Forecast Combinations with an Application on Realized Volatility Forecasting

Chao Cheng[1], Man Wang[2], and Kun Chen*[3]

[1]Department of Mathmatical Sciences, Tsinghua University, Beijing, China

[2]Department of Finance, Donghua University, Shanghai, China

[3]School of Statistics, Southwestern University of Finance and Economics, Chengdu, China

Abstract

In order to improve the forecasting performance in the presence of structural breaks, we propose an exponential weighted forecast combination method which averages forecast based on different estimation windows in this paper. Comparing with previous forecast combination methods, our proposed method places heavier weights on the most recent estimation windows and the information regarding to the last break date is utilized to calculate its parameter. For a linear regression model with a single break in the slope parameter, our proposed method has a lower mean square forecast error compared with the initial regression model based all the sample and equally weighted forecast combination method. A series of simulation studies and empirical analysis show that the forecast combination method based on the exponential weighting scheme produces more accurate forecasts than those by using the ordinary least square method based on the whole observations, or those by using the same forecast combination method with other weighting schemes.

Key words: Exponential weight, forecast combination, model confidence set, realized volatility, volatility forecasting.

1 Introduction

In macroeconomics and finance, structural breaks are unexpected changes in the parameters over various time points of the underlying data generating process (DGP), which may result in large forecasting unreliability and errors in general [1, 2]. That is because the presence of structural breaks, which are usually difficult to detect and measure by statistial methods, could damage the forecasts. Recently, It is widely acknowledged that the structural break is an important source of forecasting failure in empirial researches [3, 4, 5, 6]. How

*The authorship may change.

to improve the forecasting results in the presence of structural breaks is an open problem waiting to be solved. And now, there are two popular strategies to deal with the occurrence of structural breaks.

The first popular strategy to conduct forecasting under structural breaks is a two-step procedure. One first estimate the break points and then forecast according to the model that estimated based on the after-last-break data. Accordingly, plenty of statistical tests have been proposed to detect the occurrence, number and locations the structural breaks, such as CUSUM test [7], MOSUM test [8], Chow test [9], etc. However, the estimation of the break points by these statistical tests can be inaccurate; especially for the situation of occurrence of multiple structural breaks, the structural-break detection procedures may ignore the last break points and mistakenly consider a previous break point as the last one. Consequently, the forecasting bias based on this two-step method may be highly biased. Moreover, even if the the last-break point is correctly specified, the after-last-break sample size is usually not large enough, which could lead to a large forecasting variance. In fact, Pesaran and Timmermann (2007) [10] shows that incorporating some pre-last-break data to conduct forecasting can reduce the out-of-sample mean squared forecasting error (MSFE), which implies the pre-break information could be useful.

An alternative strategy is the forecast combination across estimation windows with different lengths, which make use of the pre-break information to achieve a bias-variance trade-off. Specifically, the final forecast is a weighted sum of forecasts which are from the same model but estimated over different estimation windows. Since some estimation windows could include some pre-break data if breaks have occurred, the forecasting by forecast combination is typically biased, but incorporating the pre-break data could also reduce the forecasting variance, thus achieve a bias-variance trade-off. One can reduce the MSFE by approperately choosing the weighting schemes. The weighting schemes can be determined by some fixed functions, without incorporating any information from the target time series, such as equally weight and location weight; or some structural-break information can be utilized to proposed the weighting schemes, such as ROC weight [11], which arranges the weights based on the statistics of the reversed odered CUSUM test.

In this article, we propose a new weighting scheme called exponential weighting scheme, which assigns exponentially decaying weights to forecast components. Our proposed method has two distingushed features. First, as its name demonstrated, the exponential weighting scheme places more weight on forecasts which are derived from more recent samples. Typically, more recent information is usually more relevant for forecasting, thus our proposed method is expected to be less biased. Second, the information regarding structural break dates is utilized to determine the decaying parameter in the exponential weighting scheme, in order to reduce the forecating variance. A series of simulated experiments were developed to compare proposed exponential weighted forecast combination method with other existing weighting schemes and two-stage method. In particular, the simulation experiments indicate that the forecast combination methods could result in a lower MSFE in comparision with the single model based on the whole time horizon and the two stage method. And among the forecast combination method, our proposed exponentially weighting scheme generally minimize

the forecast error. Next, we used our proposed methods to forecast realized volatility series in Chinese stock market, where similar results were achieved.

The rest of the paper is organized as follows. Section 2 briefly describes the forecast combination and proposes a new exponential weighting scheme. Next, these forecast combination methods are examined by Monte Carlo simulation experiments in Section 3. In Section 4, we conduct an out-of-sample forecasting exercise of realized volatility series based on the heterogeneous autoregressive (HAR) model for Chinese stock market. Section 6 concludes the paper.

2 Methodology

2.1 Forecast combination across estimation windows

Consider the following regression model defined over the sample period $t = 1, \ldots, T+1$:

$$y_t = \boldsymbol{x}_t^{'}\boldsymbol{\beta}_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0, \sigma^2), \tag{1}$$

where the coefficients, $\boldsymbol{\beta}_t$, may subject to m ($m \geq 0$) structural breaks at time $t = T_1, \ldots, T_m$. Specifically,

$$\begin{cases} \boldsymbol{\beta}_t = \boldsymbol{\beta}^{(1)}, & t = 1, \ldots, T_1, \\ \vdots & \\ \boldsymbol{\beta}_t = \boldsymbol{\beta}^{(m)}, & t = T_{m-1}+1, \ldots, T_m, \\ \boldsymbol{\beta}_t = \boldsymbol{\beta}^{(m+1)}, & t = T_m+1, \ldots, T+1. \end{cases} \tag{2}$$

Note that the number of the structural breaks, m, and the locations, $\{T_1, \ldots, T_m\}$, are all unknown. The aim is to forecast y_{T+1} based on the whole observations, $\{y_t, \boldsymbol{x}_t\}_{t=1}^{T}$. With no breaks, $\boldsymbol{\beta}_t$ becomes a constant $\boldsymbol{\beta}$ and can be estimated by the ordinary least square (OLS) method based on $\{y_t, \boldsymbol{x}_t\}_{t=1}^{T}$. Denote this OLS estimate as $\hat{\boldsymbol{\beta}}$; then the forecast of y_{T+1} can be obtained by $\hat{y}_{T+1} = \boldsymbol{x}_{t+1}^{'}\hat{\boldsymbol{\beta}}$. However, the existing structural breaks may make $\hat{y}_{T+1}$ biased with a high MFSE.

To reduce the forecasting bias and achieve a small MSFE, we propose to use the forecast combination approach, which produces a weighted average of the forecast estimates generated from different estimation windows that strat from $t = \eta$ and end at $t = T$. For simplicity, we abbreviate these estimation windows as $[\eta : T]$. The OLS estimates based on the observations in above estimation window, i.e $\{y_t, \boldsymbol{x}_t\}_{t=\eta}^{T}$, can be denoted as $\hat{\boldsymbol{\beta}}(\eta) = (\boldsymbol{X}^{'}(\eta)\boldsymbol{X}(\eta))^{-1}\boldsymbol{X}^{'}(\eta)\boldsymbol{y}(\eta)$, where $\boldsymbol{y}(\eta) = (y_\eta, y_{\eta+1}, \ldots, y_T)^{'}$ and $\boldsymbol{X}(\eta) = (\boldsymbol{x}_\eta, \boldsymbol{x}_{\eta+1}, \ldots, \boldsymbol{x}_T)^{'}$. The forecast combination method assumes that the forecast of y_{T+1} is a weighted sum of the following forecast components:

$$\hat{y}_{T+1} = \sum_{\eta=1}^{T-\omega+1} \kappa_\eta \hat{y}_{T+1}(\eta)\,, \tag{3}$$

where $\hat{y}_{T+1}(\eta) \triangleq \boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}(\eta)$ are the forecast components, κ_η's are the weights satisfying $\sum_{\eta=1}^{T-\omega+1}\kappa_\eta = 1$. The time η in the original sample $[1:T]$ is subject to a minimum acceptable estimation window size, so that each forecast component is estimated based on at least ω observations. Different choices of weights lead to different weighting schemes. Now, we introduce several commonly used weighting schemes.

The simplest weighting scheme is the **equally weight** (**EW**), with the weights $\kappa_\eta^{(EW)} = (T-\omega+1)^{-1}$ for $\eta = 1, \ldots, T-\omega+1$. As a result, the forecast of y_{T+1} becomes

$$\hat{y}_{T+1} = \sum_{\eta=1}^{T-\omega+1} \frac{\hat{y}_{T+1}(\eta)}{T-\omega+1}. \tag{4}$$

Pesaran and Timmermann (2007) [10] pointed out that compared to the forecast derived from a full sample OLS estimate (i.e $\hat{y}_{T+1}(1)$, denoted as full sample forecast henceforth), the forecast combination with equal weight can lead to a bias and variance reduction. The EW could present a better performace because the forecast componments that depend on the after-break observations are consistent estimators of y_{T+1}, which could help the EW reduce the forecasting bias. In contrast, the full sample forecast, which is based on the whole observations, is typically with large bias in the presence of structural breaks.

Another popular weighting scheme is the **location weight** (**LW**), which means each forecast component $\hat{y}_{T+1}(\eta)$ has a weight that is proportional to the starting location η, i.e $\kappa_\eta^{(LW)} = \frac{\eta}{\sum_{i=\eta}^{T-\omega+1}\eta}$. The one-step-ahead forecast with LW becomes

$$\hat{y}_{T+1} = \sum_{\eta=1}^{T-\omega+1} \frac{\eta}{\sum_{\eta=1}^{T-\omega+1}\eta}\hat{y}_{T+1}(\eta). \tag{5}$$

Compared to the EW, the LW put larger weights on forecast components, obtained by using more recent data.

Drawback of the EW and LW is lack of using break information. Tian and Anderson (2014) [11] introduced the **ROC weight** (**ROCW**), which is calculated based on the reverse order CUSUM (ROC) test statistics to incorporate information of breaks. They argued that although the estimation of structural break may be imprecise, the break information is still useful for constructing the weights and obtain more accurate forecast. The idea is that the final forecast should depend more on the forecast components which are computed by using more after-break data. Specifically, the ROC test statistic is

$$s_\eta = \frac{\sum_{t=\eta}^{T-\omega+1}(e_t^{t+1:T})^2}{\sum_{t=1}^{T-\omega+1}(e_t^{t+1:T})^2}, \quad \text{for} \quad \eta = T-\omega+1, \ldots, 1, \tag{6}$$

where the scaled residuals, $e_t^{t+1:T}$'s, are defined as $e_t^{t+1:T} = \frac{y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}(t+1)}{\sqrt{1+\boldsymbol{x}_t'(\boldsymbol{X}'(t+1)\boldsymbol{X}(t+1))^{-1}\boldsymbol{x}_t}}$. If there are no structrual breaks, $\mathrm{E}(s_\eta) = \frac{T-\omega+1-\eta}{T-\omega}$ and s_η should be close to $\frac{T-\omega+1-\eta}{T-\omega}$. Correspondingly, the larger $|s_\eta - \frac{T-\omega+1-\eta}{T-\omega}|$ is, the more possible that η is the breakpoint and the more weight should be given to $\hat{y}_{T+1}(\eta)$. Based on this

idea, Tian and Anderson (2014) [11] defines the ROC weighted forecast combination as

$$\hat{y}_{T+1} = \sum_{i=1}^{T-\omega+1} \kappa_{\eta}^{(ROCW)} \hat{y}_{T+1}(\eta), \text{ with } \kappa_{\eta}^{(ROCW)} = \frac{\left| s_{\eta} - \frac{T-\omega-\eta+1}{T-\omega} \right| l_{\eta}}{\sum_{i=1}^{T-\omega+1} \left| s_i - \frac{T-\omega-i+1}{T-\omega} \right| l_i}, \tag{7}$$

where l_η is the weight adjustment function. In particular, l_η can be fixed at 1 or η, which are denoted as **ROCW1** and **ROCW2** in the following.

2.2 Exponential Weight

From the previous discussion on EW and LW weighting scheme, it can be noted that one can improve the performance of forecast combination by using a weighting scheme which relies on more on recent observations. With this motivation, we propose a new weighting scheme called **exponential weight** (**ExpW**), which assigns exponentially decaying weights to forecast components. This idea is analog to the exponentially weighted regression method proposed by Harvey (1989) [12], which is an another popular forecasting approach in the presence of structural breaks.

2.2.1 Weighting Scheme

The exponentially weighted forecast combination assigns exponentially decaying weights to forecast components $\hat{y}_{T+1}(\eta)$, i.e.,

$$\begin{cases} \kappa_{\eta}^{(ExpW)} = \frac{\gamma^{T-\omega+1-\eta}}{\sum_{i=1}^{T-\omega+1} \gamma^{T-\omega+1-i}} = \frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta}, \ \eta = 1, \ldots, T-\omega+1, \\ \hat{y}_{T+1} = \sum_{\eta=1}^{T-\omega+1} k_{\eta}^{(ExpW)} \hat{y}_{T+1}(\eta), \end{cases} \tag{8}$$

where $\gamma \in (0, 1)$ is a decaying parameter and controls the decaying rate. The smaller the value of γ is, the faster the weight decays.

In this method, the forecast component $\hat{y}_{T+1}(T-\omega+1)$ carries the largest weight. The weight decays exponentially as η decreases. The approach has two important advantages over the EW and the LW. First, the forecast combination reacts faster to structural breaks. This is because as more distant observations are included in the estimation window, the exponentially decaying weights makes the combination relies more on recent observations than the EW and LW methods. Second, as T increases, more after-break observations are included into the forecast combination and the near-last-break observations become distant. Then the break influence on the forecast combination decays rapidly. We can use the information of the break point to determine γ, and the choice of γ will be given in section 2.2.3.

Now, to compare EW, ExpW and the full sample forecast, we develop a theoretical result for a linear regression model with a single break in the slope parameter.

2.2.2 Theoretical Analysis for Single Break in the Slope Parameter

In this section, we compare EW, ExpW and full sample forecast for DGPs with a single break in the slope parameter. Specifically, we consider following data generation process:

$$(y_t - \mu_y) = \beta_t(x_t - \mu_x) + \sigma\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1), \tag{9}$$

where $t = 1, 2, \ldots, T+1$. The exogenous variable, x_t, is assumed to follow a covariance stationary process with mean μ_x. We further assume that the slope parameter, β_t, is subject to a single break at time $t = T_b$ $(1 < T_b < T)$,

$$\beta_t = \begin{cases} \beta^{(1)} & \forall t \leq T_b \\ \beta^{(2)} & \forall t > T_b \end{cases},$$

where $\beta^{(1)} \neq \beta^{(2)}$. The one-step-ahead single forecast of y_{T+1} based on the estimation window $[\eta : T]$ is

$$\hat{y}_{T+1}(\eta) = \bar{y}(\eta) + \hat{\beta}(\eta)\big(x_{T+1} - \bar{x}(\eta)\big),$$

where $\bar{x}(\eta) = \frac{1}{T-\eta+1}\sum_{t=\eta}^{T} x_t$, $\bar{y}(\eta) = \frac{1}{T-\eta+1}\sum_{t=\eta}^{T} y_t$, and

$$\hat{\beta}(\eta) = \frac{\sum_{t=\eta}^{T}\big(x_t - \bar{x}(\eta)\big)\big(y_t - \bar{y}(\eta)\big)}{\sum_{t=\eta}^{T}\big(x_t - \bar{x}(\eta)\big)^2}.$$

According to Pesaran and Pick (2007, lemma 1) [10], the scaled forecast error based on $[\eta : T]$ is

$$\begin{aligned}\sigma^{-1}\xi_{T+1}(\eta) &= \sigma^{-1}\big(y_{T+1} - \hat{y}_{T+1}(\eta)\big) \\ &= \epsilon_{t+1} + \Big(\frac{T_b - \eta}{T - \eta + 1}\Big)\mathrm{I}(T_b - \eta + 1)\lambda(x_{T+1} - \mu_x) + O_p\Big(\frac{1}{\sqrt{T - \eta + 1}}\Big),\end{aligned}$$

where $\mathrm{I}(c)$ is an indicator function that is taking value 1 if $c > 0$ and 0 otherwise, and $\lambda = (\beta^{(2)} - \beta^{(1)})/\sigma$. Now consider the forecast by equal weights, $\hat{y}_{\omega,T+1}^{(EW)} = \sum_{\eta=1}^{T-\omega+1}\frac{\hat{y}_{T+1}(\eta)}{T-\omega+1}$. It follows that the error of the EW forecast is $\xi_{\omega,T+1}^{(EW)} = \sum_{\eta=1}^{T-\omega+1}\frac{\xi_{T+1}(\eta)}{T-\omega+1}$. Let $\zeta(\eta) = [(T_b - \eta)/(T - \eta + 1)]\mathrm{I}(T_b - \eta)$. For fixed $R_\omega = \frac{\omega}{T}$ as $T \to \infty$, the scaled EW forecast error conditional on x_{T+1} is

$$\sigma^{-1}\xi_{\omega,T+1}^{(EW)} = \epsilon_{t+1} + B_{T+1}^{(EW)} + O_p\Big(\frac{1}{\sqrt{T}}\Big),$$

where $B_{\omega,T+1}^{(EW)} = \lambda(x_{T+1} - \mu_x)[1/(T - \omega + 1)\sum_{\eta=1}^{T-\omega+1}\zeta(\eta)]$. It follows directly from Pesaran and Pick (2007, lemma 2) [10]. Now focus on the forecast based on exponential weight, $\hat{y}_{\omega,T+1}^{(ExpW)}$. By tedious derivation (see

Appendix), the scaled ExpW forecast error is

$$\begin{aligned}\sigma^{-1}\xi_{\omega,T+1}^{(ExpW)} &= \frac{1-\gamma}{\sigma(1-\gamma^{T-\omega+1})}\sum_{\eta=1}^{T-\omega+1}\gamma^{T-\omega+1-\eta}\xi_{T+1}(\eta) \\ &= \epsilon_{t+1} + B_{T+1}^{(ExpW)} + O_p(\frac{1}{\sqrt{T}}),\end{aligned} \tag{10}$$

where

$$B_{\omega,T+1}^{(ExpW)} = \lambda(x_{T+1}-\mu_x)\Big[\frac{1-\gamma}{1-\gamma^{T-\omega+1}}\sum_{\eta=1}^{T-\omega+1}\gamma^{T-\omega+1-\eta}\zeta_{T+1}(\eta)\Big]. \tag{11}$$

We are now in a position to compare the MSFEs of the full sample forecast, EW and ExpW. Now, consider the case where T_b/T and ω/T as fixed $T \to \infty$, we have the following proposition.

Proposition 1. Suppose that the data generation process in (9) holds and is subject to a single break in β_t at T_b. For large T but fixed T_b/T and ω/T, the MSFEs of the full sample forecast ($\hat{y}_{t+1}(1)$), EW forecast ($\hat{y}_{\omega,t+1}^{(EW)}$), and the ExpW forecast ($\hat{y}_{\omega,t+1}^{(ExpW)}$) have the following relationship:

$$\text{MSFE}\Big(\hat{y}_{t+1}(1)\Big) > \text{MSFE}\Big(\hat{y}_{\omega,t+1}^{(EW)}\Big) > \text{MSFE}\Big(\hat{y}_{\omega,t+1}^{(ExpW)}\Big),$$

as $T \to \infty$.

Proposition 1 shows that our proposed ExpW presents the smallest MSFE compared with EW and single model when a single break occures in the slope parameter in process (9) as $T \to \infty$.

2.2.3 Decaying parameter γ

The exponential weight requires choosing an appropriate value of the decaying parameter γ. Small values of γ means one puts more weights on forecast components that depend on after-break observations, which result in smaller forecasting bias but larger forecasting variance; in contrast, large values of γ means one puts more weights on forecast components that depend on before-break observations, which can lead to larger forecasting bias but smaller forecasting variance.

Since $\gamma \in (0,1)$, the weight κ_η decreases to 0 as η decreases to 1. Therefore, a forecast component $\hat{y}_{T+1}(\eta)$ has weak influence on the forecast combination if η is small, while $\hat{y}_{T+1}(\eta)$ has strong influence and dominates the forecast combination if η is large. We call these dominating ones the effective forecast components. Given a domination level $\alpha \in (0,1)$, the total number of effective forecast components K can be computed by

$$K = \underset{H=1,2,\ldots T-\omega+1}{\text{argmin}} \sum_{i=1}^{H}\kappa_{T-\omega+2-i} > \alpha. \tag{12}$$

By direct calculation, it follows that

$$\sum_{i=1}^{K} \kappa_{T-\omega+2-i} = \frac{1}{\sum_{i=1}^{T-\omega+1} \gamma^{T-\omega+1-i}} \sum_{i=0}^{K-1} \gamma^i = \frac{1-\gamma^K}{1-\gamma^{T-\omega+1}} \approx 1-\gamma^K, \tag{13}$$

as $T-\omega+1$ is large. By substituting (13) into equation (12), we get $K = \lfloor \frac{\ln(1-\alpha)}{\ln\gamma} \rfloor$. Here, $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x; and correspondingly, we use $\lceil x \rceil$ denots the least integer greater than or equal to x. Note that K is a non-decreasing function of α and γ. In practice, one can first decide K and α according to prior information about data, then compute γ by

$$\gamma = (1-\alpha)^{1/K}, \tag{14}$$

which is derived from $K = \lfloor \frac{\ln(1-\alpha)}{\ln\gamma} \rfloor$. In practice, if we learn the last breakpoint is around point $\widehat{T}_b$, then we can choose K as the number of forecast componments depend on the after-break observations, i.e $K = T-\omega-\widehat{T}_b$, thus the after-break information could dominate our ExpW forecast. In practice, the MOSUM test, which will be introduced in next part, can be adopted to decide $\widehat{T}_b$.

2.2.4 Detecting the last break point T_b

Now, we introduce MOSUM test [8], which can be applied to detect whether structural breaks exist, and offer information that the number and locations of the possible structural breaks. The null hypothesis is given by

$$H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}\,, \quad t = 1, \ldots, T, \tag{15}$$

against the alternative that there exists at least one coefficient change, i.e., $\boldsymbol{\beta}_{t_1} \neq \boldsymbol{\beta}_{t_2}$ for any given t_1 and t_2 in the range.

The MOSUM test utilize the moving sums of the scaled residuals to detect the structural breaks. Specifically, the scaled residuals used are $\{e_t^{1:t-1}\}_{t=k+1}^{T}$, where k is the number of covariates (including the intercept) and $e_t^{1:t-1} = \frac{y_t - \boldsymbol{x}_t' \hat{\boldsymbol{\beta}}_{1:t-1}}{\sqrt{1+\boldsymbol{x}_t'(\boldsymbol{X}'_{1:t-1}\boldsymbol{X}_{1:t-1})^{-1}\boldsymbol{x}_t}}$. Here, $\boldsymbol{X}_{1:t-1} = (x_1, \ldots, x_{t-1})'$ and $\hat{\boldsymbol{\beta}}_{1:t-1}$ is the OLS estimate based on $[1 : t-1]$. Note that the total number of residuals is $\tau = T-k$. Denote $\omega^* = \lfloor h\tau \rfloor$ as the bandwidth of moving windows, which is based on the bandwith parameter $h \in (0,1)$. In this article, we choose $h = 0.1$. The moving sums at point t is given by

$$MS(t|h) = \frac{1}{\hat{\sigma}\sqrt{\tau}} \left| \sum_{j=t-\lfloor \frac{\omega^*}{2} \rfloor + 1}^{t+\lceil \frac{\omega^*}{2} \rceil} e_j^{1:j-1} \right|, \quad t = k + \lfloor \frac{\omega^*}{2} \rfloor, \ldots, T - \lceil \frac{\omega^*}{2} \rceil,$$

where $\hat{\sigma} = \sqrt{\frac{1}{\tau}\sum_{t=k+1}^{T}(e_t^{1:t-1} - \bar{e}_t)^2}$ with $\bar{e}_t = \frac{1}{\tau}\sum_{t=k+1}^{T} e_t^{1:t-1}$. The final MOSUM test statistic is given by

$$MS_h = \max_{k+\lfloor \frac{\omega^*}{2} \rfloor \leq t \leq T-\lceil \frac{\omega^*}{2} \rceil} MS(t|h) \tag{16}$$

Heuristically, the moving sums, $MS(t|h)$'s, should near 0 under H_0 and a large value of MS_h implies the existence of structural breaks. Under the null hypothesis (15), Chu (1995) [8] proved that

$$MS_h \Rightarrow \max_{0 \leq s \leq 1-h} \left| h^{-1/2} \left(W(s+h) - W(s) \right) \right|, \tag{17}$$

where $\Rightarrow$ denotes convergence in distribution and $W(\cdot)$ is the Wiener process. To conduct the MOSUM test based on an $\alpha \times 100\%$ significance level, one first computes the critical value $B(\alpha)$ based on (17) and Theorem 2.2 in Chu et al. (1995) [8], then observes whether the MOSUM test statistic MS_h is greater than $B(\alpha)$. If MS_h is great than $B(\alpha)$, the null hypothesis (15) should be rejected.

Additionally to testing for breaks, Kirch, Muhsal and Ombao (2015) [13] suggests the moving sums can be used to estimate the number and locations of the structural breaks. In particular, the local maxima beyond the boundary can be good estimates of the locations of the structural breaks. Specifically, we define estimators for the number of change points as well as their locations as follows: Consider all pairs of time points $(v_m; w_m)$ such that

$$\begin{cases} MS(t|h) \geq B(\alpha), & \text{for } t = v_m, \ldots, w_m, \\ MS(t|h) < B(\alpha), & \text{for } t = v_{m-1}, \ldots, w_{m+1}, \\ w_m - v_m \geq \zeta, \end{cases} \tag{18}$$

where ζ is a small integer which is set to 3 in the following simulation experiments and real data analysis. Then the number of pairs $(v_k; w_k)$ can be an estimator of the number of structural breaks, and $\hat{t}_m = \underset{v_m \leq t \leq w_m}{\operatorname{argmax}} MS(t|h)$, , can be estimators for the locations of the structural breaks. As a result, the last break point can be estimated as $\widehat{T}_b = \max\{\hat{t}_m\}$.

3 Simulation experiments

In this section, we develop a series of simulation experiments in order to evaluate our proposed forecast methods in a regression setting. Specifically, we focus on the DGPs following the HAR process proposed by Corsi (2009) [14]:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 \bar{y}_{t-5,t-1} + \beta_3 \bar{y}_{t-22,t-1} + \epsilon_t,$$

where $\bar{y}_{t-m,t-1} = \frac{1}{m}\sum_{i=1}^{m} y_{t-i}$, the error term ϵ_t follows $N(0,1)$, and $t = 1, 2, \ldots, T$ for $T = 500$. Theoretically, this HAR process is a speicial autoregressive model where its response variable y_t depends on its own 22 past values $\{y_{t-1}, \ldots, y_{t-22}\}$. We consider the HAR process because our empirical analysis in the next section

forecast the realized volatility series in Chinese stock market based on the HAR model. The forecast target is y_{501}, and we use the mean squared forecasting error to measure the forecast loss. Our simulations focus on structural breaks in the persistence of intercept (β_0) and coefficients, β_1, β_2 and β_3.

We consider both single break and multiple break scenarios with three possible break points occuring $t = 125$, 250 and 375. Thus, the estimation sample can cover up to four possible regimes, R_1, R_2, R_3 and R_4, which relate to observations $\{1, \ldots, 125\}$, $\{126, \ldots, 250\}$, $\{251, \ldots, 375\}$ and $\{376, \ldots, 500\}$. For the two-break scenarios, both structural breaks occure at $t = 125$ and 375 respectively; for experiments with three breaks, one addtional intermediate break occurs at $t = 250$. We conduct a total of 28 experiments, and examine the performance of the full sample forecast, two-stage approach, and five forecasting combination methods, including EW, LW, RocW1, RocW2 and ExpW. The decaying parameter γ in ExpW are determined by equation (14) with $\alpha = 0.9$. The estimation procedure for the two-stage approach is that (1), Detect the last structure break, T_b, based on MOSUM test with 95% significant level; (2), Forecast $\hat{y}_{501} = \boldsymbol{x}_{501}^{'}\hat{\beta}_b$ with $\hat{\beta}_b$ estimated by the OLS method based on the observations $\{y_t, \boldsymbol{x}_t\}_{t=T_b+1}^{T}$, where $\boldsymbol{x}_t = (y_{t-1}, \bar{y}_{t-5,t-1}, \bar{y}_{t-22,t-1})^{'}$. The minmium acceptable estimation window size for all forecast combinations is set to $\omega = \lfloor 0.1T \rfloor = 50$.

We conduct a total of 28 experiments and report the ratios of the MSFEs relative to the benchmark, i.e full sample forecast, in Table 1. Experiment 1 represents the case when there are no structural breaks, where we can see all the MFSE ratios are great than one, implying the full sample method performs better than all other forecasting methods because the full sample method utilizes all available information if there are no structural breaks. Now focus on the experiments with structural breaks.

Next, we consider the single-break experiments. Specifically, Experiments 2–13 relate to single-break DGPs, in which Experiments 2–5 consider the situations in intercept shifts, Experiments 6–11 considers coefficients change, and Experiments 12–13 represent multiple changes in the coefficients and intercepts at one breakpoint. The results present clear evidence that all of our forecast combination methods considered could improve the forecasing performance compared to the full sample method. Among the forecast combination methods, the two-stage method presents larger MSFEs. This is mainly because the two-stage method only depends on observations after the estimated last break point, which requires high estimation accuracy for the last break point; once the estimated last break point is inaccurate, this two-stage method may lead to large forecasting bias. On the contrary, the forecast combination method combines the information before and after the last break through a certain weighting scheme. The after-last-break information can help reduce the bias and the before-last-break information could reduce the variance. Hence, one can achieve a smaller MSFE.

Among the forecast combination methods, the proposed ExpW method performs best in most situations, followd by the LW method, which reduce the MSFE relative to the benchmark by about 25% and 21% respectively. In addition, when the structural break is early in the sample, the ROC weights slightly outperform the ExpW and LW (see Experiments 2, 6 and 7), whereas when the break occurs late in the sample, the ExpW and LW perform better than the ROC weights (see experiments 4, 5, 9, 10 and 11). Figure 1 intuitively

Table 1: The MSFEs relative to the benchmark of the forecasting methods

Parameter	β_0	β_1	β_2	β_3	Forecasting Methods						
Exp/Regime	R_1, R_2, R_3, R_4	R_1, R_2, R_3, R_4	R_1, R_2, R_3, R_4	R_1, R_2, R_3, R_4	HAR	EW	LW	RocW1	RocW2	TS	ExpW
1	1,1,1,1	0.6,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	**1.000**	1.017	1.031	1.015	1.025	1.038	1.067
2	1,4,4,4	0.6,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.687	0.667	**0.655**	**0.655**	0.901	0.673
3	1,1,4,4	0.6,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.770	0.700	0.760	0.709	0.837	**0.681**
4	1,1,1,4	0.6,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.903	0.827	0.981	0.961	0.841	**0.680**
5	1,1,1,2	0.6,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.968	0.926	1.021	1.008	0.892	**0.838**
6	1,1,1,1	0.6,0.1,0.1,0.1	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.646	0.623	**0.615**	**0.613**	0.757	0.627
7	1,1,1,1	0.1,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.876	0.872	**0.857**	**0.859**	0.895	0.881
8	1,1,1,1	0.1,0.1,0.1,0.1	0.6,0.6,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.753	0.672	0.741	0.683	0.789	**0.646**
9	1,1,1,1	0.1,0.1,0.1,0.1	0.1,0.1,0.6,0.6	0.1,0.1,0.1,0.1	1.000	0.908	0.881	0.880	0.866	**0.848**	0.884
10	1,1,1,1	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	0.6,0.6,0.6,0.1	1.000	0.924	0.843	1.024	1.004	0.702	**0.658**
11	1,1,1,1	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.6	1.000	0.946	0.906	0.977	0.961	0.931	**0.831**
12	1,1,1,1	0.6,0.6,0.1,0.1	0.1,0.1,-0.3,-0.3	0.1,0.1,0.1,0.1	1.000	0.686	0.592	0.697	0.623	0.848	**0.561**
13	3,3,1,1	0.1,0.1,0.6,0.6	0.6,0.6,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.817	0.756	0.770	**0.739**	0.842	**0.743**
14	1,2,2,4	0.6,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.915	0.847	0.972	0.947	0.783	**0.722**
15	4,2,2,1	0.6,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.865	0.806	0.879	0.837	0.976	**0.772**
16	1,1,1,1	0.6,0.3,0.3,0.1	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.743	0.690	0.740	0.717	0.870	**0.672**
17	1,1,1,1	0.6,0.2,0.2,-0.2	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.635	0.555	0.634	0.594	0.810	**0.528**
18	1,1,1,1	0.6,0.4,0.4,-0.2	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.766	0.668	0.793	0.735	0.915	**0.611**
19	1,1,1,1	0.6,0,0,-0.2	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.573	0.515	0.548	0.529	0.739	**0.505**
20	1,1,1,1	-0.2,0.2,0.2,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.966	0.945	0.957	0.936	1.016	**0.904**
21	1,1,1,1	0.6,-0.2,-0.2,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.958	0.933	0.981	0.971	1.003	**0.907**
22	1,1,1,1	0.1,0.1,0.1,-0.2	0.6,0.2,0.2,0.1	0.1,0.1,0.1,0.1	1.000	0.670	0.594	0.666	0.632	0.800	**0.568**
23	1,1,1,1	0.1,0.1,0.1,0.6	-0.2,0.2,0.2,0.1	0.1,0.1,0.1,0.1	1.000	0.966	0.945	0.966	0.950	1.024	**0.906**
24	3,2,2,4	0.4,0.6,0.6,0.1	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.869	0.773	0.964	0.937	**0.592**	**0.584**
25	1,2,3,4	0.6,0.6,0.6,0.6	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.907	0.848	0.899	0.838	0.999	**0.807**
26	1,1,1,1	0.6,0.4,0.2,0.1	0.1,0.1,0.1,0.1	0.1,0.1,0.1,0.1	1.000	0.757	0.706	0.753	0.723	0.892	**0.693**
27	3,2,1,0	0.1,0.2,0.3,0.6	0.6,0.3,0.2,0.1	0.1,0.1,0.1,0.1	1.000	0.858	0.814	0.872	0.853	0.939	**0.794**
28	1,4,3,2	0.6,0.3,0,0.4	0.1,-0.3,-0.4,0.1	0.1,0.2,0.2,0.4	1.000	0.940	0.886	0.972	0.930	**0.739**	0.763
Average relative MSFE values of Experiments 2–13					1.000	0.839	0.792	0.846	0.823	0.855	0.752
Average relative MSFE values of Experiments 14–28					1.000	0.826	0.768	0.840	0.809	0.873	0.716
Average relative MSFE values of Experiments 1–28					1.000	0.832	0.779	0.842	0.815	0.865	0.732

[a] The experiments are based on $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 \bar{y}_{t-5,t-1} + \beta_3 \bar{y}_{t-22,t-1} + \epsilon_t$, where the error term follows $N(0,1)$, and $t = 1, 2, \ldots, T$ for $T = 501$. The forecast target is y_{501}, and we use the mean squared forecasting error to measure the forecast loss. The four possible regimes, denoted by R1, R2, R3 and R4 respectively, relate to observations $\{1, \ldots, 125\}$, $\{126, \ldots, 250\}$, $\{251, \ldots, 375\}$ and $\{376, \ldots, 500\}$.

[b] The smallest MSFE in each row of the table is highlighted by underscoring, and the numbers in bold denote the corresponding methods that are included in MCS for the significant level of 10% and the semi-quadratic statistic is used.

compares the ExpW, LW and ROC weights through three situations that the structural break occurs at the early, middle and late stage. It can be seen that when the break point is in the early stage (see Experiment 2 in Figure 1), ROC weights assign most weights on estimation windows after the strucrural break. And as the break point moves toward the later samples, the weights assigned on the estimation windows before the break point increased gradually (see experiments 3–4 in Figure 1). Particularily in Experiment 4 in Figure 1, nearly 90% of the weights are placed on the before-break estimation windows by the ROC weights, which leads to large forecasting bias and poor forecasting accuracy. Different with ROC weights, ExpW and LW always place more weights on the most recent estimation windows, whether the structural break occurs in the early or late period (see the blue and black lines in Experiments 2–4 in Figure 1), ensuring a relatively small forecasting bias. Figure 1 also shows that the ExpW can dynamically adjust the weight according to the location of the break point, which assigns about 90% of weights on the estimation windows starting after the structural break by adjusting the decaying parameter γ based on the MOSUM test.

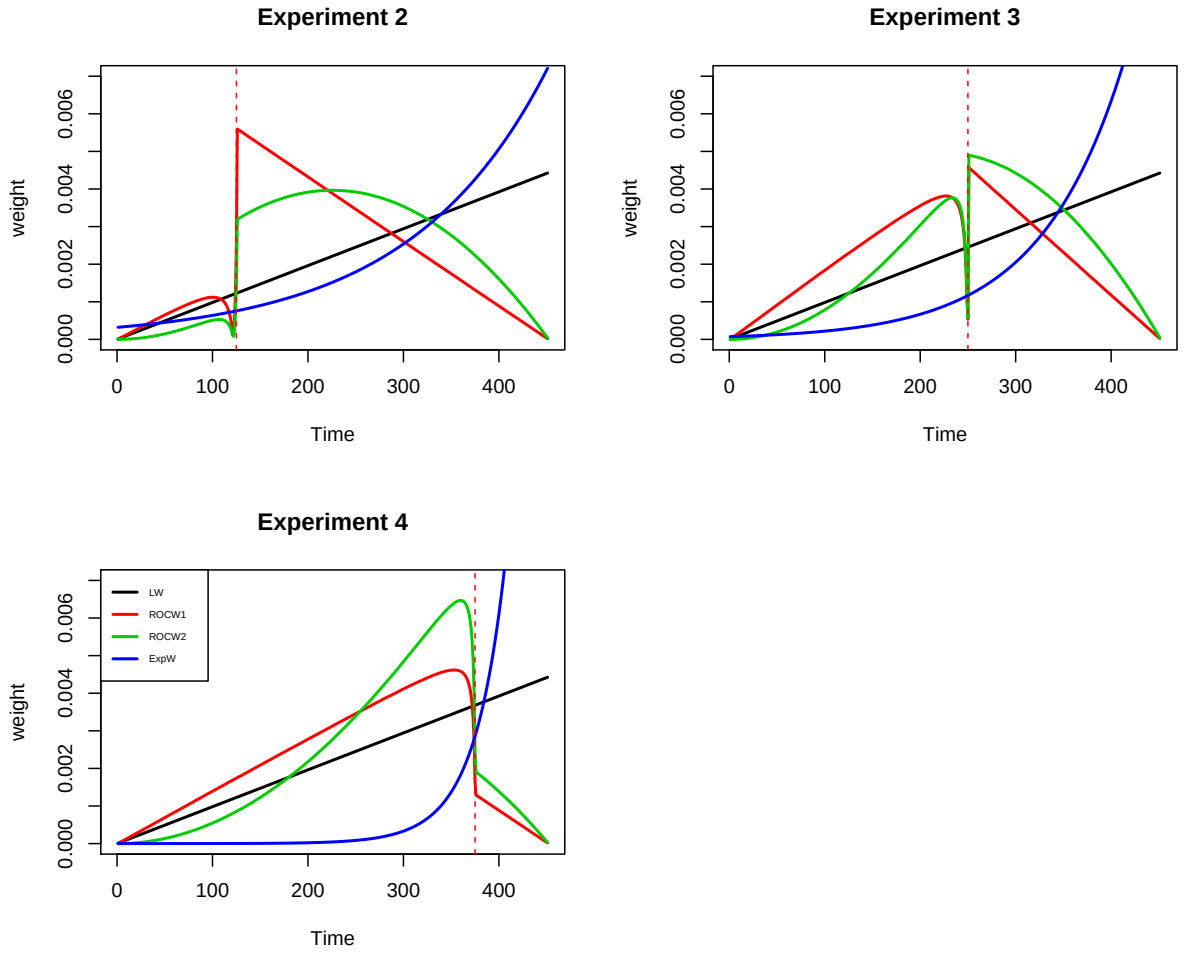

Figure 1: Combination weights for the estimation windows in experiments 2–4. Notes: The plots show the averages over all simulations, and the vertical dotted lines indicate the break points.

Now, we look at the results of multiple-break experiments. Specifically, experiments 14–24 focus on the situations with two structural breaks, while experiments 25–28 relat to the cases with three structural breaks. Compared with the single-break scenario, the multiple-break scenario can examine the forecasting effect of the methods more comprehensively. That is becasue (1), the time series in applied situations usually contains a considerable longer time span, thus may include several structural breaks; (2) a multiple-break scenario enables us to evaluate the abilities of various forecasting methods to account for a recent change, when some other (and potentially less relevant) changes have occurred in the past. The simulation results are similar with the single-break scenario, where all our forecasting methods, including the two-stage methods and forecast combination methods, improve the forecast ability compared with the benchmark. And generally, forecast combination methods (reducing the MSFE by over 16%) works better than two-stage method (reducing the MSFE by about 13%). The ExpW typically outperforms the two-stage method, which further illustrates that one can make a better use of the information of the MOSUM test than simply estimating the most recent structural break and conducting the forecast based on this estimate. Focusing our attention on the forecast combination methods, the ExpW performs best (reducing the MSFE by over 28%), followed by LW (reducing the MSFE by over 23%), the ROC weights (reducing the MSFE by about 17%) and EW (reducing the MSFE by about 16%) generally have weaker performances. The ROC weights perform worse in most situations since all the most recent structural breaks are late in the sample among our multiply-break experiments.

In addition to the above discussion, we also note the following observations: (1) forecast combination methods perform well when the changes in the parameters are quite large. For example, the intercept in experiment 4 has increased over 400%, the forecast combination methods has reduce the MSFE over 13% compared with the single model in average; but in experiment 5, which intercept increased just 200%, the forecast combinations have reduced the MSFE just 6% in average. (2) If the persistance (i.e β_1, β_1 and β_2) in the paramaters decreased (see experiments 6, 8 and 10), forecast combination methods perform particularly well; however, when the persistance inceased (see experiments 7, 9 and 11), the forecast combination methods perform relatively worse. (3) Generally, ROCW1 forecasts are dominated by that of ROCW2 since the prior function of ROCW2, $l_\eta = \eta$, gives more wights to the recent estimation windows compared with ROCW1 whose l_η is fixed at one. (4) If the parameters change in the same direction over time, the forecast combination methods perform better when the last break is small. (see experimrnts 17–19)

In conclusion, the forecast combinations are suited to dealing with situations in which the resulting persistence by the structural breaks is low, or when the change in the coefficients is quite large. Among the forecast combination methods, the ExpW and LW do better performances when the most recent structural break occurs quite late in the sample, and ROC weights perform better for a early structural break.

4 Forecasting Realized Volatility

4.1 Realized Volatility and HAR model

Realized volatility forecasting has attract much interest in financal and economical reasearches because it provides a relatively accurate measure of volatility which is critical to portfolio optimization, derivatives pricing and risk management [15, 16, 17, 18]. Blair et al. (2001) [19] introduced realized volatility as the sum of squared intraday returns and squared overnight log-return. Specifically, Let $\{P_{t,j}\}_{j=1}^{J}$ be the j-th intraday price observed at day t, $t = 1, \ldots, T$, the realized volatility in the t-th trading day is given as follows

$$\mathrm{RV}_t = \sum_{j=1}^{J} r_{t,j}^2 + r_{t,0}^2, \ t = 1, \ldots, T, \tag{19}$$

where $r_{t,j}$, the j-th intraday log-return, is given by $r_{t,j} = (\log P_{t,j} - \log P_{t,j-1}) \times 100\%$, $r_{t,0} = (\log P_{t,0} - \log P_{t-1,J}) \times 100\%$ as the overnight log-return, reflecting the difference between the closing price of time $t-1$ and the opening price of time t. With Hansen and Lunde (2006) [20] and Bandi and Russell's (2008) [21] suggestion, we adopt a five-min sampling frequency in the following analysis. Among various models for realized volatility forecasting, Heterogeneous Auto-Regressive (HAR) model, proposed by Corsi (2009) [14], has attracted more attention nowadays [22, 23]. The HAR model could reflect the main empirical features of financial returns, inlcuding fat tails, long memory and self-similarity, in a very tractable and parsimonious way [14]. Specifically, the HAR model is defined as

$$\mathrm{RV}_t = \beta_0 + \beta_1 \mathrm{RV}_{t-1} + \beta_2 \mathrm{RV}_{t-1}^{w} + \beta_3 \mathrm{RV}_{t-1}^{m} + \epsilon_t, \tag{20}$$

where $t = 1, \ldots T$, ϵ_t are independent and identically distributed (i.i.d.) noise with zero mean and finite variance $\sigma^2 < \infty$, $\mathrm{RV}_t^{w} = \frac{1}{5}\sum_{i=1}^{5} \mathrm{RV}_{t-i}$ and $\mathrm{RV}_t^{m} = \frac{1}{22}\sum_{i=1}^{22} \mathrm{RV}_{t-i}$ are respectively the weekly and monthly averages of daily log realized volatility. The tractable estimation and superior forecast performance of the HAR model make it a widely used model for forecasting the realized volatility. Similar to the extensions of the ordinary GARCH(1,1) model, the standard HAR-RV model has been extended in various ways, such as HAR model with jump [24] and the leverage HAR model [25]. All these extensions introduce extra explanatory variables to equation (20) and can be rewritten as

$$y_t = \boldsymbol{x}_t^{'}\boldsymbol{\beta} + \epsilon_t, \tag{21}$$

where $y_t = \mathrm{RV}_{t+1}$, $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tk})^{'}$ is the k-dimensional vector of regressors, including the lagged daily, weekly and monthly realized volatilities and extra explanatory variables, $\boldsymbol{\beta}$ is the k-dimensional coefficients and ϵ_t is the same as (20). Based on the form of (21), one can use the ordinary least squares estimator for $\boldsymbol{\beta}$. In this paper, our analysis is based on the original HAR-RV model with $\boldsymbol{x}_t = (1, \mathrm{RV}_t, \mathrm{RV}_t^{w}, \mathrm{RV}_t^{m})^{'}$.

4.2 Data and Evidence of Structural Breaks

In this study, we use price data of Shanghai Securities Composite Index (SSCI) from the February 6th, 2007 to January 17th, 2018 with a total 2665 trading days. SSCI is one of the most important stock indices that Chinese investors pay most attention on. The trading hours are from 9:00 a.m. to 11:30 a.m. and from 13:30 p.m. to 15:00 p.m. with a two-hour break between 11:30 and 13:30, so there are $J = 48$ data in each trading day when 5-min frequency is adopted. All of the data are taken from the Wind database. During the sample period, the Chinese stock market has experienced the creation and burst of speculative bubbles twice, which happened during 2007–2008 and 2014–2015 respectively. These events have great influence on the investigators' trading behavior that may lead to structural breaks in realized volatility. The time series plot of RV_t, $t = 1, \ldots, 2665$ is displayed in Figure 2. The associated summary statistics are shown in Table 2. Figure 2 shows that the RV has the feature of clustering: the volatility during 2007–2008 and 2014–2015 are very high, which coincides with the global financial crises and Chinese stock market bubble. From Table 2, the daily, weekly and monthly RV are right skewed and have large excess kurtosis, indicating that the RV data are heavy-tailed non-Gaussian distributed. This is further confirmed by the results of the Jarque-Bera test, which rejects the null hypothesis of normal distribution at 1% significance level.

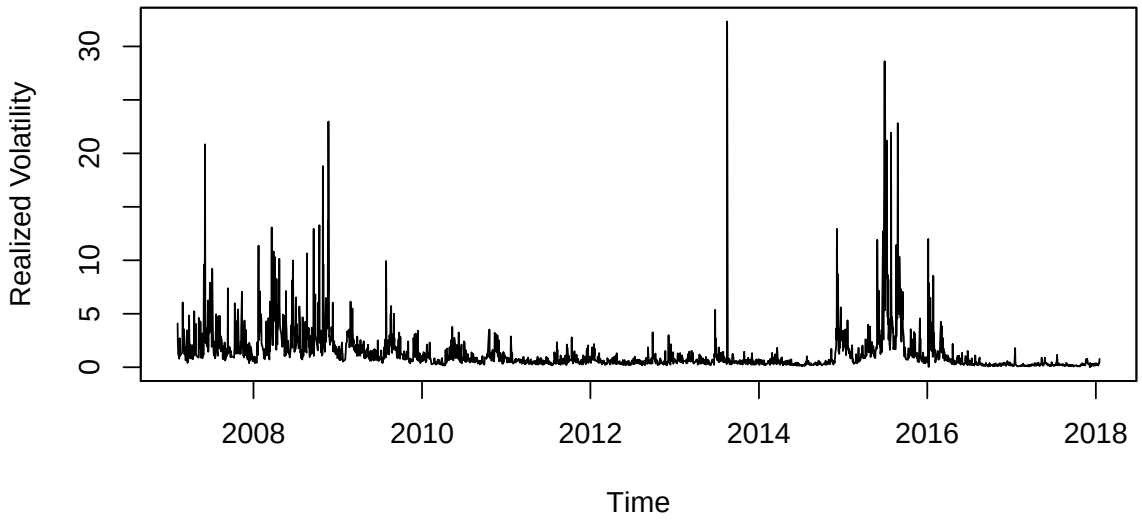

Figure 2: Realized volatility series of Shanghai Securities Composite Index from February 6th, 2007 to January 17th, 2018.

Next we use the MOSUM test for structural breaks in RV of SSCI. The regresssion process used in this MOSUM test is HAR model. Figure 3 plots the empirical MOSUM process defined in (16), with bandwidth parameter fixed at $h = 0.1$. The red horizontal line represents the confidence boundary at significant level

Table 2: Summary statistics for realized volatility of Shanghai Securities Composite Index from February 6th, 2007 to January 17th, 2018. (I will change this table later)

	RV_t	RV_t^w	RV_t^m
Mean	1.4472	1.4497	1.4612
Median	0.7751	0.8097	0.8921
Minimum	0.0255	0.1024	0.1362
Maximum	32.3302	19.0123	10.2309
Std.Dev.	2.2153	1.7243	1.4401
Skewness	5.8776	3.6289	2.3558
Kurtosis	51.6150	20.2895	7.7555
Jarque-Bera	2.88e5	4.77e4	8.47e3

0.05. It is found that the empirical MOSUM process has crossed the boundary in June and September, 2009, implying the occurence of structural breaks. This shows the fact that the Chinese market were experiencing a severe financial crisis, followed by a long period of adjustment. In addition, the right vertical line shows a local maximum of the empirical MOSUM process occurs on February 13th, 2015. The empirical MOSUM process at this time is near the confidence boundary, although has not crossed, implying another possible structural break. This verifies a well known Chinese bull period that the SSCI began to boom since the beginning of 2015 and reached its local highest point on June 12th, 2015. In conclusion, the MOSUM test detects two structural breaks in the RV of SSCI in June and September 2009, and one possible structural break in February 2015.

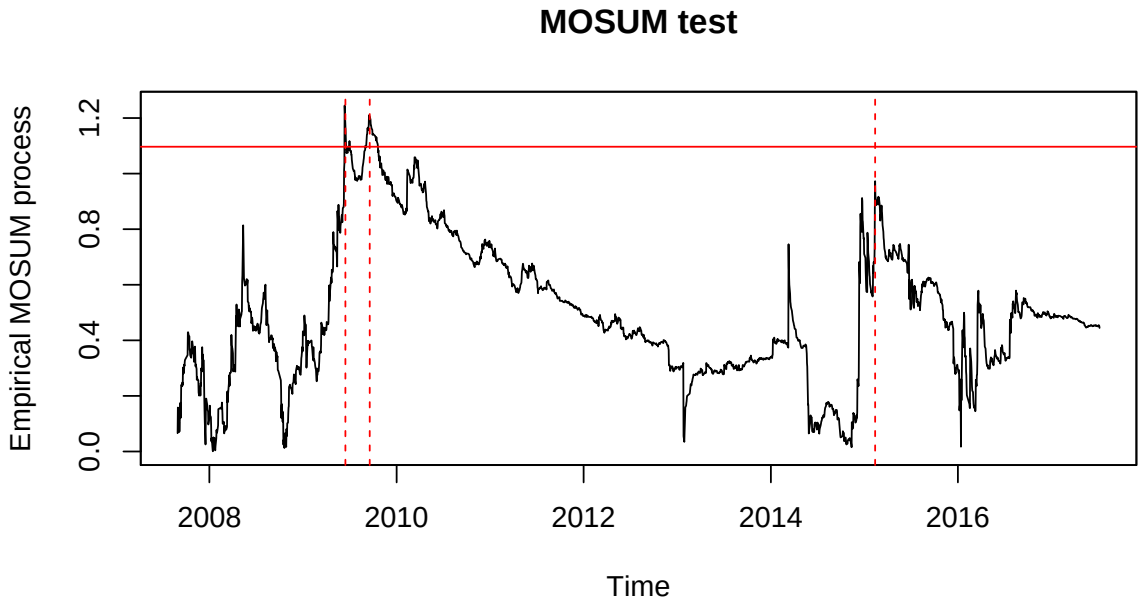

Figure 3: The empirical Rec-MOSUM process for testing structural breaks. The red horizontal line represents the confidence boundary at significant level 0.05.

4.3 Out-of-sample forecasting results

The out-of-sample performance of a forecasting method is critical to market participants, because they can adjust their investment strategies according to the expected realized volatility. To investigate the out-of-sample performance of our proposed exponentially weighted forecast combination, we conduct an out-of-sample analysis which is based on the subsamples starting from March 29, 2017 to January 17 2018 (a total of 200 observations). All the analysis were based on the HAR model (20).

We start by using data up to March 28, 2017 to estimate equation (20) and forecast the next realized volatility. The five forecast combination methods introduced in Section 2 (EW, LW, ROCW1, ROCW2, ExpW) and the two-stage method (denoted as TS) were applied to make a one-step-ahead prediction of the realized volatility on March 29, 2017. Note that the last break point, T_b, used in two-stage method to determine the estimation window and in ExpW to calculate the decaying parameter γ, were estimated by MOSUM test based on the samples up to March 28, 2017. We also included a full sample forecast as the benchmark. We then expand the estimation sample to contain March 29, 2017, and make our next one-step-ahead predictions for the realized volatility on March 30, 2017. We continue expanding the sample and forecasting in this way until we have a full sect of 200 predictions for every forecasting method.

To evaluate the T_h-ahead forecasting accuracy of each method at forecast horizon T, we use the following evaluation criteria,

$$\begin{aligned}
\text{L1: MSFE} &= \frac{1}{T_h}\sum_{t=1}^{T_h}(\text{RV}_{T+t} - \widehat{\text{RV}}_{T+t})^2,\\
\text{L2: MAFE} &= \frac{1}{T_h}\sum_{t=1}^{T_h}\left|\text{RV}_{T+t} - \widehat{\text{RV}}_{T+t}\right|,\\
\text{L3: MSPE} &= \frac{1}{T_h}\sum_{t=1}^{T_h}\left(1 - \frac{\widehat{\text{RV}}_{T+t}}{\text{RV}_{T+t}}\right)^2,\\
\text{L4: MAPE} &= \frac{1}{T_h}\sum_{t=1}^{T_h}\left|1 - \frac{\widehat{\text{RV}}_{T+t}}{\text{RV}_{T+t}}\right|,
\end{aligned}$$

where $\widehat{\text{RV}}_{T+t}$ is the forecasting value of RV_{T+t}. MSFE and MAFE denote the mean squared forecast error and the mean absolute forecast error of RV forecasts, respectively. MSPE and MAPE denote the mean squared percentage error and the mean absolute percentage error, respectively. In empirical studies, smaller values of MSFE, MAFE, MSPE and MAPE indicate better forecasting accuracies. Table 3 presents the evaluation results. It shows that all values less than 1, implying all forecating methods outperform the benchmark (i.e the full sample forecast). Specifically, ExpW generally minimizes the loss functions, followed by two-stage method, LW and EW. ROC weighting forecast combinations (ROCW1 and ROCW2) perform worses among all loss functions, that may because it places much weights on the estimation windows before Feburary 2015, i.e a possible structural detected by MOSUM test (see Figure 4). Consider the forecast on Jan 17, 2018 (the

last observation) as an example. Figure 4 presents the corresponding weights used on this forecasting date. It is evident that ROC weighting schemes (ROCW1 and ROCW2) place too less weight on the estimation windows in the right dotted vertical line, which are high reliable data that after the last possible break points. However, ROCW1 and ROCW2 put large weights on the first 700 estimation windows, which is obviously before the first break point, which damage the forecast of ROC weighting schemes. In contrast, LW and our proposed ExpW assign enough weights on the most recent forecast components, ensureing a small forecasting bias and a more robust prediction.

Table 3: Values of different loss functions realitive to the benchmark.

Loss functions	HAR	EW	LW	ROCW1	ROCW2	TS	ExpW
MSFE	1.000 (0.000)	0.782 (0.000)	0.748 (0.000)	0.837 (0.000)	0.853 (0.000)	0.732 (**0.107**)	<u>0.716</u> (**1.000**)
MAFE	1.000 (0.000)	0.840 (0.000)	0.811 (0.000)	0.880 (0.000)	0.891 (0.000)	0.798 (0.032)	<u>0.785</u> (**1.000**)
MSPE	1.000 (0.080)	0.793 (0.080)	0.743 (0.096)	0.819 (0.080)	0.809 (0.080)	0.780 (0.096)	<u>0.705</u> (**1.000**)
MSAE	1.000 (0.080)	0.890 (0.080)	0.861 (0.082)	0.905 (0.080)	0.899 (0.080)	0.882 (0.082)	<u>0.839</u> (**1.000**)

[a] The smallest relative MSFE in each row of the table is highlighted by underscoring. The values in the brackets denotes the MCS test P-values for the corresponding forecasting methods and loss functions, and the numbers in bold denote the corresponding methods that are included in MCS for the significant level of 10% and the semi-quadratic statistic is used.

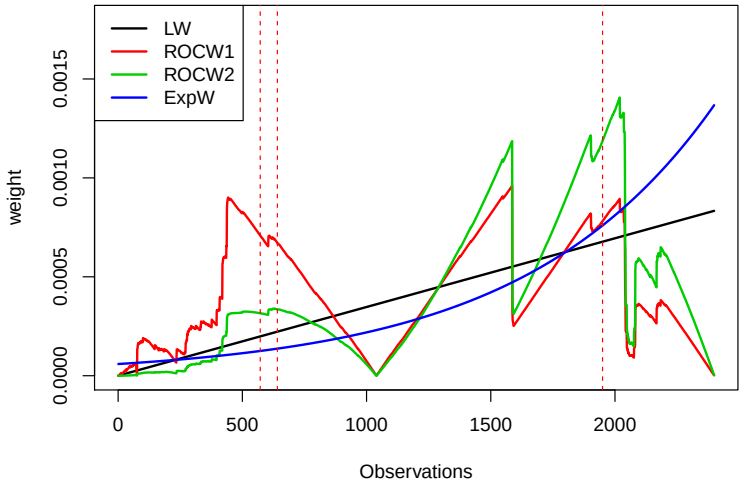

Figure 4: Combination weights for estimation windows in the RV series. The red vertical lines are the three time points that structural breaks may happen as noted in Section 4.2.

However, ranking forecast methods by a particular loss criterion is affected by the sample choice. To solve this problem, we utilize the model confidence set (MCS) test developed by Hansen et al. (2011) [26] to compare the point forecast accuracy among different forecast combination methods and the benchmark ordinary HAR-

RV model. Given a set of candidate forecast models, M_0, the goal of MCS is to identify $M^*_{1-\alpha} \subset M_0$, where $M^*_{1-\alpha}$ is the set of the models that contains the "best" forecast model given a level of confidence α. In particular, in terms of the MSPE loss function, the MCS procedures start with the full set M_0 and repeatedly test the null hypothesis of equal forecast accuracy:

$$H_0 : \quad \mathrm{E}(d_{ij,t}) = 0 \quad \forall\, i, j \in M_0\,, \tag{22}$$

where $d_{ij,t} = \mathrm{MSPE}_{i,t} - \mathrm{MSPE}_{j,t}$ is the loss differential between models i and j in the set. Then the MCS procedure sequentially eliminates the worst performance model from M_0 if the null is rejected at the significance level of α. The trimming of of models is repeated until the null is not rejected any more and the survival models form the MCS $M^*_{1-\alpha}$. Note that the number of models in the MCS increases as we decrease α just like the size of a confidence interval.

Following Hansen et al. (2011) [26], we employ the semi-quadratic statistic

$$T_{SQ} = \sum_{i,j \in M_0} \frac{\bar{d}^2_{ij}}{\widehat{\mathrm{Var}}(\bar{d}_{ij})}\,,$$

where $\widehat{\mathrm{Var}}(\bar{d}_{ij})$ is an estimate of $\mathrm{Var}(\bar{d}_{ij})$, which can be implemented through block bootstrap procedure [28].

In our study, we set $\alpha = 0.1$ meaning that the model with a p-value smaller than 0.1 will be excluded from the MCS. The values in the brackets of Table 3 report the p-values for each forecasting method and loss function based on 10,000 block bootstraps. The forecast combination methods with weights ROCW1, ROCW2, EW and LW are excluded from the MCS under all loss functions. On the other hand, our proposed exponential weight schemes are included in the MCS for all loss functions. This fact shows that different weighting schemes lead to different out-of-sample performances and our proposed exponential weighting scheme is significantly better than other weights in forecasting RV of SSCI. Note that the ordinary HAR model is excluded from the MCS under all loss functions, indicating this simple model can not provide more accurate forecasts than the forecast combination method with traditional weights or our proposed exponential weights.

5 Conclusion

In this paper, we have investigated the problem of forecasting in the presence of structural breaks and proposed a exponential weighting forecast combination method (ExpW), which could effectively reduce the MSFE in the presence of single and multiple structural breaks. We have analyzed the theoretical properties of ExpW, and compared it with forecast combinations with other weighting schemes, two-stage method, and a benchmark method which assumes that no structural breaks have occured. By extensive empirical studies based on different loss functions and the formal MCS test and a series of simulation studies, we demonstrate that the forecast combination method with our proposed weighting scheme outperforms other models have mentioned

before.

References

[1] Hansen, B. E. (2001). The new econometrics of structural change: dating breaks in US labour productivity. Journal of Economic perspectives, 15(4), 117-128.

[2] Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. Journal of applied econometrics, 18(1), 1–22.

[3] Pesaran, M. H., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. The Review of Economic Studies, 73(4), 1057-1084.

[4] Clements, M. P., & Hendry, D. F. (2006). Forecasting with breaks. Handbook of economic forecasting, 1, 605–657.

[5] Giacomini, R., & Rossi, B. (2009). Detecting and predicting forecast breakdowns. The Review of Economic Studies, 76(2), 669-705.

[6] Pesaran, M. H., Pick, A., & Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. Journal of Econometrics, 177(2), 134–152.

[7] Brown, R. L., Durbin J., & Evans J. M., (1975). Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society, B 37, 149–163.

[8] Chu, C. J., Hornik, K., & Kaun, C. (1995). Mosum tests for parameter constancy. Biometrika, 82(3), 603-617.

[9] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica, 28(3), 591-605.

[10] Pesaran, M. H., & Timmermann, A., (2007). Selection of estimation window in the presence of breaks. Journal of Econometrics 137, 134–161.

[11] Tian, J., & Anderson, H.M. (2014). Forecast combinations under structural break uncertainty. International Journal of Forecasting, 30, 161–175.

[12] Harvey, A.C., 1989. Changes in expected security returns, risk, and the level of interest rates. Journal of Finance 44, 1191–1217.

[13] Kirch, C., Muhsal, B., & Ombao, H. (2015). Detection of changes in multivariate time series with application to EEG data. Journal of the American Statistical Association, 110(511), 1197-1216.

[14] Corsi, F., 2009. A simple approximate memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

[15] Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905

[16] Andersen, T.G., Bollerslev, T., Diebod, F. X., 2003. Modelling and forecasting realized volatility. Econometrica 71, 579–625.

[17] Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realized and implied volatility measurements. Journal of Empirical Finance 12, 445–475.

[18] Asai, M., McAleer, M., Medeiros, M.C., 2012. Modelling and forecasting noisy realized volatility. Computational Statistics & Data Analysis 56, 217–230.

[19] Blair, B.J., Poon, S.H., Taylor, S.J., 2001. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. Journal of Econometrics 105, 5–26.

[20] Hansen,P.R., Lunde, A., 2006. Realized variance and market microstructure noise: rejoinder. Journal of Business and Economics Statistics 24, 208–218.

[21] Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. Review of Economics Study 75, 339–369.

[22] Zhuo, H., Hao, L., & Wang, T. (2016). Modeling long memory volatility using realized measures of volatility: a realized HAR Garch model. Economic Modelling, 52, 812-821.

[23] Tian, F., Ke, Y., & Chen, L. (2017). Realized volatility forecasting of agricultural commodity futures using the har model with time-varying sparsity. International Journal of Forecasting, 33(1), 132-152.

[24] Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Review of Economics and Statistics 89, 701–720.

[25] Corsi, F. Renó, R., 2010. HAR volatility modelling with heterogeneous leverage and jumps. Available at SSRN 1316953.

[26] Hansen, P.R., Lunde, A., James, M.N., 2011. The model confidence set. Econometrica 79, 453–497.

[27] Pesaran, M.H., Pick, A., 2011. Forecast combination across estimation windows. Journal of Business and Economic Statistics 29, 307–318.

[28] Politis, D.N., Romano, J.P. 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Appendix

Proof of Equation (10)

Based on section 3.2 in Pesaran and Pick (2011) [27], we know the bias in estimating $\beta^{(2)}$ by $\hat{\beta}(\eta)$ is given by

$$\beta^{(2)} - \hat{\beta}(\eta) = (\beta^{(2)} - \beta^{(1)})\Big(\frac{T_b - \eta}{T - \eta + 1}\Big)\mathrm{I}(T_b - \eta). \tag{23}$$

Furthermore, $\mu_y - \bar{y}(\eta) = O_p(\frac{1}{\sqrt{T-\eta+1}})$ and $\mu_x - \bar{x}(\eta) = O_p(\frac{1}{\sqrt{T-\eta+1}})$. Conditional on x_{T+1}, the ExpW forcast error is

$$\begin{aligned}
\xi_{T+1}^{(ExpW)} &= \frac{1-\gamma}{1-\gamma^{T-\omega+1}} \sum_{\eta=1}^{T-\omega+1} \gamma^{T-\omega+1-\eta}\Big[y_{T+1} - \bar{y}(\eta) - \hat{\beta}(\eta)(x_{T+1} - \bar{x}(\eta))\Big] \\
&= \frac{1-\gamma}{1-\gamma^{T-\omega+1}} \sum_{\eta=1}^{T-\omega+1} \gamma^{T-\omega+1-\eta}\Big\{\sigma\epsilon_{T+1} + [\beta^{(2)} - \hat{\beta}(\eta)](x_{T+1} - \mu_x) \\
&\quad + \mu_y - \bar{y}(\eta) + \hat{\beta}(\eta)[\bar{x}(\eta) - \mu_x]\Big\} \\
&= \sigma\epsilon_{T+1} + \frac{(1-\gamma)(x_{T+1} - \mu_x)}{1-\gamma^{T-\omega+1}}\sigma\lambda \sum_{\eta=1}^{T-\omega+1} \gamma^{T-\omega+1-\eta}\Big(\frac{T_b - \eta}{T - \eta + 1}\Big)\mathrm{I}(T - \eta + 1) \\
&\quad + \frac{1-\gamma}{1-\gamma^{T-\omega+1}} \sum_{\eta=1}^{T-\omega+1} \gamma^{T-\omega+1-\eta}\Big\{\mu_y - \bar{y}(\eta) + \hat{\beta}(\eta)[\bar{x}(\eta) - \mu_x]\Big\} \\
&= \sigma\epsilon_{T+1} + \sigma B_{\omega,T+1}^{(ExpW)} + \Delta_3,
\end{aligned} \tag{24}$$

where Δ_3 denotes the last term in the right part. Now consider Δ_3 for fixed $R_\omega = \frac{\omega}{T}$ and $R_b = \frac{T_b}{T}$ as $T \to \infty$. Noticing $\mu_y - \bar{y}(\eta) = O_p(\frac{1}{\sqrt{T-\eta+1}})$ and $\mu_x - \bar{x}(\eta) = O_p(\frac{1}{\sqrt{T-\eta+1}})$, and denoting $T - \eta + 1$ by v, we have

$$\begin{aligned}
\Delta_3 &= \Big|\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \sum_{\eta=1}^{T-\omega+1} \gamma^{T-\omega+1-\eta}\Big\{\mu_y - \bar{y}(\eta) + \hat{\beta}(\eta)[\bar{x}(\eta) - \mu_x]\Big\}\Big| \\
&\le \frac{K(1-\gamma)}{1-\gamma^{T-\omega+1}} \sum_{\eta=1}^{T-\omega+1} \gamma^{T-\omega+1-\eta}\Big\{\frac{1}{\sqrt{T-\eta-1}} + \frac{T_b - \eta}{(T-\eta-1)^{3/2}}\mathrm{I}(T_b - \eta)\Big\} \\
&= \frac{K(1-\gamma)}{\sqrt{T}(1-\gamma^{T-\omega+1})} \sum_{v=\omega}^{T} \gamma^{v-\omega}\Big\{\sqrt{\frac{T}{v}} + \sqrt{\frac{T}{v}}\frac{v + T_b - T - 1}{v}\mathrm{I}(v + T_b - T - 1)\Big\},
\end{aligned} \tag{25}$$

where K is a positive constant. Noting $v \ge \omega$, we have $\sqrt{T/v} \le \sqrt{T/\omega} = 1/\sqrt{R_\omega}$; and, noting that $(v + T_b - T - 1)/v^{3/2}$ is maximized at $v^* = 3(T + 1 - T_b)$, we can show

$$\sqrt{\frac{T}{v}}\frac{v + T_b - T - 1}{v}\mathrm{I}(v + T_b - T - 1) \le \frac{2}{3\sqrt{3}}\sqrt{\frac{T}{T + 1 - T_b}} = \frac{2}{3\sqrt{3}}\sqrt{\frac{1}{1 + 1/T - R_b}},$$

which is bounded as $T \to \infty$. Therefore,

$$\Delta_3 < \frac{K(1-\gamma)}{\sqrt{T}(1-\gamma^{T-\omega+1})} \sum_{v=\omega}^{T} \gamma^{v-\omega} \Big\{ \sqrt{\frac{1}{R_\omega}} + \frac{2}{3\sqrt{3}} \sqrt{\frac{1}{1-R_b}} \Big\} = O_p(\frac{1}{\sqrt{T}}).$$

In summary, $\sigma^{-1}\xi_{\omega,T+1}^{(ExpW)} = \epsilon_{T+1} + B_{\omega,T+1}^{(ExpW)} + O_p(\frac{1}{\sqrt{T}})$.

□

Proof of Proposition 1

$\text{MSFE}\Big(\hat{y}_{t+1}(1)\Big) > \text{MSFE}\Big(\hat{y}_{\omega,t+1}^{(EW)}\Big)$ follows directly from proposition 3 in Pesaran and Pick (2011) [27], thus we just prove $\text{MSFE}\Big(\hat{y}_{\omega,t+1}^{(EW)}\Big) > \text{MSFE}\Big(\hat{y}_{\omega,t+1}^{(ExpW)}\Big)$. The difference of the scaled forecast error of $\hat{y}_{\omega,t+1}^{(ExpW)}$ and $\hat{y}_{\omega,t+1}^{(EW)}$ is

$$\sigma^{-1}\Big(\xi_{\omega,T+1}^{(ExpW)} - \xi_{\omega,T+1}^{(EW)}\Big) = \lambda(x_{T+1} - \mu_x) \sum_{\eta=1}^{T-\omega+1} \zeta(\eta) \Big[\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} - \frac{1}{T-\omega+1} \Big] + O_p(\frac{1}{\sqrt{T}}), \quad (26)$$

where $\zeta(\eta) = \frac{T_b-\eta}{T-\eta+1}\text{I}(T_b - \eta)$. It is evident that $\zeta(\eta)$ is non-negative and decreasing among η. Moreover, noting that the weight of ExpW strictly increases as η increases and the average weight is $1/(T-\omega+1)$, thus there exists a interger m $(1 < m < T-\omega+1)$ satisfying

$$\begin{cases} \frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} \geq \frac{1}{T-\omega+1} & ,\forall \eta \geq m \\ \frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} < \frac{1}{T-\omega+1} & ,\forall \eta < m \end{cases} \quad (27)$$

For $\eta < m$, we have $\zeta(\eta) > \zeta(m)$ and $\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} - \frac{1}{T-\omega+1} < 0$, which imply

$$\sum_{\eta=1}^{m-1} \zeta(\eta) \Big[\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} - \frac{1}{T-\omega+1} \Big] < \sum_{\eta=1}^{m-1} \zeta(m) \Big[\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} - \frac{1}{T-\omega+1} \Big]. \quad (28)$$

Now consider $\eta \geq m$, we have $\zeta(\eta) \leq \zeta(m)$ and $\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} - \frac{1}{T-\omega+1} \geq 0$, which imply

$$\sum_{\eta=m}^{T-\omega-1} \zeta(\eta) \Big[\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} - \frac{1}{T-\omega+1} \Big] \leq \sum_{\eta=m}^{T-\omega-1} \zeta(m) \Big[\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} - \frac{1}{T-\omega+1} \Big]. \quad (29)$$

Combining equation (28) and (29) together, we know

$$\sum_{\eta=1}^{T-\omega+1} \zeta(\eta) \Big[\frac{1-\gamma}{1-\gamma^{T-\omega+1}} \gamma^{T-\omega+1-\eta} - \frac{1}{T-\omega+1} \Big] < 0,$$

which imple the abosulte bias of EW is bigger than that of ExpW; and noticing $Var(\hat{y}_{\omega,T+1}^{(ExpW)}) = Var(\hat{y}_{\omega,T+1}^{(EW)}) = \sigma^2$ as $T \to \infty$, the $\text{MSFE}\Big(\hat{y}_{\omega,t+1}^{(EW)}\Big) > \text{MSFE}\Big(\hat{y}_{\omega,t+1}^{(ExpW)}\Big)$

□