).

Next, we consider $m = 2$ with two breaks at $T_1 = \frac{T}{3}$ and $T_2 = \frac{2T}{3}$. The data generating process follows

$$y_t = \begin{cases} x_t'\beta_{(1)} + \sigma_{(1)}\varepsilon_t & \text{if} \quad 1 \ < t \leq T_1 \\ x_t'\beta_{(2)} + \sigma_{(2)}\varepsilon_t & \text{if} \quad T_1 < t \leq T_2 \\ x_t'\beta_{(3)} + \sigma_{(3)}\varepsilon_t & \text{if} \quad T_2 < t \leq T. \end{cases} \tag{35}$$

In order to consider different possibilities under the two breaks, we conduct various experiments, as summarized in Table 1. Experiment 1 has no breaks. Moderate breaks and large breaks are considered in Experiments 2 to 5. Also, we consider the cases that the direction of breaks changes from decreasing to increasing or vice versa (Experiments 6 to 9). Experiments 10 and 11 consider partial changes in the coefficients, i.e., $\beta_{(1)} = \beta_{(2)} \neq \beta_{(3)}$ and $\beta_{(1)} \neq \beta_{(2)} = \beta_{(3)}$, respectively. Finally, Experiments 12 and 13 represent the higher and lower post-break volatility, respectively.

We compare one-period ahead forecasts in MSFE, with forecasts based on the post-break observations as a benchmark. We report ratios of MSFEs relative to that of the forecast using post-break observations, so that for method i we have $RMSFE_i = MSFE_i/MSFE_{\mathrm{Postbk}}$. Thus, an RMSFE less than one has a lower MSFE than the post-break forecast, and an RMSFE exceeding

[9] We have also implemented these methods imposing the true break point and the results are qualitatively similar to those with the estimated break point. Because imposing the true break date is infeasible in practice, we only report the results with estimated break point here to save space.

one has a higher MSFE than the benchmark. The number of Monte Carlo replications is 5,000.

4.2 Simulation results

Figure 2 shows the Monte Carlo results of the single break case with different break sizes and different q. In this figure, we compare the MSFE of the Stein-like combined estimator with "Postbk", "Full", "PPP", "Troff" , "CV", "WA", and "Pooled".[10] The vertical axis shows the RMSFE for different methods while the horizontal axis shows the break size in the mean.

When $q = 0.5$, the error variance of the pre-break data is less than that of the post-break data. In such a case, using the pre-break data in forecasting can improve the forecast accuracy. When $q = 1$, there is no break in the error variance. When $q = 2$, the pre-break data is more volatile than the post-break data, and one may gain less from using the pre-break data. The results in Figure 2 confirm Theorems 2 and 3 that the Stein-like combined estimator is uniformly better than the post-break estimator. When the break size gets larger, the Stein-like combined estimator gets closer to the post-break estimator, but never performs worse than that. This is expected because as the break size gets larger, the Stein-like combined estimator assigns more weight to the post-break estimator and less weight to the full-sample estimator. This confirms that there is no cost in using the Stein-like combined estimator even under a large break size. The advantage of the Stein estimator relative to the post-break estimator is even more apparent when $q = 0.5$ than when $q = 1$ or $q = 2$.

The out-performance of the Stein-like combined estimator over the PPP estimator is also clear in Figure 2. The performance of the PPP estimator deteriorates under large break sizes, often even worse than the post-break estimator. The trade-off (Troff) method performs almost in line with the post-break method for $q = 2$, while it often performs worse than that for $q = 0.5$ and a large break. The performance of the cross validation (CV) approach deteriorates when T_1 gets close to T. The weighted average (WA) and pooled forecast combination (Pooled) methods perform better than the post-break estimator only for small break sizes. However, they perform poorly for large breaks. The full-sample method performs good under small break sizes, but as the break size increases, it

[10]See sections 3.2-3.5 in Pesaran and Timmermann (2007) for details of the Troff, CV, WA, and Pooled methods. For the last three methods, as used in Pesaran and Timmermann (2007), we set the size of the forecast evaluation window to $0.25T$ and size of the minimum estimation window to $0.1T$, in the simulation study in this section and the empirical study in the next section.

performs much worse than the post-break estimator. We note that the Stein, CV, Troff, WA, and Pooled methods generally perform better than the full-sample estimator, except for a small break size as expected. This mirrors the findings in Pesaran and Timmermann (2007).[11]

Table 2 shows the results for the two break cases under the experiments specified in Table 1, with $T = \{100, 200\}$ and $k = 5$. The results with $k \in \{3, 8\}$ are available in the supplementary online appendix. For comparison, we report the relative MSFE for different methods with estimated break dates under the label *estimated break dates* in columns 2-8 of Table 2. Besides, as the "CV", "WA" and "Pooled" methods can also be implemented without an estimation of the break dates, we report the MSFE of these methods without estimating the break dates as well. The results without estimating break dates, treating the break date as unknown, are reported under the label *unknown break dates* in columns 9-11 of Table 2.

The results in Table 2 follow the similar pattern of those for the single break case shown in Figure 2. We find that the Stein-like combined estimator has a lower MSFE than the post-break estimator and the PPP estimator in all experiments. Note that, from Theorem 3, the risk of the Stein-like combined estimator is a function of the bias term, θ. No matter whether the break size has an increasing or decreasing pattern, in all Experiments #2 to #11, the Stein estimator outperforms the post-break estimator. Similarly to the single break case, when $q < 1$ and so the pre-break sample is less volatile than the post-break sample (Experiment #12), the dominance of the Stein forecast is clearer relative to the post-break estimator.

When there is no break (Experiment #1), the full-sample estimator performs best as expected. However, other methods also perform quite well under the no break case. In other experiments, the full-sample estimator performs worst among all the methods, followed by the WA method that conditions on an estimate of the break dates. The WA method based on the estimated break dates performs rather poorly when $q = 1$. However, when the break only affects the variance (Experiments #12 and #13), this method performs quite well. The CV method performs better than the other methods from Pesaran and Timmermann (2007), however it performs worse than the post-break estimator in some cases.

[11] We note that the CV, WA and Pooled methods can also be implemented without estimating the break point, treating the break date as unknown. See Sections 3.3-3.5 in Pesaran and Timmermann (2007) for details. The results are similar to those reported in Figure 2, and figures for comparing these methods are available upon request. Table 2 reports such cases for multiple breaks.

In summary, the Stein-like combined estimator uniformly outperforms the post-break estimator (and most often other methods) in all experiments, while other methods performs worse than the post-break estimator in many experiments. The pattern of the results remains the same as we increase the sample size T from 100 to 200. The MSFE of the post-break estimator gets improved under the larger sample size, as there are more observations in the post-break sample. We note that, even with a very large sample size, the Stein-like combined estimator never performs worse than the post-break estimator. We have also compared the methods with $k = 3$ and $k = 8$ regressors in the supplementary online appendix, which shows that as the number of regressors increases, the RMSFE for the Stein-like combined estimator decreases even more.

5 Empirical analysis

This section presents an empirical application of our method. We consider forecasting output growth rate of various countries using a quarterly data set (2016 vintage) available with the GVAR toolbox, Mohaddes and Raissi (2018). To predict the output growth rate, we consider the following eight macroeconomics and financial predictors for each country: the lag of the real GDP (y_{t-1}), real equity prices (eq_t), real short term interest rate (r_t), the difference between the real long term interest rate and the real short term interest rate ($l_t - r_t$), and the corresponding country-specific foreign variables for each of the predictors. The foreign variables are constructed using rolling three year moving averages of the annual trade weights which are computed as shares of exports and imports for each country. The data set starts in 1979:Q2 and ends in 2016:Q4, and we focus on the following nine industrialized economies in which all have long term bond markets: Australia, Canada, France, Germany, Italy, Japan, Spain, UK, and USA.[12] Therefore, the h-step ahead linear forecasting model for output growth is

$$y_{t+h}^{(h)} = \mu_t + \beta_t^{(h)\prime} x_t + u_{t+h}^{(h)}, \tag{36}$$

where $x_t = (y_{t-1}, eq_t, r_t, l_t - r_t, y_{t-1}^*, eq_t^*, r_t^*, l_t^* - r_t^*)'$, in which a "star" indicates foreign variables.

We compute h-step-ahead forecasts ($h = 1, 4$) for different forecasting methods described in this paper, using both expanding and rolling windows. Each time that we expand the estimation

[12] These are the same countries studied in Pesaran et al. (2013).

window, we apply the Schwarz's Bayesian Information Criteria (BIC) to select predictors out of the eight predictors. In other words, we select a forecasting model using the BIC criterion from all 2^8 possible specifications. For the methods that require the knowledge of the break, we use the Bai and Perron (1998, 2003) method based on the significance level of 5% and trimming rate of 0.2. The rolling window forecasts is based on the most recent 10 years (40 quarters) of observations, the same window size used by Stock and Watson (2003) and Inoue et al. (2017). We also report the MSFE using the optimal window size proposed by Inoue et al. (2017), which is designed for smoothly time-varying parameters. Moreover, we present the forecasting results using the approach proposed by Clark and McCracken (2010) (labeled as "CM" in Table 3), which is the equally weighted average of the rolling and expanding window forecasts. In the supplementary online appendix, we compare the MSFEs for different methods based on the selected optimal window size using Inoue et al. (2017) method.[13]

In order to evaluate the performance of our proposed estimator, we compute its MSFE and compare it with those from the existing methods: the method proposed by Pesaran et al. (2013) ("PPP"), the five methods used in Pesaran and Timmermann (2007), namely, "Postbk", "Troff", "Pooled", "WA", "CV", the full-sample forecast ("Full"), the average window forecast proposed by Pesaran and Pick (2011) ("AveW"), and the forecast using the optimal window size proposed by Inoue et al. (2017) ("IJR"). We also test for equal forecast performance of different methods compared to the post-break forecast using a generalization of the panel version of the Diebold–Mariano test proposed by Pesaran et al. (2009).[14]

Table 3 gives cross-country averages of MSFEs for different methods. In the aggregation of the individual country MSFEs, we use both GDP with Purchasing Power Parity based weights (GDP-PPP) and the equal weights. The results of cross country averages with GDP-PPP scheme are presented in Panels A-C of Table 3 while the results for the equal weights are in Panels D-F. The first column of the table shows the forecast horizon, h. The table also shows the results of the panel version of Diebold and Mariano test statistic, indicated by asterisks. The 1% and 5% significance levels are denoted by ** and *, respectively.

Based on the results, forecasts using the Stein method provide improved forecasts over the post-

[13] See also Hong et al. (2021) regarding the optimal window size using nonparametric smoothing techniques.

[14] See also Appendix of Pesaran et al. (2013) for details of this test for cross-country aggregation weights.

break forecast for rolling, expanding, and CM approaches over different forecast horizons, no matter which cross country aggregation scheme is used. The improvement ranges from 3 to 8.7 percent for the rolling window, from 6.5 to 17.8 for the expanding window, and from 5.2 to 12.2 percent for the CM method. This improvement is statistically significant at 1% or 5% levels. Besides, the Stein forecasts most often perform better than other existing forecasting methods. With fixed rolling windows, other methods perform poorly relative to the Stein method, and sometimes even worse than the post-break forecasts. However, under expanding windows, their performance gets improved and we do not see them under-performing the post-break forecasts, except the full-sample forecasts for $h = 4$.

We have also applied our method in another empirical application, forecasting equity premium. The results are available in the supplementary online appendix, where it is shown that the Stein method outperforms the post-break forecast and other methods in forecasting equity premium using the data of Welch and Goyal (2008).

6 Conclusion

In this paper we introduce the Stein-like combined estimator of the full-sample estimator (using all the observations in the sample) and the post-break estimator (using the observations after the most recent break point). A common practice for forecasting under structural breaks may be to use the post-break estimator. We show that using the pre-break observations can improve the post-break estimator. The combined estimator uses the pre-break observations to reduce the variance of the forecast error at the cost of adding some bias. We show that the Stein-like combined estimator has a lower asymptotic risk than the post-break estimator. We compare the performance of our proposed Stein-like combined estimator with a range of alternative methods existing in the literature, in the simulation study and the empirical study. Our simulation results show that the Stein-like combined estimator uniformly outperforms the post-break estimator, for any break sizes and break points. The results from the empirical application for forecasting output growth rate of nine industrialized economies confirms that the Stein-like combined estimator performs significantly better than the post-break estimator, and most often better than other alternative methods.

References

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

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

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

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

Boot, T. and Pick A. (2020). "Does modeling a structural break improve forecast accuracy?" *Journal of Econometrics* 215, 35-59.

Castle, Jennifer L. and David F. Hendry (2019). *Modelling Our Changing World*. Palgrave Texts in Econometrics, Chapter 5 Detectives of Change: Indicator Saturation.

Clark, T. E. and McCracken, M. W. (2010). "Averaging forecasts from VARs with uncertain instabilities." *Journal of Applied Econometrics* 25, 5-29.

Clements, M.P. and Hendry D.F. (1998). *Forecasting Economic Time Series*. Cambridge University Press.

Clements, M.P. and Hendry D.F. (1999). *Forecasting Non-Stationary Economic Time Series*. The MIT Press.

Clements, M.P. and Hendry D.F. (2006). "Forecasting with breaks." In Elliott, G., C.W.J. Granger and Timmermann A. (Eds.), *Handbook of Economic Forecasting*, Vol. 1, 605-658.

Diebold, F.X. and Pauly P. (1987). "Structural change and the combination of forecasts." *Journal of Forecasting* 6, 21-40.

Hansen, B. (2009). "Averaging estimators for regressions with a possible structural break." *Econometric Theory* 25, 1498-1514.

Hansen, B. (2014). "Model averaging, asymptotic risk, and regressor groups." *Quantitative Economics* 5, 495-530.

Hansen, B. (2016). "Efficient shrinkage in parametric models." *Journal of Econometrics* 190, 115-132.

Hansen, B. (2017). "Stein-like 2SLS estimator." *Econometric Reviews* 36, 840-852.

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

Hendry, D. F., Johansen, S., and Santos, C. (2008). "Automatic selection of indicators in a fully saturated regression." *Computational Statistics* 33, 317–335. Erratum, 337–339.

Hong, Y., Sun, Y. and Wang, S. (2021) "Selection of an optimal rolling window in time-varying predictive regression." Working paper, Chinese Academy of Sciences.

Inoue, A., Jin, L. and Rossi, B. (2017). "Rolling window selection for out-of-sample forecasting with time-varying parameters." *Journal of Econometrics* 196, 55-67.

James, W. and Stein C. M. (1961). "Estimation with quadratic loss." *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 361–380.

Judge, G.G. and Bock M.E. (1978). *The Statistical Implications of Pretest and Stein-rule Estimators in Econometrics*. North-Holland, Amsterdam.

Lebedev, N. N. (1972). *Special Functions and their Applications*. Dover Publications, New York.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. 2nd ed., New York, Springer.

Mohaddes, K., and Raissi M. (2018). "Global VAR (GVAR) database, 1979Q2-2016Q4". https://doi.org/10.17863/CAM.27337.

Pesaran, M.H. and Timmermann A. (2002). "Market timing and return prediction under model instability." *Journal of Empirical Finance* 9, 495–510.

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

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

Pesaran, M.H. and Pick A. (2011). "Forecast combination across estimation windows." *Journal of Business and Economic Statistics* 29, 307–318.

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

Pesaran, M.H., Schuermann, T. and Smith, L.Vanessa, (2009). "Forecasting economic and financial variables with global VARs." *International Journal of Forecasting* 25, 642-675.

Rossi, B. (2013). "Advances in forecasting under instability." In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Vol. 2, Part B, Chapter 21, 1203 - 1324. Elsevier.

Saleh, A. K. Md. Ehsanses (2006). *Theory of Preliminary Test and Stein-type Estimation with Applications.* Wiley, Hoboken.

Stein, C. 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, 197–206.

Stock, J. H. and Watson, M. W. (2003). "Forecasting output and inflation: The role of asset prices." *Journal of Economic Literature* 41, 788–829.

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.

Timmerman, A. (2006). "Forecast combinations." In G. Elliott, C. W. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Vol. 1, 135-196. Elsevier.

Ullah, A. (1974). "On the sampling distribution of improved estimators for coefficients in linear regression." *Journal of Econometrics* 2, 143-150.

Welch, I. and Goyal A. (2008). "A Comprehensive look at the empirical performance of equity premium prediction." *Review of Financial Studies* 21, 1455-1508.

A Appendix: Mathematical details

A.1 Proof of Theorem 1

For the proof of Theorem 1, we derive the distributions of the full-sample and post-break estimators. First, the full-sample estimator is written as

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

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

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

Next, the distribution of the post-break estimator is

$$\begin{aligned}
\sqrt{T}\left(\widehat{\beta}_{(2)} - \beta_{(2)}\right) &= \left(\frac{X_{(2)}'\Omega_{(2)}^{-1}X_{(2)}}{T - Tb_1}\right)^{-1} \left(\frac{X_{(2)}'\Omega_{(2)}^{-1}U_{(2)}}{\sqrt{1-b_1}\sqrt{T - Tb_1}}\right) \\
&\xrightarrow{d} N\left(0, \frac{1}{1-b_1} Q_2^{-1}\right),
\end{aligned} \tag{A.3}$$

where $U_{(2)} = \sigma_{(2)}\left(\varepsilon_{T_1+1}, \ldots, \varepsilon_T\right)'$. Having these distributions, we can write the joint distribution and the proof of Theorem 1 is complete. ∎

A.2 Proof for Theorem 2

The asymptotic risk of the Stein-like combined estimator is

$$\begin{aligned}
\rho(\widehat{\beta}_\alpha, \mathbb{W}) &= \mathbb{E}\left[T\left(\widehat{\beta}_\alpha - \beta_{(2)}\right)' \mathbb{W} \; \left(\widehat{\beta}_\alpha - \beta_{(2)}\right)\right] \\
&= T\,\mathbb{E}\left[\left(\widehat{\beta}_{(2)} - \beta_{(2)}\right) - \alpha\left(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}\right)\right]' \mathbb{W} \left[\left(\widehat{\beta}_{(2)} - \beta_{(2)}\right) - \alpha\left(\widehat{\beta}_{(2)} - \widehat{\beta}_{Full}\right)\right] \\
&= \rho(\widehat{\beta}_{(2)}, \mathbb{W}) + \tau^2\,\mathbb{E}\left[(Z'MZ)^{-2} Z'AZ\right] - 2\tau\,\mathbb{E}\left[(Z'MZ)^{-1} Z'BZ\right],
\end{aligned} \tag{A.4}$$

where $A \equiv V^{1/2} G\, \mathbb{W}\, G' V^{1/2}$ and $B \equiv V^{1/2} G\, \mathbb{W}\, G_2' V^{1/2}$. The challenging part for calculation of this risk function is to take expectation from the noncentral chi-squared distribution, with noncentrality parameter equal to $\theta' M \theta / 2$. For calculating these expectations, we use the following Lemmas.

Lemma 1: Let $\chi_p^2(\mu)$ denote a noncentral chi-square random variable with the noncentral parameter μ and the degrees of freedom p. Besides, let p denote a positive integer such that $p > 2r$. Then

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

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 $(a)_n = a(a+1)\ldots(a+n-1)$ and $(a)_0 = 1$. See Ullah (1974). $\square$

Lemma 2: 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), pp. 262. $\square$

Lemma 3: Let the $T \times 1$ vector Z be normally distributed with mean vector θ and covariance matrix I_T, M be any $T \times T$ idempotent matrix with rank r, and A be any $T \times T$ matrix. We assume $\phi(\cdot)$ is a Borel measurable function. Then:

$$\begin{aligned}\mathbb{E}\left[\phi\big(Z'MZ\big)Z'AZ\right] &= \mathbb{E}\left[\phi\Big(\chi^2_{r+2}(\mu)\Big)\right]\mathrm{tr}(AM) + \mathbb{E}\left[\phi\Big(\chi^2_{r+4}(\mu)\Big)\right]\theta' MAM\theta \\ &+ \mathbb{E}\left[\phi\Big(\chi^2_{r}(\mu)\Big)\right]\mathrm{tr}\big(A - AM\big) + \mathbb{E}\left[\phi\Big(\chi^2_{r}(\mu)\Big)\right]\theta'\big(I_T - M\big)A\big(I_T - M\big)\theta \\ &+ \mathbb{E}\left[\phi\Big(\chi^2_{r+2}(\mu)\Big)\right]\Big(\theta' AM\theta + \theta' MA\theta - 2\theta' MAM\theta\Big),\end{aligned}$$

where $\mu \equiv \frac{\theta' M \theta}{2}$ is the non-centrality parameter.

Proof: Let P be an orthogonal matrix such that

$$PMP' = D = \begin{bmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & & \\ & & \vdots & \\ 0 & \ldots & 0 & d_T \end{bmatrix} = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{T-r} \end{bmatrix}; \quad d_i \in \{0, 1\}.$$

Define the $T \times 1$ vector $\omega = PZ$, which has a $N(P\theta, I_T)$ distribution. Therefore

$$\mathbb{E}\left[\phi(Z'MZ)Z'AZ\right] = \mathbb{E}\left[\phi(\omega'D\omega)\omega'C\omega\right],$$

where $C \equiv PAP' = [c_{ij}]$ is a $T \times T$ matrix. Let $\omega' = \begin{bmatrix}\omega_1 & \dots & \omega_T\end{bmatrix}$. The diagonal and off-diagonal elements of $\mathbb{E}\left[\phi(\omega'D\omega)\omega'C\omega\right]$ are

$$\begin{aligned}
\sum_{i=1}^{T}\sum_{j=1}^{T}\mathbb{E}\left[\phi\Big(\sum_{s=1}^{T} d_s\omega_s^2\Big)\omega_i\omega_j c_{ij}\right] &= \sum_{i=1}^{T}\mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\omega_i^2 + \sum_{\substack{s=1,\\ s\neq i}}^{T} d_s\omega_s^2\Big)\omega_i^2\; c_{ii}\Big|\omega_s, s\neq i\right]\right] \\
&+ \sum_{i=1}^{T}\sum_{\substack{j=1,\\ j\neq i}}^{T}\mathbb{E}\left[\mathbb{E}\left[\phi\Big(d_i\omega_i^2 + d_j\omega_j^2 + \sum_{\substack{s=1,\\ s\neq i,j}}^{T} d_s\omega_s^2\Big)\omega_i\omega_j\; c_{ij}\Big|\omega_s, s\neq i, s\neq j\right]\right] \\
&= \sum_{i=1}^{T} c_{ii}\begin{cases}\mathbb{E}\left[\phi(\chi^2_{r+2}(\mu))\right] + (P_i'\theta)^2\,\mathbb{E}\left[\phi(\chi^2_{r+4}(\mu))\right], & \text{if } d_i = 1\\ \mathbb{E}\left[\phi(\chi^2_{r}(\mu))\right] + (P_i'\theta)^2\,\mathbb{E}\left[\phi(\chi^2_{r}(\mu))\right], & \text{if } d_i = 0.\end{cases} \\
&+ \sum_{i=1}^{T}\sum_{j=1}^{T} P_i'\theta P_j'\theta c_{ij}\begin{cases}\mathbb{E}\left[\phi(\chi_{r+4}(\mu))\right], & \text{if } d_i = d_j = 1\\ \mathbb{E}\left[\phi(\chi_{r}(\mu))\right], & \text{if } d_i = d_j = 0\\ \mathbb{E}\left[\phi(\chi_{r+2}(\mu))\right], & \text{if } d_i = 1 \text{ and } d_j = 0\end{cases} \\
&= \mathbb{E}\left[\phi\big(\chi^2_{r+2}(\mu)\big)\right]\mathrm{tr}(CD) + \mathbb{E}\left[\phi\big(\chi^2_{r+4}(\mu)\big)\right]\;\left[\theta'P'DCDP\theta\right] \\
&+ \mathbb{E}\left[\phi\big(\chi^2_{r}(\mu)\big)\right]\mathrm{tr}\big[C(I_T - D)\big] + \mathbb{E}\left[\phi\big(\chi^2_{r}(\mu)\big)\right]\left[\theta'P'(I_T - D)C(I_T - D)P\theta\right] \\
&+ \mathbb{E}\left[\phi(\chi^2_{r+2}(\mu))\right]\left[\theta'P'D(C - \mathrm{diag}(C))(I_T - D)P\theta\right].
\end{aligned}$$

where the third equality holds by Lemma 1 of Appendix B.1 in Judge and Bock (1978). Substituting for D and C, Lemma 3 is proved. □

Using Lemmas 1 - 3, we calculate the expectations in (A.4). Let the non-centrality parameter of the chi-squared distribution based on Lemma 3 be defined as $\mu \equiv \frac{\theta'M\theta}{2}$. For clarity, we focus on the second and third terms in equation (A.4) one by one. The second term can be simplified as

$$\begin{aligned}
\mathbb{E}\left[(Z'MZ)^{-2}Z'AZ\right] &= \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-2}\mathrm{tr}\Big(AM\Big) + \mathbb{E}\left[\chi^2_{k+4}(\mu)\right]^{-2}\Big(\theta'MAM\theta\Big) \\
&+ \mathbb{E}\left[\chi^2_{k}(\mu)\right]^{-2}\mathrm{tr}\Big(A - AM\Big) + \mathbb{E}\left[\chi^2_{k}(\mu)\right]^{-2}\Big(\theta'(I_{2k} - M)A(I_{2k} - M)\theta\Big) \\
&+ \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-2}\Big(\theta'AM\theta + \theta'MA\theta - 2\theta'MAM\theta\Big) \\
&= \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-2}\mathrm{tr}(A) + \mathbb{E}\left[\chi^2_{k+4}(\mu)\right]^{-2}(\theta'A\theta) \\
&= \left[\frac{1}{4}\,e^{-\mu}\,\frac{\Gamma(\frac{k}{2} - 1)}{\Gamma(\frac{k}{2} + 1)}\;{}_1F_1\Big(\frac{k}{2} - 1; \frac{k}{2} + 1; \mu\Big)\right]\mathrm{tr}(A)
\end{aligned}$$

$$
\begin{aligned}
&+ \left[\frac{1}{4}\, e^{-\mu}\, \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+2)}\; {}_1F_1\Big(\frac{k}{2}; \frac{k}{2}+2; \mu\Big)\right] (\theta' A\theta) \\
&= \left[\frac{\theta' A\theta}{(k-2)\theta' M\theta}\right] \left[e^{-\mu}\; {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}; \mu\Big)\right] \\
&- \left[\frac{\theta' A\theta}{(k-2)\theta' M\theta} - \frac{\mathrm{tr}\big(\,\mathbb{W}(V_{(2)} - V_{Full})\big)}{k(k-2)}\right] \left[e^{-\mu}\; {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}+1; \mu\Big)\right], \quad \text{(A.5)}
\end{aligned}
$$

where $\mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-2}$ can be obtained for $k > 2$ from Lemma 1, and the last equality follows from using Lemma 2 several times, $AM = MA = A$, and $MAM = A$. Note that M is an idempotent matrix. The third term in equation (A.4) can be simplified as

$$
\begin{aligned}
\mathbb{E}\Big[(Z'MZ)^{-1} Z'BZ\Big] &= \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-1} \mathrm{tr}\Big(BM\Big) + \mathbb{E}\left[\chi^2_{k+4}(\mu)\right]^{-1} \Big(\theta' MBM\theta\Big) \\
&+ \mathbb{E}\left[\chi^2_{k}(\mu)\right]^{-1} \mathrm{tr}\Big(B - BM\Big) + \mathbb{E}\left[\chi^2_{k}(\mu)\right]^{-1} \Big(\theta'(I_{2k} - M)B(I_{2k} - M)\theta\Big) \\
&+ \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-1} \Big(\theta' BM\theta + \theta' MB\theta - 2\theta' MBM\theta\Big) \\
&= \mathbb{E}\left[\chi^2_{k+2}(\mu)\right]^{-1} \Big(\mathrm{tr}\big(B\big) - \theta' A\theta\Big) + \mathbb{E}\left[\chi^2_{k+4}(\mu)\right]^{-1} \Big(\theta' A\theta\Big) \\
&= \left[\frac{1}{2}\, e^{-\mu}\, \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+1)}\; {}_1F_1\Big(\frac{k}{2}; \frac{k}{2}+1; \mu\Big)\right] \Big(\mathrm{tr}\big(\,\mathbb{W}(V_{(2)} - V_{Full})\big) - \theta' A\theta\Big) \\
&+ \left[\frac{1}{2}\, e^{-\mu}\, \frac{\Gamma(\frac{k}{2}+1)}{\Gamma(\frac{k}{2}+2)}\; {}_1F_1\Big(\frac{k}{2}+1; \frac{k}{2}+2; \mu\Big)\right] \Big(\theta' A\theta\Big) \\
&= e^{-\mu} \left[\frac{\mathrm{tr}\big(\,\mathbb{W}(V_{(2)} - V_{Full})\big)}{k-2} - \frac{2\big(\theta' A\theta\big)}{\big(\theta' M\theta\big)\big(k-2\big)}\right] \\
&\times \left[{}_1F_1(\frac{k}{2}-1; \frac{k}{2}; \mu) - {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}+1; \mu\Big)\right], \qquad \text{(A.6)}
\end{aligned}
$$

where the last equality is obtained by using Lemma 2 several times, $BM = A$, $MB = B$, $B\theta = 0$, and $MBM = A$. Finally, plugging (A.5) and (A.6) into (A.4) produces

$$
\begin{aligned}
\rho\big(\widehat{\beta}_\alpha, \mathbb{W}\,\big) = \rho\big(\widehat{\beta}_{(2)}, \mathbb{W}\,\big) &+ \frac{\tau^2}{k-2}\left[\frac{\theta' A\theta}{\theta' M\theta}\right] \left[e^{-\mu}\; {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}; \mu\Big) - e^{-\mu}\; {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}+1; \mu\Big)\right] \\
&+ \frac{\tau^2\, \mathrm{tr}\big(\,\mathbb{W}(V_{(2)} - V_{Full})\big)}{k(k-2)} \left[e^{-\mu}\; {}_1F_1\Big(\frac{k}{2}-1; \frac{k}{2}+1; \mu\Big)\right]
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{2\ \tau}{k-2}\left[\mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)-\frac{2\ \theta' A\theta}{\theta' M\theta}\right]\left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big)\right]\\
&+\frac{4\ \tau}{k-2}\left[\frac{\mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)}{k}-\frac{\theta' A\theta}{\theta' M\theta}\right]\left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right]\\
&=\rho(\widehat{\beta}_{(2)},\mathbb{W})+\frac{\tau^2}{k-2}\left[\frac{\theta' A\theta}{\theta' M\theta}\right]\left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big)-e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right]\\
&+\frac{\tau^2\ \mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)}{k(k-2)}\left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right]\\
&-\frac{2\ \tau}{k-2}\left[\mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)-\frac{2\ \theta' A\theta}{\theta' M\theta}\right]\left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu\Big)-e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right]\\
&-\frac{2\ \tau}{k-2}\left[\frac{(k-2)\ \mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)}{k}\right]\left[e^{-\mu}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right]\\
&=\rho(\widehat{\beta}_{(2)},\mathbb{W})\\
&+\frac{2\ \tau\ \theta' A\theta}{k(k+2)\ \theta' M\theta}\left[\tau-2\left(\frac{\mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)\ \theta' M\theta}{\theta' A\theta}-2\right)\right]\left[\mu\ e^{-\mu}{}_1F_1\Big(\frac{k}{2};\frac{k}{2}+2;\mu\Big)\right]\\
&+\frac{\tau\ \mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)}{k(k-2)}\Big[\tau-2(k-2)\Big]\left[e^{-\mu}{}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu\Big)\right],
\end{aligned}
\tag{A.7}
$$

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:

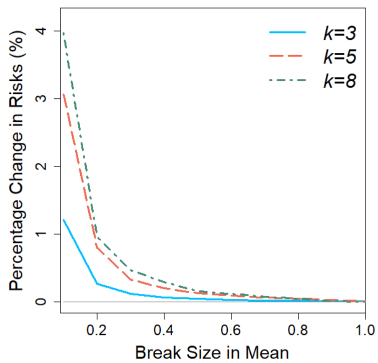
(I) $0\le\tau<2\left(\frac{\mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)\ \theta' M\theta}{\theta' A\theta}-2\right)$,

(II) $0\le\tau<2(k-2)$,

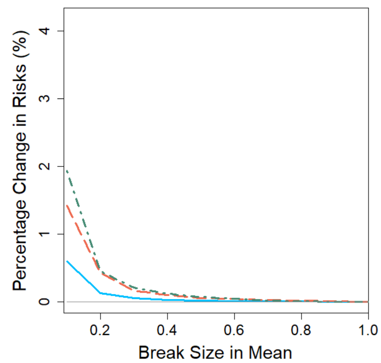
where the upper bound in condition (I) is greater than zero if

$$
\begin{aligned}
&\mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)>\operatorname*{Sup}_{\vartheta}\ \frac{2\ \theta' A\theta}{\theta' M\theta},\\
&\mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)>\operatorname*{Sup}_{\vartheta}\ \frac{2\ \vartheta'(V_{(2)}-V_{Full})^{1/2}\,\mathbb{W}(V_{(2)}-V_{Full})^{1/2}\vartheta}{\vartheta'\vartheta},\\
&\mathrm{tr}\big(\ \mathbb{W}(V_{(2)}-V_{Full})\big)>\lambda_{\max}\big((V_{(2)}-V_{Full})^{1/2}\,\mathbb{W}(V_{(2)}-V_{Full})^{1/2}\big),
\end{aligned}
\tag{A.8}
$$

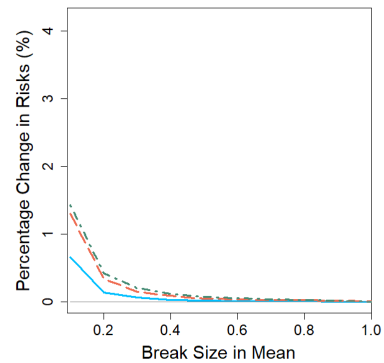
where $\vartheta\equiv(V_{(2)}-V_{Full})^{-1/2}G'V^{1/2}\theta$. Besides, the upper bound in condition (II) is positive if the number of regressors is greater than 2, i.e., $k>2$. This completes the proof of Theorem 2. ∎



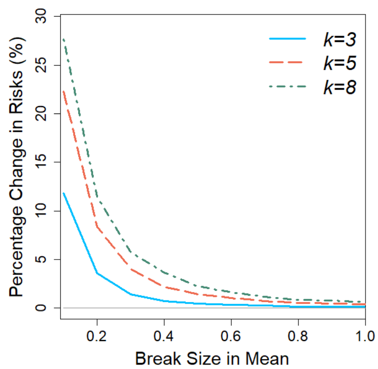
(a) $T_1 = 20, \sigma_{(1)} = 1 < \sigma_{(2)} = 2$



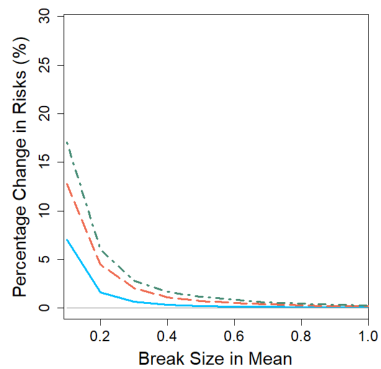
(b) $T_1 = 20, \sigma_{(1)} = \sigma_{(2)} = 1$



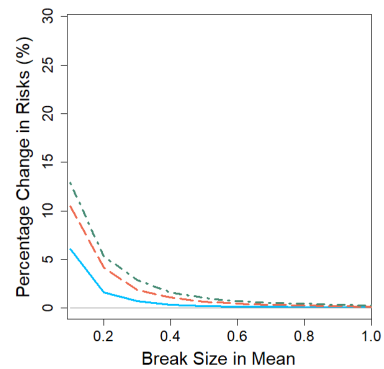
(c) $T_1 = 20, \sigma_{(1)} = 2 > \sigma_{(2)} = 1$



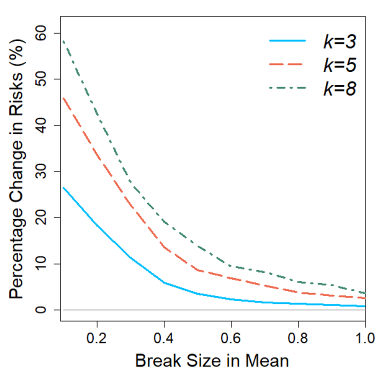
(d) $T_1 = 50, \sigma_{(1)} = 1 < \sigma_{(2)} = 2$



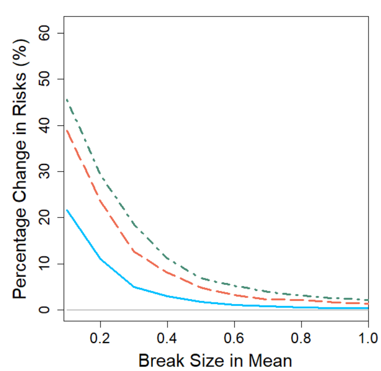
(e) $T_1 = 50, \sigma_{(1)} = \sigma_{(2)} = 1$



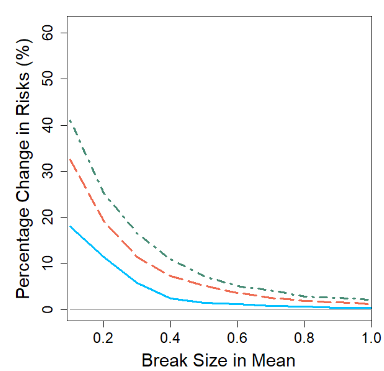
(f) $T_1 = 50, \sigma_{(1)} = 2 > \sigma_{(2)} = 1$



(g) $T_1 = 80, \sigma_{(1)} = 1 < \sigma_{(2)} = 2$

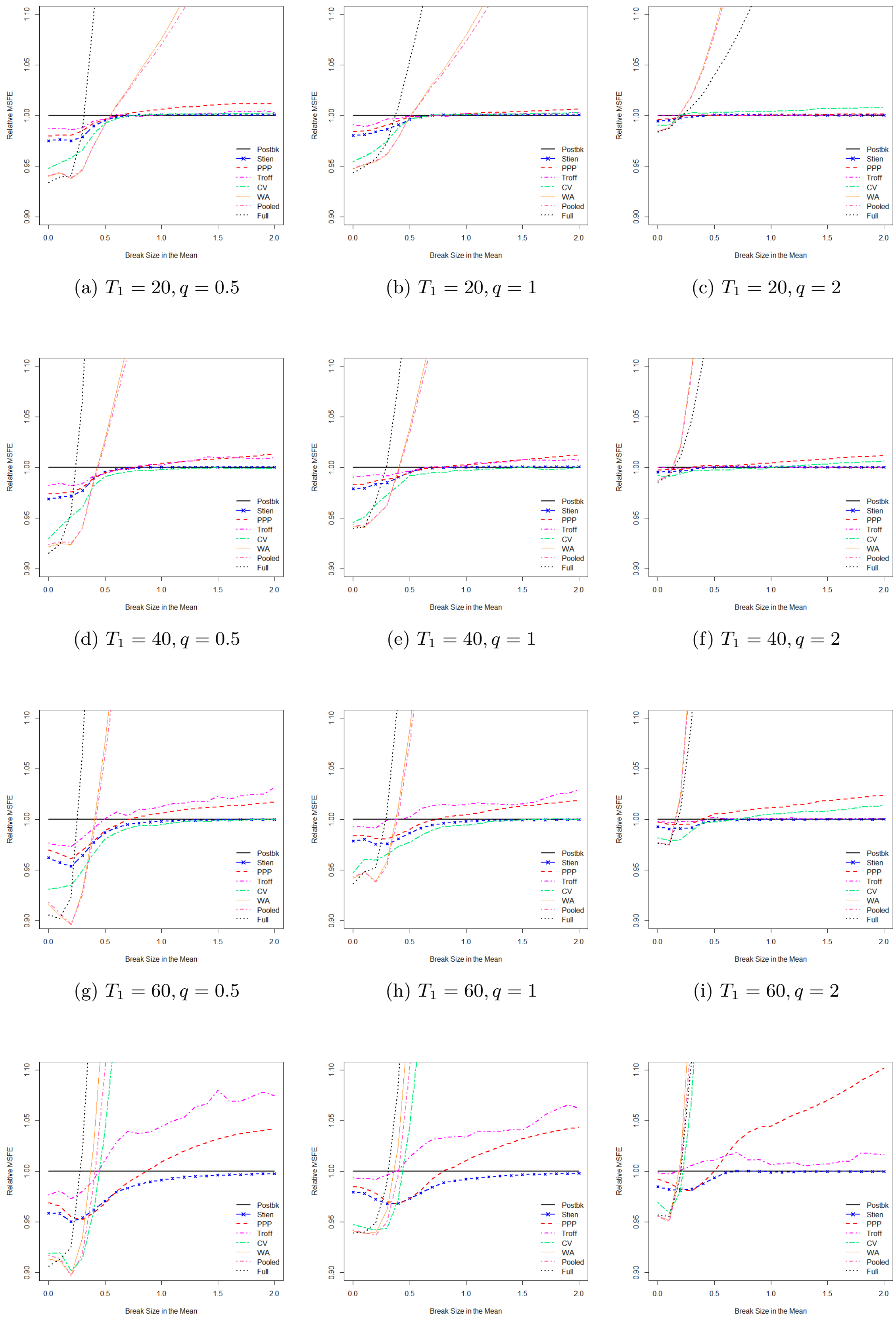


(h) $T_1 = 80, \sigma_{(1)} = \sigma_{(2)} = 1$



(i) $T_1 = 80, \sigma_{(1)} = 2 > \sigma_{(2)} = 1$

Figure 1: Risk-gain(%) between the Stein-like combined estimator and the post break estimator, when $T = 100$.



(a) $T_1 = 20, q = 0.5$

(b) $T_1 = 20, q = 1$

(c) $T_1 = 20, q = 2$

(d) $T_1 = 40, q = 0.5$

(e) $T_1 = 40, q = 1$

(f) $T_1 = 40, q = 2$

(g) $T_1 = 60, q = 0.5$

(h) $T_1 = 60, q = 1$

(i) $T_1 = 60, q = 2$

(j) $T_1 = 80, q = 0.5$

(k) $T_1 = 80, q = 1$

(l) $T_1 = 80, q = 2$

Figure 2: Simulation results with $T = 100, k = 5$.

Table 1: Break specifications ($m = 2$)

Experiments	$\beta_{(1)}$	$\beta_{(2)}$	$\beta_{(3)}$	$\sigma_{(1)}$	$\sigma_{(2)}$	$\sigma_{(3)}$
#1 : No break	1	1	1	1	1	1
#2 : Moderate break in coefficients (decrease)	1.3	1	0.7	1	1	1
#3 : Large break in coefficients (decrease)	2	1	0	1	1	1
#4 : Moderate break in coefficients (increase)	0.7	1	1.3	1	1	1
#5 : Large break in coefficients (increase)	0	1	2	1	1	1
#6 : Moderate decreasing and increasing breaks	1.6	1	1.6	1	1	1
#7 : Large decreasing and increasing breaks	2	1	2	1	1	1
#8 : Moderate increasing and decreasing breaks	0.4	1	0.4	1	1	1
#9 : Large increasing and decreasing breaks	0	1	0	1	1	1
#10 : No break, then increasing break	1	1	1.5	1	1	1
#11 : Decreasing, then no break	1.5	1	1	1	1	1
#12 : Higher post-break volatility	1	1	1	0.5	1	2
#13 : Lower post-break volatility	1	1	1	2	1	0.5

Table 2: Simulation results for multiple breaks with $k = 5$

	Estimated break dates							Unknown break dates		
Experiments	Stein	PPP	Troff	CV	WA	Pooled	Full	CV	WA	Pooled
$T = 100$										
#1	0.991	0.993	0.998	0.976	0.972	0.972	0.970	1.023	0.988	1.008
#2	0.995	0.997	1.002	0.980	1.080	1.065	1.286	0.961	1.028	0.970
#3	0.999	1.247	1.019	1.050	3.028	2.243	5.239	1.043	3.009	1.640
#4	0.988	0.991	0.996	1.192	1.016	1.017	1.197	1.193	0.968	0.938
#5	0.999	1.277	1.019	4.941	2.070	2.148	5.020	4.941	2.042	1.576
#6	0.972	1.016	0.998	1.139	1.022	1.014	0.981	1.199	1.032	0.961
#7	0.982	1.045	1.016	1.883	1.417	1.395	1.316	1.900	1.419	1.181
#8	0.975	1.012	1.002	0.973	1.054	1.031	1.011	0.939	1.071	0.970
#9	0.986	1.042	1.014	1.051	1.557	1.452	1.394	1.042	1.560	1.219
#10	0.978	0.982	1.001	1.245	1.058	1.050	1.232	1.247	1.042	0.967
#11	0.995	0.999	0.999	0.986	1.016	1.014	1.123	1.010	0.985	0.995
#12	0.988	0.991	0.996	0.965	0.962	0.964	0.958	1.005	0.990	1.016
#13	0.998	0.999	1.000	0.997	0.993	0.992	0.992	0.944	0.854	0.837
$T = 200$										
#1	0.998	0.998	0.998	0.992	0.991	0.991	0.990	1.014	0.998	1.006
#2	1.000	1.008	1.004	0.992	1.118	1.104	1.341	0.974	1.077	1.011
#3	1.000	1.110	1.011	1.036	3.209	2.366	5.606	1.034	3.204	1.734
#4	1.000	1.002	1.003	1.335	1.097	1.098	1.337	1.335	1.055	1.006
#5	1.000	1.108	1.009	5.610	2.290	2.383	5.637	5.610	2.278	1.741
#6	0.991	1.009	1.001	1.375	1.170	1.159	1.128	1.381	1.171	1.071
#7	0.997	1.018	1.007	2.175	1.594	1.574	1.493	2.175	1.594	1.307
#8	0.991	1.008	1.004	0.996	1.184	1.153	1.130	0.991	1.185	1.069
#9	0.997	1.014	1.007	1.036	1.681	1.567	1.499	1.036	1.681	1.306
#10	0.996	1.000	1.009	1.464	1.199	1.193	1.456	1.464	1.197	1.090
#11	0.999	1.002	1.002	0.997	1.030	1.028	1.123	1.012	1.003	1.005
#12	0.996	0.997	0.996	0.989	0.987	0.988	0.985	1.009	0.999	1.011
#13	0.999	0.999	1.000	1.003	1.000	1.000	0.999	0.976	0.926	0.914

Note: This table reports the results of the relative MSFE for different methods. All MSFEs are reported relative to the associated MSFE based on the post-break sample. The first column shows the experiment numbers which represents the specification of breaks based on Table 1.

Table 3: Empirical results for forecasting real output growth averaged across countries

	Estimated break dates								Unknown break dates				
h	Postbk	Stein	PPP	Troff	CV	WA	Pooled	Full	CV	WA	Pooled	AveW	IJR
	Panel A: Rolling, GDP weighted average												
1	0.062	0.057**	0.060**	0.063	0.065	0.058	0.058	0.060	0.078	0.058	0.128	0.097	0.060
4	0.239	0.232*	0.258	0.251	0.253	0.306	0.293	0.332	0.215*	0.273	3.387	0.296	0.329
	Panel B: Expanding, GDP weighted average												
1	0.066	0.056**	0.063	0.064	0.059	0.051*	0.052*	0.053*	0.059	0.050*	0.049**	0.049**	-
4	0.406	0.381**	0.395	0.395	0.397	0.380	0.373	0.423	0.375	0.360*	0.303**	0.303**	-
	Panel C: CM, GDP weighted average												
1	0.064	0.057*	0.061	0.064	0.062	0.055	0.055	0.056	0.068	0.054	0.089	0.073	-
4	0.322	0.306*	0.327	0.323	0.325	0.343	0.333	0.377	0.295**	0.317	1.845	0.299	-
	Panel D: Rolling, Equally weighted average												
1	0.055	0.051*	0.054	0.057	0.057	0.053	0.053	0.055	0.068	0.053	0.140	0.107	0.051
4	0.269	0.254**	0.284	0.280	0.280	0.314	0.303	0.323	0.243	0.284	2.913	0.388	0.304
	Panel E: Expanding, Equally weighted average												
1	0.057	0.049**	0.055**	0.057	0.053	0.046*	0.046*	0.047	0.055	0.045*	0.044**	0.044**	-
4	0.376	0.344**	0.377	0.374	0.365	0.346	0.338*	0.381	0.348	0.333*	0.279**	0.279**	-
	Panel F: CM, Equally weighted average												
1	0.056	0.050*	0.055*	0.057	0.055	0.049	0.050	0.051	0.062	0.049	0.092	0.075	-
4	0.323	0.299*	0.331	0.327	0.323	0.330	0.320	0.352	0.296**	0.308	1.596	0.334	-

Note: This table reports the $100 \times$ MSFE for different forecasting methods. The out-of-sample forecast period is 1994:Q1-2016:Q4. h in the first column shows the forecast horizon. In the heading of the table, Stein shows the results for our proposed Stein-like combined estimator, PPP is the one proposed by Pesaran et al. (2013), Postbk, Troff, CV, WA, and Pooled are the five methods used in Pesaran and Timmermann (2007), AveW is the method proposed by Pesaran and Pick (2011) with $T(1 - w_{\min}) + 1$ windows and $w_{\min} = 0.1$, and IJR is the optimal window method proposed by Inoue et al. (2017) with $\underline{R} = \max(1.5T^{2/3}, 20)$ and $\overline{R} = \min(4T^{2/3}, T - h)$. Panels A-C report the results based on the GDP weighted average with rolling window of most recent 40 observations, expanding window, and the Clark and McCracken (2010) method (CM), respectively, while Panels D-F report the results of equally weighted average across countries. The GDP weighted average uses weights $w_i = Y_i/(\sum_{j=1}^{s} Y_j)$, where Y_i is the 2008 GDP in purchasing power terms for country i available from the GVAR data base and $s = 9$ is the number of countries. The equal weights average uses $w_i = 1/s$. An asterisk denote forecast that is significantly better than that obtained from the post-break forecasts according to the panel Diebold–Mariano test statistic. The 1% and 5% significance levels are denoted by ** and *, respectively.