Forecasting Random Walks Under Drift Instability[*]

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

Andreas Pick
University of Cambridge, CIMF

March 11, 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 monthly inflation series of 21 OECD countries 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

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 (2007) 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 (2007) 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. To allow for the uncertainty of γ, we shall also consider a mixed procedure where ExpW(γ) forecasts obtained for different values of γ are also averaged to give a new average forecast, which

we refer to as AveExpW.

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, ExpW, and AveExpW procedures to forecasting monthly inflation series in a number of OECD countries.

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

3

2   Average Window Forecasts

2.1   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, $y_1, y_2, \ldots, y_T$. 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_i is

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

where

$$\hat{\mu}_{T+1}(w_i) = \frac{1}{Tw_i} \sum_{t=T_i}^{T} y_t = \frac{1}{Tw_i} \sum_{t=T(1-w_i)+1}^{T} y_t.$$

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 object of interest in this paper is to compare the single-window and the AveW forecasts, $\hat{y}_{T+1}(w_i)$ and $\hat{y}_{T+1}(\text{AveW})$, respectively, 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 not an alternative to rolling forecasts and can be used irrespective of whether a rolling or an expanding window is used in recursive forecasting.

2.2 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

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

where $I(c)$ 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

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

Hence, the forecast bias is

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

and since $(w-d)\,I(w-d) > 0$, 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

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

$\lambda = (\mu_2 - \mu_1)/\sigma$, 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 and depends on Tw. The (scaled) mean squared forecast error (MSFE) for a window of size w is given

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

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, $w_0 > 0$, 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

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

6

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

$$\sigma^{-1} e_{T+1}(\text{AveW}) \;=\; u_{T+1} + \frac{\lambda}{m+1} \sum_{i=0}^{m} \left(\frac{w_i - d}{w_i} \right) \text{I}\,(w_i - d)$$

$$- \frac{1}{m+1} \sum_{i=0}^{m} \frac{1}{Tw_i} \sum_{t=T(1-w_i)+1}^{T} u_t.$$

Hence, the bias of the AveW forecast is given by

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

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

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

and note that

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

As a result it is easily verified that

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

Therefore, the scaled MSFE in this case is given by

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

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

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

$$\text{MSFE}(w_a; \lambda, d) - \text{MSFE}(m, w_{\min}; \lambda, d) \;=\;$$

$$\lambda^2 \left(\frac{w_a - d}{w_a} \right)^2 \text{I}(w_a - d) + \frac{1}{Tw_a} \tag{14}$$

$$- \left[\frac{\lambda}{m+1} \sum_{i=0}^{m} \frac{w_i - d}{w_i} \text{I}(w_i - d) \right]^2 - \frac{1}{(m+1)^2} \sum_{i=0}^{m} \frac{1 + 2i}{Tw_i},$$

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

$$
\frac{1}{T} \sum_{i=0}^{m} \left(\frac{w_i - d}{w_i} \right) \mathrm{I}(w_i - d) \overset{T \to \infty}{\longrightarrow} \int_{d}^{1} \left(\frac{x - d}{x} \right) dx,
$$

$$
= (1 - d) + d \ln(d) \geq 0,
$$

where the lower boundary of the intregral, 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

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

$$
= 1 - w_{\min} + w_{\min} \ln w_{\min}.
$$

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

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

$$
- \frac{\ln(w_{\min})}{(m + 1)^2} + \frac{2m [1 - w_{\min} + w_{\min} \ln(w_{\min})]}{(1 - w_{\min})(m + 1)^2} + 1.
$$

The first term is asymptotic bias due to the break, the next two terms capture the sampling effects of the variances and the covariances of the

8

forecast errors across the different windows, and the final term is the error variance of the forecast period. For m sufficiently large only the bias term remains. Note that the size of the break only enters the bias term and that the dominant term of the sampling error part only depends on $w_{\min}$, which is a choice parameter. This term, $[1 - w_{\min} + w_{\min} \ln(w_{\min})] / (1 - w_{\min})$, is monotonically declining over the range $(0, 1)$, and since $w_{\min} \leq d$ by assumption, then the optimal choice of $w_{\min}$ is to set it equal to the distance to break. In practice, since d is unknown one needs to select $w_{\min}$ so that it is sufficiently large, but not too large so that $w_{\min} \leq d$ is not violated.

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

$$\text{MSFE}(w; \lambda, d) - \text{MSFE}(m, w_{\min}; \lambda, d) \tag{16}$$
$$\approx \lambda^2 \left(\frac{w_a - d}{w_a} \right)^2 \text{I}(w_a - d) + \frac{1 - w_a}{m w_a}$$
$$- \frac{m^2}{(m + 1)^2} \frac{\lambda^2}{(1 - w_{\min})^2} \left[(1 - d) + d \ln(d) \right]^2$$
$$- \frac{2m \left(1 - w_{\min} + w_{\min} \ln(w_{\min}) \right)}{(1 - w_{\min})(m + 1)^2} + \frac{\ln(w_{\min})}{(m + 1)^2}$$

The difference depends on the length of the single window forecast, w_a, the number of windows used for averaging, m, 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 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 bias, 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 bias only if $d = 1$, namely if there has not been a break.

Next consider the difference in variances (up to order m^{-1}) for values of $w_{\min}$ in the range $(0, 1)$, which is proportional to

$$\xi(w_{\min}) = \frac{2 \left(1 - w_{\min} + w_{\min} \ln w_{\min} \right)}{1 - w_{\min}} - (1 - w_{\min})$$

9

Figure 1: Asymptotic difference in variance of AveW and single window forecast, $\xi(w_{\min})$
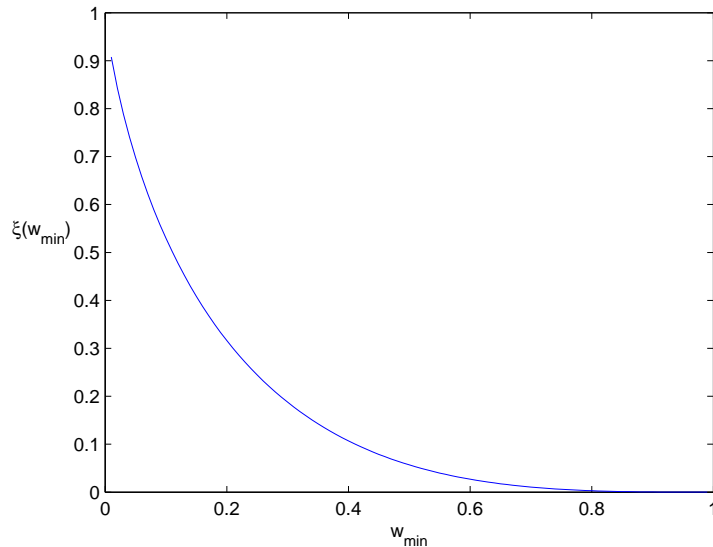

Figure 1 plots this function, which is monotonically declining.

Therefore, a trade-off exists between the lower bias and the higher variance of the AveW forecast relative to the singe window forecast. 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 2 plots the exact and the asymptotic (approximate) differences in MSFE of the two forecast procedures in (14) and (16) for $T = 1000$, where the triangular shape of the surface is due to the fact that $w_i \leq d$. It can be seen that the asymptotic MSFE in the right column of Figure 2 and the exact MSFE in the left column of Figure 2 are very similar. Furthermore, for breaks as small as $\lambda = 0.05$ the MSFE of the AveW forecast is smaller than that of the single window forecast of size $w = 1$ for most of values of d and $w_{\min}$.

Figure 3 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 and asymptotic difference in MSFE with $T = 1000$
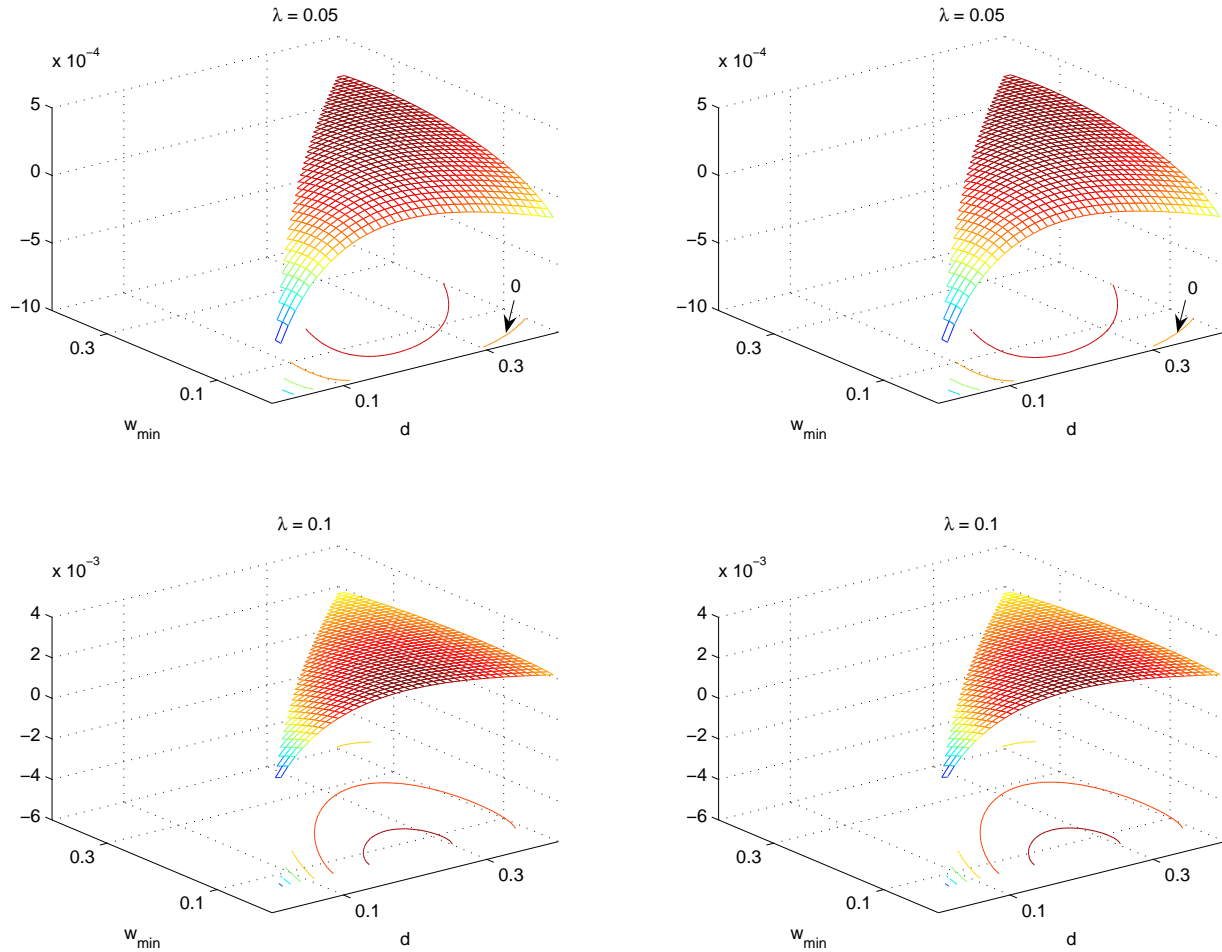
The two plots on the left show the exact difference in MSFE in (14)
and the two plots on the right show the asymptotic difference in MSFE
in (16). The arrows in the top plots point to the zero-isoquant; the
surfaces in the plots in the second row are always positive.

2.3   A generalization to multiple breaks in drift

Consider the following random walk model where the drift term is subject
to n different breaks. Denote the break points by d_i, $i = 1, 2, ..., n$, such that
$d_1 > d_2 > ... > d_n$, and let the means of the process over these segments be
μ_1, μ_2,...and μ_{n+1}. Specifically,

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

11

Figure 3: Exact difference in MSFE with $T = 100$
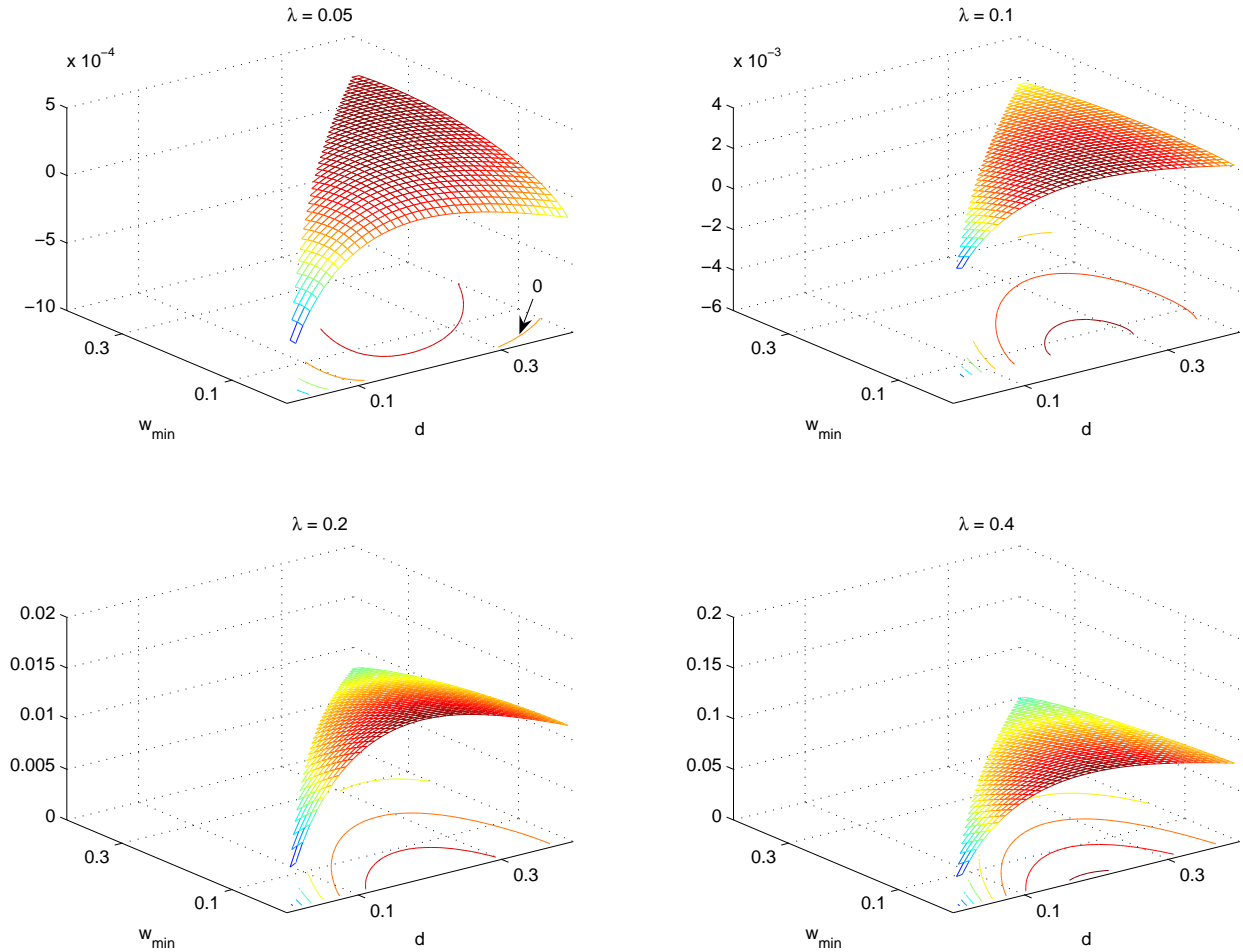
The plots show the exact difference in MSFE in (14). The arrow in the
left upper plot points to the zero-isoquant; the surfaces in the other
plots are always positive

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

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

The one-step ahead forecast error is

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

which after some algebra, and noting that $I(w - d_1)I(w - d_2) = I(w - d_1)$, 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

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

with

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

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

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

where

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

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

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

For AveW forecast the bias per break will be

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

13

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

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

$$\lim_{n \to \infty} B_F(n) = B_F = \mathrm{E}(\lambda_i)(1 - \mathrm{E}(d_i))$$

$$\lim_{n \to \infty} B_{\mathrm{AveW}}(n) = B_{\mathrm{AveW}} = \frac{m}{m+1} \frac{\mathrm{E}(\lambda_i)}{1 - w_{\min}} \{1 - \mathrm{E}(d_i) + \mathrm{E}[d_i \ln(d_i)]\},$$

and

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

$$= [E(\lambda_i)]^2 \left\{ (1 - \mathrm{E}(d_i))^2 \left[1 - \left(\frac{m}{m+1} \right)^2 \left(\frac{1}{1 - w_{\min}} \right)^2 \right] \right.$$

$$\left. - \left(\frac{m}{m+1} \right)^2 \left(\frac{1}{1 - w_{\min}} \right)^2 \mathrm{E}[d_i \ln(d_i)] \{2 - 2\mathrm{E}(d_i) + \mathrm{E}[d_i \ln(d_i)]\} \right\}.$$

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

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

Furthermore, as $d_i \ln(d_i) \leq 0$ for all $d_i \in (0, 1)$, then $-\mathrm{E}[d_i \ln(d_i)] \geq 0$.[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

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

Hence,

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

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

The magnitude of $\lim_{n \to \infty} \left[B_F^2(n) - B_{\mathrm{AveW}}^2(n) \right]$ 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 $E(d_i) = 1/2$,

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

and

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

---

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

Hence, we have

$$\lim_{n\to\infty}\left[B_F^2(n)-B_{\text{AveW}}^2(n)\right]=\frac{3}{16}\left[\text{E}(\lambda_i)\right]^2\geq0.$$

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

2.4 Break in drift and volatility

Here we assume that there is only one break point, but now also allow the volatility to change, that is, in model (1) we consider the case where $\sigma_1\neq\sigma_2$ and $\mu_1\neq\mu_2$. Note that the bias terms in the MSFE of the single window forecast and the AveW forecast are independent of the variance terms and will be identical to the case of the break in drift only. We can therefore 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. Due to the separability of the MSFE the extension to different break dates is straightforward.

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

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

The scaled MSFE for the single window forecast is

$$
\begin{aligned}
\text{MSFE}(w;\kappa,d) &= \text{E}\left[\sigma_2^{-2}e_{T+1|T}(w)^2\right]\\
&= \frac{(w-d)}{Tw^2}\text{I}(w-d)\,\kappa^2+\frac{\min(w,d)}{Tw^2}+1 \qquad (17)
\end{aligned}
$$

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

The forecast error for the AveW forecast is

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

and the scaled MSFE of the AveW forecast is

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

15

For m sufficiently large the scaled MSFE in (18) can be approximated similarly to the break in drift case considered above. The asymptotic scaled MSFE of the AveW forecast is

$$
\text{MSFE}(m, w_{\min}; \kappa, d) \overset{T \to \infty}{\longrightarrow} \frac{1}{(m+1)^2} \left\{ \kappa^2 [-\ln(d) + d - 1] \right. \tag{19}
$$
$$
+ \ln(d) - \ln(w_{\min}) + 1 - d \}
$$
$$
+ \frac{2m}{(1 - w_{\min})(m+1)^2} \left\{ \kappa^2 \left[1 + d\ln(d) - d - \frac{d}{2}\ln(d)^2 \right] \right.
$$
$$
\left. + w_{\min}\ln(w_{\min}) - w_{\min} - d\ln(d) + d + \frac{d}{2}\ln(d)^2 \right\}.
$$

The derivation and some further results are provided in Appendix A.

We can now combine these results with those of the break in drift to obtain the scaled MSFE for the single window forecast

$$
\text{E}\big(\sigma_2^{-2} e_{T+1}(w)^2\big) = \left(\frac{w-d}{w}\right)^2 \lambda^2 \text{I}(w-d)
$$
$$
+ \frac{w-d}{Tw^2}\text{I}(w-d)\,\kappa^2 + \frac{\min(w,d)}{Tw^2} + 1, \tag{20}
$$

where $\lambda = |\mu_2 - \mu_1| / \sigma_2$.

For the AveW forecasts over $m+1$ windows, the scaled MSFE is

$$
\text{E}\big(\sigma_2^{-2} e_{T+1}(\text{AveW})^2\big) = \frac{1}{(m+1)^2} \left\{ \left[\sum_{T_i \in \mathbb{T}(d)} \left(\frac{w_i - d}{w_i}\right) \lambda \text{I}(w_i - d) \right]^2 \right.
$$
$$
+ \kappa^2 \sum_{i=0}^{m} \frac{w_i - d}{Tw_i^2} \text{I}(w_i - d) + \sum_{i=0}^{m} \frac{\min(w_i, d)}{Tw_i^2}
$$
$$
+ 2\kappa^2 \sum_{i=0}^{m-1} \frac{w_i - d}{w_i} \text{I}(w_i - d) \sum_{j=i+1}^{m} \frac{1}{Tw_j}
$$
$$
\left. + 2 \sum_{i=0}^{m} \frac{\min(w_i, d)}{w_i} \sum_{j=i+1}^{m} \frac{1}{Tw_j} \right\} + 1. \tag{21}
$$

The difference between (20) and (21) is

$$\mathrm{MSFE}\,(w;\lambda,d) - \mathrm{MSFE}\,(m,s,w_{\min};\lambda,d) \;=$$

$$
\begin{aligned}
=\;\; & \lambda^2 \left(\frac{w-d}{w}\right)^2 \mathrm{I}(w-d) - \left[\frac{\lambda}{m+1}\sum_{i=0}^{m}\frac{w_i-d}{w_i}\mathrm{I}(w_i-d)\right]^2 \\
& +\frac{w-d}{Tw^2}\mathrm{I}(w-d)\,\kappa^2 + \frac{\min(w,d)}{Tw^2} \\
& -\frac{1}{(m+1)^2}\left[\kappa^2\sum_{i=0}^{m}\frac{w_i-d}{Tw_i^2}\mathrm{I}(w_i-d) + \sum_{i=0}^{m}\frac{\min(w_i,d)}{Tw_i^2}\right. \\
& +2\kappa^2\sum_{i=0}^{m-1}\frac{w_i-d}{w_i}\mathrm{I}(w_i-d)\sum_{j=i+1}^{m}\frac{1}{Tw_j} \\
& +2\sum_{i=0}^{m-1}\frac{\min(w_i,d)}{w_i}\left.\sum_{j=i+1}^{m}\frac{1}{Tw_j}\right],
\end{aligned}
\tag{22}
$$

which can be approximated by

$$
\begin{aligned}
& \mathrm{MSFE}(w;\lambda,\kappa,d) - \mathrm{MSFE}(m,w_{\min};\lambda,\kappa,d) \\
& \stackrel{T\to\infty}{\longrightarrow}\;\; \lambda^2\left(\frac{w-d}{w}\right)^2\mathrm{I}(w-d) + \frac{1-w_{\min}}{m}\left[\frac{w-d}{w^2}\mathrm{I}(w-d)\,\kappa^2 + \frac{\min(w,d)}{w^2}\right] \\
& -\frac{m^2}{(m+1)^2}\frac{\lambda^2}{(1-w_{\min})^2}[1-d+d\ln(d)]^2 \\
& -\frac{1}{(m+1)^2}\left\{\sigma_1^2[-\ln(d)+d-1] + \ln(d)-\ln(w_{\min})+1-d\right\} \\
& -\frac{2m}{(1-w_{\min})(m+1)^2}\left\{\sigma_1^2\left[1+d\ln(d)-d-\frac{d}{2}\ln(d)^2\right]\right. \\
& \left. + w_{\min}\ln(w_{\min}) - w_{\min} - d\ln(d) + d + \frac{d}{2}\ln(d)^2\right\},
\end{aligned}
\tag{23}
$$

where it has been established above that the bias part is always positive and the variance component is only positive for large values of κ.

Figure 4 plots the exact differences in scaled MSFEs of the forecast procedures. When comparing the plots to those for the break in drift only in Figure 3, 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 $w_{\min}$.

17

Figure 4: Exact difference in MSFE for a break in drift and volatility with $T = 100$
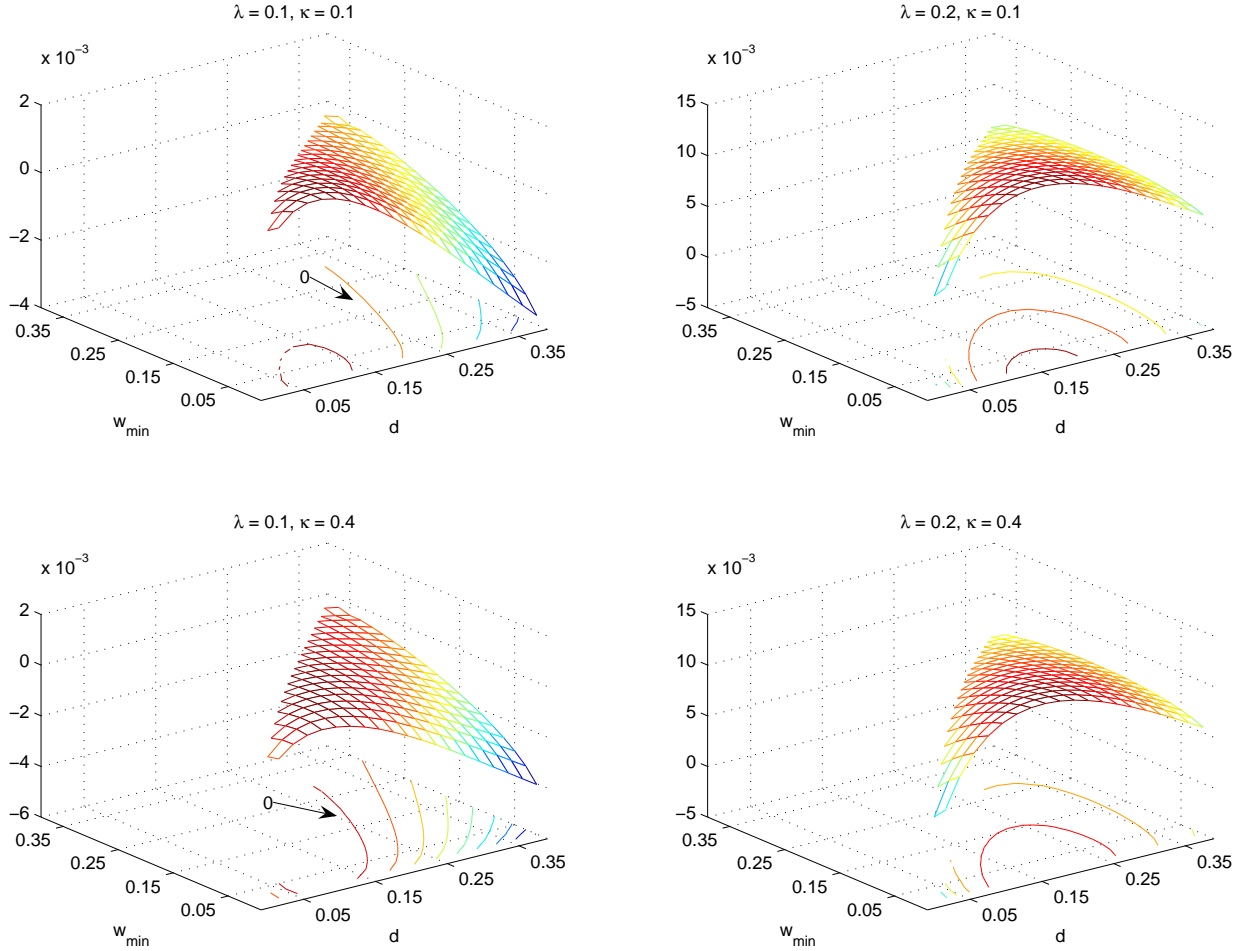
The plots show the exact different in MSFE in (22). The arrows point to the zero-isoquants.

3 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 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

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

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

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

19

But

$$\sum_{j=1}^{T_b-1} \gamma^{T-j} \mu_1 = \mu_1 \left(\frac{\gamma^{T-T_b+1} - \gamma^T}{1-\gamma} \right)$$

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

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

$$\begin{aligned}
\text{MSFE}\left[\hat{y}_{T+1}(\text{ExpW}, \gamma)\right] &= f(\gamma) \qquad\qquad\qquad (24) \\
&= 1 + \lambda^2 \left(\frac{\gamma^{1+T\,d} - \gamma^T}{1-\gamma^T} \right)^2 \\
&\quad + \left(\frac{1-\gamma}{1-\gamma^T} \right)^2 \left(\frac{1-\gamma^{2T}}{1-\gamma^2} \right),
\end{aligned}$$

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

Figure 5 compares the MSFE of the single window forecast with $w = 1$ to that of the ExpW forecast given in (24) 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

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

Figure 5: Exact difference of MSFEs of single window and ExpW for $T = 100$
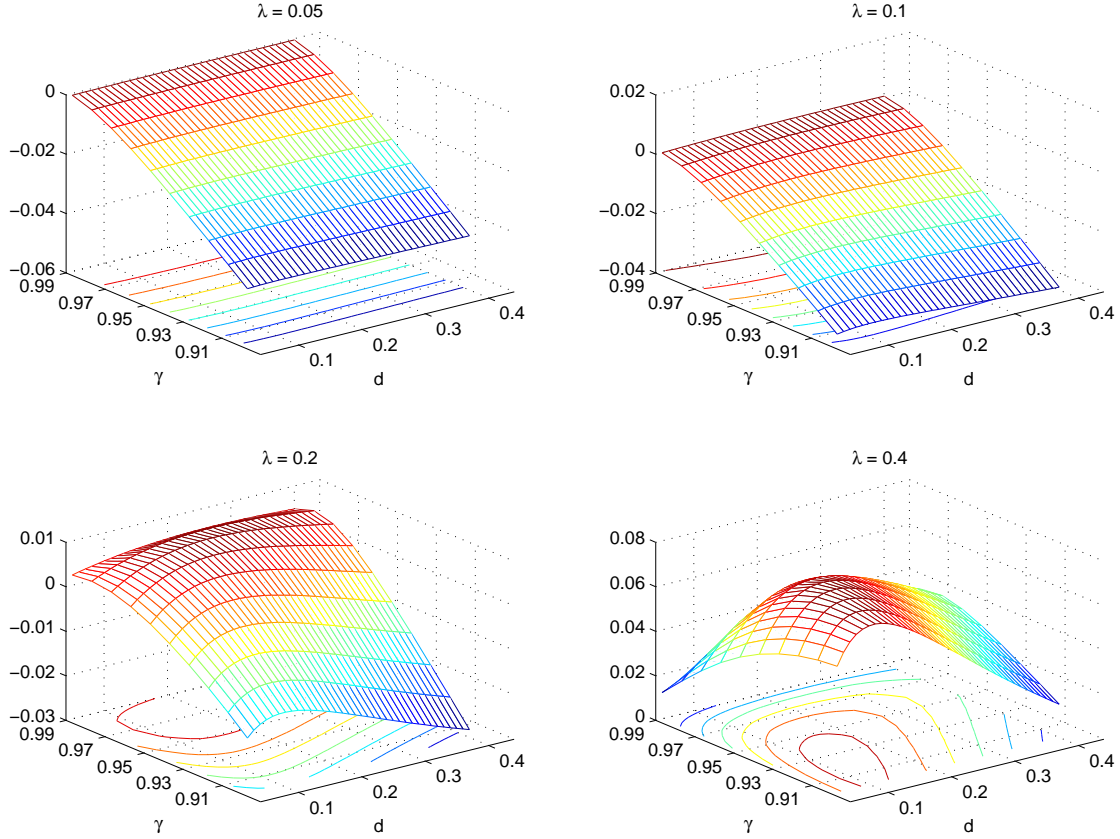
The graphs plot the exact difference between the single window and ExpW forecasts, that is, [MSFE(Single window) − MSFE(ExpW$(\gamma; \lambda, d)$)]$/\sigma^2$.

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

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

21

Figure 6: Exact difference of MSFEs of AveW and ExpW for $T = 100$
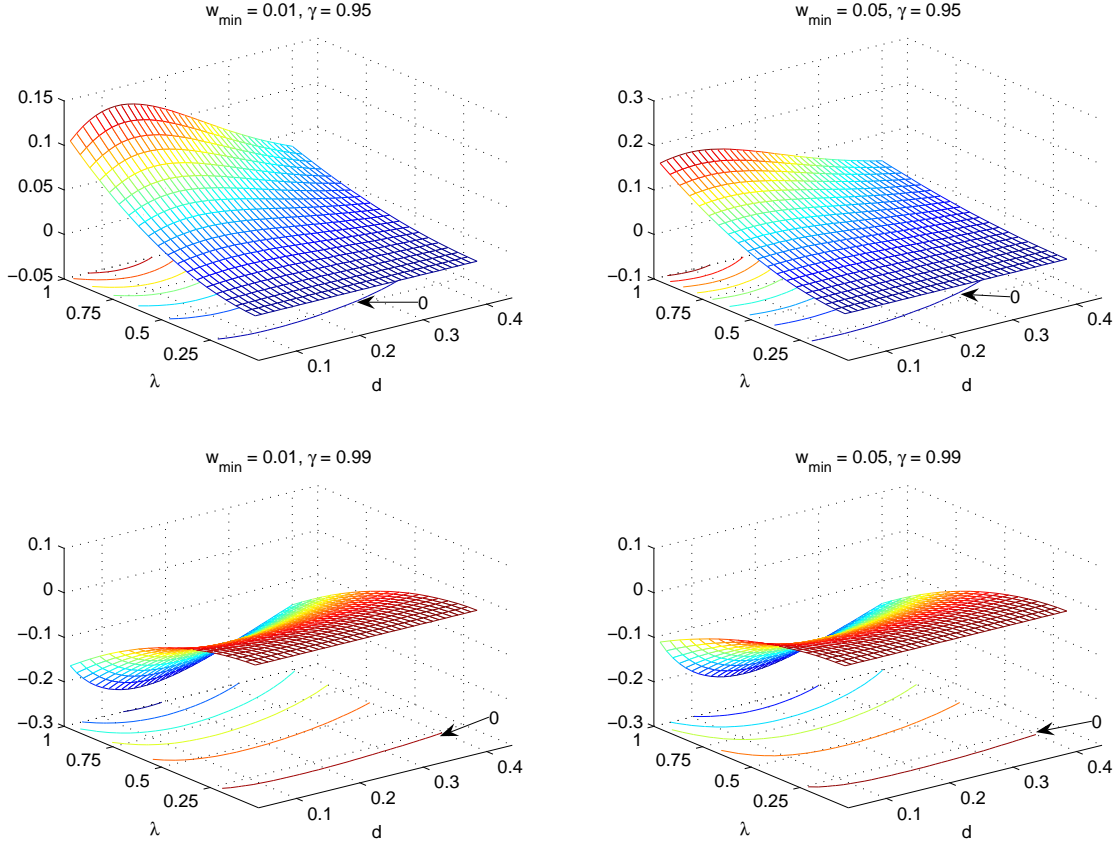
The graphs plot the exact difference between the AveW and ExpW forecasts, that is, $[\text{MSFE}(\text{AveW}(w_{\min}; \lambda, d)) - \text{MSFE}(\text{ExpW}(\gamma; \lambda, d))]/\sigma^2$. The arrows point to the zero-isoquant.

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 6–8 compare the AveW forecast with the ExpW forecasts for different values of T, d, and λ, and for different choices of γ. Figure 6 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

Figure 7: Exact difference of MSFEs of AveW and ExpW for $T = 1000$
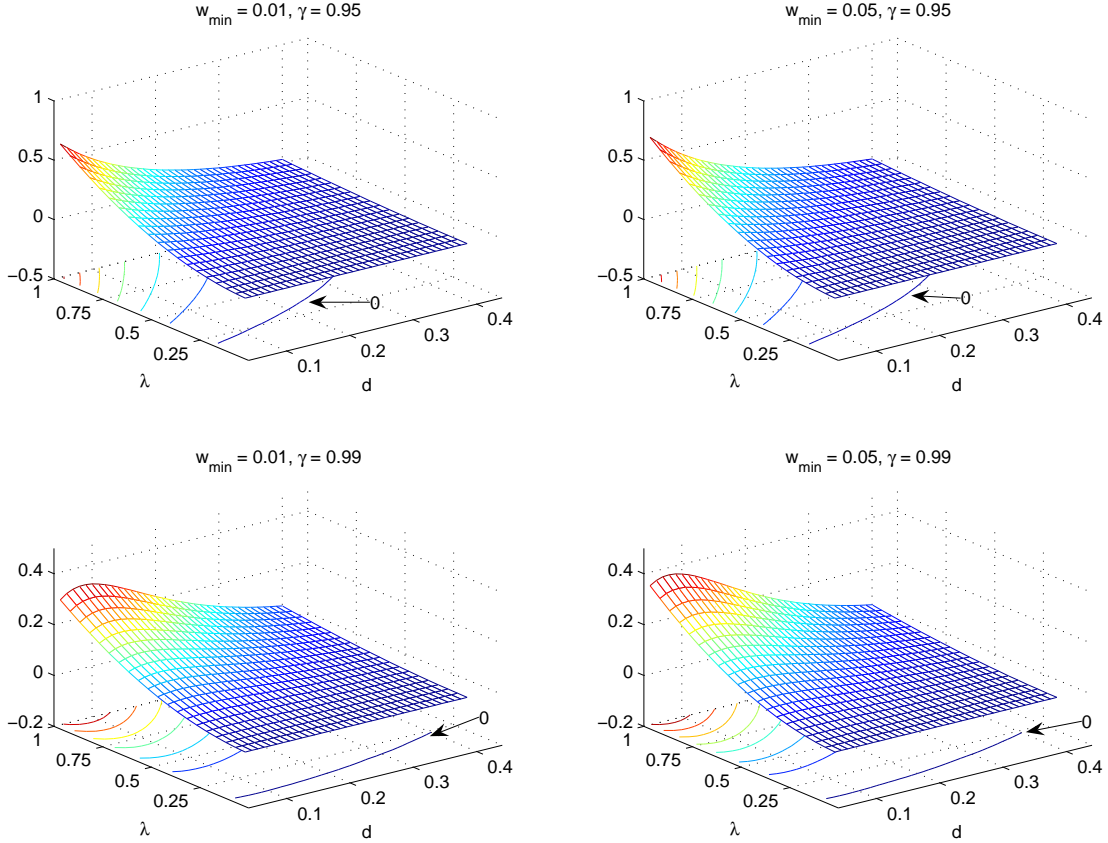
See footnote to Figure 6.

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 7, the choice of γ is less important. The ExpW forecasts have a smaller MSFE except for relatively small values of λ and large values of d.

Figure 8 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

23

Figure 8: Difference in MSFEs between AveW and ExpW forecasts with fixed break point

D = 10; T = 100; Tw$_{min}$ = 10

D = 20; T = 100; Tw$_{min}$ = 10

D = 10; T = 1000; Tw$_{min}$ = 10

D = 20; T = 1000; Tw$_{min}$ = 10
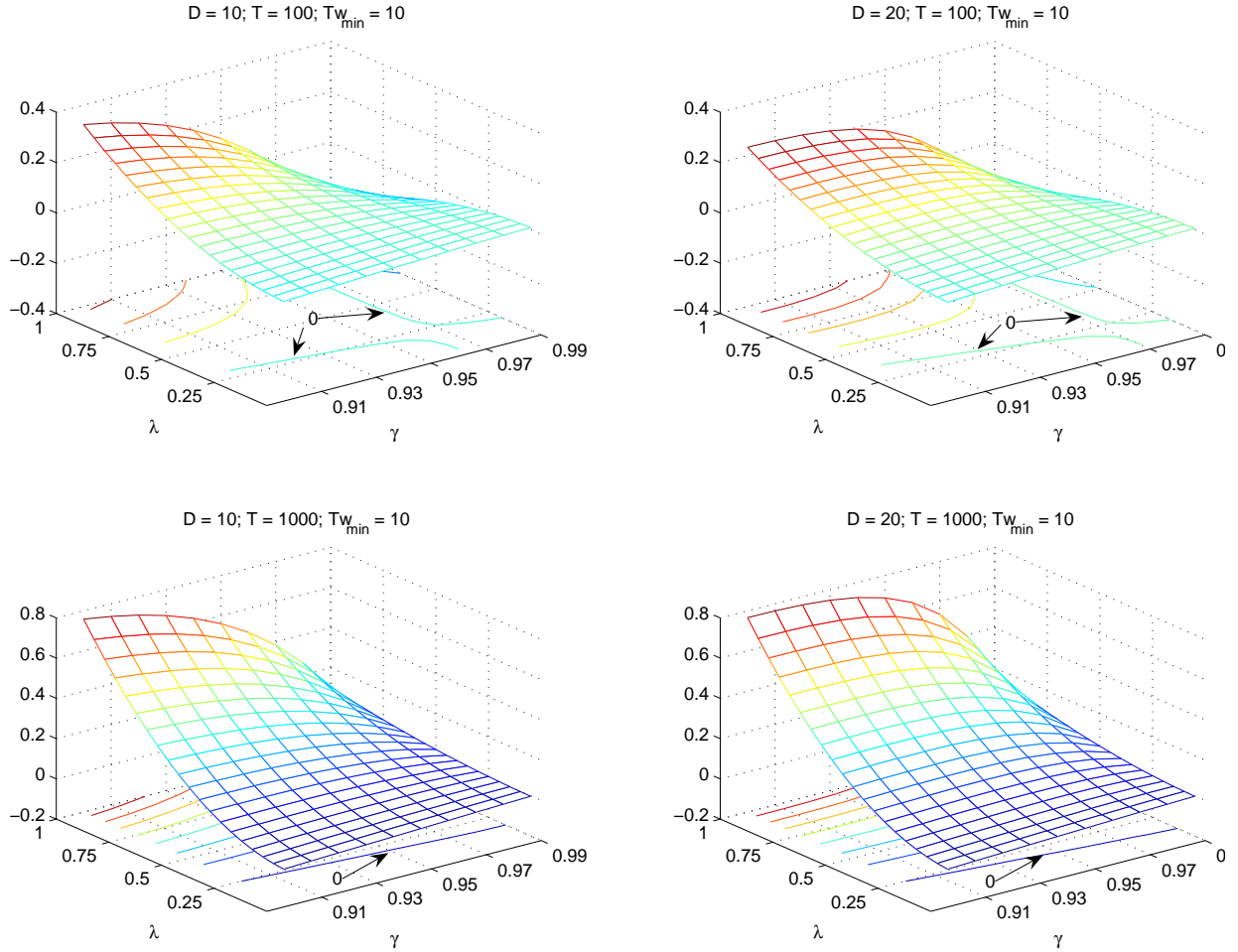
See footnote to Figure 6. Here, however, the break point $D = dT$ and the minimum window Tw_{min} are fixed and not fractions of T.

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 9. 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

24

Figure 9: Weights attached to the observations in the AveW and ExpW forecasts for $T = 100$
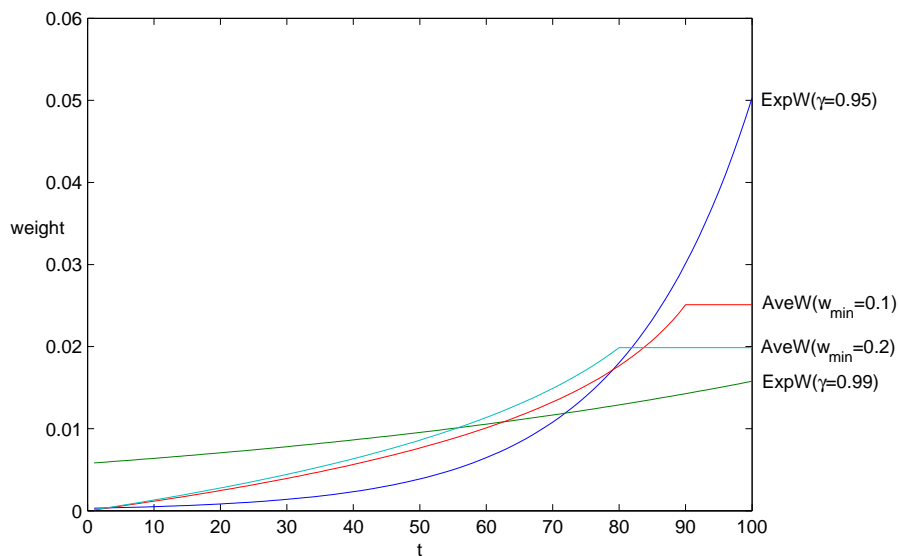
interesting observation is that the AveW weights do not differ as much as those of ExpW between the different weighting schemes. This suggests that ExpW forecasts will depend considerably more on the choice of γ than AveW forecasts depend on the choice of $w_{\min}$.

3.1 A Mixed Approach – AveExpW Forecasts

It is clear that the optimal choice of γ critically depends on the relative values of λ, d and T. For T sufficiently large the bias term vanishes if γ is not too close to unity - but the variance will then tend to $(1 - \gamma)/(1 + \gamma)$ which does not vanish. This suggests using a value of γ between 0.95 and 0.99. One possibility would be to compute the ExpW forecasts for different values of γ in the range $0.95 - 1$, say, and then combine these forecasts by simple averaging.

4 Applications to Forecasting Inflation

In this section we will apply the AveW, the ExpW and AveExpW procedures to forecast monthly inflation of 21 OECD countries using CPI data from 1985M1 to 2007M10. Details of the data sources are given in Appendix B. By using (1) we assume that inflation is a stationary variable with breaks in its mean, which reflects the recent findings in the literature, for example

Burdekin and Siklos (1999) and Robalo Marques (2004).

We recursively compute one-step ahead forecasts using rolling windows of 8 years, which leads to 177 forecasts for each time series, where the first forecast is for 1993M2. We also computed two-step ahead forecasts, which give similar results. In order to save space we don't report these results here.

The baseline forecasts are based on eight-years rolling regressions. We compare these forecasts with AveW rolling forecasts based on the same eight-year window. Recursive forecasts are then obtained by rolling the eight year window forward by one month and then repeating the forecasting exercise for the next month and so on. We compute AveW forecasts for two choices of the minimum window, $w_{\min} = 0.1$ and 0.2. Additionally, we compute ExpW forecasts using $\gamma = 0.95$ and 0.98, and the AveExpW forecasts where the different ExpW(γ) forecasts are averaged over two set of values, namely $\gamma \in \{0.95, 0.952, 0.954, \ldots, 1\}$ and $\gamma \in \{0.98, 0.982, 0.984, \ldots, 1\}$.

We report the bias, the root mean square forecast error (RMSFE) and the tests for predictive performance proposed by Diebold and Mariano (1995) (DM). We has also computed the test for predictive ability by Harvey, Leybourne and Newbold (1997). The results turned out to be very similar and therefore only the DM test statistics are reported. More precisely, assume that we want to forecast y_{t+h}, and the forecast is based on the observations up to t, that is, $y_{t+h|t}$. Then the h-step ahead forecast error is defined as $e_{t,h} = \frac{1}{h}(y_{t+h} - y_{t+h|t})$. The RMSFE is then computed as

$$\text{RMSFE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} e_{t,h}^2}.$$

We report the RMSFE and the relative RMSFE, that is for, say, the AveW forecast we report

$$\frac{\text{RMSFE(AveW)}}{\text{RMSFE(SW)}},$$

where here and in the tables SW denotes the forecast from the single rolling window. Values smaller than one therefore imply that the single window forecast has a larger RMSFE than the AveW forecast. The test 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 results are reported in Tables 1 to 6. The relative RMSFEs are smaller than one in 86% of the cases. However, not in all those cases is the difference significant. Using the test of Diebold and Mariano (1995) we find that the single window produces a significantly (at the 5% level) worse forecast in 43% of forecasts and, importantly, it is not significantly better in forecasting inflation in any of the 21 countries considered.

26

	SW	AveW		ExpW		AveExpW	
		0.1	0.2	0.95	0.98	(0.95, 1)	(0.98, 1)
Austria							
Bias	0.0255	0.0177	0.0194	0.0131	0.0203	0.0191	0.0230
RMSFE	0.2990	0.2981	0.2983	0.2991	0.2980	0.2980	0.2983
SW	1.000	0.3070	0.2671	−0.0347	0.4763	0.4464	0.6821
AveW(0.1)		0.9970	−0.3981	−0.8091	0.0841	0.1573	−0.0918
AveW(0.2)			0.9976	−0.5406	0.2998	0.3699	−0.0075
ExpW(0.95)				1.0005	0.5559	0.6693	0.2950
ExpW(0.98)					0.9968	0.1255	−0.2516
AveExpW(0.95, 1)						0.9967	−0.2333
AveExpW(0.98, 1)							0.9976
Belgium							
Bias	0.0153	0.0107	0.0117	0.0078	0.0122	0.0115	0.0138
RMSFE	0.2717	0.2722	0.2720	0.2741	0.2720	0.2723	0.2717
SW	1.000	−0.2843	−0.2237	−0.9533	−0.2454	−0.4185	−0.0103
AveW(0.1)		1.0019	0.3978	−2.0418	0.3614	−0.1514	0.4293
AveW(0.2)			1.0013	−1.5985	0.1314	−0.4434	0.3566
ExpW(0.95)				1.0087	1.5155	1.5583	1.2349
ExpW(0.98)					1.0011	−1.2110	0.4918
AveExpW(0.95, 1)						1.0021	0.7365
AveExpW(0.98, 1)							1.0000
Canada							
Bias	0.0384	0.0175	0.0194	0.0108	0.0242	0.0225	0.0311
RMSFE	0.3265	0.3198	0.3200	0.3197	0.3212	0.3207	0.3234
SW	1.000	2.0371	2.3003	1.4190	2.3801	2.1992	2.7831
AveW(0.1)		0.9794	−0.3390	0.0639	−1.2929	−1.2844	−1.6405
AveW(0.2)			0.9801	0.1383	−1.7674	−1.3062	−1.9648
ExpW(0.95)				0.9791	−0.5669	−0.4692	−0.9892
ExpW(0.98)					0.9837	1.0535	−1.9627
AveExpW(0.95, 1)						0.9823	−1.7413
AveExpW(0.98, 1)							0.9904
Switzerland							
Bias	0.0650	0.0383	0.0425	0.0260	0.0470	0.0438	0.0560
RMSFE	0.3394	0.3326	0.3334	0.3321	0.3340	0.3334	0.3362
SW	1.000	2.4362	2.3338	1.9219	2.7532	2.7404	3.0940
AveW(0.1)		0.9801	−1.3173	0.3656	−1.6546	−1.1482	−2.0245
AveW(0.2)			0.9823	0.6406	−0.8890	−0.0417	−1.8081
ExpW(0.95)				0.9785	−0.9291	−0.7586	−1.4177
ExpW(0.98)					0.9843	1.6249	−2.3461
AveExpW(0.95, 1)						0.9824	−2.3226
AveExpW(0.98, 1)							0.9907

See note of Table 6.

Table 2: Relative performance in forecasting inflation, $h = 1$

	SW	AveW		ExpW		AveExpW	
		0.1	0.2	0.95	0.98	(0.95, 1)	(0.98, 1)
Germany							
Bias	0.0275	0.0221	0.0239	0.0179	0.0240	0.0228	0.0259
RMSFE	0.2018	0.1968	0.1978	0.1959	0.1974	0.1967	0.1991
SW	1.000	1.3647	1.1986	1.1899	1.7524	1.8106	2.1088
AveW(0.1)		0.9750	−2.0878	0.5734	−0.5122	0.0978	−0.9489
AveW(0.2)			0.9800	0.9668	0.4404	1.4691	−0.6159
ExpW(0.95)				0.9709	−0.5696	−0.3378	−0.8422
ExpW(0.98)					0.9780	1.9345	−1.3626
AveExpW(0.95, 1)						0.9746	−1.5333
AveExpW(0.98, 1)							0.9864
Denmark							
Bias	0.0210	0.0052	0.0063	0.0010	0.0102	0.0091	0.0154
RMSFE	0.3185	0.3200	0.3193	0.3230	0.3193	0.3199	0.3187
SW	1.000	−0.6774	−0.4170	−1.3892	−0.5368	−0.7610	−0.2390
AveW(0.1)		1.0046	1.8402	−2.4394	0.9874	0.2780	0.9164
AveW(0.2)			1.0025	−2.3431	−0.0142	−1.1110	0.5311
ExpW(0.95)				1.0140	2.0605	2.0922	1.7243
ExpW(0.98)					1.0026	−1.8190	0.8431
AveExpW(0.95, 1)						1.0042	1.1540
AveExpW(0.98, 1)							1.0006
Spain							
Bias	0.0590	0.0303	0.0336	0.0201	0.0395	0.0369	0.0490
RMSFE	0.3959	0.3903	0.3903	0.3916	0.3917	0.3916	0.3934
SW	1.000	2.4339	2.7981	1.3281	2.6725	2.3622	3.1053
AveW(0.1)		0.9859	0.0419	−1.1292	−1.8579	−2.3618	−2.0472
AveW(0.2)			0.9858	−0.8701	−2.8266	−2.8130	−2.5591
ExpW(0.95)				0.9891	−0.0254	0.0102	−0.7070
ExpW(0.98)					0.9892	0.2051	−2.1973
AveExpW(0.95, 1)						0.9891	−1.7206
AveExpW(0.98, 1)							0.9935
Finland							
Bias	0.0641	0.0274	0.0308	0.0154	0.0391	0.0361	0.0513
RMSFE	0.3208	0.3076	0.3095	0.3042	0.3106	0.3091	0.3151
SW	1.000	3.0124	2.9291	2.6552	3.3687	3.3707	3.7030
AveW(0.1)		0.9587	−2.5439	1.5637	−2.1859	−1.5134	−2.6112
AveW(0.2)			0.9645	1.8270	−1.2540	0.4823	−2.3946
ExpW(0.95)				0.9482	−1.8831	−1.6822	−2.2569
ExpW(0.98)					0.9682	2.8459	−2.9941
AveExpW(0.95, 1)						0.9632	−3.0453
AveExpW(0.98, 1)							0.9821

See note of Table 6.

Table 3: Relative performance in forecasting inflation, $h = 1$

	SW	AveW		ExpW		AveExpW	
		0.1	0.2	0.95	0.98	(0.95, 1)	(0.98, 1)
France							
Bias	0.0229	0.0108	0.0123	0.0065	0.0148	0.0136	0.0187
RMSFE	0.2281	0.2253	0.2254	0.2261	0.2260	0.2259	0.2268
SW	1.000	1.3923	1.4622	0.7198	1.5545	1.4167	1.8480
AveW(0.1)		0.9879	−0.3065	−0.9074	−1.0301	−1.2095	−1.1401
AveW(0.2)			0.9884	−0.6055	−1.0858	−1.0293	−1.2154
ExpW(0.95)				0.9915	0.1177	0.1805	−0.3189
ExpW(0.98)					0.9908	0.2283	−1.2383
AveExpW(0.95, 1)						0.9906	−1.0323
AveExpW(0.98, 1)							0.9944
U.K.							
Bias	0.0647	0.0284	0.0326	0.0151	0.0401	0.0367	0.0522
RMSFE	0.3745	0.3675	0.3678	0.3686	0.3692	0.3689	0.3714
SW	1.000	1.7058	1.8064	1.0582	1.9110	1.7877	2.2135
AveW(0.1)		0.9813	−0.3831	−0.5537	−1.2569	−1.3588	−1.4242
AveW(0.2)			0.9820	−0.3048	−1.3856	−1.1407	−1.5450
ExpW(0.95)				0.9840	−0.2092	−0.1253	−0.6484
ExpW(0.98)					0.9857	0.6307	−1.5820
AveExpW(0.95, 1)						0.9849	−1.3997
AveExpW(0.98, 1)							0.9915
Greece							
Bias	0.2817	0.1548	0.1701	0.1082	0.1956	0.1837	0.2377
RMSFE	1.1505	1.1298	1.1302	1.1323	1.1346	1.1338	1.1411
SW	1.000	1.6476	1.8107	1.0454	1.8515	1.7070	2.1439
AveW(0.1)		0.9821	−0.2700	−0.4652	−1.1971	−1.4024	−1.3758
AveW(0.2)			0.9824	−0.2962	−1.6272	−1.9425	−1.5905
ExpW(0.95)				0.9842	−0.2561	−0.1972	−0.6694
ExpW(0.98)					0.9862	0.6070	−1.5415
AveExpW(0.95, 1)						0.9855	−1.3410
AveExpW(0.98, 1)							0.9918
Ireland							
Bias	−0.0163	−0.0134	−0.0139	−0.0122	−0.0142	−0.0140	−0.0152
RMSFE	0.3882	0.3871	0.3878	0.3879	0.3873	0.3870	0.3875
SW	1.000	0.3037	0.1371	0.0579	0.4010	0.4220	0.5635
AveW(0.1)		0.9973	−0.8184	−0.3814	−0.1031	0.1226	−0.1706
AveW(0.2)			0.9989	−0.0435	0.6212	0.8242	0.1312
ExpW(0.95)				0.9992	0.2169	0.3508	0.0901
ExpW(0.98)					0.9976	0.4290	−0.2360
AveExpW(0.95, 1)						0.9971	−0.3092
AveExpW(0.98, 1)							0.9983

See note of Table 6.

Table 4: Relative performance in forecasting inflation, $h = 1$

	SW	AveW		ExpW		AveExpW	
		0.1	0.2	0.95	0.98	(0.95, 1)	(0.98, 1)
Iceland							
Bias	0.1418	0.0353	0.0410	0.0120	0.0688	0.0631	0.1036
RMSFE	0.5517	0.4398	0.4470	0.4202	0.4663	0.4583	0.5038
SW	1.000	6.2457	6.3165	5.8937	6.6289	6.5514	6.9455
AveW(0.1)		0.7971	−4.3076	3.7804	−5.0595	−4.7755	−5.7090
AveW(0.2)			0.8102	3.9557	−4.9750	−4.1499	−5.7783
ExpW(0.95)				0.7618	−4.5576	−4.3946	−5.2689
ExpW(0.98)					0.8453	5.2903	−6.1995
AveExpW(0.95, 1)						0.8307	−6.0906
AveExpW(0.98, 1)							0.9132
Italy							
Bias	0.0770	0.0428	0.0471	0.0297	0.0538	0.0505	0.0652
RMSFE	0.1709	0.1519	0.1545	0.1464	0.1568	0.1548	0.1632
SW	1.000	5.3741	5.3271	5.0522	5.7915	5.7647	6.1509
AveW(0.1)		0.8888	−4.4759	3.7438	−4.3593	−3.7082	−4.9010
AveW(0.2)			0.9042	3.9958	−3.2236	−0.5456	−4.7032
ExpW(0.95)				0.8566	−4.1280	−3.9352	−4.5863
ExpW(0.98)					0.9176	5.0055	−5.3768
AveExpW(0.95, 1)						0.9059	−5.3859
AveExpW(0.98, 1)							0.9549
Japan							
Bias	0.0453	0.0239	0.0267	0.0153	0.0308	0.0286	0.0380
RMSFE	0.2370	0.2343	0.2349	0.2340	0.2348	0.2344	0.2357
SW	1.000	0.7879	0.6430	0.6989	0.9416	1.0239	1.0877
AveW(0.1)		0.9888	−1.3861	0.2457	−0.4430	−0.0924	−0.6161
AveW(0.2)			0.9912	0.5452	0.0974	0.5044	−0.3762
ExpW(0.95)				0.9875	−0.3731	−0.2141	−0.5284
ExpW(0.98)					0.9908	1.2406	−0.7780
AveExpW(0.95, 1)						0.9891	−0.9434
AveExpW(0.98, 1)							0.9944
Korea							
Bias	0.0775	0.0428	0.0481	0.0272	0.0542	0.0501	0.0658
RMSFE	0.5024	0.5024	0.5018	0.5045	0.5018	0.5021	0.5018
SW	1.000	0.0046	0.1472	−0.3552	0.2047	0.1048	0.4008
AveW(0.1)		1.0000	0.6235	−0.8555	0.4169	0.2623	0.2126
AveW(0.2)			0.9988	−0.8113	0.0142	−0.1618	0.0119
ExpW(0.95)				1.0042	0.8148	0.8808	0.5941
ExpW(0.98)					0.9988	−0.4371	0.0081
AveExpW(0.95, 1)						0.9993	0.1468
AveExpW(0.98, 1)							0.9988

See note of Table 6.

Table 5: Relative performance in forecasting inflation, $h = 1$

	SW	AveW		ExpW		AveExpW	
		0.1	0.2	0.95	0.98	$(0.95, 1)$	$(0.98, 1)$
Netherlands							
Bias	0.0052	0.0091	0.0104	0.0064	0.0081	0.0073	0.0070
RMSFE	0.4259	0.4286	0.4278	0.4318	0.4276	0.4282	0.4266
SW	1.000	−1.0534	−0.8134	−1.6955	−0.9883	−1.1669	−0.7893
AveW(0.1)		1.0063	1.7405	−2.6327	1.1805	0.6103	1.1905
AveW(0.2)			1.0045	−2.4487	0.2886	−0.5437	0.8235
ExpW(0.95)				1.0140	2.2349	2.2872	1.9549
ExpW(0.98)					1.0041	−1.8272	1.1922
AveExpW(0.95, 1)						1.0054	1.4456
AveExpW(0.98, 1)							1.0017
Norway							
Bias	0.0588	0.0255	0.0279	0.0169	0.0360	0.0339	0.0469
RMSFE	0.3188	0.3098	0.3095	0.3099	0.3114	0.3111	0.3143
SW	1.000	2.7356	3.0947	1.9518	3.1267	2.8949	3.5583
AveW(0.1)		0.9715	0.5778	−0.1202	−1.6829	−1.8423	−2.2079
AveW(0.2)			0.9707	−0.2336	−2.6348	−2.5301	−2.7303
ExpW(0.95)				0.9721	−0.6253	−0.5869	−1.3043
ExpW(0.98)					0.9766	0.8243	−2.6218
AveExpW(0.95, 1)						0.9757	−2.2639
AveExpW(0.98, 1)							0.9859
Portugal							
Bias	0.1646	0.0875	0.0964	0.0596	0.1123	0.1052	0.1379
RMSFE	0.4170	0.3725	0.3771	0.3620	0.3840	0.3799	0.3989
SW	1.000	5.3822	5.4219	4.8761	5.7681	5.7041	6.1331
AveW(0.1)		0.8932	−4.0649	3.0787	−4.4178	−3.9992	−4.9096
AveW(0.2)			0.9043	3.3423	−3.9941	−2.1986	−4.8981
ExpW(0.95)				0.8681	−3.7657	−3.5852	−4.3331
ExpW(0.98)					0.9208	4.6398	−5.3348
AveExpW(0.95, 1)						0.9109	−5.2666
AveExpW(0.98, 1)							0.9564
Sweden							
Bias	0.1076	0.0552	0.0611	0.0358	0.0722	0.0671	0.0896
RMSFE	0.4185	0.3947	0.3972	0.3895	0.4006	0.3983	0.4085
SW	1.000	3.6392	3.5282	3.3615	3.9337	4.0011	4.1985
AveW(0.1)		0.9431	−3.0959	2.0109	−2.9173	−2.2729	−3.2899
AveW(0.2)			0.9491	2.2972	−2.1662	−0.7898	−3.0675
ExpW(0.95)				0.9307	−2.5575	−2.3515	−2.9675
ExpW(0.98)					0.9572	3.5469	−3.6169
AveExpW(0.95, 1)						0.9517	−3.7519
AveExpW(0.98, 1)							0.9761

See note of Table 6.

Table 6: Relative performance in forecasting inflation, $h = 1$

	SW	AveW		ExpW		AveExpW	
		0.1	0.2	0.95	0.98	$(0.95, 1)$	$(0.98, 1)$
U.S.A.							
Bias	0.0253	0.0109	0.0122	0.0064	0.0155	0.0144	0.0203
RMSFE	0.2261	0.2232	0.2235	0.2236	0.2239	0.2237	0.2248
SW	1.000	1.6033	1.7892	0.8253	1.8023	1.5595	2.1888
AveW(0.1)		0.9873	-0.6251	-0.2806	-1.1361	-1.4731	-1.2894
AveW(0.2)			0.9886	-0.0526	-1.1046	-0.4866	-1.4698
ExpW(0.95)				0.9890	-0.1442	-0.0732	-0.4701
ExpW(0.98)					0.9902	0.4750	-1.4122
AveExpW(0.95, 1)						0.9895	-1.1169
AveExpW(0.98, 1)							0.9941

SW denotes the single window forecast, AveW(0.1) and AveW(0.2) the AveW forecasts with $w_{\min} = 0.1$ and 0.2, ExpW(0.95) and ExpW(0.98) the ExpW forecasts with $\gamma = 0.95$ and 0.98, and AveExpW(0.95,1) and AveExpW(0.98,1) the AveExpW forecasts with γ ranging from 0.95 to 1 and 0.98 to 1 with steps of 0.002. On the diagonals of rows three to nine are the root mean square forecast error (RMSFE) as ratios where the RMSFE of the single window is in the denominator. In the upper triangle of rows three to nine are the statistics for the test of predictive ability of Diebold and Mariano (1995), where a positive value indicates that the method given in the top row has better predictive ability.

For the AveW forecasts the choice of $w_{\min} = 0.1$ leads to six significantly better forecasts compared to $w_{\min} = 0.2$, whereas the $w_{\min} = 0.2$ does not produce a significantly better forecast compared to $w_{\min} = 0.1$. Therefore, there is a preference for $w_{\min} = 0.1$

For ExpW, the choice of $\gamma = 0.95$ improves the forecast significantly over the choice of $\gamma = 0.98$ in 4 forecasts, while the reverse is true in two cases. The AveExpW with $\gamma \in (0.95, 1)$ forecasts are significantly worse than the ExpW forecasts with $\gamma = 0.95$ in five cases and better in only one case, but better that the ExpW forecasts with $\gamma = 0.98$ in five cases and never worse. AveExpW with $\gamma \in (0.98, 1)$ in contrast is never better than any of the AveW, ExpW, or AveExpW forecasts but often worse.

Overall it therefore appears that AveW with $w_{\min} = 0.1$ and ExpW with $\gamma = 0.95$ are the best choices for forecasting inflation. The difference between them is small: AveW is significantly better in three cases and ExpW in four cases. This suggests that the shorter memory implied by the shorter minimum window and the smaller decay parameter lead to better forecasts of inflation.

5   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.

Applying these techniques to international inflation data demonstrates that the AveW, ExpW and AveExpW forecasts have a lower MSFE than the single window forecast for most inflation series and that these improvements are often statistically significant. Importantly, the single window never forecasts inflation significantly better.

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.

Appendix A: Mathematical details

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

We have that

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

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

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

and, finally,

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

This results in the MSFE in (19).

The difference between the scaled MSFE of the single window forecast and that of the AveW forecast is

$$\mathrm{MSFE}(w;\sigma_1,d) - \mathrm{MSFE}(m,w_{\min};\sigma_1,d) \overset{T\to\infty}{\longrightarrow}$$
$$\frac{1-w_{\min}}{m}[(1-d)\sigma_1^2+d] - \frac{1}{(m+1)^2}\left\{\sigma_1^2[-\ln(d)+d-1]\right.$$
$$+ \ln(d)-\ln(w_{\min})+1-d\}$$
$$-2\frac{m}{(1-w_{\min})(m+1)^2}\left\{\sigma_1^2\left[1+d\ln(d)-d-\frac{d}{2}\ln(d)^2\right]\right. \tag{25}$$
$$+ w_{\min}\ln(w_{\min})-w_{\min}-d\ln(d)+d+\frac{d}{2}\ln(d)^2\bigg\}.$$

Figure 10: Minimum κ^2 for a positive difference in asymptotic MSFE
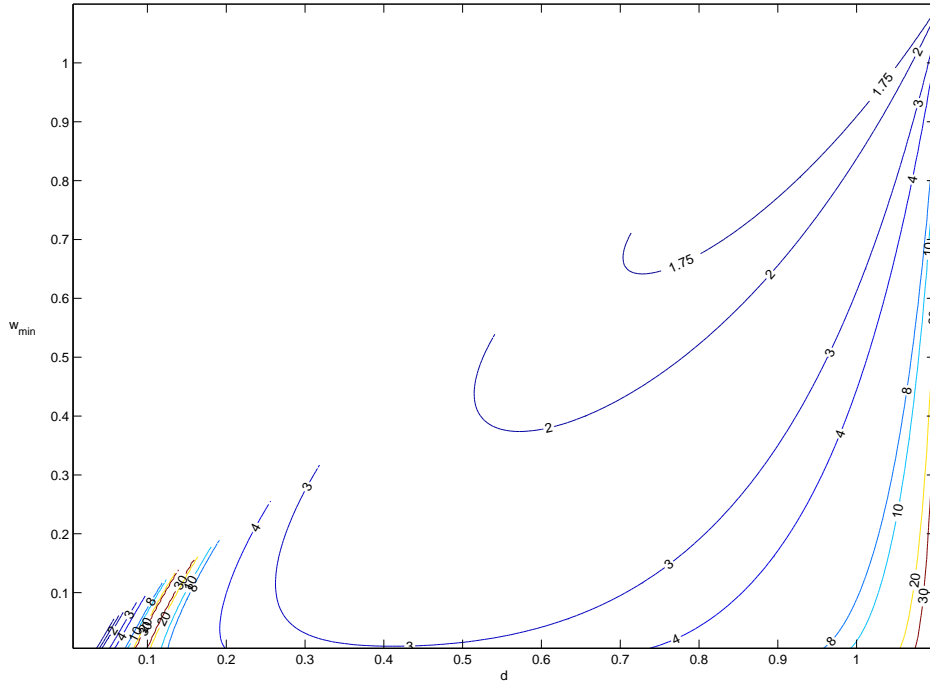
The graph shows the contours of (26), that is, the minimum values of
κ^2 that are required for the AveW forecast to have a lower MSFE than
the single window forecast.

From this we can establish the values of κ for which the difference in MSFE
is positive, which is

$$\kappa^2 > \frac{2\left\{w_{\min}[\ln(w_{\min}) - 1] + d[1 - \ln(d)] + \frac{d}{2}\ln(d)^2\right\} - (1 - w_{\min})^2 d}{(1 - w_{\min})^2(1 - d) - 2\left[1 + d\ln(d) - d - \frac{d}{2}\ln(d)^2\right]} \quad (26)$$

if the denominator is positive; if the denominator is negative the inequality
is reversed.

The contours of the function on the right hand side of (26) are plotted
in Figure 10, where the negative of (26) is plotted if the denominator is
negative so that κ^2 needs to exceed the values of the contour everywhere.
It emerges that the function is highly nonlinear and fairly large values of κ
are necessary to obtain a positive difference in scaled MSFEs.

Appendix B: Data

The data on CPI inflation are obtained from the OECD Main Economic Indicators data base and Datastream. The series are:

- Austria: OECD MEI; code: AUT.CPALTT01.IXOB

- Belgium: OECD MEI; code: BEL.CPALTT01.IXOB

- Canada: OECD MEI; code: CAN.CPALTT01.IXOB

- Switzerland: OECD MEI; code: CHE.CPALTT01.IXOB

- Germany: Datastream; code: BDCONPRCE

- Denmark: OECD MEI; code: DNK.CPALTT01.IXOB

- Spain: OECD MEI; code: ESP.CPALTT01.IXOB

- Finland: OECD MEI; code: FIN.CPALTT01.IXOB

- France: OECD MEI; code: FRA.CPALTT01.IXOB

- UK: OECD MEI; code: GBR.CPALTT01.IXOB

- Greece: OECD MEI; code: GRC.CPALTT01.IXOB

- Ireland: OECD MEI; code: IRL.CPALTT01.IXOB

- Iceland: OECD MEI; code: ISL.CPALTT01.IXOB

- Italy: OECD MEI; code: ITA.CPALTT01.IXOB

- Japan: OECD MEI; code: JPN.CPALTT01.IXOBSA

- Korea: OECD MEI; code: KOR.CPALTT01.IXOB

- Netherlands: OECD MEI; code: NLD.CPALTT01.IXOB

- Norway: Datastream; code: NWCONPRCE

- Portugal: OECD MEI; code: PRT.CPALTT01.IXOB

- Sweden: OECD MEI; code: SWE.CPALTT01.IXOB

- USA: OECD MEI; code: USA.CPALTT01.IXOBSA

For reasons of comparability we use time series of equal length and therefore start the series in 1985M1 and end in 2007M10. A number of series are not seasonally adjusted. While we do not expect price data from developed countries to exhibit a strong seasonal pattern, it could be accounted for by using only data of the same month for the forecasting process. This would, however, entail a significant loss of power due to the reduced sample size, and we do not attempt this here.

For each series we calculate the month-on-month inflation rate

$$\pi_t = 100[\log(\text{CPI}_t) - \log(\text{CPI}_{t-1})].$$

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.

Bates, J. M., and C. W. J. Granger (1969) 'The combination of forecasts.' *Operational Research Quarterly* 20, 451–468.

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

Burdekin, Richard C. K., and Pierre L. Siklos (1999) 'Exchange rate regimes and shifts in inflation persistence: Does nothing else matter?' *Journal of Money, Credit and Banking* 31, 235–247.

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 *Resarch 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.

Harvey, David, Stephen Leybourne, Paul Newbold (1997) 'Testing the equality of prediction mean squared errors.' *International Journal of Forecasting* 13, 281–291.

Pesaran, M. Hashem, Til Schuermann and L. Vanessa Smith (2007) 'Forecasting economic and financial variables with global VARs.' *mimeo* University of Cambridge.

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

Robalo Marques, Carlos (2004) 'Inflation persistence: Facts or artefacts?' ECB *Working Paper* 371.

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).