Forecasting Random Walks under Drift Instability*

M. Hashem Pesaran
University of Cambridge, CIMF, and USC

Andreas Pick
De Nederlandsche Bank and University of Cambridge, CIMF

January 20, 2008

Abstract

This paper considers forecast averaging when the same model is used but estimation is carried out over different estimation windows. It develops theoretical results for random walks when their drift and/or volatility are subject to one or more structural breaks. It is shown that compared to using forecasts based on a single estimation window, averaging over estimation windows leads to a lower bias and to a lower root mean square forecast error for all but the smallest of breaks. Similar results are also obtained when observations are exponentially down-weighted, although in this case the performance of forecasts based on exponential down-weighting critically depends on the choice of the weighting coefficient. The forecasting techniques are applied to 20 weekly series of stock market futures and it is found that average forecasting methods in general perform better than using forecasts based on a single estimation window.

Keywords Forecast combinations, averaging over estimation windows, exponentially down-weighting observations, structural breaks
JEL classifications C22, C53

*We are grateful to Todd Clark, Jiangqin Fan, Clive Granger, Alessio Sancetta, Christoph Schleicher, Ron Smith, Vanessa Smith and Jim Stock for comments. This paper was written while the second author was Sinopia Research Fellow at the University of Cambridge. He acknowledges financial support from Sinopia, quantitative specialist of HSBC Global Asset Management. The opinions expressed in this paper do not necessarily reflect those of DNB.

1 Introduction

There now exists a sizeable literature on the possible merits of combining forecasts obtained from different models, reviewed by Clemen (1989), Stock and Watson (2004), and more recently by Timmermann (2006). Bayesian and equal weighted forecast combinations are being used increasingly in macroeconomics and finance to good effects. In this literature, the different forecasts are typically obtained by estimating a number of alternative models over the same sample period. Pesaran and Timmermann (2007) argue that the forecast averaging procedure can be extended to deal with other types of model uncertainty, such as the uncertainty over the size of the estimation window, and propose the idea of averaging forecasts from the same model but obtained over different estimation windows. Using Monte Carlo experiments these authors show that this type of forecast averaging reduces the mean square forecast error (MSFE) in many cases when the underlying economic relations are subject to structural breaks.

The idea of forecast averaging over estimation windows has been fruitfully applied in macro economic forecasting. Assenmacher-Wesche and Pesaran (2008) average forecasts based on different VARX* models of the Swiss economy estimated over different estimation windows and observe that averaging forecasts across windows result in further improvements over averaging of forecasts across models. Similar results are obtained by Pesaran, Schuermann and Smith (2009) who apply the forecast averaging ideas to global VARs composed of 26 individual country/region VARX* models. It is therefore of interest to see if some theoretical insights can be gained in support of these empirical findings.

In this paper we begin by deriving theoretical results for the average windows (AveW) forecast procedure in the case of random walk models subject to breaks. The most interesting case is when the break occurs in the drift term, but we shall also consider other cases when the volatility of the random walk undergoes changes, and when the breaks in the drift and the volatility of the random walk model occur simultaneously. We consider both the case of a single break as well as when there are a multiplicity of breaks.

We also compare the AveW forecasting procedure with an alternative method sometimes employed in the literature where the past observations are down-weighted exponentially such that the most recent observations carry the largest weight in the estimation and forecasting, see Gardner (2006) for a review. We refer to this as the exponential down-weighted (ExpW) forecast. This approach is related to the random coefficient model and its performance in practice crucially depends on the parameter, γ, used to down-weight the past observations.

Restricting attention to random walk models allows us to simplify the problem and attain exact theoretical results that shed light on the properties of these forecasting methods. In particular, we show that in the presence of breaks AveW and ExpW forecasts always have a lower bias than forecasts based on a single estimation window. The forecast variance depends on the size and the time of the break. For all but the smallest break sizes, however, the MSFE of the AveW and ExpW forecasts are also smaller than those of the single window forecasts.

An attractive feature of these methods is that no exact information about the structural break is necessary. This contrasts with the conventional approach of estimating the break point using methods such as those of Bai and Perron (1998, 2003) before incorporating them into the modeling process or incorporate the break process into the estimation procedure using methods such as that of Hamilton (1989); see Clements and Hendry (2006) for a review of the recent literature. As argued in Pesaran and Timmermann (2007), to optimally exploit break information in forecasting one needs to know the point as well as the size of the break(s). Even if the point of the break can be estimated with some degree of confidence, it is unlikely that the size of the break can be estimated accurately, since it involves estimating the model over the pre- as well as the post-break periods. If the distance to break (measured from the date at which forecasts are made) is short the post-break parameters are likely to be rather poorly estimated relative to the ones obtained using pre-break data. If the pre- and post-break samples are both relatively large, it might be possible to estimate the size of the break reasonably accurately, but in such cases the break information might not be all that important.

Clark and McCracken (2006) argue that averaging over different models can improve forecasts in the presence of model instability, and our approach is complementary to this. More closely related to our approach is the suggestion by Clark and McCracken (2004) that averaging expanding and rolling windows can be useful for forecasting when faced with structural breaks. This can be seen as a limited version of AveW forecasts where only two different windows are combined.

A further reason for considering the random walk model with drift and volatility instability is that it is generally thought to describe the stochastic properties of many macroeconomic and financial time series. In this paper we apply the AveW and the ExpW procedures to forecasting weekly returns on futures contracts in twenty world equity markets. We find that the AveW forecasts outperform the single window forecasts in the root mean squared sense in 18 out of the 20 equity markets. Although, the results did not prove to be statistically significant in the case of individual equity returns, which could be due to the high volatility of equity returns, particularly over the past two years. The sample period being considered differs across equity indices due to differences in the start dates of the equity futures markets. But the forecasts for all the 20 equity futures cover the past two years and end on November 24, 2008.

The rest of the paper is organized as follows: Section 2 sets out the model and Section 3 develops the AveW forecasting procedure and its properties. Section 4 considers the ExpW forecast procedure. Section 5 reports the results of the applications to stock market futures and, finally, Section 6 draws some conclusions.

2 Basic model and notations

Consider the following random walk model with drift

$$x_t = x_{t-1} + \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.}\left(0, \sigma_t^2\right).$$

Define $y_t = x_t - x_{t-1}$, then we have the model

$$y_t = \mu_t + \varepsilon_t, \tag{1}$$

which is defined over the sample period $t = 1, 2, \ldots, T$, and where it is believed that its drift coefficient, μ_t, and its volatility, σ_t, have been subject to a single break at time $t = T_b$ $(1 < T_b < T)$

$$\mu_t = \begin{cases} \mu_1, \; \forall \; t \leq T_b \\ \mu_2, \; \forall \; t > T_b \end{cases},$$

$$\sigma_t = \begin{cases} \sigma_1, \; \forall \; t \leq T_b \\ \sigma_2, \; \forall \; t > T_b \end{cases}.$$

The aim is to forecast x_{T+1}, or y_{T+1} based on the observations, $y_1, y_2, ..., y_T$. In the case where it is known with certainty that the random walk model has not been subject to any breaks, the sample mean, $\bar{y}_T = T^{-1}\Sigma_{t=1}^{T} y_t$ yields the most efficient forecast in the mean squared error sense. However, when the process is subject to break(s) more efficient forecasts could be obtained. As shown in Pesaran and Timmermann (2007) there is typically a trade off between bias and variance of forecasts. For example, when there is a break in the drift term the use of the full sample will yield a biased forecast but will continue to have the least variance. On the other hand a forecast based on the sub-sample $\{y_{T_i}, y_{T_i+1}, \ldots, y_T\}$, where $T_i > 1$ is likely to have a lower bias but could be inefficient due to a higher variance as compared to $\bar{y}_T$. Knowing the point of the break helps but cannot be exploited optimally unless a reliable estimate of the size of the break, $|\mu_2 - \mu_1| / \sigma$, can also be obtained. Often this is not possible since in most applications of interest breaks might be quite recent and $T - T_b$ too small for a reliable estimation of μ_2.

In the absence of reliable information on the point and the size of the break(s) in μ_t and σ_t, a forecasting procedure which is reasonably robust to such breaks will be of interest. One approach considered in Pesaran and Timmermann (2007) is to use different sub-windows to forecast and then average the outcomes, either by means of cross-validated weights or by simply using equal weights.

To this end consider the sample $\{y_{T_i}, y_{T_i+1}, \ldots, y_T\}$ with $T_i > 1$, which yields an observation window of size $T - T_i + 1$. It proves convenient to denote this observation window by $w_i = (T - T_i + 1)/T$, which represents the fraction w_i of the single window (from the point of the forecast) used in estimation. The estimation process could start with a minimum window $\{y_{T_{\min}}, y_{T_{\min}+1}, \ldots, y_T\}$ of size $w_{\min} = (T - T_{\min} + 1)/T$. From $w_{\min}$ other larger windows can be considered with $T_i = T_{\min}, T_{\min} - 1, \ldots, T_{\min} - m$, where $m = T_{\min} - 1$, thus yielding $m+1$ separate estimation windows. More specifically

$$w_i = w_{\min} + \frac{i}{T}, \text{ for } i = 0, 1, ..., m, \tag{2}$$

with

$$w_0 = w_{\min}, \text{ and } w_m = 1,$$

so that

$$m = T(1 - w_{\min}). \tag{3}$$

Clearly, $w_m = 1$ corresponds to the full sample.

The one-step ahead forecast based on a given window w_a is

$$\hat{y}_{T+1}(w_a) = \hat{\mu}_{T+1}(w_a), \tag{4}$$

where

$$\hat{\mu}_{T+1}(w_a) = \frac{1}{Tw_a} \sum_{t=T_a}^{T} y_t = \frac{1}{Tw_a} \sum_{t=T(1-w_a)+1}^{T} y_t.$$

3 Average window forecast

The AveW forecast is defined by the simple forecast combination rule

$$\hat{y}_{T+1}(\text{AveW}) = \frac{1}{m+1} \sum_{i=0}^{m} \left(\frac{1}{Tw_i} \sum_{t=T(1-w_i)+1}^{T} y_t \right), \tag{5}$$

where forecasts from all windows are given equal weights.

The first object of interest in this paper is to compare the single-window and the AveW forecasts, $\hat{y}_{T+1}(w_i)$ and $\hat{y}_{T+1}(\text{AveW})$, in the mean squared error sense. In the case of the single window forecast we focus on the most frequently encountered case where all observations in a given sample is used, namely we consider $\hat{\mu}_{T+1}(1) = \bar{y}_T$. In recursive estimation these alternative forecasts can be considered both under expanding and rolling windows. The AveW procedure is therefore not an alternative to rolling forecasts and can be used irrespective of whether a rolling or an expanding window is used in recursive forecasting.

3.1 Break in drift only

In the first instance assume that a single break occurs in the drift of the process at date $1 < T_b < T$, whereas the error variance is constant, that is, $\mu_1 \neq \mu_2$ but $\sigma_1 = \sigma_2 = \sigma$. The distance to the break is defined by $d = (T - T_b)/T$. In this case the one-step ahead forecast of y_{T+1} based on a given window of size wT (from $t = T$) is given by

$$\hat{y}_{T+1}(w) = \mu_2 \left[1 - \mathrm{I}\left(w - d\right)\right] + \mathrm{I}\left(w - d\right) \left[\frac{d\mu_2 + \left(w - d\right)\mu_1}{w}\right] + \frac{1}{Tw} \sum_{t=T(1-w)+1}^{T} \varepsilon_t,$$

where $\mathrm{I}(c)$ is an indicator function which is unity if $c > 0$ and zero otherwise. It is clear that if $w \leq d$ the forecast will have mean μ_2 and will be unbiased. There is, however, a bias when $w > d > 0$. The associated forecast error, $e_{T+1}(w) = y_{T+1} - \hat{y}_{T+1}(w)$, can then be written as

$$e_{T+1}(w) = (\mu_2 - \mu_1) \left(\frac{w - d}{w}\right) \mathrm{I}\left(w - d\right) + \varepsilon_{T+1} - \frac{1}{Tw} \sum_{t=T(1-w)+1}^{T} \varepsilon_t. \quad (6)$$

Hence, the forecast bias is

$$\mathrm{E}\left[e_{T+1}(w)\right] = (\mu_2 - \mu_1) \left(\frac{w - d}{w}\right) \mathrm{I}\left(w - d\right), \quad (7)$$

and since $(w - d)\,\mathrm{I}\left(w - d\right) > 0$, the direction of the bias depends on the sign of $(\mu_2 - \mu_1)$.

Scaling the forecast error by σ, we have the decomposition

$$\sigma^{-1} e_{T+1}(w) = u_{T+1} + B_{T+1}(w) - \frac{1}{Tw} \sum_{t=T(1-w)+1}^{T} u_t, \quad (8)$$

where

$$B_{T+1}(w) = \lambda \left(\frac{w - d}{w}\right) \mathrm{I}\left(w - d\right) \quad (9)$$

$\lambda = (\mu_2 - \mu_1)/\sigma$, and $u_t = \varepsilon_t/\sigma$. The first term, u_{T+1} represents the future uncertainty which is given and unavoidable, the second term is the 'bias' that depends on the size of the break, λ, and the distance to break, d, and the last term represents the estimation uncertainty that depends on Tw. The (scaled) mean squared forecast error (MSFE) for a window of size w is given

$$\mathrm{MSFE}(w) = 1 + B_{T+1}^2(w) + \frac{1}{Tw}. \quad (10)$$

Consider now the AveW forecast based on $m + 1$ successive windows of sizes from the smallest window fraction $w_{\min}$ to the largest possible one,

$w_m = 1$. While we need enough observations in the first window, $w_{\min} > 0$, we will assume that $w_{\min}$ is chosen to be sufficiently small so that $w_{\min} \leq d$. The AveW forecast constructed from these windows is then given by

$$\hat{y}_{T+1}(\text{AveW}) = \frac{1}{m+1} \sum_{i=0}^{m} \hat{y}_{T+1}(w_i).$$

The (scaled) one step ahead forecast error associated with the above average forecast is

$$\begin{aligned} \sigma^{-1} e_{T+1}(\text{AveW}) &= u_{T+1} + \frac{\lambda}{m+1} \sum_{i=0}^{m} \left(\frac{w_i - d}{w_i} \right) \mathrm{I}\left(w_i - d\right) \\ &\quad - \frac{1}{m+1} \sum_{i=0}^{m} \frac{1}{T w_i} \sum_{t=T(1-w_i)+1}^{T} u_t. \end{aligned}$$

Hence, the bias of the AveW forecast is given by

$$B_{T+1}(\text{AveW}) = \frac{\lambda}{m+1} \sum_{i=0}^{m} \left(\frac{w_i - d}{w_i} \right) \mathrm{I}\left(w_i - d\right), \tag{11}$$

and as before the sign of the bias depends on the sign of $(\mu_2 - \mu_1)$. In this case the computation of the variance of the forecast error is complicated due to the cross correlation of forecasts from the different windows. Let

$$\nu_T(w_i) = \frac{1}{T w_i} \sum_{t=T(1-w_i)+1}^{T} u_t,$$

then

$$\text{Cov}\left[\nu_T(w_i), \nu_T(w_j)\right] = \frac{\min(w_i, w_j)}{T w_i w_j}, \text{ for all } i, j = 0, 1, ..., m.$$

As a result it is easily verified that

$$\text{Var}\left[\hat{y}_{T+1}(\text{AveW})\right] = 1 + \left(\frac{1}{T}\right) \left(\frac{1}{m+1}\right)^2 \left[\sum_{i=0}^{m} \frac{1}{w_i} + 2 \sum_{i=0}^{m} \frac{i}{w_i} \right]. \tag{12}$$

Therefore, the scaled MSFE in this case is given by

$$\text{MSFE}(\text{AveW}) = 1 + B_{T+1}^2(\text{AveW}) + \text{Var}\left[\hat{y}_{T+1}(\text{AveW})\right], \tag{13}$$

with $B_{T+1}(\text{AveW})$ and $\text{Var}\left[\hat{y}_{T+1}(\text{AveW})\right]$ as defined above.

The difference between the scaled MSFE of the single window forecast (10) and that of the AveW Forecast (13) is

$$
\begin{aligned}
\mathrm{MSFE}(w_a;\lambda,d)-\mathrm{MSFE}(m,w_{\min};\lambda,d) &= \\
\lambda^2\left(\frac{w_a-d}{w_a}\right)^2 \mathrm{I}(w_a-d)+\frac{1}{Tw_a} & \\
-\left[\frac{\lambda}{m+1}\sum_{i=0}^{m}\frac{w_i-d}{w_i}\mathrm{I}(w_i-d)\right]^2-\frac{1}{(m+1)^2}\sum_{i=0}^{m}\frac{1+2i}{Tw_i}, &
\end{aligned} \tag{14}
$$

Since $m = T(1-w_{\min})$, for fixed values of $w_{\min}$ and d, as T becomes sufficiently large the bias and variance terms of the AveW forecast can be approximated by means of the Riemann integral. Using (2) and (3) we first note that

$$
\begin{aligned}
T &= m/(1-w_{\min}), \\
i &= T(w_i-w_{\min}) = m(w_i-w_{\min})/(1-w_{\min}).
\end{aligned}
$$

The bias term in (11) can be approximated using

$$
\begin{aligned}
\frac{1}{T}\sum_{i=0}^{m}\left(\frac{w_i-d}{w_i}\right)\mathrm{I}(w_i-d) &\overset{T\to\infty}{\longrightarrow} \int_d^1\left(\frac{x-d}{x}\right)dx, \\
&= (1-d)+d\ln(d)\geq 0,
\end{aligned}
$$

where the lower boundary of the integral, d, is due to the fact that the indicator function $\mathrm{I}(w_i-d)$ implies that values of the expression below d are zero.

Using the results in (12) we have

$$
\begin{aligned}
\frac{1}{T}\sum_{i=0}^{m}\frac{1+2i}{w_i} &= \frac{1}{T}\sum_{i=0}^{m}\frac{1+2T(w_i-w_{\min})}{w_i}, \\
&= \frac{1}{T}\sum_{i=0}^{m}\frac{1}{w_i}+\frac{2T}{T}\sum_{i=0}^{m}\frac{(w_i-w_{\min})}{w_i},
\end{aligned}
$$

which can be approximated using

$$
\frac{1}{T}\sum_{i=0}^{m}\frac{1}{w_i} \overset{T\to\infty}{\longrightarrow} \int_{w_{\min}}^1\frac{1}{x}dx = -\ln(w_{\min}),
$$

and

$$
\begin{aligned}
\frac{1}{T}\sum_{i=0}^{m}\frac{(w_i-w_{\min})}{w_i} &\overset{T\to\infty}{\longrightarrow} \int_{w_{\min}}^1\frac{x-w_{\min}}{x}dx, \\
&= 1-w_{\min}+w_{\min}\ln w_{\min}.
\end{aligned}
$$

Therefore, using the above results as $T \to \infty$ and $m \to \infty$ for a fixed $w_{\min} < d \leq 1$ and recalling that $T = m/(1 - w_{\min})$ we have

$$\mathrm{MSFE}(m, w_{\min}; \lambda, d) \approx \frac{\lambda^2}{(1 - w_{\min})^2} \left[(1 - d) + d \ln(d)\right]^2 + 1. \quad (15)$$

The first term is asymptotic bias due to the break, and the second term is the error variance of the forecast period.

Comparing the two scaled MSFEs (10) and (15) we have

$$\begin{aligned} &\mathrm{MSFE}(w; \lambda, d) - \mathrm{MSFE}(m, w_{\min}; \lambda, d) \quad (16) \\ &\approx \lambda^2 \left(\frac{w_a - d}{w_a}\right)^2 \mathrm{I}(w_a - d) - \frac{\lambda^2}{(1 - w_{\min})^2} \left[(1 - d) + d \ln(d)\right]^2 \end{aligned}$$

The difference depends on the length of the single window forecast, w_a, and the minimum window (fraction), $w_{\min}$, which are chosen by the forecaster, and the properties of the DGP, which are the size and the distance to the break, λ and d.

In the absence of any reliable knowledge of the break it would be of interest to compare the AveW forecast with the one based on the full estimation window, namely when w is set to unity. For this comparison it is readily seen that the AveW forecast is the one with the lower MSFE, since for large m

$$(1 - d) - \frac{[(1 - d) + d \ln(d)]}{1 - w_{\min}} \geq 0$$

implies that

$$w_{\min} \leq \frac{-d \ln(d)}{1 - d},$$

and since $w_{\min} \leq d$ this condition can be rewritten as

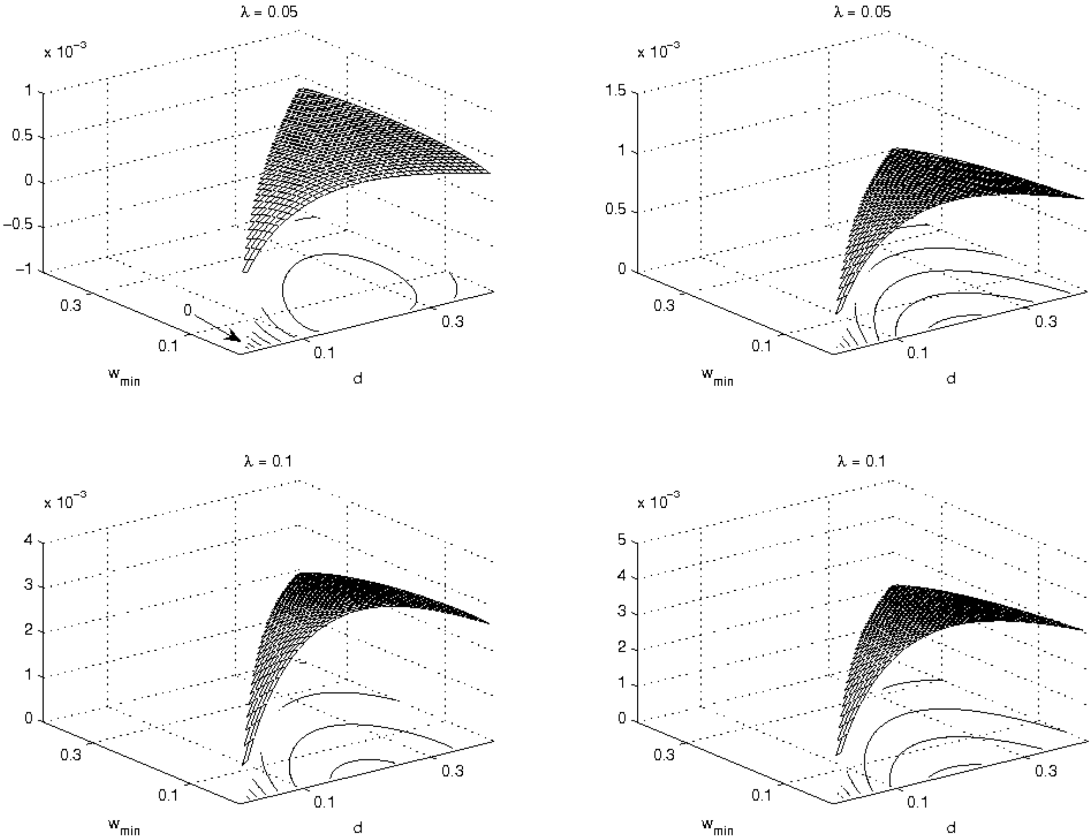
$$1 \leq d - \ln(d),$$

which is true for any $d > 0$. The two forecasts have the same MSFE only if $d = 1$, namely if there has not been a break.

While the AveW forecast asymptotically always has a lower MSFE, a trade-off exists between the lower bias and the higher variance of the AveW forecast relative to the singe window forecast in finite samples. When $\lambda = 0$, that is, there is no break in the sample, using the entire sample is most efficient estimator. As λ increases the smaller bias of the AveW forecast will start to dominate the lower variance of the single window forecast. The degree of trade-off depends on the magnitudes of λ, d, T and $w_{\min}$. The figures below shed light on the extent of these trade-offs.

Figure 1 plots the exact and the asymptotic differences in MSFE of the two forecast procedures in (14) and (16) for $T = 2000$, where the triangular shape of the surface is due to the fact that $w_i \leq d$. It can be seen that

Figure 1: Exact and asymptotic difference in MSFE with $T = 2000$

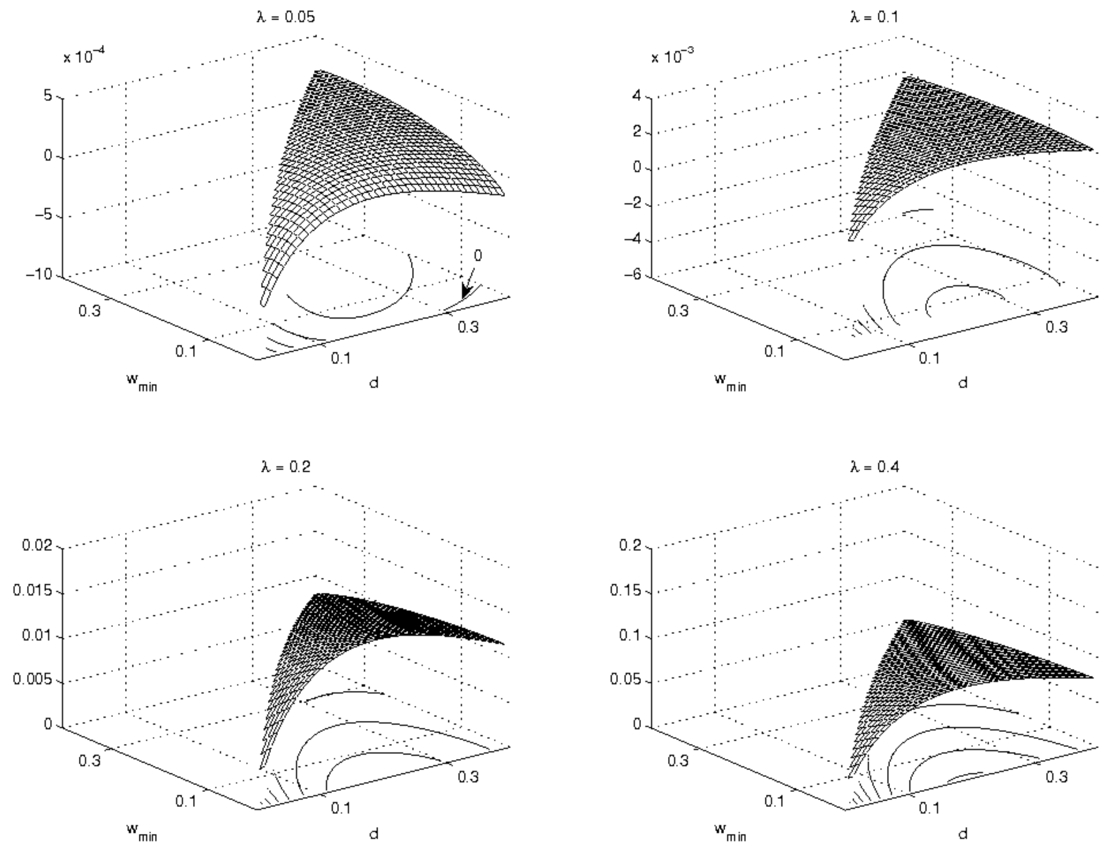


The two plots on the left show the exact difference in MSFE in (14) and the two plots on the right show the asymptotic difference in MSFE in (16). The arrows in the top plots point to the zero-isoquant; the surfaces in the plots in the second row are always positive.

the asymptotic MSFE in the right column of Figure 1 and the exact MSFE in the left column of Figure 1 are fairly similar, in particular for $\lambda = 0.1$. However, even for a data set as large as $T = 2000$ the exact difference in MSFE can be negative for very small breaks.

Figure 2 plots the differences in the exact MSFE (14) for $T = 100$. It is clear that even for this smaller sample size the difference between the RMFEs of the two procedures becomes positive even for relatively small values of λ, and the difference rises rapidly with λ.

Figure 2: Exact difference in MSFE with $T = 100$



The plots show the exact difference in MSFE in (14). The arrow in the left upper plot points to the zero-isoquant; the surfaces in the other plots are always positive.

3.2 Multiple breaks in drift

Consider a random walk model where the drift term is subject to n different breaks. Denote the break points by d_i, $i = 1, 2, \ldots, n$, such that $d_1 > d_2 > \ldots > d_n$, and let the means of the process over these segments be $\mu_1, \mu_2, \ldots, \mu_{n+1}$. Specifically,

$$y_t = \mu_t + \varepsilon_t, \text{ for } t = 1, 2, \ldots, T,$$

such that if the sample period is mapped to the unit interval the mean from $t = 1$ to $t = d_1 T$ is given by μ_1, and the mean from $t = d_1 T + 1$ to $t = d_2 T$

is μ_2, and so forth.

To simplify the analysis to begin with assume that $n = 2$, and note that the one step ahead forecast of y_{T+1} based on the window of size wT (from $t = T$) is given by

$$\begin{aligned}\hat{y}_{T+1}(w) &= [1 - \mathrm{I}(w - d_1)]\,\mu_3 + \\ &\quad \mathrm{I}(w - d_1)[1 - \mathrm{I}(w - d_2)]\left[\frac{d_2\mu_3 + (w - d_2)\mu_2}{w}\right] \\ &\quad + \mathrm{I}(w - d_2)\left[\frac{d_1\mu_3 + (d_1 - d_2)\mu_2 + (w - d_1)\mu_1}{w}\right] + \frac{1}{wT}\sum_{t=T-wT+1}^{T}\varepsilon_t.\end{aligned}$$

The one-step ahead forecast error is

$$\begin{aligned}e_{T+1}(w) &= y_{T+1} - \hat{y}_{T+1}(w) \\ &= \mu_3 + \varepsilon_{T+1} - \hat{y}_{T+1}(w),\end{aligned}$$

which after some algebra, and noting that $\mathrm{I}(w - d_1)\mathrm{I}(w - d_2) = \mathrm{I}(w - d_1)$, can be written as

$$e_{T+1}(w)/\sigma = B_{T+1}(w) + \varepsilon_{T+1}/\sigma - \frac{1}{wT}\sum_{t=T-wT+1}^{T}\varepsilon_t/\sigma,$$

where

$$B_{T+1}(w) = \lambda_1 \mathrm{I}(w - d_1)\left(\frac{w - d_1}{w}\right) + \lambda_2 \mathrm{I}(w - d_2)\left(\frac{w - d_2}{w}\right),$$

with

$$\lambda_1 = (\mu_2 - \mu_1)/\sigma,\ \lambda_2 = (\mu_3 - \mu_2)/\sigma.$$

From the above results, it is clear that for the case of n breaks we have

$$B_{T+1}(w) = \sum_{i=1}^{n}\lambda_i \mathrm{I}(w - d_i)\left(\frac{w - d_i}{w}\right),$$

where

$$\begin{aligned}\lambda_i &= (\mu_{i+1} - \mu_i)/\sigma,\ i = 1, 2, ..., n \\ n^{-1}\sum_{i=1}^{n}\lambda_i &= (\mu_{n+1} - \mu_1)/n\sigma.\end{aligned}$$

For a single window estimation with $w = 1$, the forecast bias per break will be

$$B_F(n) = B_{T+1}(1)/n = n^{-1}\sum_{i=1}^{n}\lambda_i \mathrm{I}(1 - d_i)(1 - d_i) = n^{-1}\sum_{i=1}^{n}\lambda_i (1 - d_i).$$

For AveW forecast the bias per break will be

$$B_{\text{AveW}}(n) = n^{-1} \sum_{i=1}^{n} \frac{m}{m+1} \frac{\lambda_i}{1-w_{\min}} [(1-d_i) + d_i \ln(d_i)].$$

The variance term is unaffected by the possibility of multiple breaks in the mean.

In the case where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distributed independently of the break points, $d_1, d_2, \ldots, d_n$, the bias terms can be approximated for n large as

$$\begin{aligned} \lim_{n\to\infty} B_F(n) &= B_F = \mathrm{E}(\lambda_i)(1-\mathrm{E}(d_i)) \\ \lim_{n\to\infty} B_{\text{AveW}}(n) &= B_{\text{AveW}} = \frac{m}{m+1} \frac{\mathrm{E}(\lambda_i)}{1-w_{\min}} \left\{1 - \mathrm{E}(d_i) + \mathrm{E}\left[d_i \ln(d_i)\right]\right\}, \end{aligned}$$

and

$$\begin{aligned} \lim_{n\to\infty} \left[B_F^2(n) - B_{\text{AveW}}^2(n)\right] &= \\ &= [\mathrm{E}(\lambda_i)]^2 \left\{ (1-\mathrm{E}(d_i))^2 \left[1 - \left(\frac{m}{m+1}\right)^2 \left(\frac{1}{1-w_{\min}}\right)^2 \right] \right. \\ &\quad \left. - \left(\frac{m}{m+1}\right)^2 \left(\frac{1}{1-w_{\min}}\right)^2 \mathrm{E}\left[d_i \ln(d_i)\right] \left\{2 - 2\mathrm{E}(d_i) + \mathrm{E}\left[d_i \ln(d_i)\right]\right\} \right\}. \end{aligned}$$

Since as $n \to \infty$ then $w_{\min} \to 0$, for large m we have

$$\lim_{n\to\infty} \left[B_F^2(n) - B_{\text{AveW}}^2(n)\right] = -\mathrm{E}\left[d_i \ln(d_i)\right] \left\{2 - 2\mathrm{E}(d_i) + \mathrm{E}\left[d_i \ln(d_i)\right]\right\}$$

Furthermore, as $d_i \ln(d_i) \leq 0$ for all $d_i \in (0,1)$, then $-\mathrm{E}\left[d_i \ln(d_i)\right] \geq 0$.[1] Also it is easily established that

$$f(d_i) = 2 - 2d_i + d_i \ln(d_i) > 0 \text{ for all } d_i \in (0,1),$$

and hence for all distributions of break points over the unit interval it must be that

$$2 - 2\mathrm{E}(d_i) + \mathrm{E}\left[d_i \ln(d_i)\right] > 0.$$

Hence,

$$\lim_{n\to\infty} \left[B_F^2(n) - B_{\text{AveW}}^2(n)\right] \geq 0.$$

The strict equality holds only if $\mathrm{E}(\lambda_i) = 0$.

The magnitude of $\lim_{n\to\infty} \left[B_F^2(n) - B_{\text{AveW}}^2(n)\right]$ depends on the distribution of the break points d_i. For example, if we assume that d_i is distributed uniformly over $d_i \in (0,1)$, then $\mathrm{E}(d_i) = 1/2$,

$$\mathrm{E}\left[d_i \ln(d_i)\right] = \int_0^1 x \ln(x) dx = \left[-\frac{1}{4}x^2 + \frac{1}{2}x^2 \ln(x)\right]_0^1 = -1/4,$$

[1] $d_i = 0$ is ruled out by assumption, and $d_i = 1$ refers to the case of no breaks.

and

$$2 - 2\mathrm{E}(d_i) + \mathrm{E}\left[d_i \ln(d_i)\right] = 1 - 1/4 = 3/4 > 0.$$

Hence, we have

$$\lim_{n\to\infty}\left[B_F^2(n) - B_{\mathrm{AveW}}^2(n)\right] = \frac{3}{16}\left[\mathrm{E}(\lambda_i)\right]^2 \geq 0.$$

Strict equality holds only if $\mathrm{E}(\lambda_i) = 0$.

3.3 Breaks in drift and volatility

For simplicity assume that there is only one break point but that the volatility also changes, that is, in model (1) $\sigma_1 \neq \sigma_2$ and $\mu_1 \neq \mu_2$. We initially proceed by analysing the effect of a structural break in volatility only, and in a second step combine the result with that of the break in drift analysed above. For simplicity of exposition assume that the drift and the volatility break at the same time—the extension to different break dates is however straightforward.

Initially ignoring the effect of a break in drift, the one-step ahead forecast error for a window of size w_a is given by

$$e_{T+1}(w_a) = \varepsilon_{T+1} - \frac{1}{Tw_a}\sum_{t=T(1-w_a)+1}^{T}\varepsilon_t.$$

The scaled MSFE for the single window forecast when the variance breaks at time T_b is

$$\begin{aligned}\mathrm{MSFE}(w_a;\kappa,d) &= \mathrm{E}\big[\sigma_2^{-2}e_{T+1|T}(w_a)^2\big] \\ &= \frac{(w_a-d)}{Tw_a^2}\mathrm{I}(w_a-d)\,\kappa^2 + \frac{\min(w_a,d)}{Tw_a^2} + 1 \qquad (17)\end{aligned}$$

where $\kappa = \sigma_1/\sigma_2$.

The forecast error for the AveW forecast is

$$e_{T+1}(\mathrm{AveW}) = \varepsilon_{T+1} - \frac{1}{m+1}\sum_{i=0}^{m}\left(\frac{1}{Tw_i}\sum_{t=Tw_{\min}-i}^{T}\varepsilon_t\right),$$

and the scaled MSFE of the AveW forecast is

$$\begin{aligned}\mathrm{MSFE}(m,w_{\min};\kappa,d) &= \mathrm{E}\big(\sigma_2^{-2}[e_{T+1}(\mathrm{AveW})]^2\big) \\ &= \frac{1}{(m+1)^2}\left(\kappa^2\sum_{i=0}^{m}\frac{w_i-d}{Tw_i^2}\mathrm{I}(w_i-d) + \sum_{i=0}^{m}\frac{\min(w_i,d)}{Tw_i^2}\right. \\ &\quad + 2\kappa^2\sum_{i=0}^{m-1}\frac{w_i-d}{w_i}\mathrm{I}(w_i-d)\sum_{j=i+1}^{m}\frac{1}{Tw_j} \\ &\quad \left. + 2\sum_{i=0}^{m-1}\frac{\min(w_i,d)}{w_i}\sum_{j=i+1}^{m}\frac{1}{Tw_j}\right) + 1. \qquad (18)\end{aligned}$$

The derivation in Appendix A show that the asymptotic MSFE for the AveW forecast in (18) is zero, and the same is true for the single window MSFE as can readily be seen in (17).

Combining these results with those of the break in drift yields the scaled MSFE for the single window forecast

$$\begin{aligned} \mathrm{E}\big(\sigma_2^{-2} e_{T+1}(w_a)^2\big) &= \left(\frac{w_a - d}{w_a}\right)^2 \lambda^2 \mathrm{I}(w_a - d) \\ &\quad + \frac{w_a - d}{T w_a^2} \mathrm{I}(w_a - d)\, \kappa^2 + \frac{\min(w_a, d)}{T w_a^2} + 1, \end{aligned} \tag{19}$$

where $\lambda = |\mu_2 - \mu_1| / \sigma_2$. For the AveW forecasts over $m+1$ windows, the scaled MSFE is

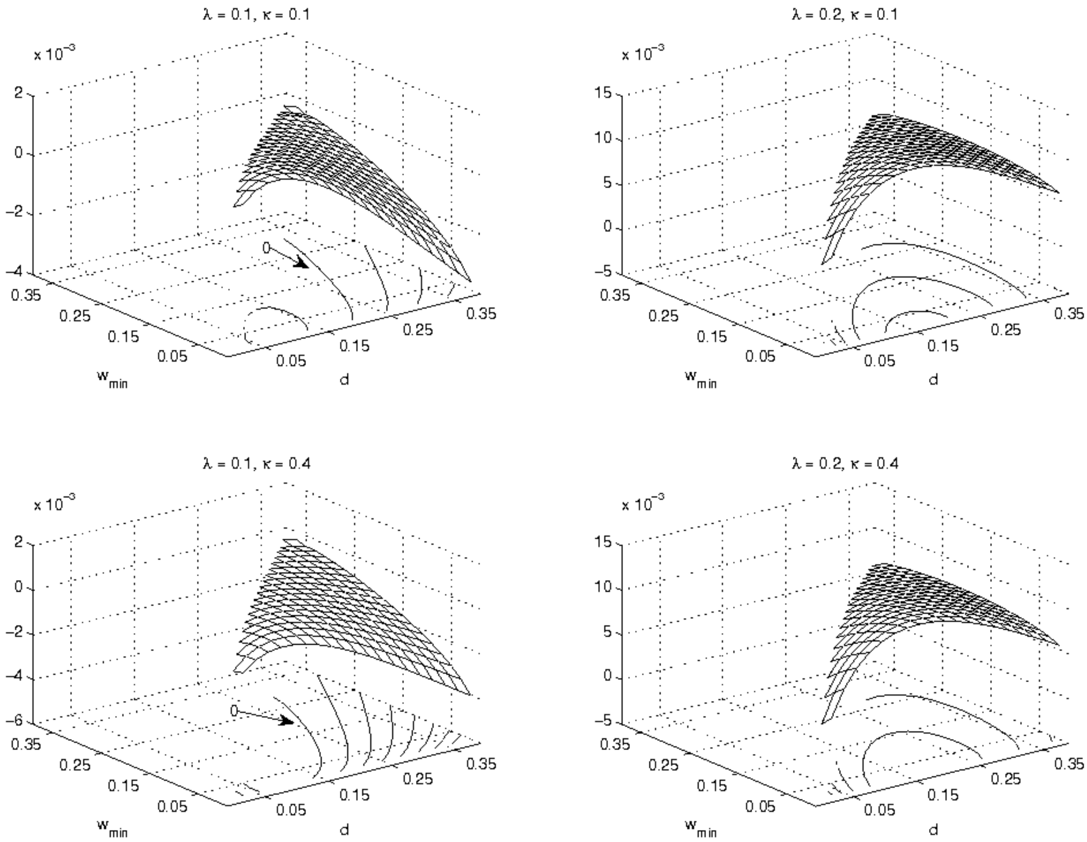
$$\begin{aligned} \mathrm{E}\big(\sigma_2^{-2} e_{T+1}(\mathrm{AveW})^2\big) &= \frac{1}{(m+1)^2} \left\{ \left[\sum_{i=0}^{m} \left(\frac{w_i - d}{w_i}\right) \lambda \mathrm{I}(w_i - d) \right]^2 \right. \\ &\quad + \kappa^2 \sum_{i=0}^{m} \frac{w_i - d}{T w_i^2} \mathrm{I}(w_i - d) + \sum_{i=0}^{m} \frac{\min(w_i, d)}{T w_i^2} \\ &\quad + 2\kappa^2 \sum_{i=0}^{m-1} \frac{w_i - d}{w_i} \mathrm{I}(w_i - d) \sum_{j=i+1}^{m} \frac{1}{T w_j} \\ &\quad \left. + 2 \sum_{i=0}^{m-1} \frac{\min(w_i, d)}{w_i} \sum_{j=i+1}^{m} \frac{1}{T w_j} \right\} + 1. \end{aligned} \tag{20}$$

Figure 3 plots the exact differences between (19) and (20) in scaled MSFEs of the forecast procedures. When comparing the plots to those for the break in drift only in Figure 2, it becomes obvious that the break in volatility tilts the surface downwards as d is increased and $w_{\min}$ remains small. However, when the break in drift increases it quickly dominates the break in volatility and the difference in scaled MSFEs become positive over the whole range of d and $w_{\min}$.

4 Recursive forecasts for time-varying parameter models

As an alternative to averaging forecasts over estimation windows we consider time varying parameter models. A number of time-varying parameter models have been considered in the forecasting literature in which the unknown parameters are assumed to follow random walks, see, for example, Harvey (1989). Recently, Branch and Evans (2006) consider a number of variations on this class of models and show that a particularly simple form, known as

Figure 3: Exact difference in MSFE for a break in drift and volatility with $T = 100$



The plots show the difference of the MSFE in (19) and that in (20).
The arrows point to the zero-isoquants.

the 'constant gain least squares', works reasonably well in forecasting US inflation and GDP growth.

The time varying parameter regression model is defined by

$$\begin{aligned} y_t &= \beta_t' \mathbf{x}_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.}\left(0, \sigma_t^2\right), \\ \beta_t &= \beta_{t-1} + \mathbf{v}_t, \end{aligned}$$

where it is assumed that ε_t and $\mathbf{v}_t$ are mutually and serially independent with zero means and variances, σ_t^2 and Ω_t, respectively. For given values of these variances the optimal one-step ahead forecast of y_{T+1}, formed at time

T using Kalman Filters is given by

$$\hat{y}_{T+1}(KF) = \hat{\beta}'_T \mathbf{x}_T,$$

where

$$\hat{\beta}_T = \hat{\beta}_{T-1} + \mathbf{G}_T(y_T - \hat{\beta}_{T-1}\mathbf{x}_{T-1}),$$
$$\mathbf{G}_T = (\sigma_T^2 + \mathbf{x}'_{T-1}\mathbf{P}_T\mathbf{x}_{T-1})^{-1}\mathbf{P}_T\mathbf{x}_{T-1},$$

and

$$\mathbf{P}_T = \mathbf{P}_{T-1} - (\sigma_T^2 + \mathbf{x}'_{T-1}\mathbf{P}_{T-1}\mathbf{x}_{T-1})^{-1}(\mathbf{P}_{T-1}\mathbf{x}_{T-1}\mathbf{x}'_{T-1}\mathbf{P}_{T-1}) + \Omega_T.$$

Many different estimators proposed in the literature are special cases of the above recursive expressions for different choices of σ_T^2 and Ω_T, and the initialization of $\mathbf{P}_t$, $t = 1, 2, ..., T$.

In what follows we focus on a very simple application where $\mathbf{x}_t = 1$, and only consider the constant gain least squares, which is equivalent to discounting past observations at a geometric rate, γ, see Branch and Evans (2006, p.160). We denote this forecast by

$$\hat{y}_{T+1}(\text{ExpW}, \gamma) = \hat{y}_{T+1}(\gamma) = \left(\frac{1-\gamma}{1-\gamma^T}\right)\sum_{j=1}^{T}\gamma^{T-j}y_j.$$

It is clear that for $\gamma = 1$, $\hat{y}_{T+1}(1) = T^{-1}\sum_{j=1}^{T} y_j = \bar{y}_T$.

Consider now the case where the mean of y_t is subject to a single break in mean at date $1 < T_b < T$, with $\mu_1 \neq \mu_2$ but $\sigma_1 = \sigma_2 = \sigma$. The error of the one-step ahead forecast in this case is given by

$$\begin{aligned} e_{T+1}(\gamma) &= y_{T+1} - \left(\frac{1-\gamma}{1-\gamma^T}\right)\sum_{j=1}^{T}\gamma^{T-j}y_j \\ &= \varepsilon_{T+1} - \left(\frac{1-\gamma}{1-\gamma^T}\right)\sum_{j=1}^{T}\gamma^{T-j}\varepsilon_j + \mu_2 \\ &\quad - \left(\frac{1-\gamma}{1-\gamma^T}\right)\sum_{j=1}^{T_b-1}\gamma^{T-j}\mu_1 - \left(\frac{1-\gamma}{1-\gamma^T}\right)\sum_{j=T_b}^{T}\gamma^{T-j}\mu_2. \end{aligned}$$

But

$$\begin{aligned} \sum_{j=1}^{T_b-1}\gamma^{T-j}\mu_1 &= \mu_1\left(\frac{\gamma^{T-T_b+1}-\gamma^T}{1-\gamma}\right) \\ \sum_{j=T_b}^{T}\gamma^{T-j}\mu_2 &= \mu_2\left(\frac{1-\gamma^{T-T_b+1}}{1-\gamma}\right), \end{aligned}$$

and hence

$$\text{Bias}\left[\hat{y}_{T+1}(\text{ExpW}, \gamma)\right] = (\mu_2 - \mu_1)\left(\frac{\gamma^{T-T_b+1} - \gamma^T}{1-\gamma^T}\right).$$

Since, $0 < \gamma < 1$, the sign of the forecast bias is the same as the sign of $(\mu_2 - \mu_1)$. The forecast error variance is given by

$$\text{Var}\left[e_{T+1}(\gamma)\right] = \sigma^2\left[1 + \left(\frac{1-\gamma}{1-\gamma^T}\right)^2\left(\frac{1-\gamma^{2T}}{1-\gamma^2}\right)\right].$$

It is interesting to note that for all values of $0 < \gamma < 1$ the sampling variance of the forecast - the second part in the [ ], does not vanish even for T sufficiently large. Therefore, the exponential decay-weighting of the past observations would work only through bias reduction. As before, let $d = (T - T_b)/T$ denote the distance to the beak, and note that the scaled one-step ahead MSFE in this case is given by

$$\begin{aligned}\text{MSFE}\left[\hat{y}_{T+1}(\text{ExpW}, \gamma)\right] &= f(\gamma) \qquad (21)\\ &= 1 + \lambda^2\left(\frac{\gamma^{1+T\ d} - \gamma^T}{1-\gamma^T}\right)^2 \\ &\quad + \left(\frac{1-\gamma}{1-\gamma^T}\right)^2\left(\frac{1-\gamma^{2T}}{1-\gamma^2}\right),\end{aligned}$$

and as before, $\lambda = |\mu_2 - \mu_1|/\sigma$.

Figure 4 compares the MSFE of the single window forecast with $w = 1$ to that of the ExpW forecast given in (21) for different values of γ. It can be seen that for small values of λ the ExpW forecast has a higher MSFE but that as the size of the break increases the MSFE of single $w = 1$ window forecast increases above that of the ExpW forecast. The ExpW procedure begins to dominate the single window forecasts when λ is increased to 0.4 for all values of d and γ.

For large T and small d, $f(\gamma)$ can be approximated by

$$f(\gamma) = 1 + \lambda^2\gamma^{2+2T\ d} + \frac{1-\gamma}{1+\gamma} + O(\gamma^T).$$

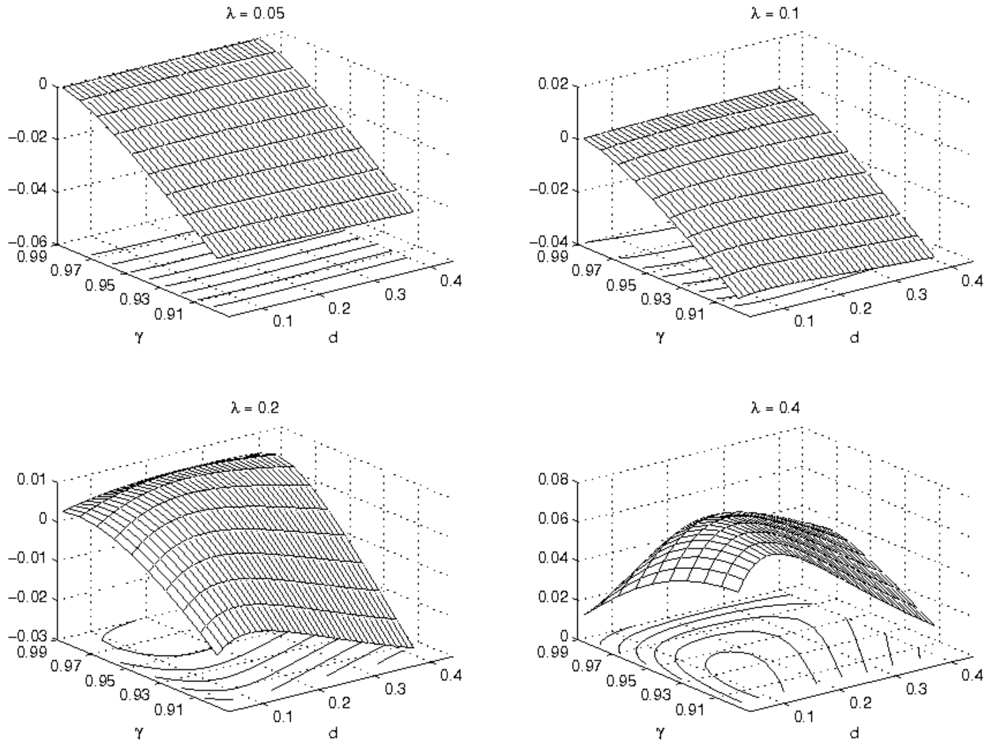
It is easily seen that

$$\frac{1}{2}f'(\gamma) = \lambda^2(1+Td)\gamma^{1+2Td} - \frac{1}{(1+\gamma)^2} + O(\gamma^T),$$

and

$$\frac{1}{2}f''(\gamma) = \lambda^2(1+Td)\left(1+2Td\right)\gamma^{2Td} + \frac{2}{(1+\gamma)^3} + O(\gamma^T) > 0$$

Figure 4: Exact difference of MSFEs of single window and ExpW for $T = 100$



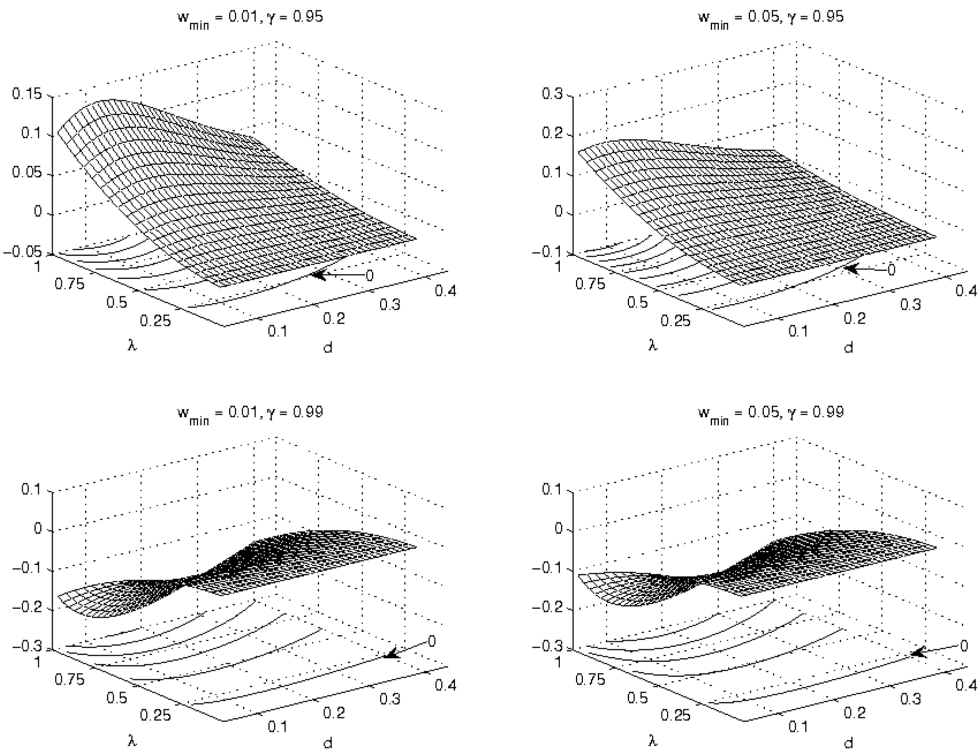
The graphs plot the exact difference between the single window and ExpW forecasts, that is, [MSFE(Single window) − MSFE(ExpW($\gamma; \lambda, d$))]/σ^2.

for all $0 < \gamma < 1$. Hence, $f(\gamma) = 0$ has a unique solution in terms of d and λ for a sufficiently large T.

Figures 5–7 compare the AveW forecast with the ExpW forecasts for different values of T, d, and λ, and for different choices of γ. Figure 5 plots the difference in MSFE between the AveW and the ExpW forecasts for $T = 100$ for different values of λ and d. The difference across values of λ dominates that of different values of d and depends crucially on the choice of γ. While the ExpW forecasts have a smaller MSFE for $\gamma = 0.95$ except for small λ, this is reversed for $\gamma = 0.99$, where the AveW forecasts have a smaller MSFE for most values of λ.

In the case of $T = 1000$, which is plotted in Figure 6, the choice of γ is less important. The ExpW forecasts have a smaller MSFE except for

Figure 5: Exact difference of MSFEs of AveW and ExpW for $T = 100$



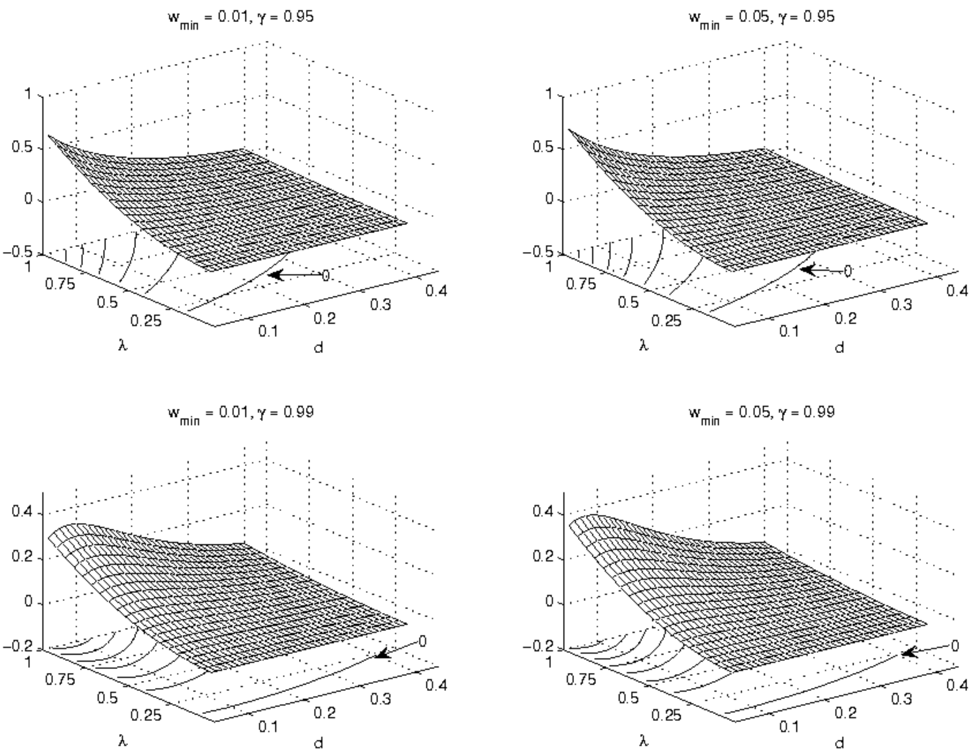
The graphs plot the exact difference between the AveW and ExpW forecasts, that is, $[\text{MSFE}(\text{AveW}(w_{\min}; \lambda, d)) - \text{MSFE}(\text{ExpW}(\gamma; \lambda, d))]/\sigma^2$. The arrows point to the zero-isoquant.

relatively small values of λ and large values of d.

Figure 7 plots the difference in MSFE between the AveW and ExpW forecasts for fixed break points $D = dT$ and fixed minimum windows $Tw_{\min}$. The region where ExpW has a smaller MSFE depends on T, the size of the break, λ, and the decay parameter γ. While for $T = 100$ and large values of γ the difference becomes increasingly negative with λ, the difference grows in λ for values of 0.96 or less. For $T = 1000$ the difference is negative only for small values of λ.

In order to gain additional insight into the differences between the AveW and ExpW procedures, we plot the weights attached to the observations in a sample of $T = 100$ observations in Figure 8. It can be seen that AveW gives equal weights to the observations in the minimum window whereas the weights of these observations decline in the ExpW forecasts. Another

Figure 6: Exact difference of MSFEs of AveW and ExpW for $T = 1000$



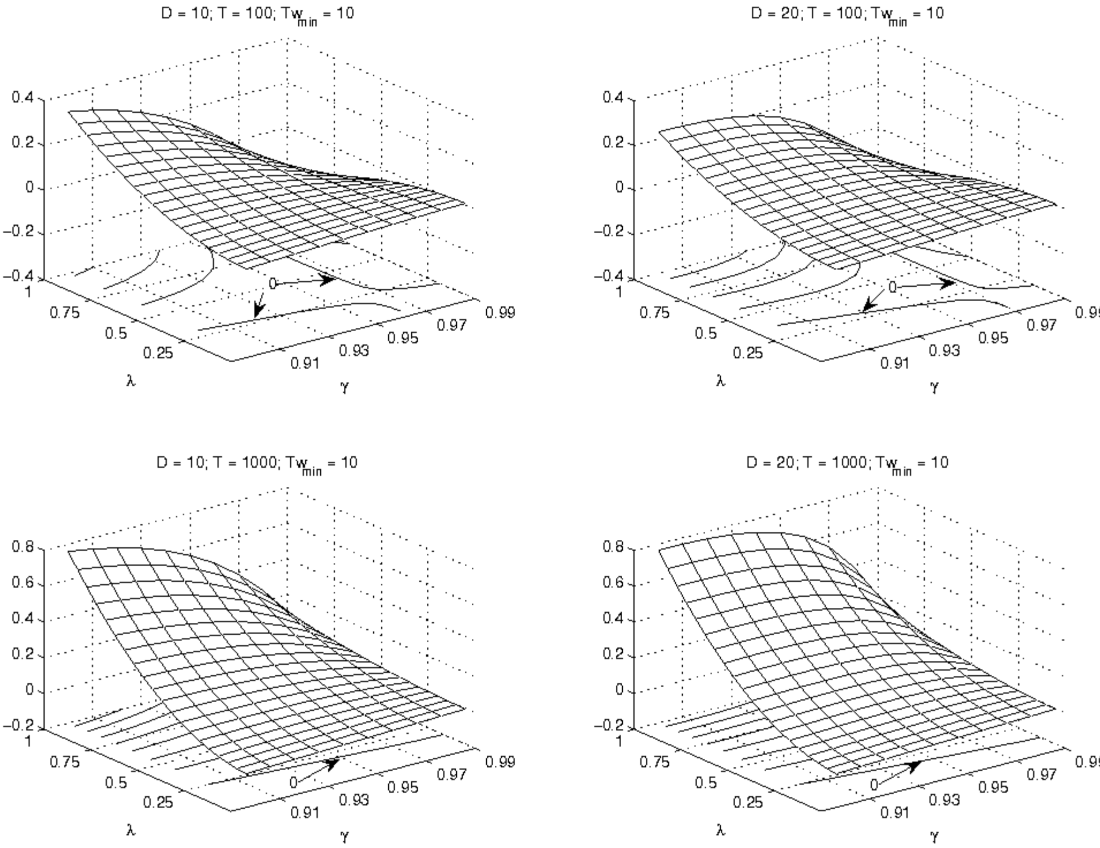
See footnote to Figure 5.

interesting observation is that the AveW weights do not differ as much as those of ExpW between the different weighting schemes. This suggests that ExpW forecasts will depend considerably more on the choice of γ than AveW forecasts depend on the choice of $w_{\min}$.

5 Applications to financial time series

In this section we will apply the AveW and the ExpW procedures to weekly returns on futures contracts in the case of twenty stock market indices. Details of the price indices and the periods over which they are observed are given in Appendix B. Note that S&P and FTSE futures go back to 1985, whilst the start dates for other futures markets are much more recent. Our sample ends on November 24, 2008 and thus covers the recent highly volatile

Figure 7: Difference in MSFEs between AveW and ExpW forecasts with fixed break point

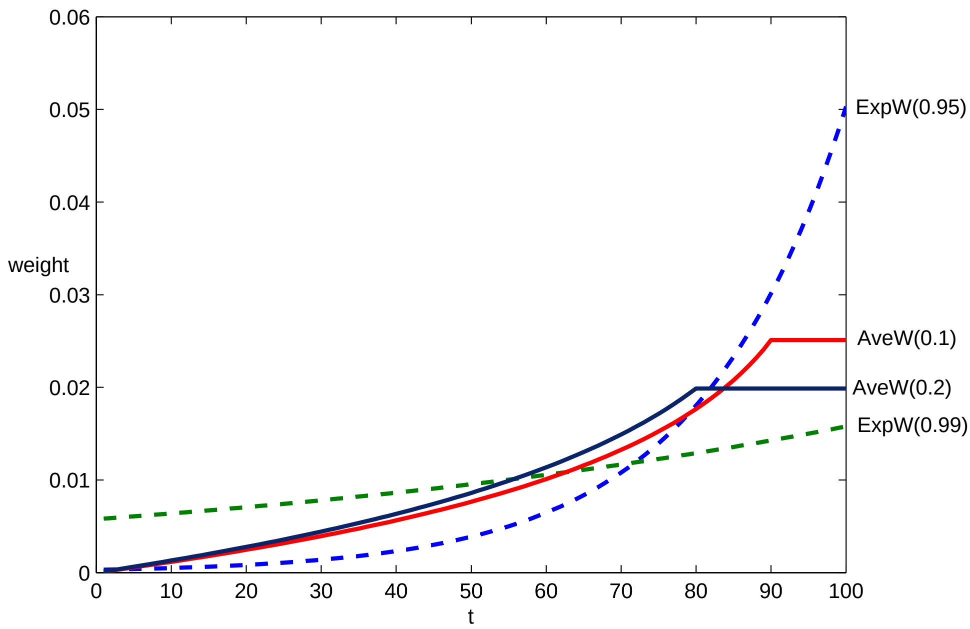


See footnote to Figure 5. Here, however, the break point $D = dT$ and the minimum window $Tw_{\min}$ are fixed and not fractions of T.

episodes associated with the credit crunch.

We recursively compute one-week ahead forecasts using rolling windows. The baseline fixed window forecasts are obtained using 156 and 260 weeks rolling regressions. We compare these forecasts with AveW rolling forecasts based on the same samples. We compute AveW forecasts for two choices of the minimum window, $w_{\min} = 16$ weeks and 32 weeks for the 156 week rolling window and $w_{\min} = 26$ weeks and 52 weeks for the 260 weeks rolling window, which correspond to about 10% and 20% percent of the observations. For example, in the case of the sample with 156 weeks the AveW

Figure 8: Weights attached to the observations in the AveW and ExpW forecasts for $T = 100$



Plotted are the weights attached to each observation in a sample of $T = 100$ observations. The number in brackets are the minimum window, $w_{\min}$, in the case of the AveW weights and the down-weighting parameter, γ, in the case of the ExpW weights.

forecast is computed as the simple average of 141 forecasts computed based on past $156, 155, \ldots, 16$ weeks. Finally, we computed ExpW forecasts using the decay parameters $\gamma = 0.95$ and 0.98.

We report the bias and the root mean square forecast error (RMSFE) and tests for predictive performance proposed by Diebold and Mariano (1995) (DM). More precisely,

$$\text{RMSFE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2},$$

where $e_t = y_{t+1} - \hat{y}_{t+1|t}$, the one-week ahead forecast, $\hat{y}_{t+1|t}$, is based on the observations up to t, and n is the number of forecasts under consideration. The DM test statistic for predictive ability are calculated for the loss differential

$$l_t(A, B) = e_{tA}^2 - e_{tB}^2,$$

where e_{tA} and e_{tB} are the forecast error for forecast methods A and B.

The bias and RMSFE for each time series is reported in Tables 1 and 2 and the DM statistics in Tables 3 to 6. When considering the 156 week rolling window, the AveW forecasts have a lower RMSFE in 18 out of the 20 series for the shorter minimum window and in 19 out of 20 series for the longer minimum window. While the difference in RMSFEs is relatively small—the average ratio of the RMSFE of the AveW forecasts to that of the single window is 0.9965 for both minimum windows. It is clear that in terms of RMSFE the AveW forecasts systematically outperform forecasts based on a single window, although the outperformance of the AveW is statistically significant only in one case. But it is interesting to note, that in no case is the AveW forecast significantly worse.

The improvement of the ExpW forecasts over the single window forecasts crucially depends on the down-weighting parameter. For $\gamma = 0.95$ the RMSFE is lower than that of the single window in only 7 out of the 20 series, and in one case the Diebold-Mariano statistic suggests that the ExpW(0.95) forecast is significantly worse than the single window forecast. Using $\gamma = 0.98$ changes the results and the ExpW(0.98) forecast has a lower RMSFE in 17 out of 20 cases, although none of the improvements are statistically significant as the forecast RMSFEs are again similar to that of the single window with average ratios of 1.007 ($\gamma = 0.95$) and 0.9966 ($\gamma = 0.98$).

When using a longer rolling window of 260 weeks, the AveW forecasts have a lower RMSFE than the single window in 19 out of the 20 series with an average ratio of the AveW RMSFE over the single window RMSFE of 0.9952 and 0.9960 for the shorter and the longer minimum window. Again, in no individual series is the improvement statistically significant.

For the ExpW forecasts the improvement again depends strongly on the down-weighting parameter, with a lower RMSFE in 9 out of 20 series when γ is set to 0.95 against 17 out of 20 when γ is set to 0.98. The average ratio of RMSFEs is 0.9984 and 0.9940 for $\gamma = 0.95$ and 0.98.

6 Conclusion

In this paper we have shown that AveW and ExpW forecasts always have a lower bias than full sample forecasts. The forecast variance of the AveW and ExpW forecasts depends, however, on the size and time of the break in the sample. For all but the smallest breaks, however, also the MSFE of the AveW and ExpW forecasts are smaller than those of the single window forecasts.

A comparison of the AveW and ExpW forecasts suggest that their relative performances depend on the size and timing of the break as well as the size of the sample. It emerges that when the break is relatively small—roughly less than a quarter of the variance of the disturbance term—the

AveW forecast has a lower MSFE. Otherwise ExpW will dominate if the sample size is small and the downweighting parameter, γ, is set below approximately 0.96, or when the sample size is large.

Extensions of the results in the paper to more general set ups is possible but analytical derivations might not be easy to achieve. This is particularly the case if we consider dynamic models with breaks. However, Monte Carlo simulations for AveW forecasts for AR(1) models, not reported here but available from the authors, suggest that the main findings of this paper are likely to hold more generally.

Table 1: Forecasting performance for stock market indices, $w = 156$ weeks rolling window

Name		SW	AveW (16 weeks)	AveW (32 weeks)	ExpW (0.95)	ExpW (0.98)
AEX	Bias	0.0130	0.0121	0.0128	0.0086	0.0118
	RMSFE	0.6131	0.6123	0.6119	0.6164	0.6125
ASX	Bias	0.0211	0.0248	0.0251	0.0229	0.0242
	RMSFE	0.5116	0.5086	0.5088	0.5102	0.5088
BEL	Bias	0.0286	0.0244	0.0258	0.0167	0.0236
	RMSFE	0.6207	0.6182	0.6181	0.6216	0.6185
CAC	Bias	0.0100	0.0086	0.0091	0.0062	0.0084
	RMSFE	0.5994	0.5983	0.5979	0.6032	0.5987
DAX	Bias	0.0144	0.0141	0.0148	0.0102	0.0136
	RMSFE	0.6791	0.6787	0.6784	0.6836	0.6790
DJE	Bias	0.0202	0.0132	0.0140	0.0078	0.0129
	RMSFE	0.6044	0.6014	0.6011	0.6052	0.6016
FOX	Bias	0.0181	0.0293	0.0297	0.0260	0.0279
	RMSFE	0.5909	0.5831	0.5841	0.5821	0.5830
FTSE	Bias	0.0073	0.0049	0.0052	0.0029	0.0048
	RMSFE	0.4711	0.4719	0.4714	0.4766	0.4723
IBE	Bias	0.0107	0.0096	0.0100	0.0074	0.0094
	RMSFE	0.6133	0.6126	0.6121	0.6169	0.6127
KFX	Bias	0.0521	0.0548	0.0573	0.0409	0.0527
	RMSFE	0.6694	0.6659	0.6660	0.6688	0.6663
MIB	Bias	0.0361	0.0297	0.0315	0.0197	0.0288
	RMSFE	0.6238	0.6221	0.6219	0.6252	0.6221
ND	Bias	0.0484	0.0377	0.0399	0.0254	0.0369
	RMSFE	0.8994	0.8962	0.8966	0.8992	0.8959
NK	Bias	0.0188	0.0142	0.0148	0.0098	0.0139
	RMSFE	0.6549	0.6540	0.6539	0.6579	0.6542
OBX	Bias	0.0331	0.0347	0.0348	0.0318	0.0340
	RMSFE	0.7572	0.7498	0.7508	0.7476	0.7492
OMX	Bias	0.0133	0.0131	0.0136	0.0101	0.0127
	RMSFE	0.7185	0.7167	0.7167	0.7208	0.7169
PSI	Bias	0.0299	0.0196	0.0210	0.0114	0.0193
	RMSFE	0.5309	0.5262	0.5270	0.5257	0.5259
SMI	Bias	0.0183	0.0144	0.0152	0.0094	0.0140
	RMSFE	0.5528	0.5524	0.5521	0.5566	0.5527
SP	Bias	0.0118	0.0090	0.0094	0.0065	0.0089
	RMSFE	0.4562	0.4564	0.4560	0.4600	0.4566
TPX	Bias	0.0044	0.0058	0.0058	0.0053	0.0056
	RMSFE	0.6755	0.6752	0.6751	0.6808	0.6757
TSX	Bias	0.0013	0.0095	0.0089	0.0132	0.0095
	RMSFE	0.4993	0.4965	0.4971	0.4945	0.4960

The column with heading SW gives the results for the single window of length w specified above, the columns with headings AveW(16 weeks) and AveW(32 weeks) those for the AveW forecasts with minimum window size of 16 and 32 weeks, the columns with headings ExpW(0.95) and ExpW(0.98) give those for the ExpW forecasts with downweighting parameters 0.95 and 0.98. The details of the series including the forecast periods are given in Appendix B.

Table 2: Forecasting performance for stock market indices, $w = 260$ weeks rolling window

Name		SW	AveW (26 weeks)	AveW (52 weeks)	ExpW (0.95)	ExpW (0.98)
AEX	Bias	0.0191	0.0201	0.0207	0.0136	0.0184
	RMSFE	0.6421	0.6401	0.6403	0.6447	0.6404
ASX	Bias	0.0505	0.0610	0.0628	0.0439	0.0574
	RMSFE	0.6159	0.6128	0.6136	0.6128	0.6114
BEL	Bias	0.0442	0.0381	0.0402	0.0201	0.0316
	RMSFE	0.6319	0.6279	0.6286	0.6295	0.6268
CAC	Bias	0.0082	0.0101	0.0104	0.0067	0.0095
	RMSFE	0.6096	0.6075	0.6078	0.6123	0.6078
DAX	Bias	0.0113	0.0139	0.0141	0.0105	0.0134
	RMSFE	0.7091	0.7073	0.7075	0.7123	0.7077
DJE	Bias	0.0017	0.0158	0.0143	0.0228	0.0213
	RMSFE	0.5140	0.5104	0.5113	0.5115	0.5092
FOX	Bias	0.0676	0.0813	0.0844	0.0507	0.0729
	RMSFE	0.6526	0.6444	0.6458	0.6391	0.6407
FTSE	Bias	0.0087	0.0081	0.0085	0.0048	0.0072
	RMSFE	0.4779	0.4773	0.4772	0.4825	0.4782
IBE	Bias	0.0267	0.0265	0.0277	0.0149	0.0233
	RMSFE	0.6401	0.6384	0.6385	0.6433	0.6387
KFX	Bias	0.1158	0.1158	0.1207	0.0706	0.1009
	RMSFE	0.8431	0.8384	0.8392	0.8385	0.8367
MIB	Bias	0.0511	0.0411	0.0431	0.0233	0.0341
	RMSFE	0.5872	0.5829	0.5837	0.5841	0.5816
ND	Bias	0.0374	0.0181	0.0191	0.0114	0.0128
	RMSFE	0.7398	0.7373	0.7380	0.7402	0.7371
NK	Bias	0.0023	0.0046	0.0044	0.0057	0.0056
	RMSFE	0.6467	0.6464	0.6464	0.6509	0.6468
OBX	Bias	0.0418	0.0681	0.0686	0.0572	0.0703
	RMSFE	0.8124	0.8048	0.8058	0.8013	0.8016
OMX	Bias	0.0177	0.0170	0.0179	0.0097	0.0145
	RMSFE	0.7458	0.7433	0.7437	0.7475	0.7433
PSI	Bias	0.0065	0.0049	0.0049	0.0045	0.0053
	RMSFE	0.5161	0.5107	0.5118	0.5089	0.5088
SMI	Bias	0.0178	0.0164	0.0170	0.0107	0.0146
	RMSFE	0.5752	0.5738	0.5739	0.5787	0.5743
SP	Bias	0.0135	0.0112	0.0117	0.0066	0.0096
	RMSFE	0.4637	0.4628	0.4628	0.4665	0.4632
TPX	Bias	0.0008	0.0084	0.0085	0.0065	0.0091
	RMSFE	0.6585	0.6585	0.6585	0.6647	0.6595
TSX	Bias	0.0361	0.0528	0.0531	0.0476	0.0536
	RMSFE	0.5430	0.5384	0.5391	0.5307	0.5351

See Table 1 for details.

Table 3: Tests of forecasting performance for stock market indices, $w = 156$ weeks rolling window

Name		AveW (16 weeks)	AveW (32 weeks)	ExpW (0.95)	ExpW (0.98)
AEX	SW	0.3950	0.7312	−0.5933	0.2481
	AveW(0.1)		0.6673	−1.1622	−0.6185
	AveW(0.2)			−1.1054	−0.6958
	ExpW(0.95)				1.2180
ASX	SW	0.9285	1.1187	0.1809	0.7633
	AveW(0.1)		−0.2035	−0.3155	−0.3942
	AveW(0.2)			−0.2485	−0.0437
	ExpW(0.95)				0.3049
BEL	SW	0.9532	1.2483	−0.1315	0.7415
	AveW(0.1)		0.0950	−0.8022	−0.6457
	AveW(0.2)			−0.7131	−0.3360
	ExpW(0.95)				0.8161
CAC	SW	0.5412	0.9134	−0.8217	0.3237
	AveW(0.1)		0.9911	−1.7585	−1.2633
	AveW(0.2)			−1.6689	−1.2043
	ExpW(0.95)				1.7899
DAX	SW	0.1623	0.3441	−0.7404	0.0398
	AveW(0.1)		0.4429	−1.2659	−0.6622
	AveW(0.2)			−1.1640	−0.5760
	ExpW(0.95)				1.3256
DJE	SW	0.7985	1.0380	−0.1023	0.6678
	AveW(0.1)		0.2839	−0.7910	−0.5426
	AveW(0.2)			−0.7282	−0.4179
	ExpW(0.95)				0.8085
FOX	SW	1.8463	1.9936	0.9557	1.7304
	AveW(0.1)		−1.0683	0.1854	0.2053
	AveW(0.2)			0.3191	0.8048
	ExpW(0.95)				−0.1818
FTSE	SW	−0.8485	−0.4669	−2.2743	−1.1564
	AveW(0.1)		1.7345	−2.9807	−2.3531
	AveW(0.2)			−2.8343	−2.1484
	ExpW(0.95)				3.0210
IBE	SW	0.3426	0.7059	−0.6805	0.2476
	AveW(0.1)		0.8736	−1.2781	−0.3861
	AveW(0.2)			−1.2307	−0.7152
	ExpW(0.95)				1.3763

See the footnote of Table 1 for details on the forecast methods and time series. This table reports the test statistics for predictive ability of Diebold and Mariano (1995) against the single window forecast, where a positive value indicates that the method given in the top row has better predictive ability.

Table 4: Tests of forecasting performance for stock market indices, $w = 156$ weeks rolling window

Name		AveW (16 weeks)	AveW (32 weeks)	ExpW (0.95)	ExpW (0.98)
KFX	SW	0.6723	0.8895	0.0370	0.4894
	AveW(0.1)		−0.0653	−0.2800	−0.3346
	AveW(0.2)			−0.2387	−0.1180
	ExpW(0.95)				0.2726
MIB	SW	0.6892	0.8641	−0.2437	0.6343
	AveW(0.1)		0.2165	−0.8844	−0.0734
	AveW(0.2)			−0.7960	−0.1708
	ExpW(0.95)				0.9770
ND	SW	0.5913	0.6341	0.0140	0.5934
	AveW(0.1)		−0.3158	−0.4154	0.3782
	AveW(0.2)			−0.3113	0.3578
	ExpW(0.95)				0.5015
NK	SW	0.5383	0.7121	−0.8181	0.3927
	AveW(0.1)		0.2741	−1.7605	−0.8069
	AveW(0.2)			−1.5561	−0.5185
	ExpW(0.95)				1.8471
OBX	SW	1.4827	1.7202	0.7133	1.3860
	AveW(0.1)		−0.6938	0.2517	0.6199
	AveW(0.2)			0.3145	0.6809
	ExpW(0.95)				−0.2070
OMX	SW	0.7402	0.8872	−0.4703	0.6102
	AveW(0.1)		−0.0261	−1.3788	−0.7418
	AveW(0.2)			−1.1721	−0.2802
	ExpW(0.95)				1.4210
PSI	SW	1.3191	1.3483	0.6387	1.2859
	AveW(0.1)		−0.9847	0.0909	0.6084
	AveW(0.2)			0.2211	0.8860
	ExpW(0.95)				−0.0348
SMI	SW	0.2202	0.4800	−0.8663	0.0415
	AveW(0.1)		0.6243	−1.4161	−0.8187
	AveW(0.2)			−1.3231	−0.7541
	ExpW(0.95)				1.4827
SP	SW	−0.1080	0.2191	−1.3024	−0.3052
	AveW(0.1)		1.1253	−1.9757	−1.3215
	AveW(0.2)			−1.8761	−1.2806
	ExpW(0.95)				2.0379
TPX	SW	0.0909	0.2075	−0.9507	−0.1121
	AveW(0.1)		0.3764	−1.6034	−1.4221
	AveW(0.2)			−1.4583	−0.8417
	ExpW(0.95)				1.6146
TSX	SW	0.8669	0.9358	0.5712	0.8909
	AveW(0.1)		−0.6309	0.3677	0.8538
	AveW(0.2)			0.4105	0.7259
	ExpW(0.95)				−0.3150

See footnote of Table 3.

Table 5: Tests of forecasting performance for stock market indices, $w = 260$ weeks rolling window

Name		AveW (26 weeks)	AveW (52 weeks)	ExpW (0.95)	ExpW (0.98)
AEX	SW	0.9299	1.0331	−0.3705	0.4701
	AveW(0.1)		−0.3447	−0.8630	−0.1380
	AveW(0.2)			−0.7764	−0.0393
	ExpW(0.95)				1.1857
ASX	SW	1.0323	1.0374	0.2680	0.7676
	AveW(0.1)		−0.9247	−0.0017	0.4594
	AveW(0.2)			0.0781	0.5653
	ExpW(0.95)				0.2239
BEL	SW	1.5107	1.5451	0.2576	1.0644
	AveW(0.1)		−1.2314	−0.2330	0.4891
	AveW(0.2)			−0.1196	0.6457
	ExpW(0.95)				0.5707
CAC	SW	0.9975	1.0292	−0.4419	0.5307
	AveW(0.1)		−0.6614	−1.0901	−0.1662
	AveW(0.2)			−0.9542	0.0060
	ExpW(0.95)				1.5304
DAX	SW	0.7366	0.8053	−0.4078	0.3474
	AveW(0.1)		−0.3458	−0.8563	−0.1839
	AveW(0.2)			−0.7624	−0.0683
	ExpW(0.95)				1.1568
DJE	SW	1.1235	1.0537	0.2160	0.8262
	AveW(0.1)		−1.1189	−0.1182	0.4444
	AveW(0.2)			−0.0165	0.5936
	ExpW(0.95)				0.3836
FOX	SW	1.8285	1.8540	0.9584	1.5396
	AveW(0.1)		−1.4823	0.5167	1.0481
	AveW(0.2)			0.6035	1.1470
	ExpW(0.95)				−0.2397
FTSE	SW	0.6192	0.8829	−1.4901	−0.1931
	AveW(0.1)		0.5338	−2.2519	−1.2354
	AveW(0.2)			−2.1249	−1.0909
	ExpW(0.95)				2.6903
IBE	SW	0.7889	0.8964	−0.4667	0.3956
	AveW(0.1)		−0.1689	−0.9658	−0.1755
	AveW(0.2)			−0.8813	−0.1026
	ExpW(0.95)				1.3182

See footnote of Table 3.

Table 6: Tests of forecasting performance for stock market indices, $w = 260$ weeks rolling window

Name		AveW (26 weeks)	AveW (52 weeks)	ExpW (0.95)	ExpW (0.98)
KFX	SW	0.7748	0.8455	0.1767	0.5342
	AveW(0.1)		−0.5232	−0.0037	0.2885
	AveW(0.2)			0.0316	0.3338
	ExpW(0.95)				0.1313
MIB	SW	1.4544	1.3839	0.4029	1.2001
	AveW(0.1)		−1.4976	−0.2115	0.7215
	AveW(0.2)			−0.0540	0.9013
	ExpW(0.95)				0.7123
ND	SW	0.5623	0.4801	−0.0343	0.3898
	AveW(0.1)		−0.7484	−0.3701	0.0847
	AveW(0.2)			−0.2637	0.2378
	ExpW(0.95)				0.6152
NK	SW	0.2709	0.2614	−1.0255	−0.0574
	AveW(0.1)		−0.2039	−1.4708	−0.4868
	AveW(0.2)			−1.3326	−0.3289
	ExpW(0.95)				1.9018
OBX	SW	1.4890	1.6069	0.5499	1.1021
	AveW(0.1)		−0.9013	0.2241	0.6446
	AveW(0.2)			0.2701	0.6937
	ExpW(0.95)				−0.0328
OMX	SW	1.1045	1.1465	−0.2660	0.6885
	AveW(0.1)		−0.7020	−0.8986	0.0498
	AveW(0.2)			−0.7524	0.2131
	ExpW(0.95)				1.3531
PSI	SW	1.5226	1.5181	0.6423	1.2087
	AveW(0.1)		−1.2425	0.2229	0.7125
	AveW(0.2)			0.3183	0.8401
	ExpW(0.95)				0.0196
SMI	SW	0.8360	0.9315	−0.6343	0.3382
	AveW(0.1)		−0.1997	−1.1441	−0.3429
	AveW(0.2)			−1.0558	−0.2335
	ExpW(0.95)				1.4681
SP	SW	0.7681	0.8392	−0.7321	0.2547
	AveW(0.1)		−0.3114	−1.3585	−0.4707
	AveW(0.2)			−1.2385	−0.3142
	ExpW(0.95)				1.7626
TPX	SW	0.0151	0.0384	−0.9174	−0.2831
	AveW(0.1)		0.0700	−1.1978	−0.5868
	AveW(0.2)			−1.1238	−0.4904
	ExpW(0.95)				1.4680
TSX	SW	1.6726	1.8061	1.0127	1.3858
	AveW(0.1)		−1.1163	0.7973	1.0713
	AveW(0.2)			0.8198	1.0810
	ExpW(0.95)				−0.6696

See footnote of Table 3.

Appendix A: Mathematical details

This appendix gives the mathematical details for the AveW forecast with a break in volatility.

We have that

$$\frac{1}{T}\sum_{i=0}^{m}\frac{w_i-d}{w_i^2}\mathrm{I}(w_i-d) \overset{T\to\infty}{\longrightarrow} \int_d^1 \frac{x-d}{x^2}dx = -\ln(d)+d-1,$$

$$\begin{aligned}\frac{1}{T}\sum_{i=0}^{m}\frac{\min(w_i,d)}{w_i^2} &\overset{T\to\infty}{\longrightarrow} \int_{w_{\min}}^{d}\frac{1}{x}dx+\int_d^1\frac{d}{x^2}dx \\ &= \ln(d)-\ln(w_{\min})+1-d,\end{aligned}$$

$$\begin{aligned}\frac{1}{T}\sum_{i=0}^{m-1}\frac{w_i-d}{w_i}\mathrm{I}(w_i-d)\,\frac{1}{T}\sum_{j=i+1}^{m}\frac{1}{w_j} &\overset{T\to\infty}{\longrightarrow} \int_d^1\left(1-\frac{d}{x}\right)\int_x^1\frac{1}{y}dydx \\ &= \int_d^1\left(\frac{d}{x}-1\right)\ln(x)dx \\ &= 1+d\ln(d)-d-\frac{d}{2}\ln(d)^2,\end{aligned}$$

and, finally,

$$\begin{aligned}&\frac{1}{T}\sum_{i=0}^{m-1}\frac{\min(w_i,d)}{w_i}\frac{1}{T}\sum_{j=i+1}^{m}\frac{1}{w_j} \\ &\overset{T\to\infty}{\longrightarrow} \int_{w_{\min}}^{d}\int_x^1\frac{1}{y}dydx+\int_d^1\frac{d}{x}\int_x^1\frac{1}{y}dydx \\ &= -\int_{w_{\min}}^{d}\ln(x)dx-\int_d^1\frac{d}{x}\ln(x)dx \\ &= w_{\min}\ln(w_{\min})-w_{\min}-d\ln(d)+d+\frac{d}{2}\ln(d)^2.\end{aligned}$$

This results in the MSFE

$$\begin{aligned}\mathrm{MSFE}(m,w_{\min};\kappa,d) \overset{T\to\infty}{\longrightarrow}\; &\frac{1}{(m+1)^2}\left\{\kappa^2[-\ln(d)+d-1]\right. \\ &\left.+\ln(d)-\ln(w_{\min})+1-d\right\} \\ &+\frac{2m}{(1-w_{\min})(m+1)^2}\left\{\kappa^2\left[1+d\ln(d)-d-\frac{d}{2}\ln(d)^2\right]\right. \\ &\left.+w_{\min}\ln(w_{\min})-w_{\min}-d\ln(d)+d+\frac{d}{2}\ln(d)^2\right\}.\end{aligned} \tag{22}$$

and as m increases the MSFE converges to zero.

Appendix B: Equity Futures and Sample Periods

The equity series refer to futures contracts taken from Datastream and cover the different periods as set out below. The start of the samples generally coincide with the start dates of the futures markets in question.

AEX: Amsterdam Exchange Index, Netherlands
$w = 156$ – Number of forecasts: 864 (01-Jun-1989 to 24-Nov-2008)
$w = 260$ – Number of forecasts: 760 (25-Oct-1989 to 24-Nov-2008)

ASX: Australian Securities Exchange Index
$w = 156$ – Number of forecasts: 279 (06-Dec-2000 to 19-Nov-2008)
$w = 260$ – Number of forecasts: 175 (02-May-2001 to 19-Nov-2008)

BEL: BEL 20 Index, Belgium
$w = 156$ – Number of forecasts: 603 (07-Jun-1994 to 24-Nov-2008)
$w = 260$ – Number of forecasts: 499 (31-Oct-1994 to 24-Nov-2008)

CAC: CAC40 index, France
$w = 156$ – Number of forecasts: 868 (24-Mar-1989 to 24-Nov-2008)
$w = 260$ – Number of forecasts: 764 (17-Aug-1989 to 24-Nov-2008)

DAX: DAX 30 index, Germany
$w = 156$ – Number of forecasts: 753 (02-Jul-1991 to 24-Nov-2008)
$w = 260$ – Number of forecasts: 649 (25-Nov-1991 to 24-Nov-2008)

DJE: DJ EURO STOXX 50, DJ euro index
$w = 156$ – Number of forecasts: 375 (27-Jan-1999 to 25-Nov-2008)
$w = 260$ – Number of forecasts: 271 (22-Jun-1999 to 25-Nov-2008)

FTSE: FTSE 100, U.K.
$w = 156$ – Number of forecasts: 1054 (09-Aug-1985 to 19-Nov-2008)
$w = 260$ – Number of forecasts: 950 (06-Jan-1986 to 19-Nov-2008)

FOX: FOX Index, Finland
$w = 156$ – Number of forecasts: 283 (02-May-2000 to 19-Nov-2008)
$w = 260$ – Number of forecasts: 179 (25-Sep-2000 to 19-Nov-2008)

IBE: IBEX 35, Spain
$w = 156$ – Number of forecasts: 672 (25-Nov-1992 to 24-Nov-2008)
$w = 260$ – Number of forecasts: 568 (21-Apr-1993 to 24-Nov-2008)

KFX: KFX Index, Denmark
$w = 156$ – Number of forecasts: 233 (14-Aug-2001 to 25-Nov-2008)
$w = 260$ – Number of forecasts: 129 (08-Jan-2002 to 25-Nov-2008)

MIB: Milan index, Italy
$w = 156$ – Number of forecasts: 551 (04-Jul-1995 to 20-Nov-2008)
$w = 260$ – Number of forecasts: 447 (27-Nov-1995 to 20-Nov-2008)

ND: NASDAQ 100 index, U.S.A.
$w = 156$ – Number of forecasts: 480 (14-Nov-1996 to 21-Nov-2008)
$w = 260$ – Number of forecasts: 376 (10-Apr-1997 to 21-Nov-2008)

NK: NIKKEI 225, Japan
$w = 156$ – Number of forecasts: 938 (30-Apr-1987 to 20-Nov-2008)
$w = 260$ – Number of forecasts: 834 (23-Sep-1987 to 20-Nov-2008)

OBX: OBX index, Norway
$w = 156$ – Number of forecasts: 326 (26-Aug-1999 to 24-Nov-2008)
$w = 260$ – Number of forecasts: 222 (19-Jan-2000 to 24-Nov-2008)

OMX: OMX Index, Sweden
$w = 156$ – Number of forecasts: 783 (17-Sep-1990 to 19-Nov-2008)
$w = 260$ – Number of forecasts: 679 (11-Feb-2002 to 19-Nov-2008)

PSI: PSI 20 Index, Portugal
$w = 156$ – Number of forecasts: 463 (27-Jan-1997 to 24-Nov-2008)
$w = 260$ – Number of forecasts: 359 (20-Jun-1997 to 24-Nov-2008)

SP: S&P COMP index, U.S.A.
$w = 156$ – Number of forecasts: 1050 (09-Aug-1985 to 19-Nov-2008)
$w = 260$ – Number of forecasts: 946 (06-Jan-1986 to 19-Nov-2008)

SMI: SWISS MI index, Switzerland
$w = 156$ – Number of forecasts: 766 (18-Jun-1991 to 20-Nov-2008)
$w = 260$ – Number of forecasts: 662 (11-Nov-1991 to 20-Nov-2008)

TPX: Topix Stock Price Index, Japan
$w = 156$ – Number of forecasts: 422 (18-Aug-1997 to 19-Nov-2008)
$w = 260$ – Number of forecasts: 204 (12-Jan-1998 to 19-Nov-2008)

TSX: Toronto Stock Exchange Index, Canada
$w = 156$ – Number of forecasts: 308 (12-Apr-2000 to 20-Nov-2008)
$w = 260$ – Number of forecasts: 204 (05-Sep-2000 to 20-Nov-2008)

References

Assenmacher-Wesche, Katrin, and M. Hashem Pesaran (2008) 'Forecasting the Swiss economy using VECX* models: An exercise in forecast combination across models and observation windows', *National Institute Economic Review* 203, 91–108.

Bai, Jushan, and Pierre Perron (1998) 'Estimating and testing linear models with multiple structural changes.' *Econometrica* 66, 47–78.

Bai, Jushan, and Pierre Perron (2003) 'Computation and analysis of multiple structural change models.' *Journal of Applied Econometrics* 18, 1–22.

Branch, William A., and George W. Evans (2006) 'A simple recursive forecasting model.' *Economic Letters* 91, 158–166.

Clark, Todd E., and Michael W. McCracken (2004) 'Improving forecast accuracy by combining recursive and rolling forecasts.' Federal Reserve Bank of Kansas City *Research Working Paper* 04-10.

Clark, Todd E., and Michael W. McCracken (2006) 'Averaging forecasts from VARs with uncertain instabilities.' Federal Reserve Bank of Kansas City *Research Working Paper* 06-12.

Clemen, Robert T. (1989) 'Combining forecasts: A review and annotated bibliography.' *International Journal of Forecasting* 5, 559–581.

Clements, Michael P. and David F. Hendry (2006) 'Forecasting with breaks.' In *Handbook of Economic Forecasting*, ed. G. Elliott, C.W.J. Granger and A. Timmermann (Elsevier) pp. 605–657.

Diebold, Francis X., and Roberto S. Mariano, (1995) 'Comparing predictive accuracy.' *Journal of Business and Economic Statistics* 13, 253–263.

Gardner Jr., Everette S. (2006) 'Exponential smoothing: The state of the art–Part II.' *International Journal of Forecasting* 22, 637–666.

Hamilton, James D. (1989) 'A new approach to the economic analysis of nonstationary times series and the business cycle.' *Econometrica* 57, 357–384.

Harvey, Andrew C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Pesaran, M. Hashem, Til Schuermann and L. Vanessa Smith (2009) 'Forecasting economic and financial variables with global VARs.' *International Journal of Forecasting* forthcoming.

Pesaran, M. Hashem, and Allan Timmermann (2007) 'Selection of estimation window in the presence of breaks.' *Journal of Econometrics* 137(1), 134–161.

Stock, James H., and Mark W. Watson (2004) 'Combination forecasts of output growth in a seven-country data set.' *Journal of Forecasting* 23, 405–430.

Timmermann, Allan (2006) 'Forecast combinations.' In *Handbook of Economic Forecasting,* ed. G. Elliot, C.W.J. Granger, and A. Timmermann (Elsevier).