Forecast Combination Across Estimation Windows

M. Hashem Pesaran
Faculty of Economics and CIMF, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DD, U.K.; and Department of Economics, USC, Kaprielian Hall 300, Los Angeles, CA 90089-0253

Andreas Pick
Econometric Institute, Erasmus University Rotterdam, Rotterdam 3000 DR, Netherlands; De Nederlandsche Bank; and CIMF (*andreas.pick@cantab.net*)

In this article we consider combining forecasts generated from the same model but over different estimation windows. We develop theoretical results for random walks with breaks in the drift and volatility and for a linear regression model with a break in the slope parameter. Averaging forecasts over different estimation windows leads to a lower bias and root mean square forecast error (RMSFE) compared with forecasts based on a single estimation window for all but the smallest breaks. An application to weekly returns on 20 equity index futures shows that averaging forecasts over estimation windows leads to a smaller RMSFE than some competing methods.

KEY WORDS: Estimation windows; Exponential down-weighting; Forecast averaging; Structural breaks.

1. INTRODUCTION

There is a sizeable literature on the 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 increasingly used in macroeconomics and finance to good effect. 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) argued 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 proposed the idea of averaging forecasts from the same model but computed over different estimation windows. Using Monte Carlo experiments these authors show that this type of forecast averaging reduces the mean squared 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 macroeconomic forecasting. Assenmacher-Wesche and Pesaran (2008) used average forecasts based on different vector autoregressive models with weakly exogenous regressors (VARX*) of the Swiss economy estimated over different estimation windows and found that averaging forecasts across windows resulted in improvements over averaging of forecasts across models. Similar results were obtained by Pesaran, Schuermann, and Smith (2009), who applied the forecast averaging ideas to global VARs composed of 26 individual country/region VARX* models. Schrimpf and Wang (2010) applied averaging over estimation windows to forecasts of GDP growth based on the yield curve. Therefore, it is of interest to see if some theoretical insights can be gained to support these empirical findings.

In this article we derive theoretical results for the average windows (AveW) forecasting procedure. First, we consider a random walk model. The most interesting case is when the break occurs in the drift term but we also allow for simultaneous breaks in the drift and volatility of the random walk. We consider single and multiple breaks. We then extend the analysis to a linear regression model where the slope coefficient is subject to a break, and show that the results from the random walk model carry over to the linear regression model.

We compare the AveW forecasting procedure with an alternative method sometimes used in the literature where the past observations are down-weighted exponentially such that the most recent observations carry the greatest weight in the estimation. Gardner (2006) provided a recent review. We refer to this as the exponential smoothing (ExpS) forecast. In practice, the performance of this approach depends crucially on the choice of the parameter to downweight the past observations.

Initially focusing on a random walk model we show that, in the presence of breaks, the AveW and ExpS forecasting methods always have a lower bias than forecasts from a single estimation window. Whereas the MSFE depends on the time and the size of the breaks, the MSFE of the AveW and ExpS forecasts are smaller than those of the single-window forecasts for all but the smallest break sizes.

An attractive feature of these methods is that no exact information about the structural break is needed. This contrasts with the conventional approach of estimating the break points using such methods as those of Bai and Perron (1998, 2003), and then basing the forecasts on the post-break observations. However, as pointed out by Pesaran and Timmermann (2007), it is not always optimal to base forecasts only on the post-break data. Using pre-break data biases the forecast, but also reduces the forecast error variance. The overall effect of using pre-break

Journal of Business & Economic Statistics
April 2011, Vol. 29, No. 2
DOI: 10.1198/jbes.2010.09018

data on the MSFE is ambiguous and depends on the size and the point of the break. To optimally exploit information concerning parameter breaks in forecasting requires knowing the point and the size of the latest break. Even if the point of the last break can be estimated with some degree of confidence, it is unlikely that the size of the break can be estimated accurately, because it involves estimating the model over the pre-break and post-break periods. If the distance to break (measured from the date on which forecasts are made) is short, then the post-break parameters are likely to be poorly estimated relative to those obtained using pre-break data. In contrast, if the pre-break and post-break samples are both relatively large, then it might be possible to estimate the size of the break reasonably accurately, but in such cases the break information might not be that important. Results from Monte Carlo experiments and from the application to financial time series confirm this intuition.

Closely related to our approach is the suggestion by Clark and McCracken (2009) that averaging expanding and rolling windows can be useful for forecasting in the presence of structural breaks. This can be seen as a limited version of AveW forecasts where forecasts from only two different windows are combined.

Another 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 article we apply the AveW procedure to forecasting weekly returns on futures contracts for 20 world equity markets. Compared with a range of competing approaches, such as forecasts from rolling windows, expanding windows, ExpS forecasts, and forecasts based on post-break observations with breaks estimated by the sequential procedure of Bai and Perron (1998, 2003), the AveW forecast has the lowest RMSFE on average. However, in many cases the differences were not statistically significant, largely reflecting the highly volatile nature of weekly returns.

The rest of the article is organized as follows. Section 2 sets out the model, and Section 3 develops the AveW forecasting procedure and establishes its properties. Section 4 considers the ExpS forecast procedure. Section 5 reports the results of the applications to weekly returns on equity futures, and Section 6 concludes. Mathematical details are provided in Appendix A.

2. BASIC MODEL AND NOTATIONS

Consider the following time-varying regression model:

$$(y_t-\mu_y)=\beta_t(x_t-\mu_x)+\sigma_t\varepsilon_t, \qquad \varepsilon_t\sim \text{iid }(0,1), \quad (1)$$

which is defined over the sample period $t=1,2,\ldots,T+1$ and where the exogenous variable, x_t, is assumed to follow a covariance stationary process with mean μ_x and autocovariances, $\gamma_x(s)$, that are absolute summable, $\sum_{s=0}^{\infty}|\gamma_x(s)|<K<\infty$. Further assume that the slope parameter, β_t, and the standard deviation, σ_t, are subject to a break at time $t=T_b$ $(1<T_b<T)$,

$$\beta_t=\begin{cases}\beta^{(1)} & \forall t\le T_b\\ \beta^{(2)} & \forall t>T_b,\end{cases} \qquad \sigma_t=\begin{cases}\sigma^{(1)} & \forall t\le T_b\\ \sigma^{(2)} & \forall t>T_b.\end{cases}$$

The aim is to forecast y_{T+1} based on the observations $(y_1,y_2,\ldots,y_T)$ and $(x_1,x_2,\ldots,x_T,x_{T+1})$. When it is known with certainty that the parameters have not been subject to breaks, the forecast based on the ordinary least squares (OLS) estimates using all of the available observations is most efficient in the mean squared error sense. However, when the parameters are subject to breaks, more efficient forecasts can be obtained. As pointed out earlier, Pesaran and Timmermann (2007) showed that there is typically a trade-off between bias and variance of forecast errors. For example, when there is a break in the slope parameter, 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 using parameter estimates based on the post-break sample, $\{y_t,x_t\}_{t=T_b+1}^{T}$, is unbiased but for recent breaks could be inefficient due to a higher variance compared with the full-sample estimate. A third option is to use the optimal window length as suggested by Pesaran and Timmermann (2007). But calculating the optimal window relies on the time and size of the last break. If the break is close to the point of forecast, then reliable estimates of the size of the break cannot be obtained even if the time of the break can be determined accurately. Thus the estimated window length is likely to be suboptimal.

In the absence of reliable information on the point and size of the break(s) in β_t and σ_t, a forecasting procedure that is reasonably robust to such breaks will be of interest. In similar fashion to model averaging, which improves forecasts when the optimal model is uncertain, Pesaran and Timmermann (2007) considered the use of different sub-windows to forecast and then to average the outcomes, either by using cross-validated weights or simply using equal weights.

Toward this end, consider the sample $\{y_t,x_t\}_{t=T_i+1}^{T}$, with $0\le T_i<T$, which yields an observation window of size $W_i=T-T_i$. It is convenient to denote the fraction of observations in the single window (from the time when the forecast is formed) by $w_i=(T-T_i)/T$. The estimation process could start with a minimum window, $\{y_t,x_t\}_{t=T_{\min}+1}^{T}$, of size $w_{\min}=(T-T_{\min})/T$. From $w_{\min}$, other, larger windows can be considered with $T_i=T_{\min},T_{\min}-j,\ldots,T_{\min}-j(m-1)$, thus yielding m separate estimation windows with j observations apart. More specifically, we have

$$w_i=w_{\min}+\left(\frac{i-1}{m-1}\right)(1-w_{\min}) \quad \text{for } i=1,2,\ldots,m, \quad (2)$$

so that $w_i\in[w_{\min},1]$. Clearly, $w_m=1$ corresponds to using the full sample. The number of estimation windows, m, can be kept fixed as T changes or can be allowed to increase with T. In both cases we must have $m\le T(1-w_{\min})+1$. The maximum number of possible windows is set by $m=T(1-w_{\min})+1$. For this choice of m, we have

$$w_i=w_{\min}+\frac{i-1}{T}, \qquad i=1,2,\ldots,T(1-w_{\min})+1. \quad (3)$$

Similar to the window size, define the distance to the break by $b=(T-T_b)/T$, with $b\in(0,1)$. The forecast outcomes depend on whether b is a fixed fraction or changes with T. In the former case, $W_b=T-T_b\to\infty$ as $T\to\infty$; that is, the number of post-break observations is large when T is large. In this case, the point and size of the break can be estimated consistently, as was shown by Bai (1997). Under the latter, we consider the case where $b\to 0$ as $T\to\infty$, such that $W_b=T-T_b$ is small even when T is large. In this case, which is the focus of this article, the small number of post-break data is likely to lead to imprecise estimates of the point and size of the break.

Given that we consider one-step-ahead forecasts, we assume that no structural breaks occur in the forecast period. (For forecasting with structural breaks over the forecast period see Pesaran, Pettenuzzo and Timmermann 2006 and Maheu and Gordon 2008.)

3. AVERAGE WINDOW FORECAST

The AveW forecast is defined by the simple forecast combination rule

$$\hat{y}_{m,T+1} = \frac{1}{m}\sum_{i=1}^{m}\hat{y}_{T+1}(w_i), \tag{4}$$

where $\hat{y}_{T+1}(w_i)$ is the forecast from a given estimation window w_i, and forecasts from all windows are given equal weight.

The first object of interest in this article is comparing the single-window forecast, $\hat{y}_{T+1}(w)$, and the AveW forecasts, $\hat{y}_{m,T+1}$, 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 are used. In recursive estimation, the single window can be an expanding window or a rolling window, and AveW forecasts can be obtained by averaging over sub-windows within the given expanding or rolling window. Therefore, the AveW procedure is not an alternative to rolling forecasts and can be used irrespective of whether rolling or expanding windows are used in recursive forecasting.

3.1 Random Walk With Drift

Initially, we focus on a simple version of (1), where $\mu_y = \mu_x = 0$, $x_t = 1, \forall t$, and $\beta_t = \mu_t$ is subject to a single break at time T_b, that is,

$$y_t = \mu_t + \sigma_t\varepsilon_t, \qquad \varepsilon_t \sim \text{iid } (0,1), \tag{5}$$

where

$$\mu_t = \begin{cases} \mu^{(1)} & \forall t \le T_b \\ \mu^{(2)} & \forall t > T_b \end{cases} \qquad \text{and}$$

$$\sigma_t = \begin{cases} \sigma^{(1)} & \forall t \le T_b \\ \sigma^{(2)} & \forall t > T_b. \end{cases}$$

The simplicity of this model allows us to obtain exact finite-sample results for a single break in mean, multiple breaks in mean, and joint breaks in mean and error variance. However, the model is also a forecasting tool for a random walk with drift instability, $z_t = z_{t-1} + \mu_t + \varepsilon_t$, so that $y_t = \Delta z_t$, and $\hat{z}_{T+1} = z_T + \hat{y}_{T+1}(w)$, where

$$\hat{y}_{T+1}(w) = \frac{1}{Tw}\sum_{t=T(1-w)+1}^{T} y_t. \tag{6}$$

3.1.1 Single Break in Drift and Volatility. In the first instance assume that a single break occurs at date T_b, $1 < T_b < T$, and suppose that only the mean of the process is subject to the break, namely $\mu^{(1)} \neq \mu^{(2)}$, and $\sigma^{(1)} = \sigma^{(2)} = \sigma$. In this simple case, the one-step-ahead forecast of y_{T+1} based on a given window of size wT (from $t = T$) is given by

$$\hat{y}_{T+1}(w) = \mu^{(2)}[1 - \mathrm{I}(w-b)] + \mathrm{I}(w-b) \times \left[\frac{b\mu^{(2)} + (w-b)\mu^{(1)}}{w}\right] + \frac{1}{Tw}\sum_{t=T(1-w)+1}^{T}\sigma\varepsilon_t,$$

where $\mathrm{I}(c)$ is an indicator function that is unity if $c > 0$ and 0 otherwise. Clearly, if $w \le b$, then the forecast will have mean $\mu^{(2)}$ and will be unbiased. There is, however, a bias when $w > b > 0$. The associated forecast error, $\xi_{T+1}(w) = y_{T+1} - \hat{y}_{T+1}(w)$, is

$$\xi_{T+1}(w) = \left(\mu^{(2)} - \mu^{(1)}\right)\left(\frac{w-b}{w}\right)\mathrm{I}(w-b) + \sigma\varepsilon_{T+1} - \frac{1}{Tw}\sum_{t=T(1-w)+1}^{T}\sigma\varepsilon_t. \tag{7}$$

Thus the forecast bias is $\mathrm{E}[\xi_{T+1}(w)] = (\mu^{(2)} - \mu^{(1)})[(w - b)/w]\mathrm{I}(w-d)$. Because $(w-b)\mathrm{I}(w-b) > 0$, the direction of the bias depends on the sign of $(\mu^{(2)} - \mu^{(1)})$. Scaling the forecast error by σ, we have the decomposition

$$\sigma^{-1}\xi_{T+1}(w) = \varepsilon_{T+1} + B_{T+1}(w) - \frac{1}{Tw}\sum_{t=T(1-w)+1}^{T}\varepsilon_t, \tag{8}$$

where $B_{T+1}(w) = \lambda[(w-b)/w]\mathrm{I}(w-b)$ and $\lambda = (\mu^{(2)} - \mu^{(1)})/\sigma$. The first term, ε_{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, b; and the last term represents the estimation uncertainty that depends on Tw. The (scaled) MSFE for a window of size w is given by

$$\mathrm{MSFE}(w) = 1 + B_{T+1}^2(w) + \frac{1}{Tw}. \tag{9}$$

Now consider the forecast from averaging over estimation windows based on m successive windows of sizes from the smallest window fraction $w_{\min}$ to the largest possible one, w_m, where each forecast is of the form given in (6). The (scaled) one-step-ahead forecast error associated with the average forecast is

$$\sigma^{-1}\xi_{m,T+1} = \varepsilon_{T+1} + \frac{\lambda}{m}\sum_{i=1}^{m}\left(\frac{w_i - b}{w_i}\right)\mathrm{I}(w_i - b) - \frac{1}{m}\sum_{i=1}^{m}\frac{1}{Tw_i}\sum_{t=T(1-w_i)+1}^{T}\varepsilon_t.$$

Thus the bias of the AveW forecast is given by

$$B_{m,T+1} = \frac{\lambda}{m}\sum_{i=1}^{m}\left(\frac{w_i - b}{w_i}\right)\mathrm{I}(w_i - b), \tag{10}$$

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 different windows. Let $\nu_T(w_i) =$

$(1/(Tw_i))\sum_{t=T(1-w_i)+1}^{T}\varepsilon_t$, then $\mathrm{Cov}[\nu_T(w_i),\nu_T(w_j)] = \min(w_i,w_j)/(Tw_iw_j)$, for all $i,j=1,2,\ldots,m$, and thus

$$\sigma^{-2}\,\mathrm{Var}(\hat{y}_{m,T+1}) = \frac{1}{Tm^2}\left[\sum_{i=1}^{m}\frac{1}{w_i} + 2\sum_{i=1}^{m}\frac{i-1}{w_i}\right]. \tag{11}$$

Thus in this case, the scaled MSFE is given by

$$\mathrm{MSFE}(m, w_{\min};\lambda,b) = 1 + B_{m,T+1}^2 + \sigma^{-2}\,\mathrm{Var}(\hat{y}_{m,T+1}), \tag{12}$$

with $B_{m,T+1}$ and $\mathrm{Var}(\hat{y}_{m,T+1})$ as defined earlier.

The difference between the scaled MSFE of the single-window forecast (9) and that of the AveW forecast (12) is

$$\begin{aligned}
&\mathrm{MSFE}(w_a;\lambda,b) - \mathrm{MSFE}(m,w_{\min};\lambda,b)\\
&\quad = \lambda^2\left(\frac{w_a-b}{w_a}\right)^2 \mathrm{I}(w_a-b) + \frac{1}{Tw_a}\\
&\qquad - \left[\frac{\lambda}{m}\sum_{i=1}^{m}\frac{w_i-b}{w_i}\mathrm{I}(w_i-b)\right]^2\\
&\qquad - \frac{1}{m^2}\sum_{i=1}^{m}\frac{1+2(i-1)}{Tw_i}.
\end{aligned} \tag{13}$$

This depends on a number of parameters, including the size of the single window, w_a. Consider two cases: $w_a = b$ and $w_a > b$. When $w_a = b$, the forecast from the single window is unbiased, whereas the AveW forecast with $w_m > b$ is biased. The variance of the single-window forecast, $\sigma^2/(Tb)$, will be very large when Tb is small, and forecasting from a postbeak sample might not be desirable.

Now assume that $w_a > b$. In this case we can set $w_m = w_a$; that is, the AveW forecast is constructed from sub-windows within the expanding or rolling window.

Proposition 1. For DGP (5) with given T and b, the single-window forecast with $w_a > b$ has a larger absolute bias than the AveW forecast with w_i, $i=1,2,\ldots,T$ and $w_m = w_a$. In particular,

$$\left(\frac{w_a-b}{w_a}\right)\mathrm{I}(w_a-b) > \frac{1}{m}\sum_{i=1}^{m}\left(\frac{w_i-b}{w_i}\right)\mathrm{I}(w_i-b), \tag{14}$$

if $w_i < w_a$ for at least one i.

In contrast, the difference between the variance terms is ambiguous. Thus there may be a trade-off between a reduction in the bias and an increase in the variance. Whether or not the AveW forecast has a lower MSFE depends on the length of the single-window forecast, w_a, and the minimum window, $w_{\min}$, which are chosen by the forecaster, and the break parameters, namely the size and the distance to the break, λ and b.

Table 1 illustrates the trade-off numerically. It reports $\mathrm{MSFE}(w_a;\lambda,b) - \mathrm{MSFE}(m,w_{\min};\lambda,b)$ computed for $T = 100$, $w_m = 1$, and different values of w_a, $w_{\min}$, m, λ, and b. The top two panels report the results when the single window uses all 100 observations, $w_a = 1$. In the lower two panels, the single window equals the minimum window, $w_a = w_{\min}$. The first and third panels give the results when the windows in the AveW forecast are one observation apart, the AveW forecasts in the second and fourth panels use 10 equally spaced windows.

First, consider the top two panels. The first line in each panel shows the difference between the MSFE of the single window and that of the AveW window for $\lambda = 0$, that is, in the absence of a break. In this case, as expected, the single window outperforms the AveW forecasts. However, as λ increases, the bias reduction implied by averaging over estimation windows leads to a decrease in the MSFE of the AveW forecast relative to that of the single-window forecast. The improvement is modest for small breaks, but the difference in MSFEs increases to about a third of the variance of the innovation when the break is equal to the standard deviation of the innovation.

For the range of b considered, the benefit of averaging forecasts over estimation windows for a given $w_{\min}$ increases with b, because a larger number of sub-windows over the post-break sample are used. For the same reason, the difference in the MSFEs decreases in $w_{\min}$ when $\lambda > 0$. When $\lambda = 0$, a smaller $w_{\min}$ increases the variance of the AveW forecast due to the larger number of correlated forecasts included. The results reported in the first line of the first panel for $m = T(1-\min)+1$ and those in the first line of the second panel for $m = 10$ suggest that the variance term of the AveW forecast decreases in m. When λincreases, the reduction in the bias leads to a larger reduction in the MSFE for a smaller m. However, the size of this effect depends on b and $w_{\min}$. Overall, the numerical examples in the first two panels show that the effects of b, $w_{\min}$, and m are of second-order importance compared with the gains from averaging forecasts over estimation windows.

The bottom two panels, which compare the AveW forecast using all $T = 100$ observations and the single window of length $w_{\min}$, show that for small breaks, the forecast from the short single window has a much larger MSFE than the AveW forecast due to the large estimation uncertainty associated with the small single window. Even for larger λ, a single window that is too small leads to an inferior forecast due to the large estimation uncertainty. However, when λ is large and the single window is not too small, using only post-break data can improve the forecast. But this procedure still requires *a priori* knowledge of the break point or its estimation by means of statistical techniques.

To investigate the implications of estimating the break point for the relative performance of the two forecast procedures, we carried out a Monte Carlo experiment that compares the AveW forecast with $w_{\min} = 0.02$ to forecasts obtained from using data after the break date estimated by the sequential procedure proposed by Bai and Perron (1998, 2003). We searched for up to three break points and used the observations after the last statistically significant break date to generate one-step-ahead forecasts. We set the trimming parameter to 0.05 and the significance level to 5%, and allowed for heterogeneous covariance matrices across segments. The results were robust to varying these settings. The data were generated using model (5) with $T = 100$ and $\sigma_t = 1, \forall t$, for 10,000 replications.

The results in Table 2 show that the MSFE of the AveW forecasts is smaller than that of the forecasts based on post-break observations when $\lambda < 1$, but when $\lambda = 1$, the post-break data forecasts have a lower MSFE. This contrasts with the results in the bottom two panels of Table 1 where the post-break data forecast has a lower MSFE for $\lambda = 0.75$. The uncertainty of the time of the break leads to deterioration of the forecast precision

Table 1. $\text{MSFE}(w_a; \lambda, b) - \text{MSFE}(m, w_{\min}; \lambda, b)$: Exact results for a single break in drift

λ	b:	0.05		0.1			0.2				
	$w_{\min}$:	0.02	0.05	0.02	0.05	0.1	0.02	0.05	0.1	0.15	0.2
					$w_a = 1, m = T(1 - w_{\min}) + 1$						
0		−0.009	−0.008	−0.009	−0.008	−0.007	−0.009	−0.008	−0.007	−0.006	−0.005
0.1		−0.007	−0.006	−0.006	−0.005	−0.004	−0.005	−0.004	−0.003	−0.002	−0.002
0.2		0.001	0.000	0.005	0.005	0.004	0.007	0.008	0.008	0.007	0.007
0.4		0.030	0.024	0.047	0.043	0.035	0.056	0.054	0.051	0.047	0.041
0.75		0.127	0.105	0.186	0.170	0.140	0.218	0.210	0.196	0.178	0.156
1		0.233	0.192	0.337	0.309	0.255	0.394	0.380	0.353	0.320	0.281
					$w_a = 1, m = 10$						
0		−0.013	−0.009	−0.013	−0.009	−0.007	−0.013	−0.009	−0.007	−0.006	−0.005
0.1		−0.010	−0.007	−0.010	−0.006	−0.004	−0.009	−0.005	−0.003	−0.003	−0.002
0.2		−0.001	0.002	0.002	0.005	0.005	0.003	0.007	0.008	0.008	0.007
0.4		0.034	0.035	0.048	0.046	0.043	0.053	0.054	0.053	0.049	0.045
0.75		0.154	0.148	0.201	0.187	0.167	0.219	0.214	0.204	0.188	0.172
1		0.285	0.269	0.368	0.339	0.303	0.400	0.388	0.369	0.338	0.310
					$w_a = w_{\min}, m = T(1 - w_{\min}) + 1$						
0		0.481	0.182	0.481	0.182	0.083	0.481	0.182	0.083	0.051	0.035
0.1		0.475	0.175	0.476	0.177	0.078	0.479	0.180	0.081	0.048	0.032
0.2		0.455	0.154	0.463	0.163	0.062	0.472	0.172	0.072	0.038	0.021
0.4		0.375	0.070	0.407	0.103	−0.004	0.443	0.142	0.039	0.001	−0.022
0.75		0.109	−0.213	0.220	−0.095	−0.225	0.348	0.040	−0.074	−0.126	−0.164
1		−0.180	−0.521	0.017	−0.311	−0.465	0.244	−0.070	−0.197	−0.263	−0.319
					$w_a = w_{\min}, m = 10$						
0		0.477	0.181	0.477	0.181	0.083	0.477	0.181	0.083	0.051	0.035
0.1		0.471	0.174	0.472	0.176	0.078	0.474	0.178	0.080	0.048	0.032
0.2		0.453	0.156	0.460	0.162	0.063	0.468	0.171	0.072	0.039	0.022
0.4		0.380	0.081	0.408	0.107	0.003	0.440	0.142	0.041	0.003	−0.017
0.75		0.137	−0.170	0.236	−0.079	−0.198	0.349	0.044	−0.066	−0.116	−0.148
1		−0.128	−0.443	0.048	−0.281	−0.417	0.250	−0.062	−0.181	−0.245	−0.290

NOTE: This table reports the difference in the exact MSFE of the single-window forecast for a given w_a, and the AveW forecast with $w_m = 1$ given in (13), namely $\text{MSFE}(w_a; \lambda, b) - \text{MSFE}(m, w_{\min}; \lambda, b)$, when $T = 100$ for different numbers of estimation windows, m, break sizes as a proportion of the standard deviation of the disturbance term, λ, distance to break, b, and different minimum window sizes, $w_{\min}$.

and favors the AveW forecast, which does not use estimates of the break dates.

Now also consider a break in the error variance. For simplicity of exposition, assume that drift and volatility break at the same time. The one-step-ahead forecast error for a window of size w is given by $\xi_{T+1}(w) = \sigma^{(2)}\varepsilon_{T+1} + B_{T+1}(w) - \frac{1}{Tw}\sum_{t=T(1-w)+1}^{T} \sigma_t \varepsilon_t$. The scaled MSFE for the single-window forecast is

$$\text{MSFE}(w_a; \lambda, \kappa, b) = 1 + B^2_{T+1}(w) + \kappa^2\left(\frac{w_a - b}{Tw_a^2}\right) \times \mathrm{I}(w_a - b) + \frac{\min(w_a, b)}{Tw_a^2}, \quad (15)$$

where $\lambda = (\mu^{(2)} - \mu^{(1)})/\sigma^{(2)}$ and $\kappa = \sigma^{(1)}/\sigma^{(2)}$. Similarly, for the AveW forecasts over m estimation windows, the scaled MSFE is

$$\begin{aligned}\text{MSFE}&(m, w_{\min}; \lambda, \kappa, b)\\ &= 1 + B^2_{m,T+1}\\ &\quad + \frac{1}{m^2}\left\{\kappa^2\left[\sum_{i=1}^{m} \frac{w_i - b}{Tw_i}\mathrm{I}(w_i - b)\left(\frac{1}{w_i} + 2\sum_{j=i+1}^{m}\frac{1}{w_j}\right)\right]\right.\\ &\quad \left. + \sum_{i=1}^{m}\frac{\min(w_i, b)}{Tw_i}\left(\frac{1}{w_i} + 2\sum_{j=i+1}^{m}\frac{1}{w_j}\right)\right\}. \quad (16)\end{aligned}$$

Table 3 gives numerical examples of the difference in MSFEs when the DGP contains a break in the mean and the error variance, that is, the difference of (15) and (16). Here we concentrate on forecasts with $w_a = 1$ and $m = T(1 - w_{\min}) + 1$. The

Table 2. $\text{MSFE}(\hat{w}_a(\text{BP}); \lambda, b) - \text{MSFE}(m, w_{\min}; \lambda, b)$: Monte Carlo results for a single break in drift

$\lambda \backslash b$	0.05	0.1	0.2
0.1	0.156	0.155	0.134
0.2	0.157	0.158	0.140
0.4	0.164	0.162	0.164
0.75	0.123	0.121	0.109
1	−0.040	−0.242	−0.014

NOTE: This table reports the difference between the MSFE of the forecast based on post-break data, where the break date is estimated using the sequential procedure proposed by Bai and Perron (1998, 2003), $\text{MSFE}(\hat{w}_a(\text{BP}); \lambda, b)$, and that of the AveW forecast, namely $\text{MSFE}(m, w_{\min}; \lambda, b)$. The MSFEs are computed using Monte Carlo experiments with 10,000 replications. Data were generated using DGP (5) with $\sigma_t = 1, \forall t$, and $T = 100$. The Bai and Perron test procedure was conducted with up to three breaks, trimming of 0.05, and a 5% significance level. Forecasts were then based on observations after the last detected break. The AveW forecast used $w_{\min} = 0.02$, windows separated by one observation, and $w_m = 1$.

Table 3. $\text{MSFE}(w_a; \lambda, \kappa, b) - \text{MSFE}(m, w_{\min}; \lambda, \kappa, b)$: Exact results for a single break in drift and volatility

		$\kappa = \sigma^{(1)}/\sigma^{(2)} = 0.1$					$\kappa = \sigma^{(1)}/\sigma^{(2)} = 10$				
	b:	0.1		0.2			0.1		0.2		
λ	$w_{\min}$:	0.05	0.1	0.05	0.1	0.2	0.05	0.1	0.05	0.1	0.2
0.1		−0.005	−0.003	−0.007	−0.006	−0.003	0.010	−0.088	0.312	0.260	0.122
0.2		0.005	0.005	0.005	0.005	0.005	0.020	−0.080	0.324	0.270	0.130
0.4		0.043	0.036	0.051	0.048	0.040	0.058	−0.049	0.371	0.314	0.165
0.75		0.170	0.141	0.207	0.193	0.155	0.185	0.056	0.527	0.458	0.280
1		0.309	0.255	0.377	0.350	0.280	0.324	0.170	0.696	0.615	0.405

NOTE: This table reports the difference in the exact MSFE of the single-window forecast with $w_a = 1$ given in (15) and the AveW forecast with $w_m = 1$ given in (16), namely $\text{MSFE}(w_a; \lambda, \kappa, b) - \text{MSFE}(m, w_{\min}; \lambda, \kappa, b)$, when $T = 100$, and $m = T(1 - w_{\min}) + 1$, for different break sizes, λ, in the drift term measured in terms of $\sigma^{(2)}$, break sizes in the error variances, κ, the distance to the break, b, and different minimum window sizes, $w_{\min}$.

results depend on whether the error variance increases or decreases after the break. In the former case, the MSFEs are not much affected by the break in volatility. But the outcome is very different when the error variance decreases after the break. When the distance to the break, b, is small, many of the estimation windows in the AveW procedure cover periods of high variance, resulting in large MSFEs. However, as b increases, more of the estimation windows in the AveW procedure fall in the low-variance part of the sample, and AveW offers significant improvements over the single-window forecast.

3.1.2 Multiple Breaks in Drift. Consider a random walk model where the drift term is subject to n different breaks. Denote the break points by b_i, $i = 1, 2, \ldots, n$, such that $b_1 > b_2 > \cdots > b_n$, and let the means of the process over these segments be $\mu^{(1)}, \mu^{(2)}, \ldots, \mu^{(n+1)}$. Specifically,

$$y_t = \mu_t + \sigma \varepsilon_t \quad \text{for } t = 1, 2, \ldots, T, \tag{17}$$

such that if the sample period is mapped to the unit interval, then the mean from $t = 1$ to $t = b_1 T$ is given by $\mu^{(1)}$, the mean from $t = b_1 T + 1$ to $t = b_2 T$ is $\mu^{(2)}$, and so forth.

To simplify the analysis, first 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

$$\begin{aligned}
&\hat{y}_{T+1}(w) \\
&\quad = \frac{1}{wT} \sum_{t=T-wT+1}^{T} \sigma \varepsilon_t \\
&\qquad + \mathrm{I}(w - b_2)[1 - \mathrm{I}(w - b_1)] \left[\frac{b_2 \mu^{(3)} + (w - b_2)\mu^{(2)}}{w} \right] \\
&\qquad + [1 - \mathrm{I}(w - b_2)]\mu^{(3)} \\
&\qquad + \mathrm{I}(w - b_1) \left[\frac{b_2 \mu^{(3)} + (b_1 - b_2)\mu^{(2)} + (w - b_1)\mu^{(1)}}{w} \right].
\end{aligned}$$

The one-step-ahead forecast error is $\xi_{T+1}(w) = y_{T+1} - \hat{y}_{T+1}(w) = \mu^{(3)} + \sigma \varepsilon_{T+1} - \hat{y}_{T+1}(w)$, which, after some algebra, and noting that $\mathrm{I}(w - b_1)\mathrm{I}(w - b_2) = \mathrm{I}(w - b_1)$, can be written as

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

where $B_{T+1}(w) = \lambda^{(1)} \mathrm{I}(w - b_1)(\frac{w - b_1}{w}) + \lambda^{(2)} \mathrm{I}(w - b_2)(\frac{w - b_2}{w})$, $\lambda^{(1)} = (\mu^{(2)} - \mu^{(1)})/\sigma$, and $\lambda^{(2)} = (\mu^{(3)} - \mu^{(2)})/\sigma$.

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

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

where $\lambda^{(i)} = (\mu^{(i+1)} - \mu^{(i)})/\sigma$, and $n^{-1} \sum_{i=1}^{n} \lambda^{(i)} = (\mu^{(n+1)} - \mu^{(1)})/(n\sigma)$. For a single-window estimation with $w = 1$, the forecast bias per break will be

$$\begin{aligned}
\bar{B}_{T+1}(1) &= \frac{B_{T+1}(1)}{n} = \frac{1}{n} \sum_{i=1}^{n} \lambda^{(i)} \mathrm{I}(1 - b_i)(1 - b_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} \lambda^{(i)} (1 - b_i).
\end{aligned}$$

In contrast, the bias of the AveW forecast is

$$\bar{B}_{m,T+1} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n} \sum_{j=1}^{n} \lambda^{(j)} \left(\frac{w_i - b_j}{w_i} \right) \mathrm{I}(w_i - b_j). \tag{18}$$

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, $b_1, b_2, \ldots, b_n$, with expectations $\mathrm{E}(\lambda^{(i)}) = \bar{\lambda}$ and $\mathrm{E}(b_i) = \bar{b}$, the expected bias terms are

$$\mathrm{E}[\bar{B}_{T+1}(1)] = \bar{\lambda}(1 - \bar{b}) \qquad \text{and}$$

$$\mathrm{E}(\bar{B}_{m,T+1}) = \frac{\bar{\lambda}}{m} \sum_{j=1}^{m} \mathrm{E}[\mathrm{I}(w_j - b_i)] \frac{w_j - \bar{b}}{w_j}.$$

If we further assume that the break points are uniformly distributed over the sample [i.e., $b_i \sim \mathrm{U}(0, 1)$], then we have that $\mathrm{E}[\mathrm{I}(w_j - b_i)] = \Pr(b_i < w_j) = w_j$, and $\mathrm{E}(\bar{B}_{m,T+1}) = (\bar{\lambda}/m) \sum_{j=1}^{m} (w_j - \bar{b})$. Using (2), it is easy to show that $(1/m) \times \sum_{j=1}^{m} w_j = (1 + w_{\min})/2$, and under uniform distribution of b_i, we also have $\bar{b} = 1/2$. Thus the difference between the absolute expected bias of the single-window forecast and that of the AveW forecast is $|\mathrm{E}[\bar{B}_{T+1}(1)]| - |\mathrm{E}(\bar{B}_{m,T+1})| = |\bar{\lambda}|(1 - w_{\min})/2 \geq 0$, which increases in the absolute average break size, $|\bar{\lambda}|$, and decreases in the minimum window size, $w_{\min}$. Equality holds only when $|\bar{\lambda}| = 0$.

3.2 Break in the Slope Parameter

Now consider the more general model (1) and assume that a single break occurs in the slope parameter of the process at date, T_b, $1 < T_b < T$, whereas the error variance is constant, namely $\beta^{(1)} \neq \beta^{(2)}$, and $\sigma^{(1)} = \sigma^{(2)} = \sigma$. In this case, the conditional (on x_{T+1}) one-step-ahead forecast of y_{T+1} based on a given window of size wT is

$$\hat{y}_{T+1}(w) = \bar{y}(w) + \hat{\beta}(w)[x_{T+1} - \bar{x}(w)], \tag{19}$$

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

$$\hat{\beta}(w) = \frac{\sum_{t=T(1-w)+1}^{T}[y_t - \bar{y}(w)][x_t - \bar{x}(w)]}{\sum_{t=T(1-w)+1}^{T}[x_t - \bar{x}(w)]^2}.$$

Under the assumption that x_t is a covariance stationary process with mean μ_x and absolute summable autocovariances, $\sum_{s=0}^{\infty}|\gamma_x(s)| < K < \infty$, we have

$$\bar{x}(w) - \mu_x = O_p\left(\frac{1}{\sqrt{Tw}}\right). \tag{20}$$

Similarly,

$$\bar{y}(w) - \mu_y = O_p\left(\frac{1}{\sqrt{Tw}}\right), \tag{21}$$

see Appendix A. The estimate of the slope coefficient can be written as

$$\hat{\beta}(w) = \frac{\sum_{t=T(1-w)+1}^{T}\beta_t(x_t - \mu_x)^2}{\sum_{t=T(1-w)+1}^{T}(x_t - \mu_x)^2} + \frac{\sum_{t=T(1-w)+1}^{T}(x_t - \mu_x)\varepsilon_t}{\sum_{t=T(1-w)+1}^{T}(x_t - \mu_x)^2} + O_p\left(\frac{1}{\sqrt{Tw}}\right), \tag{22}$$

where the first term on the right side of (22) can be rewritten as

$$\frac{\sum_{t=T(1-w)+1}^{T}\beta_t(x_t - \mu_x)^2}{\sum_{t=T(1-w)+1}^{T}(x_t - \mu_x)^2} = \beta^{(2)} + \left(\beta^{(1)} - \beta^{(2)}\right)\left(\frac{w-b}{w}\right)\mathrm{I}(w-b)\theta(\mathbf{x}, w, b),$$

where

$$\theta(\mathbf{x}, w, b) = \frac{[T(w-b)]^{-1}\sum_{t=T(1-w)+1}^{T(1-b)}(x_t - \mu_x)^2}{(Tw)^{-1}\sum_{t=T(1-w)+1}^{T}(x_t - \mu_x)^2} > 0 \tag{23}$$

and $\mathbf{x} = (x_1, x_2, \ldots, x_T)'$. Conditional on $\mathbf{x}$ and x_{T+1} the bias in estimating $\beta^{(2)}$ by $\hat{\beta}(w)$ using the estimation window, w, is given by

$$B_{T+1}(w) = \left(\beta^{(1)} - \beta^{(2)}\right)\left(\frac{w-b}{w}\right)\mathrm{I}(w-b)\theta(\mathbf{x}, w, b). \tag{24}$$

In general, $\theta(\mathbf{x}, w, b)$ varies with the particular set of the regressors realized over the estimation window. To simplify the analysis, we can replace $\theta(\mathbf{x}, w, b)$ by its mean computed with respect to the assumed distribution of the regressors. When $x_t \sim$ iid $\mathrm{N}(0, \sigma_x^2)$, using the results of Pesaran and Timmermann (2007, app. C), we have that $\mathrm{E}[\theta(\mathbf{x}, w, b)] = 1$. Simulations not reported here but available from the authors show that this is true for a range of distributions for x_t. In what follows, we work with $\theta(\mathbf{x}, w, b) \approx 1$. In this case, it can be seen from (24) that the bias is proportional to the size of the break, $(\beta^{(1)} - \beta^{(2)})$, and the proportion of pre-break observations in the sample, $(w-b)/w$.

Lemma 1. Denote the forecast error based on a single fixed estimation window, $w \in [w_{\min}, 1]$, and a given break point $b \in (0, 1)$, by $\xi_{T+1}(w) = y_{T+1} - \hat{y}_{T+1}(w)$, where y_{T+1} is defined by the DGP in model (1) and $\hat{y}_{T+1}(w)$ is given by (19). Define $\lambda = (\beta^{(2)} - \beta^{(1)})/\sigma$. Then, conditional on x_{T+1}, for fixed w and b, the (scaled) forecast error is

$$\sigma^{-1}\xi_{T+1}(w) = \varepsilon_{T+1} + \left(\frac{w-b}{w}\right)\mathrm{I}(w-b)\lambda^2(x_{T+1} - \mu_x) + O_p\left(\frac{1}{\sqrt{Tw}}\right). \tag{25}$$

Using the foregoing result, we also note that $\sigma^{-1}\xi_{T+1}(b) = \varepsilon_{T+1} + O_p(1/\sqrt{Tb})$. Now consider the forecast based on averaging the forecasts over the different windows, $w_1, w_2, \ldots, w_m$,

$$\hat{y}_{m,T+1} = \frac{1}{m}\sum_{i=1}^{m}\hat{y}_{T+1}(w_i). \tag{26}$$

It follows that the error of the AveW forecast is $\xi_{m,T+1} = \frac{1}{m}\sum_{i=1}^{m}\xi_{T+1}(w_i)$.

Lemma 2. Suppose that the DGP in (1) holds with β_t subject to a single break. Consider the forecast error of the AveW forecasts based on m estimation windows, defined by (26) and (19). Let $\zeta(w_i) = [(w_i - b)/w_i]\mathrm{I}(w_i - b)$, and $\lambda = (\beta^{(2)} - \beta^{(1)})/\sigma$. Then, conditional on x_{T+1}, for fixed m, $w_{\min}$ and given b as $T \to \infty$, the scaled AveW forecast error is

$$\sigma^{-1}\xi_{m,T+1} = \varepsilon_{T+1} + B_{m,T+1} + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{27}$$

where

$$B_{m,T+1} = \lambda(x_{T+1} - \mu_x)\left[\frac{1}{m}\sum_{i=1}^{m}\zeta(w_i)\right]. \tag{28}$$

We are now in a position to compare the MSFE of the standard forecasts based on a single window with the AveW forecasts. First, consider the case where b is fixed as $T \to \infty$.

Proposition 2. Consider the DGP given by (1) with a single break in β_t. For large T but a fixed b such that $W_b \to \infty$, the MSFE of the forecast from a single window of length b will be unbiased and will have the lowest MSFE.

This follows directly from the arguments of Bai (1997). Clearly, under such circumstances, averaging over estimation windows will not improve the forecast accuracy.

However, our focus is on the case where W_b remains small as $T \to \infty$. In this case, the forecast using only post-break data will still be unbiased, but the terms of order $O_p(\frac{1}{\sqrt{W_b}})$ will be large when W_b is small, and the variance of the forecast error might be quite high. As shown by Pesaran and Timmermann (2007), in such circumstances a larger estimation window might be more appropriate. Accordingly, in what follows we compare a single-window forecast with window size $w_a > b$ to the AveW forecast based on m windows starting with w_1 and ending with $w_m = w_a$. In this setup, we have

$$\sigma^{-1}\left(\xi_{T+1}(1) - \xi_{m,T+1}\right) = \lambda(x_{T+1} - \mu_x)\left[\zeta(w_a) - \frac{1}{m}\sum_{i=1}^{m}\zeta(w_i)\right] + O_p\left(\frac{1}{\sqrt{T}}\right). \quad (29)$$

Proposition 3. Suppose that the DGP in (1) holds and is subject to a single break in β_t at b. For large T but a small W_b, the MSFE of the forecast from a single window of length $w_a > b$ will be larger than that of the AveW forecast with $w_m = w_a$ and a fixed number of windows, $m > 1$.

This follows because the difference in square brackets in (29) is positive, which follows from Proposition 1.

4. FORECASTS FROM TIME–VARYING PARAMETER MODELS

As an alternative to averaging forecasts over estimation windows, we consider time-varying parameter models. Recently, Branch and Evans (2006) considered a number of variations on this class of models and showed that a particularly simple form, known as the "constant gain least squares," works reasonably well in forecasting U.S. inflation and GDP growth.

Constant gain least squares is equivalent to discounting past observations at a geometric rate, γ (Branch and Evans 2006, p. 160). To analyze this forecasting method, we return to the simple model (5) with a break in mean. We denote the constant gain least squares or exponential smoothing (ExpS) forecast by

$$\hat{y}_{T+1}(\gamma) = \left(\frac{1-\gamma}{1-\gamma^T}\right)\sum_{j=1}^{T}\gamma^{T-j}y_j. \quad (30)$$

Now consider the case where the mean of y_t is subject to a single break at date T_b, $1 < T_b < T$, with $\mu^{(1)} \neq \mu^{(2)}$ and $\sigma^{(1)} = \sigma^{(2)} = \sigma$. The bias of the one-step-ahead forecast error is $\mathrm{Bias}[\hat{y}_{T+1}(\gamma)] = (\mu^{(2)} - \mu^{(1)})(\frac{\gamma^{T-T_b+1}-\gamma^T}{1-\gamma^T})$ (Pesaran and Pick 2008). Because $0 < \gamma < 1$, the sign of the forecast bias is the same as the sign of $(\mu^{(2)} - \mu^{(1)})$. The forecast error variance is given by $\mathrm{Var}[\xi_{T+1}(\gamma)] = \sigma^2[1 + (\frac{1-\gamma}{1-\gamma^T})^2(\frac{1-\gamma^{2T}}{1-\gamma^2})]$. It is interesting to note that for all values of $\gamma \in (0, 1)$, the sampling variance of the forecast error (the second part in square brackets) does not vanish even for T sufficiently large. Therefore, the exponential down-weighting of the past observations can work only through bias reduction.

The scaled one-step-ahead MSFE in then given by

$$\mathrm{MSFE}(\gamma; \lambda, b) = 1 + \lambda^2\left(\frac{\gamma^{1+Tb} - \gamma^T}{1-\gamma^T}\right)^2 + \left(\frac{1-\gamma}{1-\gamma^T}\right)^2\left(\frac{1-\gamma^{2T}}{1-\gamma^2}\right), \quad (31)$$

where $\lambda = |\mu^{(2)} - \mu^{(1)}|/\sigma$. It can be shown that for a sufficiently large T, there is a unique γ that minimizes the MSFE. However, choosing the optimal down-weighting parameter γ will depend on λ and b, which typically are unknown.

Table 4 gives a numerical illustration of the difference in the MSFE of the ExpS forecast and that of the AveW forecast, where the AveW forecast uses estimation windows one observation apart. The ExpS forecasts are based on two different choices of the down-weighting parameter, namely $\gamma = 0.95$ and 0.99. The results suggest that whereas b and $w_{\min}$ have some influence on the final outcomes, the choice of the down-weighting parameter dominates the results. When $\gamma = 0.95$, the AveW forecast has a lower MSFE for small breaks, whereas the ExpS forecast has a lower MSFE for larger breaks. This comparison is reversed when $\gamma = 0.99$.

To understand these numerical results, we can express the AveW model as a "forgetting factor" model. Forgetting factor models weigh observations $\{y_t\}_{t=1}^T$ by factors $\{k_{T-t}\}_{t=1}^T$ (Hannan and Deistler 1988; Brailsford, Penm, and Terrell 2002). The ExpS model fits naturally into this framework. Using (3) and (4), the AveW forecast can be expressed as

$$\hat{y}_{m,T+1} = \frac{1}{T(1-w_{\min})+1} \times \sum_{i=1}^{T(1-w_{\min})+1}\frac{1}{Tw_{\min}+i-1}\sum_{t=T(1-w_{\min})-i+2}^{T}y_t,$$

where we use the AveW forecast with windows increasing by one observation. Thus each observation y_t, $t = 1, 2, \ldots, T$, receives the weight

$$k(T, t, w_{\min}) = \frac{1}{T(1-w_{\min})+1} \times \sum_{i=1}^{t}\frac{1}{T+1-i}\mathrm{I}[T(1-w_{\min})+1-i]. \quad (32)$$

Table 4. $\mathrm{MSFE}(\gamma; \lambda, b) - \mathrm{MSFE}(m, w_{\min}; \lambda, b)$: Exact results for a single break in drift

		$\gamma = 0.95$					$\gamma = 0.99$				
	b:	0.1		0.2			0.1		0.2		
λ	$w_{\min}$:	0.05	0.1	0.05	0.1	0.2	0.05	0.1	0.05	0.1	0.2
0.1		0.006	0.007	0.007	0.008	0.009	−0.005	−0.004	−0.005	−0.004	−0.003
0.2		0.001	0.000	0.003	0.003	0.001	0.001	0.000	0.003	0.003	0.001
0.4		−0.020	−0.027	−0.014	−0.017	−0.028	0.026	0.018	0.031	0.028	0.018
0.75		−0.089	−0.119	−0.070	−0.085	−0.125	0.108	0.078	0.127	0.112	0.072
1		−0.164	−0.219	−0.131	−0.158	−0.230	0.197	0.143	0.231	0.203	0.132

NOTE: This table reports the difference in the exact MSFE of the ExpS forecast given in (31) and the AveW forecast with $w_m = 1$ given in (12), namely $\mathrm{MSFE}(\gamma; \lambda, b) - \mathrm{MSFE}(m, w_{\min}; \lambda, b)$, when $T = 100$, $m = T(1 - w_{\min}) + 1$, for different break sizes, λ, defined as a proportion of the standard deviation of the disturbance term, the proportion of post-break data, b, the minimum window sizes, $w_{\min}$, and the down-weighting parameter, γ.

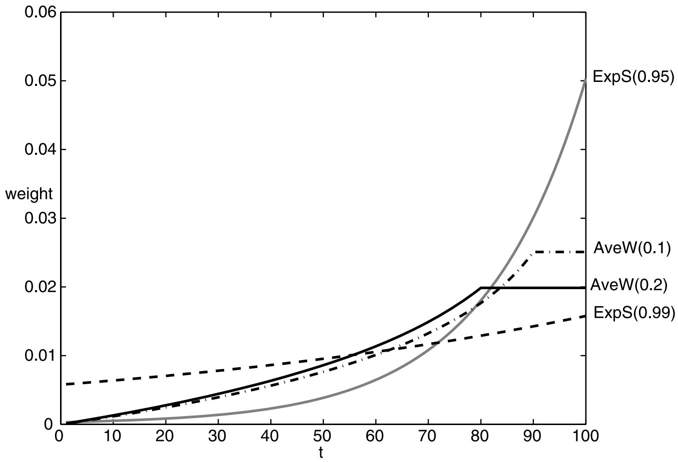

Figure 1. Weights attached to observations in AveW and ExpS forecasts for $T = 100$. This figure plots the weights attached to each observation in a sample of size $T = 100$. The numbers in brackets are the minimum window, $w_{\min}$, in the case of the AveW weights and the down-weighting parameter, γ, in the case of the ExpS weights.

Figure 1 plots the weights attached to each observation in AveW and ExpS forecasts using the minimum windows and down-weighting parameters used in the foregoing numerical example. We consider two different choices of the minimum windows, $w_{\min} = 0.1$ and 0.2, in constructing the weights for the AveW forecast. The weights implied by the AveW forecasts vary much less than the weights implied by the ExpS forecasts. When $\gamma = 0.99$, the observations are weighted more evenly than the weights of AveW for both minimum windows, but when $\gamma = 0.95$, past observations are discounted much more heavily. This largely explains the results in Table 4.

5. APPLICATIONS TO FINANCIAL TIME SERIES

We now consider the application of the AveW forecasting procedure to weekly returns on futures contracts for 20 equity indices. Our sample ends on November 24, 2008, and thus covers the highly volatile episodes associated with the credit crunch. Details of the data are given in Appendix B.

We recursively compute one-week-ahead forecasts using various forecasting methods for the mean model (5). The baseline forecast uses the observations after the last break identified by the sequential procedure of Bai and Perron (1998, 2003), designated BP, where we search for up to eight breaks and set the trimming parameter to 0.1 and the significance level to 5%. Whereas Pesaran and Timmermann (2007) showed that forecast accuracy can be improved by using some pre-break observations, we use only post-break observations because this is the more common procedure followed in practice, and exploiting the bias-variance trade-off requires knowledge of the break size, which would introduce further complications into the comparative forecasting exercise.

We compare the BP post-break forecasts with two versions of the AveW forecasts. The first version averages forecasts from sub-windows within a rolling window of 156 weeks (equal to three years) using $w_{\min} = 0.1$. This yields $W_{\min} = 15$. The second AveW forecast averages forecasts from sub-windows in an expanding window using the same number of minimum observations, $W_{\min} = 15$. We use $m = 10$ windows. The results are qualitatively similar when a larger number of estimation windows is used. We also included forecasts from expanding and rolling windows in our comparisons. For the rolling windows, we considered a rolling window of size $W_a = 156$ and a minimum rolling window of size $W_{\min} = 15$. Also, as it could be argued that the AveW forecasts are performing better because they are effectively based on a smaller average window (compared with W_a), we considered a third rolling-window forecast based on an (average) effective window size of $\overline{W} = 85$, computed as the integer part of $W_a(1/10 + 2/10 + \cdots + 10/10)/10$. Finally, we computed ExpS forecasts using two down-weighting parameters, $\gamma = 0.95$ and 0.99.

For each series, we calculate the absolute bias, the RMSFE, and tests for predictive performance of Diebold and Mariano (1995). More precisely, $\text{RMSFE} = (\frac{1}{n}\sum_{t=1}^{n}\xi_t^2)^{1/2}$, where $\xi_t = y_{t+1} - \hat{y}_{t+1|t}$, 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. We also report the RMSFE and the relative RMSFE; that is for, say, the AveW($W_{\min}$) forecast, we report RMSFE[AveW($W_{\min}$)]/ RMSFE(BP), where BP denotes the forecast from the baseline forecast using the observations after the break date estimated by the Bai and Perron procedure. Values smaller than 1 indicate that the baseline forecast has a larger RMSFE than the AveW forecast. The Diebold–Mariano test statistics for predictive ability are calculated for the loss differential $l_t(A, B) = \xi_{tA}^2 - \xi_{tB}^2$, where ξ_{tA} and ξ_{tB} are the forecast errors for two forecast methods, A and B.

The results are reported in Table 5. The first line reports the (absolute) average bias ($\times 100$) across the 20 time series, the second line gives the results for the average RMSFE ($\times 100$), and the third line presents RMSFE as a ratio of the RMSFE from the forecasts based on the post-break observations. The lower panel of Table 5 shows the fraction of series where the test of Diebold and Mariano (1995) rejects equal predictive accuracy and the forecast method in the respective column has the lower RMSFE.

The results indicate that the forecasts based on the post-break sample have a smaller average bias than the AveW forecasts but that the average RMSFE is larger than that of the AveW forecasts. Using DM tests, we find that the AveW forecasts are statistically significantly more accurate in 40% of the series when the AveW forecasts are computed within rolling windows and in 45% of the series if the AveW forecasts are based on expanding windows.

Comparing the AveW forecasts with the forecasts based on the corresponding single windows, we find that the AveW forecasts have a lower bias and RMSFE, as predicted by our theory. In contrast, the forecasts from the single rolling window of length $W_{\min}$ have a lower bias than the AveW forecasts, because they are less likely to include breaks in the estimation window. However, due to the small number of observations used in the estimation, the RMSFE is larger that that of the AveW forecasts. The AveW forecasts are significantly more accurate in about half of the series, whereas the short single rolling window is never significantly more accurate than the AveW forecasts. Comparing the AveW forecasts with the forecasts based on rolling windows of size $\overline{W}$ shows that averaging over the

Table 5. Predictive accuracy for alternative forecasts of returns of 20 equity index futures

	BP	AveW($W_{\min}$)		Expanding	Rolling windows			ExpS(γ)	
	post-break	Rolling	Expanding	windows	$W_{\min}=15$	$\overline{W}=85$	$W_a=156$	$\gamma=0.95$	0.99
Averages									
Bias	1.668	1.874	1.896	2.108	1.065	2.103	2.054	1.460	1.887
RMSFE	63.546	61.483	61.531	61.602	62.661	61.512	61.707	61.765	61.530
Rel. RMSFE	1	0.968	0.969	0.970	0.987	0.968	0.971	0.972	0.969
Diebold–Mariano tests									
Post-break	–	0.40	0.45	0.35	0.00	0.30	0.25	0.15	0.50
AveW: rolling	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AveW: expand.	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00	0.00
Expanding	0.00	0.00	0.00	–	0.00	0.00	0.00	0.00	0.00
Rolling $W_{\min}$	0.00	0.50	0.40	0.20	–	0.45	0.20	0.70	0.40
Rolling $\bar{W}$	0.00	0.00	0.00	0.00	0.00	–	0.00	0.00	0.15
Rolling W_a	0.00	0.00	0.10	0.05	0.00	0.05	–	0.00	0.00
ExpS($\gamma=0.95$)	0.00	0.10	0.05	0.00	0.00	0.05	0.05	–	0.05
ExpS($\gamma=0.99$)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	–

NOTE: The forecast methods are (i) using the observations after the last break point estimated by the procedure of Bai and Perron (1998, 2003); AveW forecasts with the minimum number of observations $W_{\min}=15$ weeks and $m=10$ sub-windows within (ii) a rolling window of length $W_a=156$ weeks and (iii) an expanding window; (iv) an expanding window; single rolling windows of size (v) $W_{\min}=15$, (vi) $\overline{W}=85$, (vii) $W_a=156$ weeks; ExpS forecasts with (viii) $\gamma=0.95$ and (ix) $\gamma=0.99$. The results in the top panel are the absolute average of the bias across the 20 time series, those in the second row are the average of the RMSFE, and those in the third row are the average of the RMSFE as a ratio of the average RMSFE of the post-break window forecast. The results are multiplied by 100. The lower panel reports the proportion of rejection of predictive accuracy using the test of Diebold and Mariano (1995) across the 20 series. We report the fraction of the series where equal forecast accuracy was rejected and the forecasting method in the respective column had the lower RMSFE than the forecasting method in the respective row. Details of the data are given in Appendix B.

different sub-windows leads to a reduction in bias beyond the implied reduction in sample size. The average RMSFE is reduced even if this difference is not statistically significant.

The ExpS forecast with $\gamma=0.95$, which discounts past observations at a faster rate compared with the ExpS forecasts with $\gamma=0.99$, has a lower average bias than the AveW forecasts and—with the exception of the shortest rolling window—all other forecast procedures. However, the rapid discounting leads to a larger RMSFE than the AveW forecasts and all other forecasting procedures with the exception of the shortest rolling window and the post-break window forecast. The ExpS forecast with $\gamma=0.99$ has a smaller average bias than the AveW forecast within the expanding window and most of the other forecast methods but a larger bias than the AveW forecast within the rolling window. Although the RMSFE is larger than that of the AveW forecasts within the rolling window, it is smaller than that of most other forecast methods.

Overall, it appears that the large variances of the series relative to the size of possible breaks implies that break points are difficult to estimate and forecasts based on such estimates are less precise. Equally, using only short rolling windows increases the estimation uncertainty, which eliminates the benefits from the reduction in forecast bias. The same is true of down-weighting observations when the weights decay too rapidly. Using more slowly decaying weights tends to improve forecast accuracy in the MSFE sense. Overall, for the data considered here, the best results are obtained from averaging forecasts over estimation windows within a rolling window.

6. CONCLUSION

We have shown that averaging forecasts over estimation windows reduces the forecast bias and, despite a potential increase in the variance, reduces the MSFE for all but the smallest breaks. We have also compared it with the forecast obtained from exponential down-weighting of past observations. Both can be cast in the framework of forgetting factor models. However, the exponential smoothing forecast is more sensitive to the down-weighting parameter than the averaged forecast is to the choice of the minimum estimation window. Monte Carlo results and the application to time series of returns on equity futures show that averaging forecasts over estimation windows can improve forecast accuracy compared with forecasts from post-break samples when the variance of the process is relatively large compared with the break size. Averaging of forecasts over different estimation windows offers a simple approach to generating forecasts that are reasonably robust to structural breaks of unknown break dates and sizes. It is likely to be particularly effective when the last break date is relatively close to the point of the forecast and the break is of moderate magnitude. Although our theoretical analysis has been confined to point forecasts for random walk and linear regression models, averaging forecasts over estimation windows is likely to improve forecast accuracy in many settings, such as richer models or density forecasts, but we leave these topics for future research.

APPENDIX A: MATHEMATICAL APPENDIX

Proof of Proposition 1

Denote $\zeta(w_i)=[(w_i-b)/w_i]\mathrm{I}(w_i-b)$, and note that $\zeta(w_i)\geq 0, \forall w_i$. Furthermore, because $\zeta(w_i)$ is increasing in w_i, $\zeta(w_a)\geq\zeta(w_i), \forall w_i\leq w_a$. Therefore $\zeta(w_a)=\frac{1}{m}\sum_{i=1}^{m}\zeta(w_a)\geq(1/m)\times\sum_{i=1}^{m}\zeta(w_i)$. Strict equality holds if one element of the last term contains at least one window for which $w_i<w_a$.

Asymptotic Equivalence of $\bar{x}(w)$ and μ_x

Under the assumptions regarding x_t in Section 3.2, $\lim_{T\to\infty}\{Tw[\bar{x}(w)-\mu_x]^2\} = \sum_{s=-\infty}^{\infty}\gamma_x(s)$, and for a given $w \in (w_{\min},1)$, $w_{\min} > 0$, $\lim_{T\to\infty}\{T[\bar{x}(w)-\mu_x]^2\} = [\sum_{s=-\infty}^{\infty}\gamma_x(s)]/w = [2\pi f_x(0)]/w$, where $f_x(0)$ is the spectral density of $\{x_t\}$ evaluated at zero frequency. Using the results of propositions 7.5 and 7.11 of Hamilton (1994), then $\sqrt{T}[\bar{x}(w)-\mu_x] \xrightarrow{L} \mathrm{N}(0, \frac{2\pi f_x(0)}{w})$, where $\xrightarrow{L}$ denotes convergence in distribution. Thus $\bar{x}-\mu_x = O_p(1/\sqrt{Tw})$.

Asymptotic Equivalence of $\bar{y}(w)$ and μ_y

Using (1) with $\sigma^{(1)}=\sigma^{(2)}=\sigma$ we have

$$\begin{aligned}\bar{y}(w) &= \mu_y + \frac{1}{Tw}\sum_{t=T(1-w)+1}^{T}\beta_t(x_t-\mu_x) + \frac{1}{Tw}\sum_{t=T(1-w)+1}^{T}\sigma\varepsilon_t\\ &= \mu_y + \frac{1}{Tw}I(w-b)\sum_{t=T(1-w)+1}^{T(1-b)}\beta^{(1)}(x_t-\mu_x)\\ &\quad + \frac{1}{Tw}\sum_{t=T(1-b)+1}^{T}\beta^{(2)}(x_t-\mu_x) + \frac{1}{Tw}\sum_{t=T(1-w)+1}^{T}\sigma\varepsilon_t\\ &= \mu_y + \beta^{(1)}I(w-b)\left(\frac{w-b}{w}\right)\bar{u}(w-b)\\ &\quad + \beta^{(2)}\frac{b}{w}\bar{u}(b) + \sigma\bar{\varepsilon}(w),\end{aligned}$$

where $\bar{u}(w-b) = [T(w-b)]^{-1}\sum_{t=T(1-w)+1}^{T(1-b)}u_t$, $\bar{u}(b) = (Tb)^{-1}\sum_{t=T(1-b)+1}^{T}u_t$, $u_t = x_t-\mu_x$, and $\bar{\varepsilon}(w) = (Tw)^{-1}\times\sum_{t=T(1-w)+1}^{T}\varepsilon_t$. Therefore, using the results for x_t above, we have that $[(w-b)/w]\bar{u}(w-b) = O_p(1/\sqrt{Tw})$, and $(b/w)\bar{u}(b) = O_p(1/\sqrt{Tw})$, and similarly (because ε_t is serially uncorrelated with a finite variance) $\bar{\varepsilon}(w) = O_p(1/\sqrt{Tw})$, which yields the result in (21).

Derivation of $\hat{\beta}(w)$ in (22)

Consider first the denominator of $\hat{\beta}(w)$,

$$\begin{aligned}&\frac{1}{Tw}\sum_{T(1-w)+1}^{T}[x_t-\bar{x}(w)]^2\\ &\quad= \frac{1}{Tw}\sum_{T(1-w)+1}^{T}(x_t-\mu_x)^2 - [\mu_x-\bar{x}(w)]^2\\ &\quad= \frac{1}{Tw}\sum_{T(1-w)+1}^{T}(x_t-\mu_x)^2 + O_p\left(\frac{1}{Tw}\right),\end{aligned}$$

where the last equality follows from the foregoing arguments. Therefore, by Slutsky's theorem, $\{\frac{1}{Tw}\sum_{T(1-w)+1}^{T}[x_t-\bar{x}(w)]^2\}^{-1} = [\frac{1}{Tw}\sum_{T(1-w)+1}^{T}(x_t-\mu_x)^2]^{-1} + O_p(1/Tw)$. For the numerator,

$$\begin{aligned}&\sum_{T(1-w)+1}^{T}y_t[x_t-\bar{x}(w)]\\ &= \sum_{T(1-w)+1}^{T}[\mu_y+\beta(x_t-\mu_x)+\sigma\varepsilon_t]\\ &\quad\times\left\{(x_t-\mu_x)+[\mu_x-\bar{x}(w)]\right\}\\ &= \sum_{T(1-w)+1}^{T}\beta_t(x_t-\mu_x)^2\\ &\quad+[\mu_x-\bar{x}(w)]\sum_{T(1-w)+1}^{T}\beta_t(x_t+\mu_x)\\ &\quad+[\mu_x-\bar{x}(w)]\sum_{T(1-w)+1}^{T}\sigma\varepsilon_t + \sum_{T(1-w)+1}^{T}\sigma\varepsilon_t(x_t-\mu_x).\end{aligned}$$

Let $u_t = x_t-\mu_x$, and note that x_t is assumed to be exogenous with respect to $\varepsilon_{t'}$ for all t and t', and the break point of β_t is also exogenously given, and thus given independently of ε_t and x_t. Then $\mathrm{Var}[1/(Tw)\sum_{T(1-w)+1}^{T}(\sigma\varepsilon_t+\beta_t u_t)] = \sigma^2/(Tw) + 1/(Tw)\,\mathrm{Var}(\sum_{T(1-w)+1}^{T}\beta_t u_t)$. Given that u_t is stationary, and because $|\beta_t| < K < \infty$, we have $(Tw)^{-1}\sum_{T(1-w)+1}^{T}(\sigma\varepsilon_t + \beta_t u_t) = O_p(1/\sqrt{Tw})$, and the result in (22) follows.

Proof of Lemma 1

Rewrite (19) as $\hat{y}_{T+1}(w) = \bar{y}(w) + \hat{\beta}(w)(x_{T+1}-\mu_x) + \hat{\beta}(w)[\mu_x-\bar{x}(w)]$ then, using the results in (20), (21), (22), and (24), the forecast error can be written as

$$\begin{aligned}&\xi_{T+1}(w)\\ &\quad= \sigma\varepsilon_{T+1} + \left[\beta^{(2)}-\hat{\beta}(w)\right](x_{T+1}-\mu_x) + O_p\left(\frac{1}{\sqrt{Tw}}\right)\\ &\quad= \sigma\varepsilon_{T+1} + \frac{w-b}{w}\mathrm{I}(w-b)\sigma\lambda(x_{T+1}-\mu_x)\\ &\qquad+\sigma\frac{\sum_{t=T(1-w)+1}^{T}u_t\varepsilon_t}{\sum_{t=T(1-w)+1}^{T}u_t^2}(x_{T+1}-\mu_x) + O_p\left(\frac{1}{\sqrt{Tw}}\right) \qquad (\mathrm{A}.1)\end{aligned}$$

With x_t being exogeneous, u_t and ε_t are uncorrelated and (25) follows, noting that $\sum_{t=T(1-w)+1}^{T}u_t\varepsilon_t/\sum_{t=T(1-w)+1}^{T}u_t^2 = O_p(1/\sqrt{Tw})$. Using (A.1), the squared forecast error is $\xi_{T+1}^2(w) = [\sigma\varepsilon_{T+1} + \frac{w-b}{w}\mathrm{I}(w-b)\sigma\lambda(x_{T+1}-\mu_x)]^2 + O_p(1/\sqrt{Tw})$.

Proof of Lemma 2

$$\begin{aligned}\xi_{m,T+1} &= \frac{1}{m}\sum_{i=1}^{m}\left\{\mu_y+\bar{y}(w_i)+\left[\beta^{(2)}-\hat{\beta}(w_i)\right](x_{T+1}-\mu_x)\right.\\ &\qquad\left.+\hat{\beta}(w_i)[\mu_x-\bar{x}(w_i)]+\sigma\varepsilon_{T+1}\right\}\\ &= \sigma\varepsilon_{T+1} + \sigma\lambda\frac{x_{T+1}-\mu_x}{m}\sum_{i=1}^{m}\mathrm{I}(w_i-b)\frac{w_i-b}{w_i}\\ &\qquad+\frac{1}{m}\sum_{i=1}^{m}\left\{\mu_y-\bar{y}(w_i)+\hat{\beta}(w_i)[\mu_x-\bar{x}(w_i)]\right\}.\end{aligned}$$

The first term does not vary with m. The second term relates to the forecast bias and is bounded in m. Now consider the last term as $T \to \infty$, for either a fixed m or as $m \to \infty$. Using (20), (21), and (22), and after some algebra (noting that $w_1 = w_{\min} < b$), we have

$$\left| \frac{1}{m} \sum_{i=1}^{m} \{ \mu_y - \bar{y}(w_i) + \hat{\beta}(w_i)[\mu_x - \bar{x}(w_i)] \} \right| < \frac{K_1}{m\sqrt{T}} \sum_{i=1}^{m} \frac{1}{\sqrt{w_i}} + \frac{K_2}{m\sqrt{T}} \sum_{i=1}^{m} \frac{1}{\sqrt{w_i}} \left(\frac{w_i - b}{w_i} \right) \mathrm{I}(w_i - b),$$

where K_1 and K_2 are positive constants. Also, $m^{-1} \times \sum_{i=1}^{m} w_i^{-1/2} < w_{\min}^{-1/2}$, and, noting that $w^{-3/2}(w-b)$ is maximized at $w^* = 3b$, $m^{-1} \sum_{i=1}^{m} w_i^{-1/2}((w_i - b)/w_i)\mathrm{I}(w_i - b) < 2/(3\sqrt{3b})$. Therefore, for $w_{\min} > 0$, $\xi_{m,T+1}$ is bounded in m irrespective of whether m is fixed as $T \to \infty$, or if $m \to \infty$ as $T \to \infty$.

APPENDIX B: EQUITY INDEX FUTURES AND SAMPLE PERIODS

The equity series refer to futures contracts obtained from Datastream and cover the different periods as set out below. The start of the samples generally coincides with the start dates of the futures markets in question. The last number in the brackets is the number of forecasts.

AEX, Amsterdam exchange index, Netherlands (01-Jun-1989 to 24-Nov-2008; 864); *ASX*, Australian securities exchange index (06-Dec-2000 to 19-Nov-2008; 279); *BEL*, BEL 20 index, Belgium (07-Jun-1994 to 24-Nov-2008; 603); *CAC*, CAC 40 index, France (24-Mar-1989 to 24-Nov-2008; 868); *DAX*, DAX 30 index, Germany (02-Jul-1991 to 24-Nov-2008; 753); *DJE*, DJ EURO STOXX 50, DJ euro index (27-Jan-1999 to 25-Nov-2008; 375); *FTSE*, FTSE 100, U.K. (09-Aug-1985 to 19-Nov-2008; 1054); *FOX*, FOX index, Finland (02-May-2000 to 19-Nov-2008; 283); *IBE*, IBEX 35, Spain (25-Nov-1992 to 24-Nov-2008; 672); *KFX*, KFX index, Denmark (14-Aug-2001 to 25-Nov-2008; 233); *MIB*, Milan index, Italy (04-Jul-1995 to 20-Nov-2008; 551); *ND*, NASDAQ 100 index, U.S.A. (14-Nov-1996 to 21-Nov-2008; 480); *NK*, NIKKEI 225, Japan (30-Apr-1987 to 20-Nov-2008; 938); *OBX*, OBX index, Norway (26-Aug-1999 to 24-Nov-2008; 326); *OMX*, OMX index, Sweden (17-Sep-1990 to 19-Nov-2008; 783); *PSI*, PSI 20 index, Portugal (27-Jan-1997 to 24-Nov-2008; 463); *SP*, S&P COMP index, U.S.A. (09-Aug-1985 to 19-Nov-2008; 1050); *SMI*, SWISS MI index, Switzerland (18-Jun-1991 to 20-Nov-2008; 766); *TPX*, Topix stock price index, Japan (06-Sep-1988 to 19-Nov-2008; 846); *TSX*, Toronto stock exchange index, Canada (12-Apr-2000 to 20-Nov-2008; 308).

ACKNOWLEDGMENTS

This article was previously circulated under the title “Forecasting Random Walks under Drift Instability.” The authors thank Todd Clark, Jiangqin Fan, (the late) Clive Granger, Alessio Sancetta, Christoph Schleicher, Ron Smith, Vanessa Smith, Jim Stock, two anonymous referees and an associate editor for helpful comments. This article was written while the second author was a 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 article do not necessarily reflect those of De Nederlandsche Bank.

[Received January 2009. Revised February 2010.]

REFERENCES

Assenmacher-Wesche, K., and Pesaran, M. H. (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. [307]

Bai, J. (1997), “Estimation of a Change Point in Multiple Regression Models,” *Review of Economics and Statistics*, 79, 551–563. [308,313]

Bai, J., and Perron, P. (1998), “Estimating and Testing Linear Models With Multiple Structural Changes,” *Econometrica*, 66, 47–78. [307,308,310,311, 315,316]

——— (2003), “Computation and Analysis of Multiple Structural Change Models,” *Journal of Applied Econometrics*, 18, 1–22. [307,308,310,311, 315,316]

Brailsford, T. J., Penm, J. H. W., and Terrell, R. D. (2002), “Selecting the Forgetting Factor in Subset Autoregressive Modelling,” *Journal of Time Series Analysis*, 23, 629–649. [314]

Branch, W. A., and Evans, G. W. (2006), “A Simple Recursive Forecasting Model,” *Economic Letters*, 91, 158–166. [314]

Clark, T. E., and McCracken, M. W. (2009), “Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts,” *International Economic Review*, 50, 363–395. [308]

Clemen, R. T. (1989), “Combining Forecasts: A Review and Annotated Bibliography,” *International Journal of Forecasting*, 5, 559–581. [307]

Diebold, F. X., and Mariano, R. S. (1995), “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13, 253–263. [315,316]

Gardner Jr., E. S. (2006), “Exponential Smoothing: The State of the Art—Part II,” *International Journal of Forecasting*, 22, 637–666. [307]

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton: Princeton University Press. [317]

Hannan, E. J., and Deistler, M. (1988), *The Statistical Analysis of Linear Models*, New York: Wiley. [314]

Maheu, J. M., and Gordon, S. (2008), “Learning, Forecasting and Structural Breaks,” *Journal of Applied Econometrics*, 23, 553–583. [309]

Pesaran, M. H., and Pick, A. (2008), “Forecasting Random Walks Under Drift Instability,” Working Papers in Economics 0814, University of Cambridge, Faculty of Economics. [314]

Pesaran, M. H., and Timmermann, A. (2007), “Selection of Estimation Window in the Presence of Breaks,” *Journal of Econometrics*, 137, 134–161. [307, 308,313,315]

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006), “Forecasting Time Series Subject to Multiple Structural Breaks,” *Review of Economic Studies*, 73, 1057–1084. [309]

Pesaran, M. H., Schuermann, T., and Smith, L. V. (2009), “Forecasting Economic and Financial Variables With Global VARs” (with discussion and rejoinder), *International Journal of Forecasting*, 25, 642–715. [307]

Schrimpf, A., and Wang, Q. (2010), “A Reappraisal of the Leading Indicator Properties of the Yield Curve Under Structural Instability,” *International Journal of Forecasting*, 26 (4), 836–857. [307]

Stock, J. H., and Watson, M. W. (2004), “Combination Forecasts of Output Growth in a Seven-Country Data Set,” *Journal of Forecasting*, 23, 405–430. [307]

Timmermann, A. (2006), “Forecast Combinations,” in *Handbook of Economic Forecasting,* eds. G. Elliot, C. W. J. Granger, and A. Timmermann, Amsterdam: Elsevier, pp. 135–196. [307]