

# Optimal Forecasts in the Presence of Structural Breaks\*

M. Hashem Pesaran  
University of Cambridge  
and USC

Andreas Pick  
Erasmus University Rotterdam  
and De Nederlandsche Bank

Mikhail Pranovich  
University of Cambridge and  
Joint Vienna Institute

October 16, 2011

## Abstract

This paper considers the problem of forecasting under continuous and discrete structural breaks and proposes weighting observations to obtain optimal forecasts in the MSFE sense. We derive optimal weights for continuous and discrete break processes. Under continuous breaks, our approach recovers exponential smoothing weights. Under discrete breaks, we provide analytical expressions for the weights in models with a single regressor and asymptotically for larger models. It is shown that in these cases the value of the optimal weight is the same across observations within a given regime and differs only across regimes. In practice, where information on structural breaks is uncertain a forecasting procedure based on robust weights is proposed. Monte Carlo experiments and an empirical application to the predictive power of the yield curve analyze the performance of our approach relative to other forecasting methods.

*Keywords:* Forecasting, structural breaks, optimal weights, robust weights, exponential smoothing

*JEL codes:* C22, C53

---

\*A preliminary version of this paper was presented as a Featured Talk at the 31<sup>st</sup> Annual International Symposium on Forecasting, held in Prague, June 26-29, 2011. Comments and suggestions by conference participants are gratefully acknowledged.

# 1 Introduction

It is now widely recognized that parameter instability is an important source of forecast failure in macroeconomics and finance as documented by Pesaran and Timmermann (2002), Pesaran, Pettenuzzo, Timmermann (2006), Koop and Potter (2007), Giacomini and Rossi (2009), Inoue and Rossi (2011), among others. Clements and Hendry (1999, 2006) and Rossi (2011) provide reviews. Parameter instability can arise as a result of changes in tastes, technology, institutional arrangements and government policy. Broadly speaking, there are two basic approaches to modeling parameter instability: parameters can be assumed to change at discrete time intervals or continuously. Under the former break dates are estimated, and forecasts are typically constructed using the post-break observations.<sup>1</sup> Assuming that the break dates are accurately estimated, the forecasts based on observations after the last break is likely to be unbiased, but as pointed out by Pesaran and Timmermann (2007) the forecasts from the post-break window may not minimize the mean square forecast error (MSFE) as the estimation uncertainty due to the relatively short post-break window may be large. For this reason Pesaran and Timmermann (2007) suggest an optimal estimation window that may include pre-break observations. When the time and size of the break is uncertain, Pesaran and Timmermann (2007) consider averaging forecasts across estimation windows (AveW), which, as Pesaran and Pick (2011) show, improves forecasts without relying on estimates of break dates and sizes.

Under the continuously changing parameter model, the breaks are assumed to occur every period, and observations are down-weighted to take account of the slowly changing nature of the parameters. Within this framework a prominent approach is the exponential smoothing (ExpS) first proposed by Holt (1957) and Brown (1959). Other approaches using Kalman filters have also been proposed as generalization of ExpS. Hyndman, Koehler, Ord, and Snyder (2008) provide a comprehensive survey. Like AveW forecasts, no estimates of the break date and size are required but exponential smoothing forecasts are highly sensitive to the down-weighting parameter.

In this paper, we develop a unified approach to obtaining optimal forecasts under both types of structural breaks. We consider forecasts based on weighted observations as in ExpS approach but derive weights that are optimal in the sense that the resulting forecasts minimize the MSFE. In the case of continuous breaks, our approach recovers the ExpS weights. But the optimal weights can differ markedly from the ExpS weights when the breaks are assumed to occur at discrete time intervals. We show that, conditional on the break size and date, the optimal weights follow a step function that allocates constant weights within regimes but different weights between regimes. A striking result emerges under multiple breaks: observations of the last regime that continues into the forecast period may not receive the highest weight. The intuition for this result is that the bias component of the MSFE can be reduced by giving the largest weights to observations in an early regime to counterbalance biases of the opposite sign in intermediate regimes.

In practice, however, the dates and sizes of the breaks are unknown and must

---

<sup>1</sup>There are many statistical procedures that can be used for detection of break dates. See, for example, Brown et al. (1975), Andrews (1993), Andrews et al. (1996), Bai and Perron (1997, 2003), and Altissimo and Corradi (2003).

be estimated. But such estimates tend to be quite imprecise and their use lead to deterioration of forecasts, sometimes quite substantially. In order to address this problem, we develop weights that are robust to the uncertainty that surrounds the dates and the sizes of the breaks. The robust weights are derived by integrating the optimal weights with respect to a uniformly distributed break dates. An interesting insight from these derivations is that the effect of uncertainty of the break size on the weights is of order  $T^{-2}$  if the break is in the slope coefficient, and of order  $T^{-3}$  if the break is in the error variances, where  $T$  is the sample size including the pre-break observations. For estimation of the robust weights knowing the break date is more important than knowing the size of the break, and that breaks in slopes are, in turn, more important than breaks in error variances.

We provide Monte Carlo experiments that compare the forecasts from optimal weights to a range of competing forecasting methods. Under a discrete break, the key factor is the size of the break. A larger break leads to more precise estimates of the break date and improves forecasts that are conditional on these estimates, which include the optimal weights forecast, post-break forecasts, and optimal window forecasts. In contrast, when the break is small relative to the noise in the DGP, the robust weights produce the best forecasts as they do not rely on the imprecise estimates of the break date and size. When the break process is continuous, ExpS forecasts that estimate the down-weighting parameter perform well. However, the forecasts from the robust weights perform similarly well and in some settings provide the best forecasts even in this setting.

We apply the different methods considered in the paper to forecasting real GDP using the slope of the yield curve across nine industrial economies over the period 1994Q1-2009Q4. The general finding is that breaks are difficult to estimate with sufficient accuracy and, similar to the Monte Carlo results, forecasts based on estimates of break dates perform poorly. Robust weights and exponential smoothing forecasts perform well and deliver large improvements over forecasts based on equal weights.

The rest of the paper is set out as follows. Using a linear regression model, derivations of optimal weights under different break processes are set out in Section 2, and the MSFE outcomes are compared across different forecasting methods. Optimal weights that are robust to the uncertainty of the break process are motivated and derived in Section 3. Monte Carlo evidence on the comparative performance of the different forecasting methods is discussed in Section 4. Empirical results are presented in Section 5. The paper ends with some concluding remarks in Section 6. A few of the less essential derivations are collected in a mathematical appendix.

## 2 Optimal weights under different break processes

Consider the linear regression model

$$y_t = \beta_t' \mathbf{x}_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim iid(0, 1), \quad t = 1, 2, \dots, T, T + 1 \quad (1)$$

where  $\mathbf{x}_t$  is a  $k \times 1$  vector of stationary regressors, and the  $k \times 1$  coefficient vector,  $\beta_t$ , and the scalar error variance,  $\sigma_t^2$ , are subject to breaks. The breaks can be continuous, that is,  $\beta_t$  changes its value in every period. A prominent example is the random walk model

$$\beta_t = \beta_{t-1} + \mathbf{S}_\beta \mathbf{v}_t, \quad \text{where } \mathbf{v}_t \sim iid(\mathbf{0}, \mathbf{I}_k),$$

where  $\mathbf{I}_k$  is the identity matrix of order  $k$ , and the break variance,  $\boldsymbol{\Sigma}_\beta = \mathbf{S}_\beta \mathbf{S}'_\beta$ , is assumed to be small relative to  $\sigma_t^2$ .<sup>2</sup> Alternatively, the breaks could be discrete where the parameters change at distinct points in time,  $T_{b,i}$ ,  $i = 1, 2, \dots, n$ ,

$$\boldsymbol{\beta}_t = \begin{cases} \boldsymbol{\beta}_{(1)} & \text{for } 1 < t \leq T_{b,1} \\ \boldsymbol{\beta}_{(2)} & \text{for } T_{b,1} < t \leq T_{b,2} \\ \vdots & \\ \boldsymbol{\beta}_{(n)} & \text{for } T_{b,n} < t \leq T \end{cases}$$

Additionally,  $\sigma_t$  may be subject to a similar break process.<sup>3</sup> In contrast to the continuously changing parameter model, the number of discrete breaks,  $n$ , is assumed to be small, although the break sizes, measured by  $\|\boldsymbol{\beta}_{(i)} - \boldsymbol{\beta}_{(i-1)}\|$  could be large relative to  $\sigma_t$ . There are merits in both specifications, and a choice between them would depend on the particular forecasting problem under consideration.

We propose a general approach to achieve a minimum mean square forecast error (MSFE) under both break processes. We weigh past observations by weights  $w_t$  in the estimation

$$\hat{\boldsymbol{\beta}}_T(\mathbf{w}) = \left( \sum_{t=1}^T w_t \mathbf{x}_t \mathbf{x}'_t \right)^{-1} \sum_{t=1}^T w_t \mathbf{x}_t \mathbf{y}_t,$$

subject to the restriction  $\sum_{t=1}^T w_t = 1$ . The weights  $\mathbf{w} = (w_1, w_2, \dots, w_T)'$  are chosen such that the resulting MSFE of the one-step ahead forecast

$$\hat{y}_{T+1} = \hat{\boldsymbol{\beta}}'_T \mathbf{x}_{T+1}$$

is minimized.

Closed form solutions under the continuous break process are only available when we simplify the model to one without regressors. In this setting the optimal weights recover the exponential smoothing forecast. For the discrete break process we derive new results for the same simple model but also for models with one or more regressors.

## 2.1 Optimal weights in a model with continuous breaks

Consider the following model

$$y_t = \beta_t + \sigma_\varepsilon \varepsilon_t, \quad (2)$$

where  $\beta_t = \beta_{t-1} + \sigma_v v_t$ , and  $\varepsilon_t$  and  $v_t$  are  $iid(0, 1)$ . The optimal weights can be found by minimizing  $E(y_{T+1} - \sum_{t=1}^T w_t y_t)^2$  with respect to  $w_t$ ,  $t = 1, 2, \dots, T$  subject to  $\sum_{t=1}^T w_t = 1$ . For a solution to this problem we first note that the forecast error is given by

$$e_{T+1} = y_{T+1} - \hat{\beta}_{T+1}(\mathbf{w}) = \beta_{T+1} - \mathbf{w}' \boldsymbol{\beta} + \sigma_\varepsilon (\varepsilon_{T+1} - \mathbf{w}' \boldsymbol{\varepsilon}),$$

<sup>2</sup>The covariance matrix  $\boldsymbol{\Sigma}_\beta$  is said to be small relative to  $\sigma_t$  if  $\|\boldsymbol{\Sigma}_\beta\|/\sigma_t$  is small, where  $\|\mathbf{A}\|^2 = \text{tr}(\mathbf{A}\mathbf{A}')$  denotes the Euclidean norm of matrix  $\mathbf{A}$ .

<sup>3</sup>Note that brackets around subscripts denote subsamples between breaks, such that  $\beta_t$  is the parameter at period  $t$  but  $\beta_{(i+1)}$  the parameter after break  $i$ .

where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_T)'$ . But using the random walk formulation of  $\boldsymbol{\beta}$  we have

$$\boldsymbol{\beta} = \beta_0 \boldsymbol{\tau}_T + \sigma_v \mathbf{H} \mathbf{v},$$

where  $\mathbf{v} = (v_1, v_2, \dots, v_T)'$  and

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}, \quad \boldsymbol{\tau}_T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}.$$

Also,  $\beta_{T+1} = \beta_0 + \sigma_v \boldsymbol{\tau}'_T \mathbf{v} + \sigma_v v_{T+1}$ . Hence,

$$\sigma_\varepsilon^{-1} e_{T+1} = [(\boldsymbol{\tau}'_T \mathbf{v} - \mathbf{w}' \mathbf{H} \mathbf{v}) \delta + \sigma_\varepsilon (\varepsilon_{T+1} - \mathbf{w}' \boldsymbol{\varepsilon}) + \delta v_{T+1}],$$

where  $\delta^2 = \sigma_v^2 / \sigma_\varepsilon^2$ . Therefore, (noting that by assumption  $\mathbf{v}$  and  $\boldsymbol{\varepsilon}$  are independently distributed)

$$E(\sigma_\varepsilon^{-2} e_{T+1}^2 | \mathbf{w}) \propto \mathbf{w}' \mathbf{H} \mathbf{H}' \mathbf{w} - 2 \mathbf{w}' \mathbf{H} \boldsymbol{\tau}_T + \delta^{-2} \mathbf{w}' \mathbf{w}.$$

The first order condition for minimization of  $E(\sigma_\varepsilon^{-2} e_{T+1}^2 | \mathbf{w})$  subject to the constraint,  $\mathbf{w}' \boldsymbol{\tau}_T = 1$ , is given by

$$\mathbf{H} \mathbf{H}' \mathbf{w} - \mathbf{H} \boldsymbol{\tau}_T + \delta^{-2} \mathbf{w} - \theta \boldsymbol{\tau}_T = 0,$$

where  $\theta$  is the Lagrangian multiplier applied to the  $\mathbf{w}' \boldsymbol{\tau}_T = 1$ . Solving for  $\mathbf{w}$  in terms of  $\theta$  we have

$$(\mathbf{H} \mathbf{H}' + \delta^{-2} \mathbf{I}_T) \mathbf{w} = (\mathbf{H} + \theta \mathbf{I}_T) \boldsymbol{\tau}_T,$$

or

$$\mathbf{w} = \delta^2 (\mathbf{H} \mathbf{H}' + \delta^{-2} \mathbf{I}_T)^{-1} (\mathbf{H} + \theta \mathbf{I}_T) \boldsymbol{\tau}_T, \quad (3)$$

Also since  $\boldsymbol{\tau}'_T \mathbf{w} = 1$ , then  $\boldsymbol{\tau}'_T (\mathbf{H} \mathbf{H}' + \delta^{-2} \mathbf{I}_T)^{-1} (\mathbf{H} + \theta \mathbf{I}_T) \boldsymbol{\tau}_T = 1$  and we have

$$\theta = \frac{1 - \boldsymbol{\tau}'_T (\mathbf{H} \mathbf{H}' + \delta^{-2} \mathbf{I}_T)^{-1} \mathbf{H} \boldsymbol{\tau}_T}{\boldsymbol{\tau}'_T (\mathbf{H} \mathbf{H}' + \delta^{-2} \mathbf{I}_T)^{-1} \boldsymbol{\tau}_T}.$$

It is easily seen that for the extreme values of  $\delta^2 = \infty$  and 0 we obtain the random walk and equal weighted solutions,  $w(\infty) = (1, 0, \dots, 0)'$  and  $w(0) = T^{-1}(1, 1, \dots, 1)'$ , respectively.

The literature on exponential smoothing has traditionally used a different solution to address the time varying  $\beta_t$ . Write the model in terms of the observables

$$y_t - y_{t-1} = \sigma_v v_t + \sigma_\varepsilon (\varepsilon_t - \varepsilon_{t-1}), \quad (4)$$

which represents an MA(1) process in  $\Delta y_t$  with the MA parameter given by  $\gamma$ , or more specifically

$$\Delta y_t = \xi_t - \gamma \xi_{t-1}, \quad (5)$$

where  $\xi_t$  is a serially uncorrelated process with mean zero and a constant variance, and by equating the first order autocorrelation in (4) and (5) we have

$$\frac{\gamma}{1 + \gamma^2} = \frac{\sigma_\varepsilon^2}{2\sigma_\varepsilon^2 + \sigma_v^2} = \frac{1}{2 + \delta^2},$$

Hence,

$$\gamma^2 - (2 + \delta^2)\gamma + 1 = 0. \quad (6)$$

This equation has two real roots given by

$$\gamma = \frac{(2 + \delta^2) \pm \delta(4 + \delta^2)^{1/2}}{2}. \quad (7)$$

Since  $\delta > 0$ , then  $\gamma = 1 + \delta^2/2 - \delta(1 + \delta^2/4)^{1/2}$  is the root that lies within the unit circle and should be used.<sup>4</sup> The optimal forecast of  $y_{T+1}$  is now given by

$$E(y_{T+1} | y_T, y_{T-1}, \dots) = y_T - \gamma \xi_T,$$

but since  $0 < \gamma < 1$  we can invert the MA process to obtain

$$\xi_T = (1 - \gamma L)^{-1}(y_T - y_{T-1}),$$

and hence

$$\begin{aligned} E(y_{T+1} | y_T, y_{T-1}, \dots) &= y_T - \gamma(1 - \gamma L)^{-1}(y_T - y_{T-1}), \\ &= (1 - \gamma)(y_T + \gamma y_{T-1} + \gamma^2 y_{T-2} + \dots). \end{aligned}$$

In practice, the infinite series must be truncated to yield the ExpS forecast

$$\hat{y}_{T+1} = \frac{1 - \gamma}{1 - \gamma^T} \sum_{j=1}^T \gamma^{T-j} y_j \quad (8)$$

and the quality of the approximation will depend on  $T$  and  $\gamma$ , and could be poor when  $T$  is relatively small and  $\gamma$  close to unity. For large  $T$  and  $\gamma$  not too close to unity the elements of  $w_t$  will be very close to  $(1 - \gamma)\gamma^{T-t}$  for  $t = T, T - 1, \dots$ . It is worth noting that the weights  $w_t$  add up to unity and adapt to the sample size  $T$ , whilst the MA weights are optimal only for large  $T$  and when  $\gamma$  is not too close unity.

Assuming that  $0 \leq \gamma < 1$ , the relationship between  $\delta$  and  $\gamma$  is given by<sup>5</sup>

$$\delta = \frac{\sigma_v}{\sigma_\varepsilon} = \frac{1 - \gamma}{\sqrt{\gamma}}. \quad (9)$$

This suggests that for values of  $\gamma$  in the range 0.95 – 0.99 used in the literature, changes in  $\beta$  (per unit of time) must be quite small relative to  $\sigma_\varepsilon$ , the size of innovations to the errors. For example, for  $\gamma = 0.95$ ,  $\sigma_v/\sigma_\varepsilon = 0.05$ , and for  $\gamma = 0.98$  we have  $\sigma_v/\sigma_\varepsilon = 0.02$ .

## 2.2 Optimal weights in a model with a single, discrete break

Again consider model (2) but now assume that  $\beta_t$  is subject to a single, discrete break at  $T_b$ ,  $1 < T_b < T$ ,

$$\beta_t = \begin{cases} \beta_{(1)} & \text{for } t \leq T_b \\ \beta_{(2)} & \text{for } T_b < t \leq T + 1 \end{cases}$$

<sup>4</sup>Since  $\delta > 0$  then it is easily seen that  $0 < \gamma = 1 + \delta^2/2 - \delta(1 + \delta^2/4)^{1/2} < 1$ .

<sup>5</sup>Using (6), note that  $(1 - \gamma)^2 = \gamma\delta^2$ .

In this case the forecast is  $\hat{y}_{T+1} = \hat{\beta}_T(\mathbf{w})$  where  $\hat{\beta}_T(\mathbf{w}) = \sum_{t=1}^T w_t y_t$  and

$$\hat{\beta}_T(\mathbf{w}) - \beta_T = (\beta_{(1)} - \beta_{(2)}) \sum_{t=1}^{T_b} w_t + \sum_{t=1}^T w_t \sigma_\varepsilon \varepsilon_t.$$

Therefore, the forecast error is given by

$$\begin{aligned} e_{T+1}(\mathbf{w}) &= y_{T+1} - \hat{\beta}_T(\mathbf{w}) \\ &= \sigma_\varepsilon \varepsilon_t - (\beta_{(1)} - \beta_{(2)}) \sum_{t=1}^{T_b} w_t - \sum_{t=1}^T w_t \sigma_\varepsilon \varepsilon_t, \end{aligned}$$

and the MSFE scaled by the error variance is

$$E[\sigma_\varepsilon^{-2} e_{T+1}^2(\mathbf{w})] = 1 - \lambda^2 \left( \sum_{t=1}^{T_b} w_t \right)^2 - \sum_{t=1}^T w_t^2, \quad (10)$$

where  $\lambda = (\beta_{(1)} - \beta_{(2)})/\sigma_\varepsilon$ .

We can now obtain the optimal weights by minimizing (10) subject to  $\sum_{t=1}^T w_t = 1$ . The first order conditions are: for  $t \leq T_b$  we have  $2\lambda^2 \sum_{t=1}^{T_b} w_t + 2w_t + \theta = 0$ , and for  $T_b < t \leq T$ ,  $2w_t + \theta = 0$ , where  $\theta$  is the Lagrange multiplier associated with  $\sum_{t=1}^T w_t = 1$ . Note that  $w_t$  for  $t \leq T_b$  does not depend on  $t$  and the same is true for  $w_t$  for  $t > T_b$ . Hence

$$w_t = \begin{cases} w_{(a)} = -\lambda^2 \sum_{t=1}^{T_b} w_t - \theta/2 & \text{for } 1 < t \leq T_b \\ w_{(b)} = -\theta/2 & \text{for } T_b < t \leq T \end{cases}$$

and

$$w_{(b)} - w_{(a)} = \lambda^2 \sum_{t=1}^{T_b} w_t = \lambda^2 T_b w_{(a)}.$$

Solving for  $w_{(b)}$  and substituting into  $\sum_{t=1}^T w_t = T_b w_{(a)} + (T - T_b) w_{(b)} = 1$  yields the optimal weights

$$w_{(a)} = \frac{1}{T} \frac{1}{1 + T_b(1 - b)\lambda^2}, \quad (11)$$

and

$$w_{(b)} = \frac{1}{T} \frac{1 + T_b\lambda^2}{1 + T_b(1 - b)\lambda^2}. \quad (12)$$

where  $b = T_b/T$ .

We can use the fact that the weights are constant in the subsamples in (10) to obtain the scaled MSFE

$$E(\sigma_\varepsilon^{-2} \hat{e}_{T+1}^2) = 1 + (T_b \lambda w_{(a)})^2 + T_b w_{(a)}^2 + (T - T_b) w_{(b)}^2$$

and using (11) and (12) it is straightforward to show that this reduces to

$$\begin{aligned} E(\sigma_\varepsilon^{-2} \hat{e}_{T+1}^2) &= 1 + \frac{1}{T} \frac{1 + T_b\lambda^2}{1 + T_b(1 - b)\lambda^2} \\ &= 1 + w_{(b)}. \end{aligned} \quad (13)$$

Namely, the MSFE varies with  $\lambda$  through the post-beak weight,  $w_{(b)}$ .

We can now compare the forecasts based on optimal weights to those from a range of competing forecasting methods: post-break window observations, the optimal estimation window, averaging across estimation windows, and exponential smoothing.

### 2.2.1 Optimal window and post-break window forecasts

We now consider the optimal window that gives equal weights to observations within the window and zero weights to preceding observations as suggested by Pesaran and Timmermann (2007). Suppose that the optimal window size contains observations  $T_v$  to  $T$  (inclusive), where  $v = (T - T_v + 1) / T$  so that  $T_v = T(1 - v) + 1$ . Then, as we show in Appendix A.1, the scaled MSFE of the post-break window is

$$\mathbb{E} \left[ \sigma_\varepsilon^{-2} \hat{e}_{T+1}^2 | v = (1 - b) \right] = 1 + \frac{1}{T(1 - b)} \quad (14)$$

and the expected, scaled MSFE for the optimal window is

$$\mathbb{E} \left( \sigma_\varepsilon^{-2} \hat{e}_{T+1}^2 | v_{v > (1-b)}^o \right) = 1 + \frac{1}{T(1 - b)} - \frac{1}{T^2} \frac{1}{4\lambda^2(1 - b)^2} \quad (15)$$

which cannot be greater than the MSFE (14) as the elements of the last fraction are non-negative. Hence, the optimal window size is

$$v^o = \begin{cases} \frac{1-b}{1 - \frac{1}{2\lambda^2(1-b)T}} & \text{if } \lambda^2 \geq \frac{T}{2(T-T_b)T_b} \\ 1 & \text{if } \lambda^2 < \frac{T}{2(T-T_b)T_b} \end{cases}$$

Comparing the MSFEs of the forecasts from the optimal window to that of the optimal weight forecast, using (15) and (13), we have

$$\begin{aligned} & \mathbb{E} \left( \sigma_\varepsilon^{-2} \hat{e}_{T+1}^2 | v_{v > (1-b)}^o \right) - \mathbb{E}(\sigma_\varepsilon^{-2} \hat{e}_{T+1}^2) \\ &= \left[ \frac{1}{T(1 - b)} - \frac{1}{T^2} \frac{1}{4\lambda^2(1 - b)^2} \right] - \frac{1}{T} \frac{1 + Tb\lambda^2}{1 + Tb(1 - b)\lambda^2} \\ &= \frac{1}{T} \frac{T\lambda^2 b(1 - b) + 2T\lambda^2 b(1 - b) - 1}{4T(1 - b)^2 \lambda^2 [1 + Tb(1 - b)\lambda^2]} > 0 \end{aligned} \quad (16)$$

where the last inequality follows since  $v^o \leq 1$ , implies that  $T\lambda^2 b(1 - b) \geq 1/2$ . Therefore, forecasts obtained from optimal weights will have a smaller MSFE than forecasts giving equal weight to observations in an optimally chosen window. In the case where  $T\lambda^2 < 1/2$ , the optimal window contains all observations, so that the comparison is between the optimal weights and equal weights. Clearly, by merit of the optimality of the weights the forecast based on optimal weights will have a lower MSFE.

While optimal weights lead to a lower MSFE, it is interesting to get a quantitative sense of the difference. Table 1 reports the ratio of MSFEs of different forecasting methods to that of the equal weight forecast using all observations for a range of values of  $\lambda$  and  $b$ . That is, for forecast method  $i$  we report  $\text{MSFE}_i / \text{MSFE}_{\text{eq}}$ . The first line gives the ratio of MSFE for the forecast using optimal weights, the



Table 1: Relative MSFE for a single break in drift for known  $b$  and  $\lambda$

	$b$		0.95		0.9		
	$\lambda$	0.5	1	2	0.5	1	2
opt. weights		0.901	0.610	0.258	0.884	0.600	0.258
post-break obs.		0.971	0.628	0.260	0.907	0.604	0.259
opt. window		0.939	0.622	0.259	0.899	0.603	0.259
AveW( $v_{\min} = 0.05$ )		0.966	0.900	0.829	0.941	0.830	0.704
ExpS( $\gamma = 0.95$ )		0.973	0.924	0.872	0.958	0.883	0.799

Note: The table reports the ratio of MSFEs of a range of forecast methods to that the equal weight forecasts using all observations,  $\text{MSFE}_i/\text{MSFE}_{\text{allobs.}}$ , where  $\text{MSFE}_i$  is forecasting method  $i$ . These are (i) using the optimal weights, (ii) using the post-break observations, (iii) forecast based on the optimal window, (iv) AveW forecasts with  $v_{\min} = 0.05$  and  $m = T(1 - v_{\min}) + 1$  windows, and (v) ExpS forecasts with  $\gamma = 0.95$ . Finally,  $T = 100$ .

second line that when using only the post-break observations, and the third line that when using the optimal window.

It can be seen that the forecast based on optimal weights has the lowest MSFE across all parameter combinations. The MSFE of the forecast based on the post-break window is relatively similar to that using optimal weights when either the break or the post-break window is large ( $b = 0.9$ ). For breaks of smaller magnitude, however, the post-break window forecast has a substantially higher MSFE. Forecasts based on the optimal window perform better than those based on the post-break window and has the second lowest MSFE.

### 2.2.2 Averaging across estimation windows

Pesaran and Pick (2011) discuss theoretical properties of averaging forecasts from sub-windows (AveW). For the case of the random walk (2) they show that the AveW forecast

$$\hat{y}_{T+1} = \sum_{i=1}^m \hat{y}_{T+1}(v_{(i)}), \text{ where } \hat{y}_{T+1}(v_{(i)}) = \frac{1}{T - T_{v_{(i)}} + 1} \sum_{s=T_{v_{(i)}}}^T y_s,$$

$v_{(i)}$  is the minimum (shortest) window, and  $m$  is the number of sub-windows, has the MSFE

$$\begin{aligned} \text{E}(\sigma_\varepsilon^{-2} \hat{e}_{T+1}^2 | v_{(i)}) &= 1 + \left[ \frac{\lambda}{m} \sum_{i=1}^m \frac{v_{(i)} - (1-b)}{v_{(i)}} \mathbf{I}[v_{(i)} - (1-b)] \right]^2 \\ &\quad + \frac{1}{m^2} \sum_{i=1}^m \frac{1 + 2(i-1)}{T v_{(i)}}. \end{aligned}$$

Pesaran and Pick (2011) show that for the case of the random walk it will improve over equal weight forecasts using all observations unless the break is very small. This is reflected in the results in the fourth line of Table 1, which report results for  $v_{\min} = 0.05$ . The AveW forecast has smaller MSFEs than the single window forecast using all observations but they have a substantially larger MSFE

than the forecasts obtained using the optimal weights. The intuition for this result is that averaging over estimation windows can be seen as a weighting observations where weights are smoothly decaying. The optimal weights (11) and (12), however, have a discrete change and will only be approximated poorly by the weights implied by the AveW forecast. Given the optimality of  $w_{(a)}$  and  $w_{(b)}$ , this means that the AveW MSFE will be necessarily larger than that of the forecasts using  $w_{(a)}$  and  $w_{(b)}$ . However, these results are not surprising as averaging forecasts is based on the idea that it will be beneficial when the break date and size are uncertain or where multiple breaks of unknown time and size may be present. We will explore such settings in the Monte Carlo experiments in Section 4.

### 2.2.3 Exponential smoothing

In Section 2.1 we have shown that under continuous breaks optimal weights recover ExpS weights. While the application of ExpS weights is not optimal under discrete breaks, it is nevertheless interesting to get a quantitative sense of the loss implied in using weights for continuous breaks when there is a single discrete break.

The MSFE of the exponential smoothing forecast can be shown to be

$$E(\sigma_\varepsilon^{-2} e_{T+1}^2 | \gamma) = 1 + \lambda^2 \left( \frac{\gamma^{1+Tb} - \gamma^T}{1 - \gamma^T} \right)^2 + \left( \frac{1 - \gamma}{1 - \gamma^Y} \right) \left( \frac{1 - \gamma^{2T}}{1 - \gamma^2} \right).$$

See Pesaran and Pick (2011). The last line in Table 1 reports results for  $\gamma = 0.95$ . Similar to the AveW forecasts, the ExpS forecasts improve on the results from the forecasts using all observations but have a larger MSFE than the forecasts based on the optimal weights. The reason is that, just as the AveW forecasts, the ExpS forecasts use smoothly decaying weights for the observations, where we have shown that discretely changing weights are optimal.

### 2.3 A single, discrete break in a multiple regression model

We now turn to the multiple regression model where the slope parameters and the error variance are subject to a single break at time  $t = T_b$

$$y_t = \begin{cases} \beta'_{(1)} \mathbf{x}_t + \sigma_{(1)} \varepsilon_t & \text{for } 1 \leq t \leq T_b, \\ \beta'_{(2)} \mathbf{x}_t + \sigma_{(2)} \varepsilon_t & \text{for } T_b + 1 \leq t \leq T \end{cases} \quad (17)$$

where  $\mathbf{x}_t$  is a  $k \times 1$  vector of exogenous regressors and  $\varepsilon_t \sim iid(0, 1)$ . Again, suppose that the slope parameter is estimated by weighting observations over the whole sample

$$\hat{\beta}_T(\mathbf{w}) = \left( \sum_{t=1}^T w_t \mathbf{x}_t \mathbf{x}'_t \right)^{-1} \sum_{t=1}^T w_t \mathbf{x}_t y_t. \quad (18)$$

Deriving and maximizing the scaled MSFE similar to the derivations in Section 2.2, we obtain the optimal weights, with the details provided in Appendix A.2. For

$t \leq T_b$  we have

$$\begin{aligned}
& [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t] q^2 w_t \tag{19} \\
&= \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 \mathbf{x}_t \mathbf{x}'_t + \sum_{t=T_b+1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t \\
&\quad - [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}],
\end{aligned}$$

and for  $t \geq T_b + 1$

$$\begin{aligned}
& [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t] w_t \tag{20} \\
&= \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 \mathbf{x}_t \mathbf{x}'_t + \sum_{t=T_b+1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t \\
&\quad + [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}].
\end{aligned}$$

where  $\boldsymbol{\lambda} = (\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(2)})/\sigma_{(2)}$  and  $q = \sigma_{(1)}/\sigma_{(2)}$ .

These optimal weights have a number of interesting properties. First, in the absence of a break, that is when  $\boldsymbol{\lambda} = \mathbf{0}$  and  $q = 1$ , then  $w_t = w$  for all  $t$ , as to be expected. To see this, note that when  $\boldsymbol{\lambda} = \mathbf{0}$  and  $q = 1$ , then for all  $t$  we have

$$w_t = \frac{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t}.$$

It is now easily seen that  $w_t = w$  (fixed) is a solution to the above. Note that for  $w_t = w$ , we have  $\mathbf{S}(\mathbf{w}) = w \mathbf{S}(1)$  and therefore

$$w_t = \frac{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(1) \left( \sum_{t=1}^T \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(1) \mathbf{x}_t}{w^{-1} \mathbf{x}'_{T+1} \mathbf{S}^{-1}(1) \mathbf{x}_t} = \frac{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(1) \mathbf{S}(1) \mathbf{S}^{-1}(1) \mathbf{x}_t}{w^{-1} \mathbf{x}'_{T+1} \mathbf{S}^{-1}(1) \mathbf{x}_t} = w.$$

Consider now the case where  $\boldsymbol{\lambda} \neq \mathbf{0}$  and  $q \neq 1$ , but suppose that  $\mathbf{x}_1 = \mathbf{x}_2$ , then using (19) and (20) the optimal weights for  $t = 1$  and  $t = 2$  we have

$$q^2 [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_1] (w_2 - w_1) = 0.$$

Hence, the weights within a given regime will be the same if the regressor values for the two points of time in that regime are identical. But the same is not true of the weights for time points in different regimes. For example, for the first regime select  $t = 1$  and for the second regime select  $t = T$ , and suppose that  $\mathbf{x}_1 = \mathbf{x}_T$ . Then from (19) and (20), and recalling that it follows that  $\mathbf{S}_1(\mathbf{w}_{(1)}) + \mathbf{S}_2(\mathbf{w}_{(2)}) = \mathbf{S}(\mathbf{w})$ , we have

$$[\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_T] (w_T - q^2 w_1) = [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] (\mathbf{x}'_T \boldsymbol{\lambda}),$$

which suggests that, in general, when  $\boldsymbol{\lambda} \neq \mathbf{0}$  and  $q \neq 1$  the weights across the two regimes differ even if the regressor values are the same. Therefore, in general, the optimal weights will differ both within and across regimes.

An exact analytical solution does not seem to be available and the unknown weights in (19) and (20) must be solved numerically. To this end let

$$\begin{aligned}\mathbf{d}(\mathbf{w}) &= \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 \mathbf{x}_t \mathbf{x}_t' + \sum_{t=T_b+1}^T w_t^2 \mathbf{x}_t \mathbf{x}_t' \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1} \\ \mathbf{p}_1(\mathbf{w}) &= \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}, \quad \text{and} \quad \mathbf{p}_2(\mathbf{w}) = \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda} \\ \gamma(\mathbf{w}) &= \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda} \\ \theta_t(\mathbf{w}) &= \begin{cases} q^2 \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t & \text{if } t \leq T_b \\ \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t & \text{if } t > T_b, \end{cases}\end{aligned}$$

where  $\mathbf{d}(\mathbf{w})$ ,  $\mathbf{p}_1(\mathbf{w})$  and  $\mathbf{p}_2(\mathbf{w})$  are  $k \times 1$  vectors, and  $\gamma(\mathbf{w})$  and  $\theta_t(\mathbf{w})$  are scalar functions of  $\mathbf{w}$ . Then, the  $T$  equations in (19) and (20) can be written as

$$\begin{aligned}\boldsymbol{\theta}(\mathbf{w}) \odot \mathbf{w} &= \mathbf{X} \mathbf{d}(\mathbf{w}) + \gamma(\mathbf{w}) \begin{pmatrix} -\mathbf{X}_{(1)} \mathbf{p}_2(\mathbf{w}) \\ \mathbf{X}_{(2)} \mathbf{p}_1(\mathbf{w}) \end{pmatrix} \\ &= \mathbf{X} \mathbf{d}(\mathbf{w}) + \gamma(\mathbf{w}) \mathbf{Z}(\mathbf{w}),\end{aligned}$$

where  $\mathbf{X}$  is the  $T \times k$  matrix of observations on  $\mathbf{x}$ ,  $\mathbf{X} = (\mathbf{X}'_{(1)}, \mathbf{X}'_{(2)})'$ ,  $\mathbf{X}_{(1)}$  and  $\mathbf{X}_{(2)}$  are  $T_b \times k$  and  $(T - T_b) \times k$  matrices of pre-break and post-break observations on  $\mathbf{x}$ . Also  $\boldsymbol{\theta}(\mathbf{w}) = (\theta_1(\mathbf{w}), \theta_2(\mathbf{w}), \dots, \theta_T(\mathbf{w}))'$  and  $\odot$  denotes element by element vector multiplication. We now need to minimize the function

$$\min_{\mathbf{w}} \mathbf{f}'(\mathbf{w}) \mathbf{f}(\mathbf{w})$$

subject to  $\boldsymbol{\iota}'_T \mathbf{w} = 1$  and  $w_t \geq 0$ , where

$$\mathbf{f}(\mathbf{w}) = \boldsymbol{\theta}(\mathbf{w}) \odot \mathbf{w} - \mathbf{X} \mathbf{q}(\mathbf{w}) - \gamma(\mathbf{w}) \mathbf{Z}(\mathbf{w}).$$

The asymptotic weights given in Section 2.3.2 below can be used as starting values for the numerical optimization.

### 2.3.1 A single, discrete break in a model with one regressor

Analytically more tractable results can be obtained when  $k = 1$ . In this case the scaled MSFE (56) simplifies to

$$\begin{aligned}E[\sigma_{(2)}^{-2} e_{T+1}^2(\mathbf{w})] &= 1 + \left[ \frac{x_{T+1} S(\mathbf{w}_{(1)}) \lambda}{S(\mathbf{w})} \right]^2 \\ &\quad + \frac{x_{T+1}^2 \left( \sum_{t=1}^{T_b} q^2 w_t^2 x_t^2 + \sum_{t=T_b+1}^T w_t^2 x_t^2 \right)}{[S(\mathbf{w})]^2}\end{aligned}\tag{21}$$

where  $\lambda = (\beta_{(1)} - \beta_{(2)})/\sigma_{(2)}$ , and the first order conditions (19) and (20) simplify to

$$w_t = \begin{cases} \frac{\sum_{t=1}^T w_t^2 x_t^2}{q^2 S(\mathbf{w})} - \lambda^2 \frac{S_1(\mathbf{w}_{(1)}) S_2(\mathbf{w}_{(2)})}{q^2 S(\mathbf{w})} & \text{for } t \leq T_b \\ \frac{\sum_{t=1}^T w_t^2 x_t^2}{S(\mathbf{w})} + \lambda^2 \frac{S_1^2(\mathbf{w}_{(1)})}{S(\mathbf{w})} & \text{for } t \geq T_b + 1 \end{cases}$$

Similar to the case of model (2),  $w_t$  for  $t \leq T_b$  does not depend on  $t$  and the same is true for  $w_t$  for  $t > T_b$ . We can, therefore, set

$$w_t = \begin{cases} w_{(a)} & \text{for } t \leq T_b \\ w_{(b)} & \text{for } t \geq T_b + 1 \end{cases}$$

Using the above results it now readily follows that  $w_{(b)} - q^2 w_{(a)} = w_{(a)} S_1(1) \lambda^2$ , where  $S_1(1) = \sum_{t=1}^{T_b} x_t^2$ . Also using the constraint  $\sum_{t=1}^T w_t = 1$  we have  $w_{(a)} T_b + (T - T_b) w_{(b)} = 1$ . Hence, for  $T_b$  reasonably large, which is not a restrictive assumption for the problem under consideration, and solving for  $w_{(a)}$  and  $w_{(b)}$  we obtain

$$w_{(a)} = \frac{1}{T} \frac{1}{b + (1-b)(q^2 + T_b \lambda^2 \omega_x^2)}, \quad (22)$$

$$w_{(b)} = \frac{1}{T} \frac{q^2 + T_b \lambda^2 \omega_x^2}{b + (1-b)(q^2 + T_b \lambda^2 \omega_x^2)}. \quad (23)$$

where  $\omega_x^2 = \text{plim}_{T_b \rightarrow \infty} (\frac{1}{T_b} \sum_{t=1}^{T_b} x_t^2)$ .

Given that for the optimal weights we have that  $w_t^{(\text{optimal})} = w_{(a)}$  for  $t \leq T_b$  and  $w_t^{(\text{optimal})} = w_{(b)}$  for  $t > T_b$ , (21) can be rewritten as

$$E \left( \sigma_{(2)}^{-2} e_{T+1}^2 | w_t^{(\text{opt.})} \right) \approx 1 + \frac{x_{T+1}^2}{T \omega_x^2} \frac{w_{(a)}^2 b (T_b \phi^2 + q^2) + w_{(b)}^2 (1-b)}{[w_{(a)} b + (1-b) w_{(b)}]^2} \quad (24)$$

From (22) and (23) it can be seen that  $w_{(b)} = (T_b \phi^2 + q^2) w_{(a)}$  and (24) simplifies to

$$E \left( \sigma_{(2)}^{-2} e_{T+1}^2 | w_t^{(\text{opt.})} \right) \approx 1 + \frac{x_{T+1}^2}{\omega_x^2} w_{(b)}.$$

Namely, the MSFE varies with  $\lambda$  and  $q$  through the post-break weight,  $w_{(b)}$ . Note that

$$\frac{\partial E \left( \sigma_{(2)}^{-2} e_{T+1}^2 | w_t^{(\text{opt.})} \right)}{\partial q^2} = \frac{x_{T+1}^2}{T \omega_x^2} \frac{b}{[b + (1-b)q^2 + T_b(1-b)\phi^2]^2} > 0.$$

In the standard case where the estimation uses the entire sample with equal weighting, that is,  $w_t^{(\text{eq.})} = 1/T$ ,

$$\begin{aligned} E \left( \sigma_{(2)}^{-2} e_{T+1}^2 | w_t^{(\text{eq.})} \right) &= 1 + b^2 \lambda^2 x_{T+1}^2 \left( \frac{T_b^{-1} \sum_{t=1}^{T_b} x_t^2}{T^{-1} \sum_{t=1}^T x_t^2} \right)^2 \\ &\quad + \frac{1}{T} \frac{b(q^2 - 1) x_{T+1}^2 \left( T_b^{-1} \sum_{t=1}^{T_b} x_t^2 \right)}{\left( T^{-1} \sum_{t=1}^T x_t^2 \right)^2} + \frac{1}{T} \left( \frac{x_{T+1}^2}{T^{-1} \sum_{t=1}^T x_t^2} \right), \end{aligned}$$

or approximately

$$E \left( \sigma_{(2)}^{-2} e_{T+1}^2 | w_t^{(\text{eq.})} \right) \approx 1 + \frac{x_{T+1}^2}{\omega_x^2} \left[ b^2 \phi^2 + \frac{bq^2 + (1-b)}{T} \right]$$

where we have used that for  $T_b$  large and  $T - T_b$  small  $\omega_x^2 = \text{plim}_{T \rightarrow \infty} (\frac{1}{T} \sum_{t=1}^T x_t^2)$ , and  $\phi^2 = \omega_x^2 \lambda^2$ .  $\phi$  and  $q$  measure the sizes of the breaks in  $\beta$  and  $\sigma$ , and  $b$  gives the proportion of pre-break observations.

Comparing MSFE of optimal weights with the one based on equal weights we have

$$\begin{aligned} \frac{\omega_x^2}{x_{T+1}^2} (MSFE_{eq.} - MSFE_{opt.}) &= \frac{b(Tb\phi^2 + q^2) + (1-b)}{T} - \frac{Tb\phi^2 + q^2}{T[b + (1-b)(Tb\phi^2 + q^2)]} \\ &= \frac{b(1-b)[1 - Tb\phi^2 - q^2]^2}{T[b + (1-b)(Tb\phi^2 + q^2)]} > 0. \end{aligned}$$

Similarly, when  $w_t$  are set independently of  $x_t$  (as in the case of exponential down-weighting) we have

$$E\left(\sigma_{(2)}^{-2} e_{T+1}^2\right) \approx 1 + \frac{x_{T+1}^2}{\omega_x^2} \left[ \phi^2 \left( \sum_{t=1}^{T_b} w_t \right)^2 + (q^2 - 1) \sum_{t=1}^{T_b} w_t^2 + \sum_{t=1}^T w_t^2 \right].$$

When only post-break observations are used, the implicit weights are  $w_t^{(\text{post})} = 0$  for  $t \leq T_b$  and  $w_t^{(\text{post})} = (T - T_b)/T$  for  $t > T_b$ . We therefore have

$$E\left(\sigma_{(2)}^{-2} e_{T+1}^2 | w_t^{(\text{post})}\right) \approx 1 + \frac{x_{T+1}^2}{\omega_x^2} \frac{1}{T(1-b)}$$

Comparing it to the MSFE based on post-break observations we have

$$\frac{\omega_x^2}{x_{T+1}^2} (MSFE_{post} - MSFE_{opt.}) = \frac{b}{T(1-b)[b + (1-b)(Tb\phi^2 + q^2)]} > 0,$$

namely, optimal weight forecasts dominate post-break forecasts for all values of  $0 < b < 1$ , but, as to be expected, the superiority of the optimal-weight forecasts diminishes as  $T(1-b) \rightarrow \infty$ .

### 2.3.2 Asymptotic results with $k \geq 1$ stationary regressors

The general solution in (19) and (20) can be simplified if we assume that  $T$  and  $T_b$  are sufficiently large with  $T - T_b$  fixed, and  $\mathbf{x}_t$  is a stationary process with  $E(\mathbf{x}_t \mathbf{x}_t') = \mathbf{\Omega}_{xx}$ , a positive definite matrix. That is we assume that  $T \rightarrow \infty$  and  $b \rightarrow 1$  but  $T(1-b) \rightarrow \tau$ , where  $\tau$  is a relatively small, constant number of post-break observations. Under these assumptions (and conditional on the weights,  $w_t$ )

$$\mathbf{S}(\mathbf{w}) \rightarrow \left( \sum_{t=1}^T w_t \right) E(\mathbf{x}_t \mathbf{x}_t') = \mathbf{\Omega}_{xx} \quad (25)$$

$$\mathbf{S}_1(\mathbf{w}_{(1)}) \rightarrow \left( \sum_{t=1}^{T_b} w_t \right) E(\mathbf{x}_t \mathbf{x}_t') = \left( \sum_{t=1}^{T_b} w_t \right) \mathbf{\Omega}_{xx}, \quad (26)$$

and

$$\sum_{t=1}^T w_t^2 \mathbf{x}_t \mathbf{x}_t' \rightarrow \left( \sum_{t=1}^T w_t^2 \right) \mathbf{\Omega}_{xx}, \quad (27)$$

and the MSFE simplifies to

$$\begin{aligned} E(\sigma_{(2)}^{-2}e_{T+1}^2) &= 1 + (\mathbf{x}'_{T+1}\boldsymbol{\lambda})^2 \left( \sum_{t=1}^{T_b} w_t \right)^2 \\ &+ (\mathbf{x}'_{T+1}\boldsymbol{\Omega}_{xx}^{-1}\mathbf{x}_{T+1}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 + \sum_{t=T_b+1}^T w_t^2 \right). \end{aligned} \quad (28)$$

The solution is similar to the case for  $k = 1$  and is given by

$$w_{(a)} = \frac{1}{T} \frac{1}{b + (1-b)(q^2 + Tb\phi^2)}, \quad (29)$$

$$w_{(b)} = \frac{1}{T} \frac{q^2 + Tb\phi^2}{b + (1-b)(q^2 + Tb\phi^2)}, \quad (30)$$

where

$$\phi = \frac{\mathbf{x}'_{T+1}\boldsymbol{\lambda}}{(\mathbf{x}'_{T+1}\boldsymbol{\Omega}_{xx}^{-1}\mathbf{x}_{T+1})^{1/2}}.$$

The above result is also in line with the result obtained for the simple case of  $k = 1$ . In that case  $\boldsymbol{\Omega}_{xx} = \omega_x^2$  and  $\phi = \lambda\omega_x$ .

## 2.4 Multiple discrete breaks in a multiple regression model

Consider now the case of multiple breaks in the slope coefficient of a linear regression model

$$y_t = \boldsymbol{\beta}'_t \mathbf{x}_t + \sigma \varepsilon_t$$

where the parameter vector  $\boldsymbol{\beta}_t$  is subject to  $n$  breaks at points  $b_i = T_{b,i}/T$ , such that  $b_1 < b_2 < \dots < b_n$ . For simplicity of exposition here we assume that the error variance is not subject to breaks. Initially, assume that  $n = 2$ , such that the parameter vector is

$$\boldsymbol{\beta}_t = \begin{cases} \boldsymbol{\beta}_{(1)} & \text{for } 1 < t \leq T_{b,1} \\ \boldsymbol{\beta}_{(2)} & \text{for } T_{b,1} < t \leq T_{b,2} \\ \boldsymbol{\beta}_{(3)} & \text{for } T_{b,2} < t \leq T \end{cases}$$

Using the weighted least squares estimator (18) we have that

$$\hat{\boldsymbol{\beta}}_T(\mathbf{w}) - \boldsymbol{\beta}_{(3)} = \mathbf{S}^{-1}(\mathbf{w}) \left[ \mathbf{S}_1(\mathbf{w}_{(1)}) (\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(3)}) + \mathbf{S}_2(\mathbf{w}_{(2)}) (\boldsymbol{\beta}_{(2)} - \boldsymbol{\beta}_{(3)}) \right] + \mathbf{S}^{-1}(\mathbf{w}) \sum_{t=1}^T w_t \mathbf{x}_t \sigma \varepsilon_t,$$

where  $\mathbf{S}_1(\mathbf{w}_{(1)}) = \sum_{t=1}^{T_{b,1}} w_t \mathbf{x}_t \mathbf{x}'_t$ ,  $\mathbf{S}_2(\mathbf{w}_{(2)}) = \sum_{t=T_{b,1}+1}^{T_{b,2}} w_t \mathbf{x}_t \mathbf{x}'_t$ , and  $\mathbf{S}(\mathbf{w}) = \sum_{t=1}^T w_t \mathbf{x}_t \mathbf{x}'_t$ . Consequently,

$$\begin{aligned} e_{T+1}(\mathbf{w}) &= y_{T+1} - \mathbf{x}'_{T+1} \hat{\boldsymbol{\beta}}_T(\mathbf{w}) \\ &= \sigma \varepsilon_{T+1} - \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left[ \mathbf{S}_1(\mathbf{w}_{(1)}) (\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(3)}) + \mathbf{S}_2(\mathbf{w}_{(2)}) (\boldsymbol{\beta}_{(2)} - \boldsymbol{\beta}_{(3)}) \right] \\ &\quad + \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \sum_{t=1}^T w_t \mathbf{x}_t \sigma \varepsilon_t \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} [\sigma^{-2} e_{T+1}^2(\mathbf{w})] &= 1 + \{ \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) [\mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}_{(1)} + \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}_{(2)}] \}^2 \\ &\quad + \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1} \end{aligned}$$

where

$$\boldsymbol{\lambda}_{(1)} = \frac{\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(3)}}{\sigma} \quad \text{and} \quad \boldsymbol{\lambda}_{(2)} = \frac{\boldsymbol{\beta}_{(2)} - \boldsymbol{\beta}_{(3)}}{\sigma}.$$

Optimal weights can therefore be obtained from

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} f(\mathbf{w}),$$

subject to  $\iota' \mathbf{w} = 1$ , where

$$\begin{aligned} f(\mathbf{w}) &= \{ \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) [\mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}_{(1)} + \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}_{(2)}] \}^2 \\ &\quad + \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1}. \end{aligned} \quad (31)$$

The first order conditions are

$$\begin{aligned} w_t [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1}] &= \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) [\mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}_{(1)} + \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}_{(2)}] \\ &\quad \times \{ \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) [\mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}_{(1)} + \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}_{(2)}] - \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \boldsymbol{\lambda}_{(i)} \} \\ &\quad + \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1} + \theta/2, \end{aligned}$$

where again  $\theta$  is the Lagrange multiplier associated with  $\iota' \mathbf{w} = 1$  and

$$\boldsymbol{\lambda} = \begin{cases} \boldsymbol{\lambda}_{(1)} & \text{if } t \leq T_{b,1} \\ \boldsymbol{\lambda}_{(2)} & \text{if } T_{b,1} < t \leq T_{b,2} \\ \mathbf{0} & \text{if } T_{b,2} < t \leq T \end{cases}$$

Again, by multiplying both sides by  $w_t$  and aggregating over  $t = 1, 2, \dots, T$  it can easily be verified that  $\theta = 0$ .

Hence, for  $\mathbf{x}_t \neq \mathbf{0}$  the optimal weights are

$$\begin{aligned} w_t &= \frac{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) [\mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}_{(1)} + \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}_{(2)}]}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t} \times \\ &\quad \times \{ \mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) [\mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}_{(1)} + \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}_{(2)}] \} \\ &\quad + \frac{\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{t+1}}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t} \\ &\quad - \frac{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) [\mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}_{(1)} + \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}_{(2)}] (\mathbf{x}'_t \boldsymbol{\lambda}_{(i)})}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t}. \end{aligned} \quad (32)$$



For  $n$  breaks it is easily seen that

$$w_t = \frac{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left[ \sum_{j=1}^{n-1} \mathbf{S}_j(\mathbf{w}_{(j)}) \boldsymbol{\lambda}_{(j)} \right] \left\{ \mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \left[ \sum_{j=1}^{n-1} \mathbf{S}_j(\mathbf{w}_{(j)}) \boldsymbol{\lambda}_{(j)} \right] \right\}}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t} + \frac{\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{t+1}}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t} - \frac{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left[ \sum_{j=1}^{n-1} \mathbf{S}_j(\mathbf{w}_{(j)}) \boldsymbol{\lambda}_{(j)} \right] (\mathbf{x}'_t \boldsymbol{\lambda}_{(i)})}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t},$$

where

$$\boldsymbol{\lambda} = \begin{cases} \boldsymbol{\lambda}_{(1)} = \frac{\beta_{(1)} - \beta_{(n+1)}}{\sigma} & \text{if } t \leq T_{b,1} \\ \boldsymbol{\lambda}_{(2)} = \frac{\beta_{(2)} - \beta_{(n+1)}}{\sigma} & \text{if } T_{b,1} < t \leq T_{b,2} \\ \vdots & \vdots \\ \boldsymbol{\lambda}_{(n)} = \frac{\beta_{(n)} - \beta_{(n+1)}}{\sigma} & \text{if } T_{b,n-1} < t \leq T_{b,n} \\ \mathbf{0} & \text{if } T_{b,n} < t \leq T \end{cases}$$

As in the case of a single break, numerical methods are necessary to obtain the weights.

#### 2.4.1 Optimal weights for multiple breaks in a simple regression model

In the case of a single regressor we can solve for the weights analytically. In this case, (32) simplifies to

$$w_t = \frac{[S_1(\mathbf{w}_{(1)}) \lambda_{(1)} + S_2(\mathbf{w}_{(2)}) \lambda_{(2)}]^2}{S(\mathbf{w})} + \frac{\sum_{t=1}^T w_t^2 x_t^2}{S(\mathbf{w})} - [S_1(\mathbf{w}_{(1)}) \lambda_{(1)} + S_2(\mathbf{w}_{(2)}) \lambda_{(2)}] \lambda_{(i)}$$

where  $\lambda_{(i)}$  is defined as above but is now a scalar. Therefore, defining  $S_1(1) = \sum_{t=1}^{T_{b,1}} x_t^2$  and  $S_2(1) = \sum_{t=T_{b,1}+1}^{T_{b,2}} x_t^2$ , solving for the optimal weights yields

$$\begin{aligned} w_{(1)} &= \frac{1}{T} \frac{1 + \lambda_{(2)}^2 S_2(1) - \lambda_{(1)} \lambda_{(2)} S_2(1)}{a_{s,2}} \\ w_{(2)} &= \frac{1}{T} \frac{1 + \lambda_{(1)}^2 S_1(1) - \lambda_{(1)} \lambda_{(2)} S_1(1)}{a_{s,2}} \\ w_{(3)} &= \frac{1}{T} \frac{1 + \lambda_{(1)}^2 S_1(1) + \lambda_{(2)}^2 S_2(1)}{a_{s,2}} \end{aligned}$$

where  $a_{s,2} = 1 + (1 - b_2) \left[ S_1(1) \lambda_{(1)}^2 + S_2(1) \lambda_{(2)}^2 \right] + [\lambda_{(1)} - \lambda_{(2)}] \left[ (b_2 - b_1) S_1(1) \lambda_{(1)} - b_1 S_2(1) \lambda_{(2)} \right]$ . This result generalizes to  $n$  breaks where

$$\begin{aligned} w_{(i)} | i \leq n &= \frac{1}{T} \frac{1 + \sum_{j=1, j \neq i}^n \lambda_{(j)}^2 S_j(1) - \lambda_{(i)} \sum_{j=1, j \neq i}^n \lambda_{(j)} S_j(1)}{a_{s,n}} \\ w_{(n+1)} &= \frac{1}{T} \frac{1 + \sum_{j=1}^n \lambda_{(j)}^2 S_j(1)}{a_{s,n}} \end{aligned}$$

and

$$a_{s,n} = 1 + \sum_{l=1}^{n+1} (b_l - b_{l-1}) \sum_{j=1, j \neq l}^n \lambda_{(j)}^2 S_j(1) - \sum_{l=1}^n \lambda_{(l)} (b_l - b_{l-1}) \sum_{j=1, j \neq l}^n \lambda_{(j)} S_j(1).$$

### 2.4.2 Asymptotic results in the multi-break case with $k \geq 1$ stationary regressors

Similar to the case with one break, we can simplify the solution when there are two or more regressors if we assume that many observations are available between breaks and  $\mathbf{x}_t$  is a stationary process with  $E(\mathbf{x}_t \mathbf{x}_t') = \boldsymbol{\Omega}_{xx}$ . Note, however, that we make no assumption about the number of observations since the last break. Initially consider the case of two breaks. In addition to (25) and (27) we have

$$\begin{aligned} \mathbf{S}_1(\mathbf{w}_{(1)}) &\rightarrow \left( \sum_{t=1}^{T_{b,1}} w_t \right) E(\mathbf{x}_t \mathbf{x}_t') = \left( \sum_{t=1}^{T_{b,1}} w_t \right) \boldsymbol{\Omega}_{xx} \\ \mathbf{S}_2(\mathbf{w}_{(2)}) &\rightarrow \left( \sum_{t=T_{b,1}+1}^{T_{b,2}} w_t \right) E(\mathbf{x}_t \mathbf{x}_t') = \left( \sum_{t=T_{b,1}+1}^{T_{b,2}} w_t \right) \boldsymbol{\Omega}_{xx} \end{aligned}$$

Then (31) simplifies to

$$f(\mathbf{w}) = \left[ \mathbf{x}'_{T+1} \left( \lambda_{(1)} \sum_{t=1}^{T_{b,1}} w_t + \lambda_{(2)} \sum_{t=T_{b,1}+1}^{T_{b,2}} w_t \right) \right]^2 + \mathbf{x}'_{T+1} \sum_{t=1}^T w_t^2 \boldsymbol{\Omega}_{xx} \mathbf{x}_{T+1}.$$

The optimal weights are therefore

$$w_{(1)} = \frac{1}{T} \frac{1 + T(b_2 - b_1)\phi_{(2)}^2 - T(b_2 - b_1)\phi_{(1)}\phi_{(2)}}{a_{a,2}} \quad (33)$$

$$w_{(2)} = \frac{1}{T} \frac{1 + T b_1 \phi_{(1)}^2 - T b_1 \phi_{(1)} \phi_{(2)}}{a_{a,2}} \quad (34)$$

$$w_{(3)} = \frac{1}{T} \frac{1 + T b_1 \phi_{(1)}^2 + T(b_2 - b_1)\phi_{(2)}^2}{a_{a,2}} \quad (35)$$

where  $a_{a,2} = 1 + T(1 - b_2)b_1\phi_{(1)}^2 + T(b_2 - b_1)(1 - b_2)\phi_{(2)}^2 + T b_1(b_2 - b_1)(\phi_{(1)} - \phi_{(2)})^2$  and

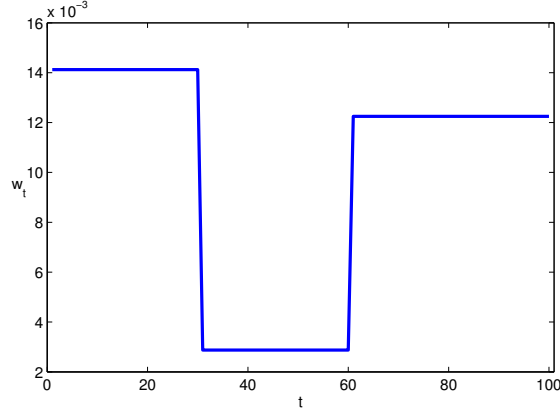
$$\phi_{(i)} = \frac{\mathbf{x}'_{T+1} \boldsymbol{\lambda}_{(i)}}{(\mathbf{x}'_{T+1} \boldsymbol{\Omega}_{xx}^{-1} \mathbf{x}_{T+1})^{1/2}}, \text{ for } i = 1, 2.$$

An interesting result is that the weights for two breaks are not necessarily decreasing in the distance from  $T$ . In particular,

- $w_{(1)} > w_{(3)} > w_{(2)}$  if  $\phi_{(1)} < 0$ ,  $\phi_{(2)} > 0$  and  $b_1 \phi_{(1)} > -(b_2 - b_1)\phi_{(2)}$
- $w_{(1)} > w_{(3)} > w_{(2)}$  if  $\phi_{(1)} > 0$ ,  $\phi_{(2)} < 0$  and  $b_1 \phi_{(1)} < -(b_2 - b_1)\phi_{(2)}$
- $w_{(2)} > w_{(3)} > w_{(1)}$  if  $\phi_{(1)} < 0$ ,  $\phi_{(2)} > 0$  and  $b_1 \phi_{(1)} < -(b_2 - b_1)\phi_{(2)}$
- $w_{(2)} > w_{(3)} > w_{(1)}$  if  $\phi_{(1)} > 0$ ,  $\phi_{(2)} < 0$  and  $b_1 \phi_{(1)} > -(b_2 - b_1)\phi_{(2)}$

Figure 1 plots the weights for  $T = 100$ ,  $b_1 = 0.3$ ,  $b_2 = 0.6$ ,  $\phi_{(1)} = -0.5$  and  $\phi_{(2)} = 1.5$ . Under this parameter constellation it is easily seen that  $w_{(1)} > w_{(3)} > w_{(2)}$ .

Figure 1: Weights for  $T = 100$ ,  $b_1 = 0.3$ ,  $b_2 = 0.6$ ,  $\phi_{(1)} = -0.5$  and  $\phi_{(2)} = 1.5$



Note also that the weights  $w_{(1)}$  and  $w_{(2)}$  can be negative. We do not restrict the weights to be positive as the weights (33), (34), and (35) give a unique minimum.

In the case of  $n$  breaks, the weights for the  $n + 1$  segments are given by

$$w_{(i)|i \leq n} = \frac{1}{T} \frac{1 + T \sum_{j=1, j \neq i}^n (b_j - b_{j-1}) \phi_{(j)}^2 - T \phi_{(i)} \sum_{j=1, j \neq i}^n (b_j - b_{j-1})}{a_{a,n}}$$

$$w_{(n+1)} = \frac{1}{T} \frac{1 + T \sum_{j=1}^n (b_j - b_{j-1}) \phi_{(j)}^2}{a_{a,n}}$$

where  $a_{a,n} = 1 + T \sum_{l=1}^{n+1} (b_l - b_{l-1}) \sum_{j=1, j \neq l}^n \phi_{(j)}^2 (b_j - b_{j-1}) - T \sum_{l=1}^n \phi_{(l)} (b_l - b_{l-1}) \sum_{j=1, j \neq l}^n \phi_{(j)}^2 (b_j - b_{j-1})$  and  $b_0 = 0$ .

### 3 Optimal weights when the time and size of the break are uncertain

So far we have assumed that the time and the size of the break are known. However, this may not be the case in many situations of practical interest. In particular, the size of the break is difficult to estimate unless a relatively large number of post-break observations is available.<sup>6</sup> It is, therefore, worthwhile to develop weights that are reasonably robust to the point and the size of the break(s). As a simple example, consider the model with a single break at time  $T_b$  both in the slopes and the error variances. Using (22) and (23) we first note that

$$Tw_{(a)} = \frac{1}{b + (1 - b)q^2 + Tb(1 - b)\phi^2},$$

$$Tw_{(b)} = \frac{q^2 + Tb\phi^2}{b + (1 - b)q^2 + Tb(1 - b)\phi^2}.$$

<sup>6</sup>Also in finite samples distribution of the estimated break point does not have a closed form expression and depends on the distribution of  $x_t$  and  $\varepsilon_t$ . (See Hinkley (1970)). Asymptotic results can be obtained that do not depend on the distribution of the regressors or the error term (e.g. Bai 1997), but such results might not be reliable in small samples.

where  $\phi^2 = \lambda^2 \hat{\omega}_x^2$ , with  $\hat{\omega}_x^2 = T^{-1} \sum_{t=1}^T x_t^2$ . The time profile of the weights can be written as

$$Tw_t(b, q^2, \phi^2) = w_{(b)} + [w_{(a)} - w_{(b)}] \mathbf{I}(T_b - t)$$

for  $t = 1, 2, \dots, T$ . Hence

$$Tw(a, b, q^2, \phi^2) = \frac{\frac{q^2}{T} + b\phi^2}{\frac{b+(1-b)q^2}{T} + b(1-b)\phi^2} + \left[ \frac{\frac{1-q^2}{T} - b\phi^2}{\frac{b+(1-b)q^2}{T} + b(1-b)\phi^2} \right] \mathbf{I}(b - a), \quad (36)$$

where  $a = t/T \in [0, 1]$ , and as before  $b = T_b/T \in [0, \bar{b}]$ , where  $\bar{b} < 1$ .

Initially, consider the case where the break is in the error variances only, namely  $\phi = 0$  and  $q^2 \neq 1$ , Then

$$Tw(a, b, q^2) = \frac{q^2}{b + (1-b)q^2} + \left[ \frac{1 - q^2}{b + (1-b)q^2} \right] \mathbf{I}(b - a),$$

or

$$Tw(a, b, q^2) = \frac{1}{1 + b\psi} + \left( \frac{\psi}{1 + b\psi} \right) \mathbf{I}(b - a).$$

where  $\psi = (1 - q^2)/q^2 = (\sigma_{(2)}^2 - \sigma_{(1)}^2)/\sigma_{(1)}^2$ . It is also worth noting that  $w_{(a)}/w_{(b)} = 1 + \psi = \sigma_{(2)}^2/\sigma_{(1)}^2$ , and more weights will be given to pre break observations if  $\sigma_{(2)}^2 > \sigma_{(1)}^2$ , and *vice versa*. This is in line with the result obtained by Pesaran and Timmermann (2007) using the concept of the optimal window.

In situations where  $b$  and  $q^2$  are uncertain their effects on the optimal weights can be integrated out with respect to a given distribution of  $b$  and  $q^2$ . Here, we assume that  $b$  and  $q^2$  are independently distributed and focus on the uncertainty of  $b$  for a given value of  $q^2$  or  $\psi$ . For  $b$  we assume that it is uniformly distributed over the range  $\underline{b}$  and  $\bar{b}$ , namely the probability density of  $b$  is given by

$$f(b) = \begin{cases} 0 & \text{if } b < \underline{b} \\ (\bar{b} - \underline{b})^{-1} & \text{if } \underline{b} \leq b < \bar{b} \\ 0 & \text{if } b \geq \bar{b}. \end{cases}$$

The expression for  $w(a, q^2)$  depends on whether  $a$  falls within the range  $[\underline{b}, \bar{b}]$  or not. Specifically we have,

$$Tw(a, q^2) = \begin{cases} (\bar{b} - \underline{b})^{-1} \int_{\underline{b}}^{\bar{b}} \frac{1+\psi}{1+b\psi} db & \text{if } a < \underline{b} \\ (\bar{b} - \underline{b})^{-1} \int_{\underline{b}}^{\bar{b}} \frac{1}{1+b\psi} db + \frac{\psi}{\bar{b}-\underline{b}} \int_a^{\bar{b}} \frac{1}{1+b\psi} db & \text{if } \underline{b} \leq a \leq \bar{b} \\ (\bar{b} - \underline{b})^{-1} \int_{\underline{b}}^{\bar{b}} \frac{1}{1+b\psi} db & \text{if } a > \bar{b} \end{cases}$$

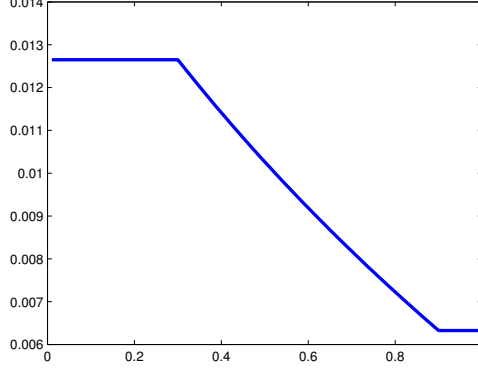
Also, it is easily seen that

$$\int_{\underline{b}}^{\bar{b}} \frac{1}{1+b\psi} db = \psi^{-1} \log \left( \frac{1 + \bar{b}\psi}{1 + \underline{b}\psi} \right),$$

and hence

$$Tw(a, q^2) = (\bar{b} - \underline{b})^{-1} \left[ \psi^{-1} \log \left( \frac{1 + \bar{b}\psi}{1 + \underline{b}\psi} \right) + \log \left( \frac{1 + \bar{b}\psi}{1 + a\psi} \right) \right], \quad \text{if } \underline{b} \leq a \leq \bar{b}.$$

Figure 2: Robust optimal weights (37) for break in variance,  $T = 100, q^2 = 0.5, \underline{b} = 0.3, \bar{b} = 0.9$



Since  $\psi = (1 - q^2)/q^2$ , we can also write

$$Tw(a, q^2) = \begin{cases} (\bar{b} - \underline{b})^{-1} \frac{1}{1-q^2} \log \left( \frac{\bar{b} + (1-\bar{b})q^2}{\underline{b} + (1-\underline{b})q^2} \right) & \text{if } a < \underline{b} \\ (\bar{b} - \underline{b})^{-1} \left[ \frac{q^2}{1-q^2} \log \left( \frac{\bar{b} + (1-\bar{b})q^2}{\underline{b} + (1-\underline{b})q^2} \right) + \log \left( \frac{\bar{b} + (1-\bar{b})q^2}{a + (1-a)q^2} \right) \right] & \text{if } \underline{b} \leq a \leq \bar{b} \\ (\bar{b} - \underline{b})^{-1} \frac{q^2}{1-q^2} \log \left( \frac{\bar{b} + (1-\bar{b})q^2}{\underline{b} + (1-\underline{b})q^2} \right) & \text{if } a > \bar{b} \end{cases} \quad (37)$$

Over the range  $\underline{b} \leq a \leq \bar{b}$

$$\frac{T \partial w(a, q^2)}{\partial a} = (\bar{b} - \underline{b})^{-1} \frac{-(1 - q^2)}{a + (1 - a)q^2} = \frac{1}{\bar{b} - \underline{b}} \frac{\sigma_{(1)}^2 - \sigma_{(2)}^2}{\sigma_{(2)}^2 [a + (1 - a)q^2]}$$

and the optimal weights  $w(a, q^2)$  monotonically rise (fall) with  $a$  if  $\sigma_{(1)}^2 > \sigma_{(2)}^2$  ( $\sigma_{(1)}^2 < \sigma_{(2)}^2$ ). In other words, more weights will be placed on more recent observations only if post break error variance is smaller than pre-break error variance. This result holds for all values of  $T$ . Figure 2 shows the optimal weights for  $T = 100, q^2 = 1/2, \underline{b} = 0.3$  and  $\bar{b} = 0.9$ .

The expression for  $w(a, q^2)$  simplifies further if we assume that the break point could be any point within the range  $(0, 1)$ . For this case we have

$$Tw(a, q^2) = \frac{-q^2}{1 - q^2} \log(q^2) - \log[a + (1 - a)q^2], \quad 0 \leq a \leq 1.$$

A discrete time approximation is given by

$$w_t(q^2) = \frac{1}{T} \frac{-q^2}{1 - q^2} \log(q^2) - \frac{1}{T} \log \left[ \frac{t}{T} + \left(1 - \frac{t}{T}\right) q^2 \right], \quad t = 1, 2, \dots, T.$$

As to be expected,  $\lim_{q \rightarrow 1} Tw(a, q^2) = 1$ . Furthermore, it is easily seen that  $w(a, q^2) = w(1 - a, 1/q^2)$ . Using these results the uncertainty regarding the value of  $q^2$  can be integrated out by assuming a suitable density for  $q^2$ . We shall not pursue this idea, since we will show below that the effects of breaks in error variances are dominated by the potential effects of breaks in slope coefficients.

Consider now the more general case where  $\phi^2 > 0$ . Using (36) we have for  $\underline{b} < a < \bar{b}$

$$Tw(a, b, q^2, \phi^2) = \frac{\frac{q^2}{\phi^2 T} + b}{\frac{b+(1-b)q^2}{T\phi^2} + b(1-b)} + \frac{\frac{1-q^2}{\phi^2 T} - b}{\frac{b+(1-b)q^2}{T\phi^2} + b(1-b)} I(b-a),$$

For given values of  $q^2$  and  $\phi^2$  and assuming that  $b$  lies in the range  $[\underline{b}, \bar{b}]$  with  $0 < \underline{b} < \bar{b} < 1$ , we have for  $\underline{b} < a < \bar{b}$

$$\begin{aligned} Tw(a | q^2, \phi^2) &= \int_{\underline{b}}^{\bar{b}} \frac{\frac{q^2}{\phi^2 T} + b}{\frac{[q^2+(1-q^2)b]}{T\phi^2} + b(1-b)} db \\ &+ \int_a^{\bar{b}} \frac{\frac{1-q^2}{\phi^2 T} - b}{\frac{[b+(1-b)q^2]}{T\phi^2} + b(1-b)} db, \end{aligned} \quad (38)$$

This result further simplifies when  $q^2 = 1$ , and we have (for  $\underline{b} < a < \bar{b}$ )

$$Tw(a | \phi^2, q^2 = 1) = \int_{\underline{b}}^{\bar{b}} \frac{1 + \phi^2 T b}{1 + b(1-b)T\phi^2} db - \phi^2 \int_a^{\bar{b}} \frac{bT\phi^2}{1 + b(1-b)T\phi^2} db.$$

It is now easily seen that

$$T \frac{\partial w(a | \phi^2, q^2 = 1)}{\partial a} = \frac{a\phi^2}{T^{-1} + a(1-a)\phi^2} > 0,$$

namely, for all values of  $\phi^2 > 0$ ,  $w(a | \phi^2, q^2 = 1)$  is a monotonically increasing function of  $a$  with the observations farthest from the end of the sample getting the smallest weights. The decay rate of the weights depends on  $T$ . When  $q^2 \neq 1$ , we have

$$T \frac{\partial w(a | \phi^2, q^2)}{\partial a} = \frac{-\left(\frac{1-q^2}{\phi^2 T} - a\right)}{\frac{[a+(1-a)q^2]}{T\phi^2} + a(1-a)}.$$

In this more general case the weights increase monotonically in  $a$  if  $a > \frac{1-q^2}{\phi^2 T}$ , which is clearly satisfied if  $q^2 > 1$ .

### 3.1 Large $T$ approximation

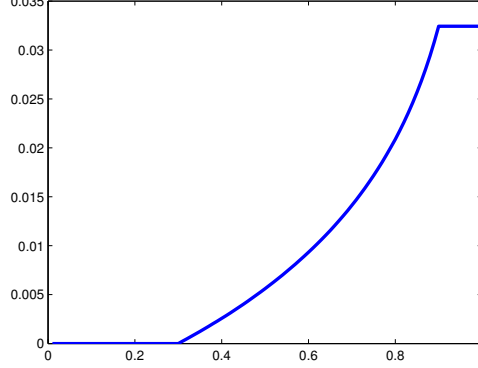
Consider now a large  $T$  approximation of the optimal weights and note that

$$Tw(a, b, q^2, \phi^2) = \frac{\frac{q^2}{\phi^2 T} + b}{b(1-b)\left(1 + \frac{\theta}{T}\right)} + \frac{\left(\frac{1-q^2}{\phi^2 T} - b\right) I(b-a)}{b(1-b)\left(1 + \frac{\theta}{T}\right)},$$

where  $\theta = [q^2 + (1-q^2)b] / \phi^2 b(1-b) > 0$ . Using  $\left(1 + \frac{\theta}{T}\right)^{-1} = 1 - \frac{\theta}{T} + O(T^{-2})$ , and replacing  $\theta$  in terms of  $b, q$ , and  $\phi$ , we have

$$\begin{aligned} Tw(a, b, q^2, \phi^2) &= \frac{1}{1-b} - \frac{1}{1-b} I(b-a) + \frac{1}{T} \left[ \frac{q^2}{\phi^2 b(1-b)} - \frac{q^2 + (1-q^2)b}{\phi^2 b(1-b)^2} \right] \\ &+ \frac{1}{T} \left[ \frac{1-q^2}{\phi^2 b(1-b)} + \frac{q^2 + (1-q^2)b}{\phi^2 b(1-b)^2} \right] I(b-a) + O(T^{-2}). \end{aligned}$$

Figure 3: Robust optimal weights (40) for break in variance,  $T = 100$ ,  $\underline{b} = 0.3$ ,  $\bar{b} = 0.9$



But since

$$\frac{q^2}{\phi^2 b(1-b)} - \frac{q^2 + (1-q^2)b}{\phi^2 b(1-b)^2} = \frac{q^2(1-b) - q^2 - (1-q^2)b}{\phi^2 b(1-b)^2} = \frac{-1}{\phi^2(1-b)^2}$$

$$\frac{1-q^2}{\phi^2 b(1-b)} + \frac{q^2 + (1-q^2)b}{\phi^2 b(1-b)^2} = \frac{(1-q^2)(1-b) + q^2 + (1-q^2)b}{\phi^2 b(1-b)^2} = \frac{1}{\phi^2 b(1-b)^2}$$

the weights profile simplifies to

$$Tw(a, b, q^2, \phi^2) = \frac{1}{1-b} - \frac{1}{1-b} \mathbb{I}(b-a) \quad (39)$$

$$- \frac{1}{T} \left[ \frac{1}{\phi^2(1-b)^2} \right] + \frac{1}{T} \left[ \frac{1}{\phi^2 b(1-b)^2} \right] \mathbb{I}(b-a) + O(T^{-2}).$$

It is interesting that the first order term in this expansion does not depend on the sizes of the breaks, and depends only on the break point,  $b$ . Also, the terms up to order  $T^{-1}$  are independent of  $q^2$  as long as  $\phi^2 > 0$ , that is, a break in the error variance is dominated by a break in the mean of the process.

Therefore, for large  $T$ , robust optimal weights are determined by the distribution of  $b$ . Here, we assuming that  $b \sim \text{Uniform}(\underline{b}, \bar{b})$  with  $0 < \underline{b} < \bar{b} < 1$ , and obtain

$$Tw(a) = \begin{cases} 0 + O(T^{-1}), & \text{for } a < \underline{b} \\ (\bar{b} - \underline{b})^{-1} \int_{\underline{b}}^{\bar{b}} \frac{1}{1-b} db - (\bar{b} - \underline{b})^{-1} \int_a^{\bar{b}} \frac{1}{1-b} db + O(T^{-1}), & \text{for } \underline{b} \leq a \leq \bar{b} \\ (\bar{b} - \underline{b})^{-1} \int_{\underline{b}}^{\bar{b}} \frac{1}{1-b} db + O(T^{-1}), & \text{for } a > \bar{b} \end{cases}$$

and

$$w(a) \approx \begin{cases} 0, & \text{if } a < \underline{b} \\ \frac{-1}{T(\bar{b}-\underline{b})} \log\left(\frac{1-a}{1-\underline{b}}\right), & \text{if } \underline{b} \leq a \leq \bar{b} \\ \frac{-1}{T(\bar{b}-\underline{b})} \log\left(\frac{1-\bar{b}}{1-\underline{b}}\right), & \text{if } a > \bar{b} \end{cases} \quad (40)$$

Figure 3 shows the approximate optimal weights for  $T = 100$ ,  $\underline{b} = 0.3$  and  $\bar{b} = 0.9$ , assuming that  $\phi^2 > 0$ .

In the case where  $\underline{b}$  and  $\bar{b}$  are close to the end points of 0 and 1, we have

$$w(a) \approx \frac{-\log(1-a)}{T}, \quad a \in [0, \bar{b}]. \quad (41)$$

A discrete time version can be obtained by setting  $T\bar{b} = T - 1$ , or  $\bar{b} = 1 - 1/T$ .<sup>7</sup> Namely,

$$w_t^* = \frac{-\log(1-t/T)}{T-1}, \quad \text{for } t = 1, 2, \dots, T-1 \quad (42)$$

$$w_T^* = \frac{-1}{T-1} \log\left(1 - \frac{T-1}{T}\right) = \frac{\log(T)}{T-1} \quad (43)$$

Due to the approximation/discretization these weights do not sum to unity, and can be scaled as

$$w_t = \frac{w_t^*}{\sum_{s=1}^T w_s^*}, \quad \text{for } t = 1, 2, \dots, T. \quad (44)$$

For  $\underline{b}$  and  $\bar{b}$  close to the end points of 0 and 1 and using calculations detailed in Appendix A.3, we can also obtain the MSFE implied by the robust weights (40) as

$$\begin{aligned} \frac{\omega_x^2}{x_{T+1}^2} \left[ E\left(\sigma_{(2)}^{-2} e_{T+1}^2\right) - 1 \right] &\approx \phi^2 [b + (1-b) \log(1-b)]^2 \\ &+ \frac{(q^2 - 1)}{T} \left[ -(1-b) [\log(1-b)]^2 + 2(1-b) \log(1-b) + 2b \right] + \frac{2}{T} \end{aligned} \quad (45)$$

Comparing this result to the equal weight MSFE we have

$$\frac{\omega_x^2}{x_{T+1}^2} (\text{MSFE}_{eq.} - \text{MSFE}_{robust}) = \phi^2 \psi_\phi(b) + \frac{(q^2 - 1)}{T} \psi_q(b) - \frac{1}{T},$$

where

$$\begin{aligned} \psi_\phi(b) &= \left[ b^2 - [b + (1-b) \log(1-b)]^2 \right] \\ &= [2b + (1-b) \log(1-b)] [-(1-b) \log(1-b)], \end{aligned}$$

and

$$\psi_q(b) = (1-b) [\log(1-b)]^2 - 2(1-b) \log(1-b) - b$$

Consider the case where  $q = 1$ . Then

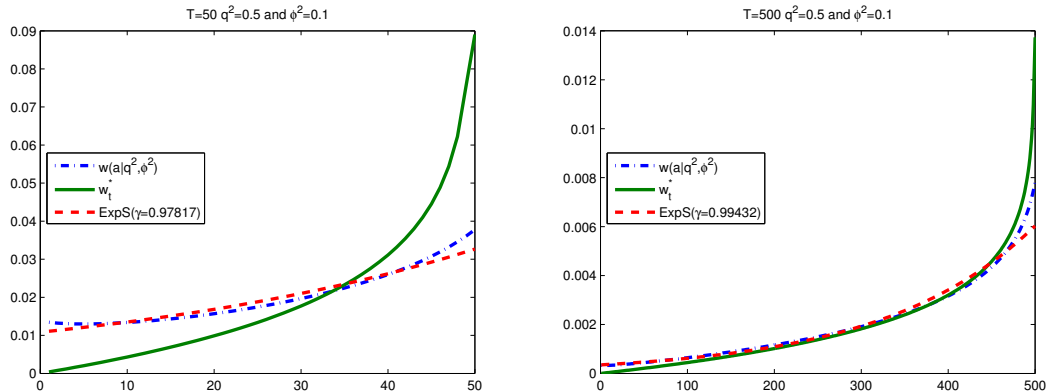
$$\frac{\omega_x^2}{x_{T+1}^2} (\text{MSFE}_{eq.} - \text{MSFE}_{robust}) = \phi^2 \psi_\phi(b) - \frac{1}{T}.$$

It is easily seen that for values of  $0 \leq b \leq \bar{b}$ ,  $\psi(b) > 0$  and attains its maximum at  $b = 0.80$  giving  $\psi(0.80) = 0.41$ . For relatively large  $T$  breaks the robust optimal weights will dominate the equal weights in MSFE. Only in cases where  $T$  and  $\phi$  are small and  $b$  close to the beginning of the sample one would expect the equal weights to perform better than the robust weights.

<sup>7</sup>Clearly, one could set  $\bar{b}$  to other values close to 1, say  $1 - 0.5/T$ . But for relatively large  $T$ , the choice of  $w_T^*$  for the forecasts is unlikely to be of great importance.



Figure 4: Comparison of optimal weights, robust optimal weights, and fitted exponential smoothing weights,  $\phi^2 = 0.1$  various  $q^2$  and  $T$



When  $q \neq 1$  the relative performance of the two sets of weights depend on the sign of  $(q^2 - 1)\psi_q(b)$ . It can be shown that  $\psi_q(b) > 0$  if  $b \leq 0.91$ , and negative otherwise. However, for reasonable values of  $q^2$  (say 1/2 or 2), the term  $\frac{(q^2-1)}{T}\psi_q(b)$  is relatively unimportant when  $T$  is 100 or more. Note that  $\max_{0 \leq b \leq 0.95} |\psi_q(b)| = 0.202$  and for  $T = 100$  the contribution of  $\frac{(q^2-1)}{T}\psi_q(b)$  to the relative performance of the two weights can be ignored, unless  $\phi$  is very small and  $b$  very close to 0 or 1.

It is also interesting to explore the fit of the robust weights and the ExpS weights to the optimal weights for a range of  $T$ ,  $q^2$  and  $\phi^2$ . Figures 4 and 5 plot the optimal weights  $w(a|q^2, \phi^2)$  in (38), the robust weights,  $w_t$ , in (44) and the exponentially smoothing (ExpS) weights used in (8), where  $\gamma$  is now set so that the distance between  $w(a|q^2, \phi^2)$  and  $w_t^\epsilon(\gamma)$  is minimized. The plots show that the accuracy of the robust optimal weights depends largely on  $\phi^2$ : for larger  $\phi^2$  the robust weights are very close to the optimal weights, for the smaller  $\phi^2$  a good approximation requires large  $T$ . The plots also show that, as predicted by our theory,  $q^2$  has a very minor influence that is only visible when  $T$  and  $\phi^2$  are both small. Finally, the down-weighting parameter  $\gamma$  in the exponential smoothing weight that best approximates the exact optimal weight varies between 0.944 and 0.994, and the ExpS weights generally give too low a weight to the most recent observations as compared to the optimal weights.

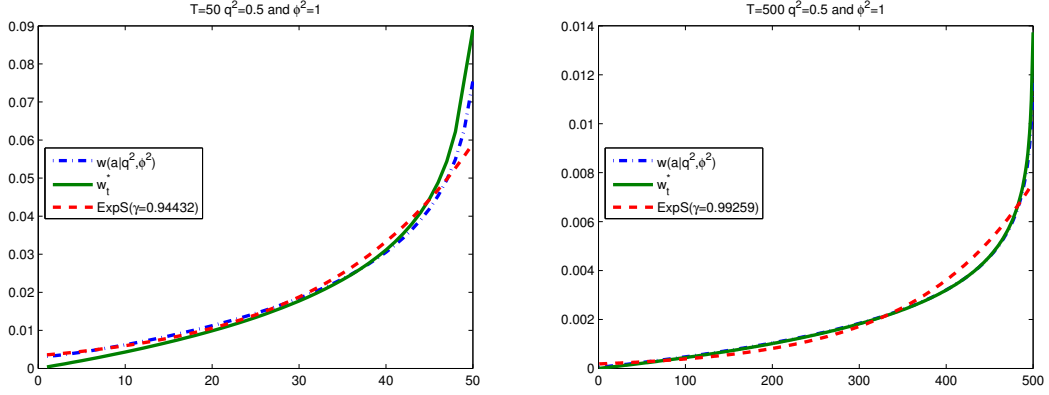
Robust weights that allow for high order terms in the expansion (39) are provided in the Appendix A.4. However, we will not pursue them further in this paper.

### 3.1.1 Robust weights for regression models with two breaks

Consider the case of two breaks, where the weights conditional on  $b$  and  $\lambda$  are given in (33) to (35). Clearly,  $\underline{b} < b_1 < b_2 < \bar{b}$  and  $\Pr(b_1, b_2) = \Pr(b_1)\Pr(b_2|b_1)$ , further  $\underline{b}_1 < b_1 < \bar{b}_1$  and  $\underline{b}_2 < b_2 < \bar{b}_2$  where  $\underline{b}_1 < \underline{b}_2$  and  $\bar{b}_1 < \bar{b}_2$ , then

$$\Pr(b_1) = \begin{cases} 0 & \text{if } b_1 < \underline{b}_1 \\ \frac{1}{\bar{b}_1 - \underline{b}_1} & \text{if } \underline{b}_1 < b_1 \leq \bar{b}_1 \\ 0 & \text{if } b_1 > \bar{b}_1 \end{cases}$$

Figure 5: Comparison of optimal weights, robust optimal weights, and fitted exponential smoothing weights,  $\phi^2 = 1$  various  $q^2$  and  $T$



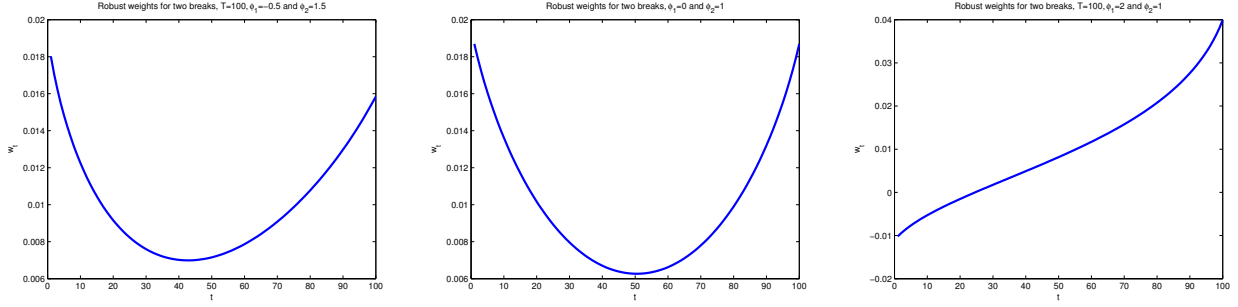
$$\Pr(b_2|b_1) = \begin{cases} 0 & \text{if } b_2 < \underline{b}_2 \\ \frac{1}{b_2 - b_1} & \text{if } \underline{b}_2 < b_2 \leq \bar{b}_2 \\ 0 & \text{if } b_2 > \bar{b}_2 \end{cases}$$

Analytic solutions for the robust weights under two breaks are not easy to obtain. However, we can obtain numerical solutions using (33) to (35) and integrating over a grid for  $b_1$  and  $b_2$  taking into account that  $b_1 < b_2$  and setting  $\underline{b}_1 = 1/T$ ,  $\underline{b}_2 = 2/T$ ,  $\bar{b}_1 = (T - 2)/T$  and, finally,  $\bar{b}_2 = (T - 1)/T$ .

Figure 6 plots the robust weights for two breaks and  $T = 100$ , where the first graph reports the weights for  $\phi_{(1)} = -0.5$  and  $\phi_{(2)} = 1.5$ , the second for  $\phi_{(1)} = 0$  and  $\phi_{(2)} = 1$ , the third for  $\phi_{(1)} = 2$  and  $\phi_{(2)} = 1$ . It can be seen that the shape of the weights depends on the parameters chosen. The first graph the parameters  $\phi_{(1)}$  and  $\phi_{(2)}$  are those that under known break dates resulted in the example in Figure 1 where the first subsample receives the largest weights. The pattern is the same with the very early observation receiving higher weights than the last observations. The second graph is for parameters that would lead to equal weights in the first and last subsample if the break dates were known. The final graph uses breaks that decrease in size, which results in continuously increasing weights.

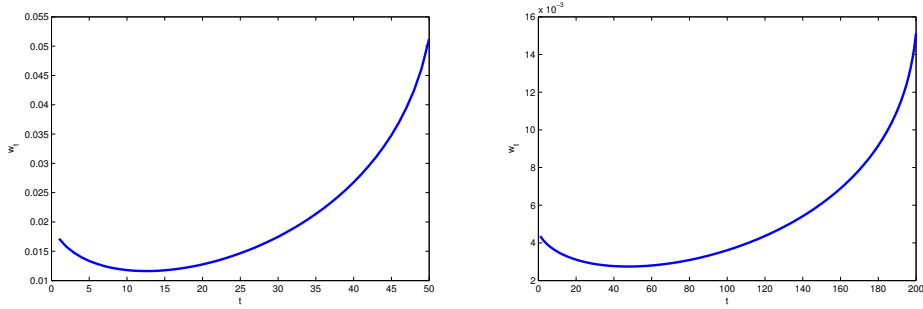
In practice, given that the break date is uncertain, the size of break is also likely to be unknown. In addition to the break date, we therefore also integrate over the break sizes in the weights. Figure 7 plots the weights when  $\phi_{(1)}$  and  $\phi_{(2)}$  are integrated with respect to a uniform distribution in the range  $-2$  to  $2$ . The first graph shows the weights for  $T = 50$  and the second for  $T = 200$ . It can be seen that the shape of the weight function is largely independent of the sample size. Most weight is given to the most recent observations. Interestingly, the first observations receive a higher weight than the observations in the middle of the sample, which reflects the possibility that the two weights could counteract each other.

Figure 6: Robust weights for two breaks,  $T = 100$ , and different values of  $\phi_{(1)}$  and  $\phi_{(2)}$



Note: The first graph plots the weights for  $\phi_{(1)} = -0.5$  and  $\phi_{(2)} = 1.5$ , the second for  $\phi_{(1)} = 0$  and  $\phi_{(2)} = 1$ , and the third for  $\phi_{(1)} = 2$  and  $\phi_{(2)} = 1$ . The weights are given in (33) to (35) and integrating uniformly over  $b_1$  and  $b_2$  over the range  $1/T$  to  $(T - 1)/T$ .

Figure 7: Robust weights for two breaks and  $\phi_{(1)}$  and  $\phi_{(2)}$  integrated out,  $T = 50$  and 200



Note: The first graph plots the weights for  $T = 50$  and the second for  $T = 200$ . The weights are given in (33) to (35) and integrating uniformly over  $b_1$  and  $b_2$  over the range  $1/T$  to  $(T - 1)/T$  and  $\phi_{(1)}$  and  $\phi_{(2)}$  over the range  $-2$  to  $2$ .

## 4 Monte Carlo evidence on the forecast performance

### 4.1 Data generating process

We now turn to evaluating the performance of the forecasting methods discussed above in a range of Monte Carlo experiments. The first series of experiments concern the continuous break model (2) in Section 2.1. A second experiment concentrates on the random walk model (2) with a discrete break discussed in Section 2.2. In this model the MSFEs of the different forecast models are known conditional on  $T_b$  and  $\lambda$  and have been reported in Table 1. The Monte Carlo experiments will show how much the uncertainty around the break date and size affects the forecasts. In a final set of experiments we add a regressor in the design using the simple linear regression model discussed in Section 2.3.1.

The first model considered is

$$y_t = \mu_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1) \quad (46)$$

where the mean follows a random walk

$$\mu_t = \mu_{t-1} + \sigma_v v_t, \quad v_t \sim N(0, 1)$$

and  $t = 1, 2, \dots, T, T + 1$  with  $T = 50, 100, 200$ , and  $\gamma = \{0.8, 0.9, 0.95, 0.98\}$ , which corresponds to  $\delta = \sigma_\varepsilon/\sigma_v \approx \{4.471, 9.487, 19.494, 49.497\}$ .

Next, we assume that the mean in (46) has a discrete break

$$\mu_t = \begin{cases} \mu_{(1)} & t \leq T_b \\ \mu_{(2)} & t > T_b \end{cases} \quad \text{and} \quad \sigma_t = \begin{cases} \sigma_{(1)} & t \leq T_b \\ \sigma_{(2)} & t > T_b \end{cases}$$

and  $t = 1, 2, \dots, T, T + 1$  with  $T = 50, 100, 200$ . We set  $b = \{0.95, 0.9\}$ ,  $\lambda = (\mu_{(1)} - \mu_{(2)})/\sigma_{(2)} = \{0.5, 1, 2\}$  and  $q = \sigma_{(1)}/\sigma_{(2)} = \{0.5, 1\}$ . We assume that  $T_b$ ,  $\lambda$  and  $q$  are unknown and have to be estimated.

The third model adds a regressor, such that

$$y_t = \beta_t x_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1)$$

where

$$\beta_t = \begin{cases} \beta_{(1)} & t \leq T_b \\ \beta_{(2)} & t > T_b \end{cases} \quad \text{and} \quad \sigma_t = \begin{cases} \sigma_{(1)} & t \leq T_b \\ \sigma_{(2)} & t > T_b \end{cases}$$

we use the same values for  $b$  and  $\lambda$  as in the second experiment but restrict attention to  $q = 1$ . Forecasts are conditional on  $x_{T+1}$ .

Forecasts based on the full estimation window with equal weights will serve as the base line to which all other forecast methods are compared. We also include forecasts based on the optimal weights that use the true parameter values of the break process as another baseline method. This baseline provides an upper bound to the performance of the feasible forecasting methods. For model (46) with the continuous breaks the weights are given by (3), for the model with discrete breaks the weights are given in (11) and (12), and for the simple regression model they are given by (22) and (23).

We then estimate  $\gamma$  from an MA(1) in first differences, which also yields an estimate of  $\delta$  via (9). The MA(1) estimation at times suffers from a relatively flat likelihood, and for this reason we restrict this estimation to the random walk model. We also report the MSE of the estimation of  $\gamma$ . We forecast the model with weights (3) using the estimated  $\hat{\delta}$  and by ExpS with weights (7) using  $\hat{\gamma}$ . For comparison we also add ExpS forecasts based on  $\gamma = 0.95$  and  $\gamma = 0.98$ .

We use the Bai and Perron (1998, 2003) procedure to estimate the break dates,  $\mathbf{b} = (b_1, b_2)'$ , and conditional on these dates the break sizes,  $\lambda = (\lambda_{(1)}, \lambda_{(2)})'$ . We then use these estimates to compute feasible forecasts based on the optimal weights (11) and (12) in the random walk model or (22) and (23) in the simple linear regression model with  $\hat{\mathbf{b}}$  and  $\hat{\lambda}$  in place of  $b$  and  $\lambda$ . For the DGP with continuous breaks we allow for two breaks, for the DGP with a discrete break we restrict attention to testing for one break.

We also make forecasts using the robust weights developed above. First, we assume that the forecaster uses the information that the break is in the last quarter but not in the last 2% of the sample. The corresponding weights are given by (40). Second, we assume that break dates in the full sample are equally likely with the weights given in (44). Finally, in the experiments with continuous break

process we use robust weights assuming two breaks, where the weights are calculated numerically integrating over  $(b_1 > b_2)$  and  $\phi_{(1)}$  and  $\phi_{(2)}$  over the range  $-2$  to  $2$ .

For comparison, we construct forecasts based on the observations after the estimated break date and using optimal estimation window based on the estimated break date and size. Given uncertainty over the break dates, we also average over estimation windows with minimum window  $v_{\min} = 0.05$ .

Using each of these methods we construct one-period ahead forecasts and base comparisons on the MSFE. We report the ratio of MSFEs relative to that of the forecasts using equal weights,  $\text{MSFE}_{eq.}$ , so that for method  $i$  we have

$$\text{rMSFE}_i = \frac{\text{MSFE}_i}{\text{MSFE}_{eq.}} \quad (47)$$

The results are based on 10,000 replications.

## 4.2 MC Results

**Continuous breaks DGP** Table 2 reports the results for the DGP with continuous breaks. The first line reports the results for the infeasible optimal weights forecasts based on the true  $\delta$ , which produces forecasts with the largest improvement in MSFE relative to the full sample equal weight forecasts.

The second and third line contains the results for the optimal and ExpS weights based on the estimated  $\hat{\delta}$  and  $\hat{\gamma}$ . As suggested by our theoretical results in Section 2.1, these two forecasts are identical. Table 3 reports the MSE of the estimation of  $\gamma$  across the experiments, and the MSE decreases foremost in  $T$  and to a lesser extent in  $\gamma$ . This is reflected in the forecast results, which are close to those for known  $\delta$  when  $T$  is large but less so when  $T = 50$ .

The next two lines report the results for  $\gamma = 0.95$  and  $0.98$ , which are set *a priori*. The benefit of using an *a priori* given value for  $\gamma$  versus estimating it depends on the deviation of the selected  $\gamma$  from the true value and on the estimation uncertainty of  $\hat{\gamma}$ , that is, on  $T$ . ExpS with fixed parameter  $\gamma = 0.95$  improves on its estimated counterparts when  $\gamma$  is at least  $0.9$  for  $T = 50$  and  $100$  and  $\gamma = 0.95$  for  $T = 200$  as it avoids estimation uncertainty but it can result in considerably worse forecasts when  $\gamma$  is set below  $0.95$ . ExpS with  $\gamma = 0.98$  improves on the estimated counterpart when true  $\gamma$  is at least  $0.95$  when  $T = 50$  and  $100$  but for larger  $T = 200$  it required the true  $\gamma$  to be  $0.98$ . These results clearly show the sensitivity of ExpS to the *a priori* choice of  $\gamma$ , even if the underlying break process is continuous as the theory of ExpS requires.

Amongst the methods that assume a discrete break, the robust weights generally perform best. For smaller  $\gamma$  the robust weights integrating  $b$  over the range  $\underline{b} = 0.75$  and  $\bar{b} = 0.98$  deliver the best forecasts. Notably, for true values of  $\gamma = 0.8$  and  $0.9$  and  $T = 50$ , and for  $\gamma = 0.9$  and  $T = 100$  they deliver the best forecasts of all feasible forecasts, including those that assume continuous breaks. For larger  $\gamma$ 's the robust weight forecast performance remains close to that of the methods that assume discrete breaks.

Forecasts based on optimal weights under the assumption of discrete breaks perform well when  $\gamma = 0.8$ , that is when  $\beta_t$  has a large variation. For larger  $\gamma$  the performance deteriorates and for the  $\gamma = 0.98$  it is generally worse than the

equally weighted forecast. The results for the forecast from the optimal window are similar to the optimal weights forecast. The post-break window forecast is the least favorable forecasting methods in this setting and often leads to the highest MSFE. In short, the robust weights proposed in the paper are likely to perform well even when the underlying break process is continuous so long as the break sizes per period are not too small.

Finally, the AveW forecasts perform well when  $\gamma = 0.95$  and  $\gamma = 0.99$ , and provide the best forecast for  $T = 50$  and  $\gamma = 0.95$  and the second best forecast for  $T = 100$  and  $\gamma = 0.98$ . However, when the true value of  $\gamma$  is small the AveW procedure performs poorly since it does not discount past observations heavily enough. It does, however, improve over ExpS forecast with fixed gamma when  $T$  is 50 and 100.

**Discrete breaks DGP** Table 4 contains the results when the break in the underlying random walk process is discrete. Amongst the feasible forecasts the relative performance depends foremost on the size of the break and then the sample size. The second line reports the results using the estimated optimal weights. When the break size,  $\lambda$ , is small, the detection of the break is difficult and using estimated optimal weights lead to forecasts with a higher MSFE than most other forecasting methods. However, when  $\lambda = 2$  the estimated optimal weights produce MSFEs that are among the smallest across all methods. The benefit of using optimal weights therefore depends on the ability to detect the break accurately.

The next two lines in the Table report the results for the robust weights. For the smaller breaks and  $T = 50$  and 100 the forecasts that use the information that the break is in the last quarter of the sample provide the best forecasts across all feasible methods. For  $T = 200$  it is second to the ExpS forecast with  $\gamma = 0.95$ . The robust forecasts that integrate  $b$  over the last quarter of the sample always perform better—and for larger breaks substantially—than those integrating over the entire sample, which shows how powerful this additional information is for the resulting forecasts. For large values of  $\lambda$  the robust weights still improve vastly over the equal weights forecast but not as much as the estimated optimal weights.

Forecasts based on post-break observations (with the break date estimated) have the highest MSFE when the break size is small, but their performance improves dramatically when  $\lambda$  is large, where the post break forecasts have MSFE's very similar to the ones obtained for the estimated optimal weights. The optimal window forecasts perform also quite similar to the estimated optimal weights forecasts, and their performance depends largely on the size of the break.

AveW forecasts perform well when  $T = 50$  and the break is small but less good when  $T$  is 100 or more. Still, in all examples, AveW offers substantial improvements over the full sample equal weight forecasts.

Forecasts that incorrectly assume the break process is continuous also reduce the MSFE relative to the full sample based forecasts, but as to be expected are generally less efficient than those based on weights derived assuming a discrete break DGP. However, as  $T$  increases these methods improve.

The results in Table 5 show that the influence of a break in the error variance is of negligible importance for the forecasts, which confirms our theoretical results.

Table 6 reports the results for the simple linear regression model. The magnitude of the relative MSFE results are now also affected by the variance of the regressor,

Table 2: Monte Carlo results for the random walk model with continuous breaks

	$\gamma$	0.8	0.9	0.95	0.98
	$\delta$	4.472	9.487	19.494	49.497
<i>T</i> = 50					
opt.weight(cont.break; $\delta$ )		0.633	0.899	0.981	0.999
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.696	0.961	1.027	1.012
ExpS( $\hat{\gamma}$ )		0.696	0.961	1.027	1.012
ExpS( $\gamma = 0.95$ )		0.778	0.922	0.980	1.003
ExpS( $\gamma = 0.98$ )		0.897	0.962	0.988	0.999
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		0.698	1.007	1.095	1.103
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.649	0.905	1.006	1.038
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.725	0.908	0.982	1.010
robust weights(two breaks)		0.829	0.938	0.983	0.999
post-break obs. ( $\hat{b}$ )		0.704	1.026	1.118	1.127
opt.window( $\hat{b}, \hat{\lambda}$ )		0.684	0.968	1.055	1.061
AveW( $w_{\min} = 0.05$ )		0.744	0.912	0.980	1.007
<i>T</i> = 100					
opt.weight(cont.break; $\delta$ )		0.444	0.772	0.956	0.999
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.455	0.794	0.995	1.022
ExpS( $\hat{\gamma}$ )		0.455	0.794	0.995	1.022
ExpS( $\gamma = 0.95$ )		0.557	0.799	0.956	1.015
ExpS( $\gamma = 0.98$ )		0.744	0.885	0.968	1.000
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		0.510	0.856	1.085	1.121
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.508	0.781	0.963	1.029
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.620	0.829	0.958	1.007
robust weights(two breaks)		0.761	0.888	0.969	1.000
post-break obs. ( $\hat{b}$ )		0.511	0.864	1.105	1.144
opt.window( $\hat{b}, \hat{\lambda}$ )		0.503	0.828	1.042	1.081
AveW( $w_{\min} = 0.05$ )		0.644	0.840	0.959	1.005
<i>T</i> = 200					
opt.weight(cont.break; $\delta$ )		0.285	0.630	0.879	0.982
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.290	0.642	0.893	0.997
ExpS( $\hat{\gamma}$ )		0.290	0.642	0.893	0.997
ExpS( $\gamma = 0.95$ )		0.351	0.647	0.879	0.995
ExpS( $\gamma = 0.98$ )		0.511	0.733	0.901	0.982
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		0.368	0.685	0.953	1.081
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.399	0.672	0.882	0.986
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.533	0.745	0.906	0.982
robust weights(two breaks)		0.699	0.833	0.937	0.987
post-break obs. ( $\hat{b}$ )		0.368	0.688	0.961	1.097
opt.window( $\hat{b}, \hat{\lambda}$ )		0.368	0.675	0.926	1.052
AveW( $w_{\min} = 0.05$ )		0.562	0.762	0.912	0.983

Note: The table reports the ratio of MSFE of forecast method *i* relative to that using equal weights,  $MSFE_i/MSFE_{\text{equal}}$ . The DGP is  $y_t = \beta_t + \sigma_\varepsilon \varepsilon_t$  where  $\beta_t = \beta_{t-1} + \sigma_v v_t$ ,  $\delta = \sigma_\varepsilon/\sigma_v$ , and  $\delta = \sqrt{\gamma}/(1 - \gamma)$ . Forecast methods: (i) infeasible optimal weights as function of  $\delta$ , (ii) optimal weights for continuous breaks where  $\delta$  is estimated from an MA(1) in the first difference of the data, (iii) ExpS with  $\gamma$  estimated from an MA(1) in the first difference of the data, (iv) ExpS with fixed  $\gamma = 0.95$  and (v)  $\gamma = 0.98$ . (vi) optimal weights based on point estimates of  $b$  and  $\lambda$  for up to two breaks, (vii) robust weights (40) with  $\underline{b} = 0.75$  and  $\bar{b} = 0.98$ , (viii) robust weights (44), (ix) robust weights for two breaks with  $\phi_{(1)}, \phi_{(2)} \in (-2, 2)$ , (x) post-break window, (xi) optimal window based on point estimates of  $b$  and  $\lambda$  for the last break, (xii) AveW forecasts with  $m = T(1 - v_{\min}) + 1$  windows and  $v_{\min} = 0.05$ . The results are based on  $R = 10,000$  repetitions.

Table 3: MSE of  $\hat{\gamma}$  in Monte Carlo experiments with continuous breaks

$T \setminus \gamma$	0.8	0.9	0.95	0.98
50	0.013	0.012	0.009	0.004
100	0.005	0.003	0.005	0.002
200	0.003	0.002	0.002	0.001

Note: The table reports the MSE of the estimation of the MA parameter  $\gamma$  in an MA(1) model in the Monte Carlo experiments reported in Table 2.

$x_t$ , although the relative ranking of the various forecasting methods is very similar to the ones reported above for the random walk model. Notable differences are that the robust weights now also deliver the best forecasts for the largest breaks when  $T = 50$  and 100 and the post-break window size is small. Also, the optimal weight forecasts now dominate the optimal window forecasts when  $\lambda = 1$ .

Overall, the Monte Carlo results suggest that when the break size is small and/or the sample is too small for an accurate estimation of the break process, the robust weights developed in this paper deliver the most precise forecasts. This is true for discrete as well as continuous break processes. When the break process is continuous and the sample large, estimated optimal weights and ExpS forecasts with estimated down-weighting parameter will result in the most precise forecasts for  $\gamma$  not too close to unity. If true  $\gamma$  is large, robust weight forecasts dominate even in large samples. ExpS forecasts can work well even when the break process is discrete, so long as the break is small and the sample large. However, this relies on *a priori* knowledge of the correct down-weighting parameter, which will not be available to the forecaster in real time. Discrete breaks that are large and easily identified imply that the optimal weight forecasts will provide the best forecasts.

## 5 Application to the yield curve as a predictor of real economic activity

### 5.1 The empirical model

The slope of the yield curve has emerged as a valuable leading indicator of GDP growth; see Stock and Watson (2003) for a survey of the literature. However, recent evidence suggests that the relationship between GDP growth and the yield curve may be subject to structural breaks (Estrella, Rodrigues and Schich 2003, Giacomini and Rossi 2006, Schrimpf and Wang 2010). We will use the forecasting methods discussed in the previous sections to investigate whether they can help in improving the forecasts of GDP growth using the slope of the yield curve as the predictor.

The forecasts are based on the regression model

$$y_{t,t+h} = \beta_0 + \beta_1 s_t + \varepsilon_t \quad (48)$$

where  $y_{t,t+h} = 100 \ln(Y_{t+h}/Y_t)$ ,  $Y_t$  is the level of real GDP in quarter  $t$ , and  $s_t = i_t^L - i_t^S$ , is the slope of the yield curve defined as difference between the long term



Table 4: Monte Carlo results for random walk model with a discrete break,  $q = 1$ 

	$b$		0.95			0.9	
	$\lambda$	0.5	1	2	0.5	1	2
$T = 50$							
opt.weight(disc.break; $b, \lambda$ )		0.923	0.653	0.284	0.910	0.634	0.276
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.040	0.873	0.428	1.040	0.842	0.342
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.948	0.782	0.604	0.927	0.717	0.480
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.956	0.857	0.751	0.940	0.810	0.662
post-break obs. ( $\hat{b}$ )		1.060	0.885	0.427	1.060	0.856	0.343
opt.window( $\hat{b}, \hat{\lambda}$ )		1.004	0.847	0.451	1.003	0.813	0.349
AveW( $w_{\min} = 0.05$ )		0.966	0.888	0.805	0.948	0.836	0.709
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.994	0.961	0.798	0.992	0.930	0.577
ExpS( $\hat{\gamma}$ )		0.994	0.961	0.798	0.992	0.930	0.577
ExpS( $\gamma = 0.95$ )		0.973	0.915	0.852	0.958	0.872	0.775
ExpS( $\gamma = 0.98$ )		0.989	0.969	0.947	0.983	0.951	0.916
$T = 100$							
opt.weight(disc.break; $b, \lambda$ )		0.893	0.603	0.256	0.875	0.592	0.257
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.022	0.826	0.320	1.014	0.737	0.263
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.934	0.796	0.648	0.901	0.705	0.480
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.953	0.867	0.775	0.931	0.805	0.662
post-break obs. ( $\hat{b}$ )		1.039	0.839	0.319	1.030	0.747	0.262
opt.window( $\hat{b}, \hat{\lambda}$ )		0.991	0.800	0.329	0.986	0.722	0.268
AveW( $w_{\min} = 0.05$ )		0.965	0.900	0.830	0.940	0.831	0.706
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.992	0.944	0.666	0.984	0.847	0.337
ExpS( $\hat{\gamma}$ )		0.992	0.944	0.666	0.984	0.847	0.337
ExpS( $\gamma = 0.95$ )		0.949	0.849	0.741	0.916	0.759	0.579
ExpS( $\gamma = 0.98$ )		0.980	0.944	0.905	0.963	0.899	0.826
$T = 200$							
opt.weight(disc.break; $b, \lambda$ )		0.869	0.571	0.238	0.862	0.577	0.248
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.010	0.711	0.245	0.984	0.618	0.249
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.924	0.788	0.643	0.892	0.697	0.474
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.949	0.863	0.771	0.928	0.802	0.658
post-break obs. ( $\hat{b}$ )		1.027	0.720	0.244	0.998	0.621	0.249
opt.window( $\hat{b}, \hat{\lambda}$ )		0.984	0.695	0.249	0.966	0.613	0.251
AveW( $w_{\min} = 0.05$ )		0.962	0.899	0.831	0.937	0.828	0.704
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.989	0.898	0.391	0.973	0.727	0.265
ExpS( $\hat{\gamma}$ )		0.989	0.898	0.391	0.973	0.727	0.265
ExpS( $\gamma = 0.95$ )		0.905	0.725	0.533	0.874	0.635	0.363
ExpS( $\gamma = 0.98$ )		0.954	0.876	0.793	0.926	0.797	0.651

Note: The table reports the relative MSFEs for the DGP  $y_t = \beta_t + \sigma_t \varepsilon_t$  with a break in  $\beta_t$  and  $\sigma_t$  at  $T_b$ . Here,  $q = \sigma_{(1)}/\sigma_{(2)} = 1$ . The first forecast method uses optimal weights for a discrete break with known  $b$  and  $\lambda$ . For the remaining forecast methods see Table 2.

Table 5: Monte Carlo results for random walk model with a discrete break,  $q = 0.5$

	$b$		0.95		0.9		
	$\lambda$	0.5	1	2	0.5	1	2
$T = 50$							
opt.weight(disc.break; $b, \lambda$ )		0.927	0.656	0.284	0.915	0.637	0.277
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.042	0.853	0.411	1.048	0.835	0.333
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.942	0.778	0.602	0.925	0.715	0.479
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.955	0.856	0.750	0.940	0.810	0.662
post-break obs. ( $\hat{b}$ )		1.065	0.864	0.410	1.073	0.850	0.334
opt.window( $\hat{b}, \hat{\lambda}$ )		1.002	0.824	0.426	1.006	0.803	0.335
AveW( $w_{\min} = 0.05$ )		0.964	0.887	0.804	0.947	0.835	0.709
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.997	0.939	0.648	0.999	0.895	0.424
ExpS( $\hat{\gamma}$ )		0.997	0.939	0.648	0.999	0.895	0.424
ExpS( $\gamma = 0.95$ )		0.972	0.914	0.852	0.958	0.872	0.774
ExpS( $\gamma = 0.98$ )		0.989	0.968	0.947	0.983	0.951	0.916
$T = 100$							
opt.weight(disc.break; $b, \lambda$ )		0.896	0.605	0.257	0.878	0.593	0.257
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.031	0.807	0.311	1.021	0.731	0.262
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.930	0.794	0.647	0.900	0.704	0.480
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.952	0.866	0.774	0.931	0.805	0.662
post-break obs. ( $\hat{b}$ )		1.054	0.821	0.311	1.042	0.742	0.261
opt.window( $\hat{b}, \hat{\lambda}$ )		0.994	0.780	0.316	0.990	0.714	0.265
AveW( $w_{\min} = 0.05$ )		0.964	0.899	0.830	0.939	0.830	0.706
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.990	0.910	0.469	0.978	0.769	0.289
ExpS( $\hat{\gamma}$ )		0.990	0.910	0.469	0.978	0.769	0.289
ExpS( $\gamma = 0.95$ )		0.946	0.847	0.740	0.916	0.758	0.579
ExpS( $\gamma = 0.98$ )		0.979	0.943	0.905	0.963	0.899	0.826
$T = 200$							
opt.weight(disc.break; $b, \lambda$ )		0.871	0.572	0.238	0.863	0.578	0.248
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.024	0.704	0.243	0.992	0.613	0.249
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.923	0.787	0.642	0.893	0.697	0.474
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.949	0.863	0.771	0.928	0.802	0.658
post-break obs. ( $\hat{b}$ )		1.047	0.714	0.243	1.010	0.616	0.248
opt.window( $\hat{b}, \hat{\lambda}$ )		0.991	0.686	0.245	0.971	0.607	0.249
AveW( $w_{\min} = 0.05$ )		0.962	0.898	0.831	0.937	0.828	0.704
estim.opt.weight(cont.break; $\hat{\delta}$ )		0.979	0.800	0.288	0.952	0.648	0.263
ExpS( $\hat{\gamma}$ )		0.979	0.800	0.288	0.952	0.648	0.263
ExpS( $\gamma = 0.95$ )		0.903	0.724	0.533	0.875	0.635	0.363
ExpS( $\gamma = 0.98$ )		0.954	0.876	0.793	0.926	0.797	0.651

Note: Here,  $q = \sigma_{(1)}/\sigma_{(2)} = 0.5$ . Otherwise see footnote of Table 4.

Table 6: Monte Carlo results for single regressor and a discrete break,  $q = 1$

	$b$		0.95			0.9	
	$\lambda$	0.5	1	2	0.5	1	2
$T = 50$							
opt.weight(disc.break; $b, \lambda$ )		0.979	0.853	0.542	0.971	0.832	0.520
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.005	0.978	0.851	1.009	0.952	0.631
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.981	0.898	0.753	0.975	0.869	0.673
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.980	0.925	0.836	0.975	0.907	0.783
post-break obs. ( $\hat{b}$ )		1.007	0.980	0.849	1.012	0.957	0.633
opt.window( $\hat{b}, \hat{\lambda}$ )		1.007	0.980	0.850	1.012	0.957	0.634
AveW( $w_{\min} = 0.05$ )		0.982	0.933	0.854	0.977	0.911	0.794
ExpS( $\gamma = 0.95$ )		0.985	0.950	0.896	0.981	0.933	0.851
ExpS( $\gamma = 0.98$ )		0.993	0.979	0.959	0.991	0.972	0.941
$T = 100$							
opt.weight(disc.break; $b, \lambda$ )		0.961	0.800	0.499	0.952	0.796	0.502
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.003	0.913	0.607	1.003	0.877	0.520
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.974	0.896	0.776	0.962	0.856	0.668
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.979	0.929	0.854	0.972	0.903	0.783
post-break obs. ( $\hat{b}$ )		1.006	0.916	0.608	1.006	0.881	0.520
opt.window( $\hat{b}, \hat{\lambda}$ )		1.006	0.916	0.608	1.006	0.881	0.520
AveW( $w_{\min} = 0.05$ )		0.983	0.941	0.880	0.974	0.911	0.800
ExpS( $\gamma = 0.95$ )		0.978	0.920	0.832	0.967	0.882	0.731
ExpS( $\gamma = 0.98$ )		0.990	0.967	0.934	0.984	0.948	0.885
$T = 200$							
opt.weight(disc.break; $b, \lambda$ )		0.955	0.786	0.473	0.945	0.793	0.485
estim.opt.weight(disc.break; $\hat{b}, \hat{\lambda}$ )		1.013	0.874	0.491	1.001	0.822	0.487
robust weights( $\underline{b} = 0.75, \bar{b} = 0.98$ )		0.972	0.894	0.757	0.957	0.853	0.649
robust weights( $\underline{b} = 0, \bar{b} = 1$ )		0.980	0.930	0.842	0.970	0.903	0.771
post-break obs. ( $\hat{b}$ )		1.018	0.878	0.491	1.006	0.823	0.486
opt.window( $\hat{b}, \hat{\lambda}$ )		1.018	0.878	0.491	1.006	0.823	0.486
AveW( $w_{\min} = 0.05$ )		0.984	0.945	0.878	0.973	0.913	0.796
ExpS( $\gamma = 0.95$ )		0.966	0.865	0.685	0.951	0.824	0.573
ExpS( $\gamma = 0.98$ )		0.982	0.936	0.856	0.969	0.901	0.766

Note: The results are for the simple linear regression model,  $y_t = \beta_t x_t + \sigma_t \varepsilon_t$  with a single break in  $\beta_t$  at  $T_b$ . For definitions and forecasting procedures see the footnote of Table 2 and 4.

interest rate,  $i_t^L$ , and the short term interest rate,  $i_t^S$ . This specification is the most common in the literature (e.g., Estrella and Hardouvelis 1991, Estrella and Mishkin 1997, and the literature cited above).

We evaluate the forecasts for horizons,  $h = 1, 2, 3, 4$  quarters. An issue involving direct forecasts with horizons greater than one is the overlap implicit in the regressions. Pesaran, Pick and Timmermann (2011) show that accounting for the overlap of observations can lead to gains in forecast accuracy but that these gains materialize at forecast horizons that are larger than the ones we consider here. In order not to complicate the forecast exercise further we restrict attention to estimation that do not account for the overlap when  $h > 1$ .

We take data on GDP, long and short term interest rates from the data set available with the GVAR toolbox (Smith and Galesi 2010). The data set contains quarterly observations for 33 countries. As not all countries have a long history in GDP and interest rate data, we restrict attention to 9 industrialized countries with long time series: Australia, Canada, France, Germany, Italy, Japan, Spain, UK, and USA. The data are quarterly, start in 1979Q1, and end in 2009Q4. Recursive out-of-sample forecasts are constructed using rolling windows of  $(60 - h)$  observations, where the first forecast uses the observations up to 1993Q4 for estimation.

We report results for the entire forecast period and for the sub-periods 1994Q1–2000Q4, 2001Q1–2006Q4, and 2007Q1–2009Q4. The first period includes the build-up of the dot-com bubble, the second contains the time after the dot-com bubble burst and the build-up of the sub-prime mortgage market, the third contains the observations following the collapse of the sub-prime mortgage market.

We will use the forecast methods outlined in Section 4. However, we do not impose knowledge of the timing of the structural break on the robust weights as such knowledge may not be available to the researcher at the time. We also do not estimate the down-weighting parameter  $\gamma$  due to the computational complexity when applied to model (48). We are also careful not to use *ex post* knowledge in the choice of  $\gamma$  and set it to 0.95 and 0.98 *a priori*. We base the comparisons on the relative MSFE as defined in (47).

## 5.2 Results for GDP growth forecasts

Table 7 reports the relative MSFE for the nine countries separately for the entire forecast period. The forecast based on optimal weights, which use estimated break dates, fail to improve on the forecasts based on equally weighted observations in the majority of countries across forecast horizons with the exception of Japan and Spain for  $h = 1$  and 2. This suggests that the breaks are difficult to estimate with sufficient accuracy.

The next two lines give the relative MSFEs for the forecasts using the robust weights assuming one and two breaks, respectively. The robust weights deliver better forecasts than equal weights in the overwhelming majority of countries and forecast horizons. The robust forecasts perform better the shorter the forecast horizon. An exception is the case of Australia, where all forecasts with the exception of the AveW forecast fare worse than the forecast based on equal weights over the four forecast horizons. For  $h = 1$  the robust weights for one break produce better forecasts than the robust weights for two breaks. For larger  $h$  this distinction is less clear cut.

With the exception of Japan forecasts based on the post-break sample (with the break date estimated) do not consistently improve over the equal weight forecasts. This contrasts with the AveW forecasts, which lead to improvements over the equal weights forecasts in many cases, even if the improvements are usually smaller than the robust weights or the ExpS forecasts. However, when these forecasts perform poorly the AveW forecasts usually perform less bad, so that AveW forecasts can be seen as a conservative forecasting strategy that leads to modest gains.

The ExpS forecasts with  $\gamma$  *a priori* set to 0.95 perform well and together with the robust forecasts for one break delivers the best forecasts for most countries and forecast horizons. However, in the case of Australia ExpS( $\gamma = 0.95$ ) produces the worst forecasts for all horizons. It is the least conservative forecasting method delivering often the best but, in particular at  $h = 4$  also often the worst forecasts. ExpS with  $\gamma = 0.98$  is more conservative with forecasts that are rarely the best but also rarely the worst forecasts. The difference between the ExpS forecasts for the two  $\gamma$ 's demonstrate the sensitivity of the forecasts to the choice of  $\gamma$ , which needs to be decided by the forecaster without the benefit of hindsight.

Tables 8 reports the results for the first subsample of forecasts, 1994Q1–2000Q4. In this subsample the gains from accounting for breaks is very large for US GDP growth. Robust weights forecasts, for example, have MSFE that is just over half as large as that of equal weights for  $h = 3$ . In contrast, none of the forecasts for France improve over those from equal weights. The cases of Germany and Italy show how inaccurate estimates of break dates can lead to highly imprecise forecasts when the forecasting method relies on the estimated breaks dates. In both cases the forecasts using optimal weights and, in particular, the post-break sample forecasts have a considerably higher MSFE than the equal weights forecast.

The forecasts for the second subsample, 2001Q1–2006Q4, which are in Table 9, show that during this period structural breaks are not a problem for forecasting GDP growth in the USA and the UK, and only small gains are made for France and Australia. Also, where gains from taking breaks into account are made, such as for Japan and Germany, they are smaller than in the first subsample. The exception is Canada, where gains increase in the second subsample.

The results for the last subsample, 2007Q1–2009Q4, in Table 10 show that in this period breaks influence the forecasts as both the robust weight forecasts and the ExpS forecasts vastly improve over the equal weights forecast. The exception is Australia where the equal weights forecast cannot be beaten systematically.

Table 11 reports averages across countries over the whole sample and over the three subsamples. For the whole sample, the results for GDP weighted averages and equally weighted averages lead to similar conclusions. The estimated optimal weight forecasts improve in only two cases over the equal weights forecasts and, in general, are more precise than the post-break forecasts but less precise than the remaining forecast methods. This reiterates that breaks are not identified with sufficient accuracy.

The robust weights deliver vastly improved forecasts compared to equal weights. In fact, they provide the best forecasts for  $h = 2, 3, 4$  and the second most precise forecast for  $h = 1$ . Allowing for one break appears sufficient as the MSFE is smaller than when allowing for two breaks, an exception is  $h = 4$  using equal weighting across countries.

Post-break window forecasts are substantially worse than equally weights fore-

Table 7: Predictive power of the yield curve: Relative forecast accuracy per country (1994Q1–2009Q4)

	USA	Japan	Ger.	UK	F	It.	Spain	Can.	Aus.
<i>h</i> = 1									
prop. breaks	0.063	1.000	0.328	1.000	0.250	1.000	0.141	0.031	0.625
estim.opt.weight	1.083	0.921	1.084	0.998	1.163	1.191	0.978	1.083	0.999
rob.weight(1 break)	0.877	0.874	0.941	0.853	0.957	0.914	0.915	0.905	0.999
rob.weight(2 breaks)	0.934	0.922	0.964	0.930	0.979	0.952	0.954	0.938	1.002
post break	1.046	0.892	1.120	0.962	1.425	1.014	1.803	1.276	1.000
AveW	1.001	0.982	0.980	0.985	0.996	0.986	0.998	1.003	0.985
ExpS( $\gamma = 0.95$ )	0.877	0.850	0.949	0.831	0.958	0.867	0.933	0.898	1.021
ExpS( $\gamma = 0.98$ )	0.932	0.925	0.971	0.921	0.985	0.946	0.979	0.948	0.992
<i>h</i> = 2									
prop. breaks	0.047	0.875	0.453	0.969	0.094	0.969	0.047	0.000	0.000
estim.opt.weight	0.999	0.953	1.152	1.031	1.039	1.141	0.999	1.000	1.000
rob.weight(1 break)	0.852	0.849	0.997	0.947	1.000	0.913	0.988	0.906	1.058
rob.weight(2 breaks)	0.923	0.904	0.977	0.977	0.995	0.947	0.988	0.932	1.025
post break	1.349	0.944	1.159	1.011	1.077	1.027	1.891	1.000	1.000
AveW	1.007	0.981	0.995	1.003	0.999	0.983	0.999	1.003	0.993
ExpS( $\gamma = 0.95$ )	0.857	0.829	1.019	0.938	1.012	0.854	1.064	0.902	1.087
ExpS( $\gamma = 0.98$ )	0.911	0.909	1.001	0.979	1.002	0.940	1.009	0.935	1.011
<i>h</i> = 3									
prop. breaks	0.063	0.095	0.587	0.048	0.095	0.968	0.016	0.000	0.000
estim.opt.weight	1.014	1.004	1.275	1.033	1.078	1.065	1.005	1.000	1.000
rob.weight(1 break)	0.896	0.844	1.041	1.003	1.035	0.938	1.009	0.921	1.066
rob.weight(2 breaks)	0.942	0.894	0.987	1.002	1.009	0.958	1.007	0.933	1.030
post break	1.187	0.983	1.423	1.022	1.105	1.039	1.006	1.000	1.000
AveW	1.001	0.979	1.000	1.004	1.001	0.993	1.000	1.003	0.996
ExpS( $\gamma = 0.95$ )	0.901	0.830	1.072	1.019	1.047	0.869	1.044	0.922	1.094
ExpS( $\gamma = 0.98$ )	0.933	0.899	1.019	1.009	1.017	0.954	1.016	0.937	1.012
<i>h</i> = 4									
prop. breaks	0.097	0.403	0.710	0.661	0.113	0.952	0.000	0.000	0.000
estim.opt.weight	0.973	1.000	1.188	1.097	0.995	1.073	1.000	1.000	1.000
rob.weight(1 break)	0.931	0.834	1.074	1.029	1.063	0.980	1.019	0.951	1.058
rob.weight(2 breaks)	0.953	0.883	0.990	1.011	1.021	0.977	1.019	0.943	1.025
post break	0.976	1.051	1.430	1.053	1.016	1.055	1.000	1.000	1.000
AveW	0.996	0.978	1.007	1.010	1.004	1.010	1.001	1.002	0.990
ExpS( $\gamma = 0.95$ )	0.940	0.823	1.106	1.048	1.077	0.911	1.024	0.962	1.101
ExpS( $\gamma = 0.98$ )	0.953	0.889	1.030	1.028	1.029	0.980	1.018	0.951	1.006

Note: The table reports the ratio of MSFE of forecast method *i* relative to that using equal weights,  $MSFE_i/MSFE_{\text{equal}}$ , for different forecast horizons, *h*. Forecast methods: (i) optimal weights for discrete breaks based on point estimates of *b* and  $\lambda$  for up to two breaks, (ii) robust weights that integrate the break date over the entire sample, (iii) robust weights for two breaks with  $\phi_{(1)}, \phi_{(2)} \in (-2, 2)$ , (iv) post-break window, (v) AveW forecasts with  $m = T(1 - v_{\min}) + 1$  windows and  $v_{\min} = 0.05$  ExpS with (vi)  $\gamma = 0.95$  and (vii)  $\gamma = 0.98$ . The line denoted “prop. break” reports the proportion of forecasts where a break was detected by the Bai and Perron (1997,2003) test. The countries are: USA, Japan, Germany, UK, France, Italy, Spain, Canada, and Australia. The date given above denotes the periods for which one-period ahead forecasts are made. The *h* = 2 forecast makes the first forecast for the observation one quarter later, the *h* = 3 forecast for that two periods later, and the *h* = 4 forecast for that three quarters later.

Table 8: Predictive power of the yield curve: Relative forecast accuracy per country (1994Q1–2000Q4)

	USA	Japan	Ger.	UK	F	It.	Spain	Can.	Aus.
<i>h = 1</i>									
prop. breaks	0.000	1.000	0.143	1.000	0.357	1.000	0.214	0.000	0.143
estim.opt.weight	1.000	0.987	1.163	0.994	1.143	1.270	1.024	1.000	0.997
rob.weight(1 break)	0.794	0.800	0.955	0.773	1.020	1.116	0.787	0.978	0.979
rob.weight(2 breaks)	0.904	0.861	0.985	0.935	1.011	1.024	0.915	0.974	0.996
post break	1.000	0.951	1.272	0.729	1.231	1.216	1.059	1.000	1.003
AveW	1.003	0.970	0.950	1.003	1.000	1.041	0.977	1.016	0.978
ExpS( $\gamma = 0.95$ )	0.811	0.802	0.982	0.786	1.053	1.091	0.796	0.993	0.989
ExpS( $\gamma = 0.98$ )	0.901	0.897	0.984	0.872	1.027	1.054	0.878	1.023	0.982
<i>h = 2</i>									
prop. breaks	0.037	0.889	0.222	1.000	0.222	1.000	0.074	0.000	0.000
estim.opt.weight	1.005	0.782	1.588	0.988	1.149	1.470	1.017	1.000	1.000
rob.weight(1 break)	0.611	0.721	1.022	0.777	1.121	1.292	0.788	0.981	1.010
rob.weight(2 breaks)	0.827	0.806	0.993	0.976	1.057	1.046	0.940	0.951	1.006
post break	1.066	0.859	1.759	0.734	1.295	1.512	1.021	1.000	1.000
AveW	1.028	0.968	0.983	0.980	1.002	1.136	0.978	1.020	0.987
ExpS( $\gamma = 0.95$ )	0.633	0.726	1.070	0.787	1.157	1.228	0.805	0.992	1.008
ExpS( $\gamma = 0.98$ )	0.832	0.866	1.022	0.868	1.071	1.119	0.872	1.009	0.978
<i>h = 3</i>									
prop. breaks	0.115	0.038	0.423	0.000	0.231	1.000	0.038	0.000	0.000
estim.opt.weight	1.050	1.104	2.193	1.000	1.278	1.976	1.023	1.000	1.000
rob.weight(1 break)	0.552	0.690	1.084	0.790	1.158	1.444	0.792	0.990	0.982
rob.weight(2 breaks)	0.819	0.768	1.010	1.001	1.070	1.070	0.967	0.937	0.993
post break	1.016	1.113	3.028	1.000	1.377	1.713	1.032	1.000	1.000
AveW	1.003	0.963	0.997	0.972	1.004	1.283	0.980	1.019	0.990
ExpS( $\gamma = 0.95$ )	0.569	0.689	1.129	0.811	1.197	1.348	0.821	1.003	0.973
ExpS( $\gamma = 0.98$ )	0.821	0.840	1.038	0.886	1.083	1.180	0.876	1.002	0.956
<i>h = 4</i>									
prop. breaks	0.240	0.840	0.440	0.320	0.280	0.960	0.000	0.000	0.000
estim.opt.weight	0.915	0.998	1.557	1.122	0.983	1.860	1.000	1.000	1.000
rob.weight(1 break)	0.570	0.697	1.147	0.820	1.184	1.506	0.819	1.019	0.922
rob.weight(2 breaks)	0.836	0.767	1.021	1.025	1.085	1.087	0.998	0.943	0.965
post break	0.925	1.168	2.965	1.059	1.054	1.703	1.000	1.000	1.000
AveW	0.974	0.969	1.018	0.978	1.006	1.331	0.981	1.017	0.979
ExpS( $\gamma = 0.95$ )	0.591	0.691	1.182	0.868	1.225	1.401	0.862	1.034	0.927
ExpS( $\gamma = 0.98$ )	0.843	0.844	1.055	0.921	1.094	1.201	0.894	1.011	0.926

Note: See footnote of Table 7. The date given above denotes the periods for which one-period ahead forecasts are made. The  $h = 2$  forecast makes the first forecast for the observation one quarter later, the  $h = 3$  forecast for that two periods later, and the  $h = 4$  forecast for that three quarters later.

Table 9: Predictive power of the yield curve: Relative forecast accuracy per country (2001Q1–2006Q4)

	USA	Japan	Ger.	UK	F	It.	Spain	KCan.	Aus.
<i>h = 1</i>									
prop. breaks	0.000	1.000	0.333	1.000	0.083	1.000	0.000	0.000	1.000
estim.opt.weight	1.000	1.099	1.108	1.054	1.182	0.824	1.000	1.000	1.008
rob.weight(1 break)	1.041	0.955	0.823	1.053	1.040	0.886	0.913	0.938	1.005
rob.weight(2 breaks)	1.010	0.950	0.903	1.014	1.020	0.978	0.925	0.961	1.002
post break	1.000	1.121	1.152	1.009	1.276	0.811	1.000	1.000	0.988
AveW	1.001	0.984	0.959	1.002	0.993	0.921	0.986	0.992	0.981
ExpS( $\gamma = 0.95$ )	1.095	0.959	0.828	1.088	1.062	0.931	0.969	0.991	1.014
ExpS( $\gamma = 0.98$ )	1.021	0.955	0.912	1.007	1.029	0.927	0.893	0.947	0.976
<i>h = 2</i>									
prop. breaks	0.000	0.958	0.542	1.000	0.000	1.000	0.000	0.000	0.000
estim.opt.weight	1.000	1.143	0.996	1.203	1.000	0.738	1.000	1.000	1.000
rob.weight(1 break)	1.184	0.980	0.885	1.087	1.070	0.831	0.942	0.892	1.090
rob.weight(2 breaks)	1.039	0.950	0.915	0.989	1.025	0.970	0.922	0.924	1.036
post break	1.000	1.107	0.947	1.072	1.000	0.724	1.000	1.000	1.000
AveW	0.999	0.982	0.987	0.980	0.996	0.847	0.979	0.993	0.996
ExpS( $\gamma = 0.95$ )	1.298	1.003	0.896	1.146	1.111	0.892	1.024	0.964	1.105
ExpS( $\gamma = 0.98$ )	1.036	0.955	0.949	0.973	1.047	0.889	0.871	0.886	1.015
<i>h = 3</i>									
prop. breaks	0.000	0.208	0.625	0.000	0.000	1.000	0.000	0.000	0.000
estim.opt.weight	1.000	0.789	0.928	1.000	1.000	0.721	1.000	1.000	1.000
rob.weight(1 break)	1.523	0.977	0.922	1.035	1.129	0.871	1.077	0.854	1.134
rob.weight(2 breaks)	1.140	0.930	0.919	0.947	1.046	1.008	0.974	0.893	1.052
post break	1.000	0.602	0.885	1.000	1.000	0.718	1.000	1.000	1.000
AveW	1.002	0.975	0.992	0.960	1.001	0.815	0.978	0.995	1.000
ExpS( $\gamma = 0.95$ )	1.695	1.020	0.926	1.082	1.195	0.955	1.160	0.916	1.151
ExpS( $\gamma = 0.98$ )	1.137	0.938	0.961	0.913	1.068	0.919	0.908	0.842	1.038
<i>h = 4</i>									
prop. breaks	0.000	0.167	0.875	0.917	0.000	1.000	0.000	0.000	0.000
estim.opt.weight	1.000	1.002	1.027	1.204	1.000	0.816	1.000	1.000	1.000
rob.weight(1 break)	1.703	0.963	0.939	0.922	1.200	1.023	1.141	0.848	1.172
rob.weight(2 breaks)	1.176	0.889	0.906	0.871	1.071	1.084	1.004	0.885	1.057
post break	1.000	0.963	0.888	0.967	1.000	0.801	1.000	1.000	1.000
AveW	1.010	0.964	1.001	0.933	1.006	0.853	0.983	0.998	1.002
ExpS( $\gamma = 0.95$ )	1.888	1.021	0.931	0.952	1.263	1.127	1.210	0.883	1.187
ExpS( $\gamma = 0.98$ )	1.156	0.899	0.963	0.828	1.090	1.016	0.931	0.830	1.056

Note: See footnote of Table 7. The date given above denotes the periods for which forecasts are made at all horizons.



Table 10: Predictive power of the yield curve: Relative forecast accuracy per country (from 2007Q1-2009Q4)

	USA	Japan	Ger.	UK	F	It.	Spain	Can.	Aus.
<i>h = 1</i>									
prop. breaks	0.333	1.000	0.750	1.000	0.333	1.000	0.250	0.167	1.000
estim.opt.weight	1.161	0.828	1.043	0.995	1.167	1.234	0.957	1.153	0.993
rob.weight(1 break)	0.866	0.899	0.964	0.854	0.898	0.843	0.964	0.856	1.059
rob.weight(2 breaks)	0.924	0.953	0.970	0.924	0.948	0.920	0.971	0.912	1.022
post break	1.089	0.790	1.046	1.001	1.573	0.977	2.176	1.506	1.003
AveW	1.000	0.990	0.998	0.980	0.995	0.977	1.007	1.000	1.009
ExpS( $\gamma = 0.95$ )	0.838	0.849	0.964	0.820	0.875	0.768	0.980	0.820	1.132
ExpS( $\gamma = 0.98$ )	0.919	0.934	0.980	0.924	0.948	0.909	1.027	0.910	1.045
<i>h = 2</i>									
prop. breaks	0.154	0.692	0.769	0.846	0.000	0.846	0.077	0.000	0.000
estim.opt.weight	0.997	1.001	1.067	1.031	1.000	1.145	0.993	1.000	1.000
rob.weight(1 break)	0.877	0.888	1.020	0.972	0.934	0.858	1.052	0.872	1.110
rob.weight(2 breaks)	0.936	0.948	0.989	0.976	0.963	0.926	1.008	0.925	1.049
post break	1.574	0.950	1.041	1.058	1.000	0.987	2.221	1.000	1.000
AveW	0.999	0.988	1.001	1.008	0.998	0.977	1.007	0.998	1.003
ExpS( $\gamma = 0.95$ )	0.844	0.842	1.037	0.956	0.931	0.782	1.145	0.836	1.207
ExpS( $\gamma = 0.98$ )	0.915	0.921	1.009	0.999	0.964	0.916	1.062	0.911	1.064
<i>h = 3</i>									
prop. breaks	0.077	0.000	0.846	0.231	0.000	0.846	0.000	0.000	0.000
estim.opt.weight	0.999	1.000	1.102	1.040	1.000	0.990	1.000	1.000	1.000
rob.weight(1 break)	0.916	0.895	1.064	1.039	0.963	0.881	1.062	0.898	1.129
rob.weight(2 breaks)	0.955	0.951	1.001	1.004	0.974	0.937	1.020	0.942	1.066
post break	1.317	1.000	1.098	1.027	1.000	0.992	1.000	1.000	1.000
AveW	1.000	0.987	1.004	1.011	1.001	0.978	1.007	0.995	1.000
ExpS( $\gamma = 0.95$ )	0.876	0.861	1.099	1.054	0.954	0.795	1.096	0.874	1.226
ExpS( $\gamma = 0.98$ )	0.940	0.920	1.030	1.034	0.978	0.929	1.060	0.924	1.072
<i>h = 4</i>									
prop. breaks	0.000	0.000	0.923	0.846	0.000	0.846	0.000	0.000	0.000
estim.opt.weight	1.000	1.000	1.123	1.088	1.000	0.998	1.000	1.000	1.000
rob.weight(1 break)	0.961	0.889	1.095	1.072	0.983	0.908	1.063	0.932	1.163
rob.weight(2 breaks)	0.969	0.948	1.008	1.013	0.982	0.951	1.025	0.960	1.087
post break	1.000	1.000	1.123	1.056	1.000	0.998	1.000	1.000	1.000
AveW	1.005	0.987	1.005	1.019	1.002	0.984	1.007	0.992	0.998
ExpS( $\gamma = 0.95$ )	0.921	0.863	1.140	1.086	0.976	0.825	1.053	0.934	1.292
ExpS( $\gamma = 0.98$ )	0.969	0.913	1.045	1.055	0.988	0.947	1.056	0.945	1.084

Note: See footnote of Table 7. The date given above denotes the periods for which forecasts are made at all horizons.

Table 11: Predictive power of the yield curve: Relative forecast accuracy averaged across countries

$h$	GDP weighted ave.				Equally weighted ave.			
	1	2	3	4	1	2	3	4
1994Q1–2009Q4								
estim.opt.weight	1.070	1.032	1.055	1.027	1.056	1.035	1.053	1.036
rob.weight(1 break)	0.906	0.912	0.945	0.970	0.915	0.946	0.973	0.993
rob.weight(2 breaks)	0.953	0.953	0.966	0.973	0.953	0.963	0.974	0.980
post break	1.103	1.226	1.144	1.056	1.171	1.162	1.085	1.065
AveW	1.005	1.010	1.009	1.008	0.991	0.996	0.997	1.000
ExpS( $\gamma = 0.95$ )	0.901	0.915	0.949	0.977	0.909	0.951	0.978	0.999
ExpS( $\gamma = 0.98$ )	0.954	0.952	0.966	0.979	0.956	0.966	0.977	0.987
1994Q1–2000Q4								
estim.opt.weight	1.051	1.076	1.239	1.084	1.064	1.111	1.292	1.160
rob.weight(1 break)	0.866	0.798	0.785	0.807	0.911	0.925	0.943	0.965
rob.weight(2 breaks)	0.939	0.906	0.903	0.916	0.956	0.956	0.960	0.970
post break	1.042	1.125	1.294	1.234	1.051	1.138	1.364	1.319
AveW	1.005	1.024	1.022	1.015	0.993	1.009	1.023	1.028
ExpS( $\gamma = 0.95$ )	0.880	0.813	0.796	0.822	0.923	0.934	0.949	0.976
ExpS( $\gamma = 0.98$ )	0.941	0.915	0.912	0.929	0.958	0.960	0.965	0.977
2001Q1–2006Q4								
estim.opt.weight	1.041	1.029	0.958	1.017	1.030	1.009	0.938	1.006
rob.weight(1 break)	1.001	1.079	1.245	1.336	0.962	0.996	1.058	1.101
rob.weight(2 breaks)	0.996	1.008	1.054	1.066	0.974	0.975	0.990	0.994
post break	1.050	1.010	0.928	0.981	1.040	0.983	0.912	0.958
AveW	0.999	0.995	0.993	0.997	0.980	0.973	0.969	0.972
ExpS( $\gamma = 0.95$ )	1.038	1.152	1.349	1.445	0.993	1.049	1.122	1.162
ExpS( $\gamma = 0.98$ )	0.997	1.002	1.046	1.052	0.963	0.958	0.969	0.974
2007Q1–2009Q4								
estim.opt.weight	1.093	1.027	1.022	1.029	1.059	1.026	1.015	1.023
rob.weight(1 break)	0.902	0.929	0.962	0.992	0.911	0.954	0.983	1.007
rob.weight(2 breaks)	0.950	0.965	0.980	0.991	0.949	0.969	0.983	0.994
post break	1.138	1.328	1.166	1.027	1.240	1.204	1.048	1.020
AveW	1.007	1.009	1.010	1.013	0.995	0.998	0.998	1.000
ExpS( $\gamma = 0.95$ )	0.876	0.910	0.942	0.976	0.894	0.953	0.982	1.010
ExpS( $\gamma = 0.98$ )	0.949	0.957	0.975	0.993	0.955	0.974	0.988	1.000

Note: The GDP weighted average uses weights  $w_i = Y_i / (\sum_{j=1}^N Y_j)$ , where  $Y_i$  is the 2008 GDP in purchasing power terms for country  $i$  available from the GVAR data base and  $N = 9$ . The equally weighted average uses  $w_i = 1/9$ . For other details see the footnote of Table 7.

casts and result in the least precise forecasts out of all methods considered. Again, the AveW forecasts improve over the equal weights forecasts but less so than the robust or the ExpS forecasts.

Finally, ExpS( $\gamma = 0.95$ ) delivers the relatively precise forecasts for all forecast horizons: for  $h = 1$  it is the most precise and for the other forecast horizons it is only second to the robust weights. ExpS( $\gamma = 0.98$ ) also performs well but is less precise overall.

When considering the subsamples separately some interesting additional patterns emerge. We see that in the first subsample with forecasts for the period 1994Q1–2000Q4 the robust forecasts with one break deliver the best forecasts for all forecast horizons irrespective of how the country results are averaged. The ExpS forecasts also perform well and the forecasts that rely on estimated break dates perform very poorly.

The second subsample 2001Q1–2006Q4 offers a different picture. Most forecasting methods cannot improve on the equal weight forecasts. The exception are the AveW forecasts and, with equal weighted averages across countries, robust weights for two breaks and ExpS( $\gamma = 0.98$ ). The reason for this difference is the poor performance of the forecasting methods for the USA. AveW forecast are the only forecasting method that delivers improvements irrespective of horizon and country weights.

The third and last subsample, 2007Q1–2009Q4 is similar to the first sub-period, in that the robust forecasts and the ExpS forecasts deliver the best results, whereas forecast that require estimates of break dates perform poorly.

Overall, forecasting methods that rely on estimates of break points perform poorly in this application. AveW leads to modest but consistent improvements over equal weighted forecasts. ExpS forecasts lead to larger improvements but depend on the choice of  $\gamma$ . Finally, robust weight forecasts also lead to improvements over equal weight forecasts, and only require the investigator to decide on the potential number of breaks.

## 6 Conclusion

This paper presents a new approach to forecasting in the presence of structural breaks. When considering continuous break processes our approach recovers the exponential smoothing weights that have long been considered in the literature. When considering discrete breaks, our approach delivers new forecasts based on optimal weights. In practice, dates and sizes of breaks are unknown and their estimates can be unreliable. For such cases we have derived robust weights that do not require *a priori* knowledge of the break dates or their sizes, but only the number of breaks.

Monte Carlo experiments and an application to forecasts of GDP growth using the slope of the yield curve revealed that the common strategy of using only post-break data leads to highly inefficient forecasts unless the break data is identifiable precisely and the post-break sample is large. The optimal weights forecasts developed in this paper also require precise knowledge of the dates and sizes of the structural breaks to deliver good forecasts. In contrast, robust weights provide forecasts that tend to perform better than other feasible alternatives such as equal or ExpS weights in a range of settings without requiring knowledge of the break or, when compared to exponential smoothing, the down-weighting parameter.

## A Appendix: Mathematical details

### A.1 MSFE of post-break and optimal window

For the window that contains  $T_v$  of  $T$  observations the one-step ahead forecast is

$$\begin{aligned}\hat{y}_{T+1} &= \frac{1}{T - T_v + 1} \sum_{s=T_v}^T y_s = \frac{1}{T - T_v + 1} \sum_{s=T_v}^{T_b} y_s + \frac{1}{T - T_v + 1} \sum_{s=T_b+1}^T y_s \\ &= \frac{(T_b - T_v + 1)\mu_{(1)} + \mu_{(2)}(T - T_b)}{T - T_v + 1} + \frac{1}{T - T_v + 1} \sum_{s=T_v}^T \sigma_\varepsilon \varepsilon_s.\end{aligned}\quad (49)$$

Set  $v = \frac{T - T_v + 1}{T}$  so that  $T_v = T(1 - v) + 1$ , and re-write the above as

$$\begin{aligned}\hat{y}_{T+1} &= \mu_{(2)}\{1 - \mathbb{I}[v - (1 - b)]\} \\ &\quad + \mathbb{I}[v - (1 - b)] \left\{ \frac{(1 - b)\mu_{(2)} + [v - (1 - b)]\mu_{(1)}}{v} \right\} + \frac{1}{T_v} \sum_{s=T_v}^T \sigma_\varepsilon \varepsilon_s\end{aligned}\quad (50)$$

where  $\mathbb{I}(c)$  is an indicator function equal to 1 if  $c > 0$  and equal to 0 otherwise. The one-step ahead forecast error is

$$\hat{e}_{T+1} = (\mu_{(2)} - \mu_{(1)}) \left[ 1 - \frac{(1 - b)}{v} \right] \mathbb{I}[v - (1 - b)] + \sigma_\varepsilon \varepsilon_{T+1} - \frac{1}{T_v} \sum_{s=T_v}^T \sigma_\varepsilon \varepsilon_s.\quad (51)$$

The expected squared error normalized by  $\sigma^2$  is

$$\begin{aligned}\mathbb{E}(\sigma_\varepsilon^{-2} \hat{e}_{T+1}^2) &= 1 + \frac{(\mu_{(2)} - \mu_{(1)})^2}{\sigma^2} \left[ 1 - \frac{(1 - b)}{v} \right]^2 \mathbb{I}[v - (1 - b)] + \frac{1}{T_v} \\ &= 1 + \lambda^2 \left[ 1 - \frac{(1 - b)}{v} \right]^2 \mathbb{I}[v - (1 - b)] + \frac{1}{T_v}\end{aligned}\quad (52)$$

Initially consider windows that do not contain the break. The window with all observations after the break will minimize the MSFE, so  $v_{v \leq (1-b)}^o = (1 - b)$  and

$$\mathbb{E}[\sigma_\varepsilon^{-2} \hat{e}_{T+1}^2 | v = (1 - b)] = 1 + \frac{1}{T(1 - b)}\quad (53)$$

This is also the MSFE of the forecast using the post-break window observations.

Now consider windows that include the break so that  $\mathbb{I}[v - (1 - b)] = 1$  in (52). The first order condition is

$$\lambda^2 \left[ \frac{2(1 - b)}{v^2} - \frac{2(1 - b)^2}{v^3} \right] - \frac{1}{T v^2} = 0,\quad (54)$$

Then from (54), the expression for the optimal window (among those containing a break) is

$$v^o = \frac{2(1 - b)^2 \lambda^2}{2(1 - b)\lambda^2 - \frac{1}{T}} = (1 - b) \frac{1}{1 - \frac{1}{2\lambda^2(1 - b)T}}\quad (55)$$

It can be seen that the optimal window is the distance to break scaled by an expression that is larger the smaller the break and the smaller the distance to break. A condition of the optimal window is that it cannot exceed 1. Therefore  $v^o = \frac{2(1 - b)^2 \lambda^2}{2(1 - b)\lambda^2 - \frac{1}{T}} \leq 1$ , if  $\lambda^2 < \frac{T}{2(T - T_b)T_b}$  the optimal window contains all observations.

Using (55) in the MSFE (52) yields the results in (15).

## A.2 Derivation of optimal weights for multiple regression model with a single break

Using  $\hat{\boldsymbol{\beta}}_T(\mathbf{w})$  in (18) we can write

$$\hat{\boldsymbol{\beta}}_T(\mathbf{w}) - \boldsymbol{\beta}_{(2)} = \mathbf{S}^{-1}(\mathbf{w})\mathbf{S}_1(\mathbf{w}_{(1)})(\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(2)}) + \mathbf{S}^{-1}(\mathbf{w}) \sum_{t=1}^T w_t \mathbf{x}_t \sigma_t \varepsilon_t,$$

where  $\mathbf{S}(\mathbf{w}) = \mathbf{S}_1(\mathbf{w}_{(1)}) + \mathbf{S}_2(\mathbf{w}_{(2)})$ ,  $\mathbf{S}_1(\mathbf{w}_{(1)}) = \sum_{t=1}^{T_b} w_t \mathbf{x}_t \mathbf{x}_t'$ ,  $\mathbf{S}_2(\mathbf{w}_{(2)}) = \sum_{t=T_b+1}^T w_t \mathbf{x}_t \mathbf{x}_t'$ .

Hence,

$$\begin{aligned} e_{T+1}(\mathbf{w}) &= \mathbf{y}_{T+1} - \mathbf{x}'_{T+1} \hat{\boldsymbol{\beta}}_T(\mathbf{w}) \\ &= -\mathbf{x}'_{T+1} \left[ \hat{\boldsymbol{\beta}}_T(\mathbf{w}) - \boldsymbol{\beta}_{(2)} \right] + \sigma \varepsilon_{T+1}, \\ &= \sigma_{T+1} \varepsilon_{T+1} - \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) (\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(2)}) \\ &\quad - \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \sum_{t=1}^T w_t \mathbf{x}_t \sigma_t \varepsilon_t \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E} \left[ \sigma_{(2)}^{-1} e_{T+1}^2(\mathbf{w}) \mid \mathbf{x}_t, t = 1, 2, \dots, T+1 \right] \\ &= 1 + [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}]^2 \\ &\quad + \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 \mathbf{x}_t \mathbf{x}_t' + \sum_{t=T_b+1}^T w_t^2 \mathbf{x}_t \mathbf{x}_t' \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1}, \end{aligned} \tag{56}$$

where  $\boldsymbol{\lambda} = (\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(2)})/\sigma_{(2)}$  and  $q = \sigma_{(1)}/\sigma_{(2)}$ .

In order to obtain the optimal weights we minimize (56) with respect to  $\mathbf{w}$  subject to  $\mathbf{1}'_T \mathbf{w} = 1$ . Using  $\theta$  as the Lagrange multiplier associated with  $\mathbf{1}'_T \mathbf{w} = 1$ , the first order conditions for the above optimization problem are given by the following. For  $t \leq T_b$

$$\begin{aligned} &[q^2 \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1}] w_t \\ &= \theta/2 + [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] \\ &\quad + \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 \mathbf{x}_t \mathbf{x}_t' + \sum_{t=T_b+1}^T w_t^2 \mathbf{x}_t \mathbf{x}_t' \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1} \\ &\quad - [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \boldsymbol{\lambda}]. \end{aligned}$$

where  $\mathbf{A}_t = \mathbf{x}_t \mathbf{x}_t'$  and for  $t \geq T_b + 1$

$$\begin{aligned} &[\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1}] w_t \\ &= \theta/2 + [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] \\ &\quad + \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{A}_t \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 \mathbf{x}_t \mathbf{x}_t' + \sum_{t=T_b+1}^T w_t^2 \mathbf{x}_t \mathbf{x}_t' \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1}. \end{aligned}$$

Multiplying both sides of the above two expressions by  $w_t$  and aggregating across  $t = 1, 2, \dots, T$  it is again easily seen that  $\theta = 0$ . The above expressions are  $T$  highly non-linear equations in the  $T$  unknown weights,  $w_t$ ,  $t = 1, 2, \dots, T$ .

If  $\mathbf{A}_t = 0$  the solution for  $w_t$  is indeterminate and without loss of generality can be set to 0. So we consider solutions where  $\mathbf{A}_t \neq 0$ , which yields for  $t \leq T_b$

$$w_t = \frac{[\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}]}{q^2 \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t} \quad (57)$$

$$+ \frac{\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 \mathbf{x}_t \mathbf{x}'_t + \sum_{t=T_b+1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1}}{q^2 \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t}$$

$$- \frac{[\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_t \boldsymbol{\lambda}]}{q^2 \mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t}.$$

and for  $t \geq T_b + 1$

$$w_t = \frac{[\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}]}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t} \quad (58)$$

$$+ \frac{\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \left( \sum_{t=1}^{T_b} q^2 w_t^2 \mathbf{x}_t \mathbf{x}'_t + \sum_{t=T_b+1}^T w_t^2 \mathbf{x}_t \mathbf{x}'_t \right) \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_{T+1}}{\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{x}_t}.$$

The last result follows since

$$\begin{aligned} & [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] - [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_t \boldsymbol{\lambda}] \\ &= - [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] \mathbf{x}'_t [\mathbf{I}_k - \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)})] \boldsymbol{\lambda} \\ &= - [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] \mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) [\mathbf{S}(\mathbf{w}) - \mathbf{S}_1(\mathbf{w}_{(1)})] \boldsymbol{\lambda} \\ &= - [\mathbf{x}'_{T+1} \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_1(\mathbf{w}_{(1)}) \boldsymbol{\lambda}] [\mathbf{x}'_t \mathbf{S}^{-1}(\mathbf{w}) \mathbf{S}_2(\mathbf{w}_{(2)}) \boldsymbol{\lambda}] \end{aligned}$$

### A.3 MSFE for robust weights

Consider the MSFE associated with the robust optimal weights defined (40). For these weights we need to compute  $\sum_{t=1}^{T_b} w_t$ ,  $\sum_{t=1}^{T_b} w_t^2$ , and  $\sum_{t=1}^T w_t^2$ . Note that when  $T$  and  $T_b$  are relatively large we can use the following approximations (noting that by assumption  $\underline{b} \leq b \leq \bar{b}$ )

$$\sum_{t=1}^{T_b} w_t \approx \frac{-1}{(\bar{b} - \underline{b})} \int_{\underline{b}}^b \log \left( \frac{1-a}{1-\underline{b}} \right) da$$

$$\sum_{t=1}^{T_b} w_t^2 \approx \frac{1}{T (\bar{b} - \underline{b})^2} \int_{\underline{b}}^b \left[ \log \left( \frac{1-a}{1-\underline{b}} \right) \right]^2 da$$

$$\sum_{t=1}^T w_t^2 \approx \frac{1}{T (\bar{b} - \underline{b})^2} \int_{\underline{b}}^{\bar{b}} \left[ \log \left( \frac{1-a}{1-\underline{b}} \right) \right]^2 da + \frac{(1-\bar{b})}{T (\bar{b} - \underline{b})^2} \left[ \log \left( \frac{1-\bar{b}}{1-\underline{b}} \right) \right]^2$$

First, note that

$$\sum_{t=1}^{T_b} w_t \approx \frac{-1}{(\bar{b} - \underline{b})} \int_{\underline{b}}^b \log(1-a) da + \frac{(b - \underline{b})}{(\bar{b} - \underline{b})} \log(1 - \underline{b}) = \frac{b - \underline{b}}{\bar{b} - \underline{b}} + \frac{1 - b}{\bar{b} - \underline{b}} \log \left( \frac{1 - b}{\bar{b} - \underline{b}} \right)$$

Also,

$$\begin{aligned}
\sum_{t=1}^{T_b} w_t^2 &\approx \frac{1}{T(\bar{b}-\underline{b})^2} \int_{\underline{b}}^{\bar{b}} \left[ \log \left( \frac{1-a}{1-\underline{b}} \right) \right]^2 da \\
&= \frac{1}{T(\bar{b}-\underline{b})^2} \int_{\underline{b}}^{\bar{b}} [\log(1-a)]^2 da - \frac{2 \log(1-\underline{b})}{T(\bar{b}-\underline{b})^2} \int_{\underline{b}}^{\bar{b}} \log(1-a) da \\
&\quad + \frac{[\log(1-\underline{b})]^2 (\bar{b}-\underline{b})}{T(\bar{b}-\underline{b})^2},
\end{aligned}$$

and

$$\begin{aligned}
\int_{\underline{b}}^{\bar{b}} \log(1-a) da &= -(1-b) \log(1-b) + (1-\underline{b}) \log(1-\underline{b}) + \underline{b} - b \\
\int_{\underline{b}}^{\bar{b}} [\log(1-a)]^2 da &= -(1-b) [\log(1-b)]^2 + 2(1-b) \log(1-b) + 2b \\
&\quad + (1-\underline{b}) [\log(1-\underline{b})]^2 - 2(1-\underline{b}) \log(1-\underline{b}) - 2\underline{b}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{t=1}^T w_t^2 &\approx \frac{1}{T(\bar{b}-\underline{b})^2} \int_{\underline{b}}^{\bar{b}} \left[ \log \left( \frac{1-a}{1-\underline{b}} \right) \right]^2 da + \frac{(1-\bar{b})}{T(\bar{b}-\underline{b})^2} \left[ \log \left( \frac{1-\bar{b}}{1-\underline{b}} \right) \right]^2 \\
&= \frac{1}{T(\bar{b}-\underline{b})^2} \int_{\underline{b}}^{\bar{b}} [\log(1-a)]^2 da - \frac{2 \log(1-\underline{b})}{T(\bar{b}-\underline{b})^2} \int_{\underline{b}}^{\bar{b}} [\log(1-a)] da \\
&\quad + \frac{[\log(1-\underline{b})]^2 (\bar{b}-\underline{b})}{T(\bar{b}-\underline{b})^2} + \frac{(1-\bar{b})}{T(\bar{b}-\underline{b})^2} \left[ \log \left( \frac{1-\bar{b}}{1-\underline{b}} \right) \right]^2
\end{aligned}$$

The above expressions simplify considerably if we set  $\underline{b} = 0$ . We have

$$\begin{aligned}
\sum_{t=1}^{T_b} w_t &\approx \frac{b}{\bar{b}} + \frac{(1-b) \log(1-b)}{\bar{b}}, \\
\sum_{t=1}^{T_b} w_t^2 &\approx \frac{-(1-b) [\log(1-b)]^2 + 2(1-b) \log(1-b) + 2b}{T\bar{b}^2}
\end{aligned}$$

and

$$\sum_{t=1}^T w_t^2 \approx \frac{2\bar{b} + 2(1-\bar{b}) \log(1-\bar{b})}{T\bar{b}^2}$$

Using these results in (28), we have

$$\begin{aligned}
\frac{\omega_x^2}{x_{T+1}^2} \left[ E \left( e_{T+1}^2 / \sigma_{(2)}^2 \right) - 1 \right] &\approx \phi^2 \left( \sum_{t=1}^{T_b} w_t \right)^2 + (q^2 - 1) \sum_{t=1}^{T_b} w_t^2 + \sum_{t=1}^T w_t^2 \\
&= \phi^2 \left[ \frac{b}{\bar{b}} + \frac{(1-b) \log(1-b)}{\bar{b}} \right]^2 \\
&\quad + (q^2 - 1) \left[ \frac{-(1-b) [\log(1-b)]^2 + 2(1-b) \log(1-b) + 2b}{T(\bar{b})^2} \right] \\
&\quad + \frac{2\bar{b} + 2(1-\bar{b}) \log(1-\bar{b})}{T\bar{b}^2}
\end{aligned}$$

In practice if we choose  $\bar{b}$  to be very close to unity but not unity then  $(1 - \bar{b}) \log(1 - \bar{b}) \approx 0$  and  $(1 - \bar{b})[\log(1 - \bar{b})]^2 \approx 0$  and the result in (45) follows.

#### A.4 Robust weights with higher order terms

Consider now the second order term in (39) and let

$$T\phi^2 H(b, a) = -\frac{1}{(1-b)^2} + \frac{1}{b(1-b)^2} I(b-a),$$

and note that for  $a < \underline{b}$

$$T\phi^2 \int_0^{\underline{b}} H(b, a) = 0 \text{ when } a < \underline{b}.$$

This is so since by assumption the probability of drawing  $b$  less than  $\underline{b}$  is zero. Consider now the value of the integral when  $\underline{b} \leq a \leq \bar{b}$ , and note that

$$\begin{aligned} T\phi^2 \int_{\underline{b}}^{\bar{b}} H(b, a) &= -\int_{\underline{b}}^{\bar{b}} \frac{1}{(1-b)^2} db + \int_{\underline{b}}^{\bar{b}} \frac{1}{b(1-b)^2} I(b-a) db \\ &= -\int_{\underline{b}}^a \frac{1}{(1-b)^2} db - \int_a^{\bar{b}} \frac{1}{(1-b)^2} db + \int_a^{\bar{b}} \frac{1}{b(1-b)^2} db \\ &= -\int_{\underline{b}}^a \frac{1}{(1-b)^2} db + \int_a^{\bar{b}} \frac{1}{b(1-b)} db \\ &= -\frac{a-\underline{b}}{(1-a)(1-\underline{b})} + \log\left(\frac{\bar{b}}{\underline{b}}\right) + \log\left(\frac{1-\bar{b}}{1-a}\right). \end{aligned}$$

Finally, for  $a > \bar{b}$  we have

$$T\phi^2 H(b, a) = -\int_{\underline{b}}^{\bar{b}} \frac{1}{(1-b)^2} db + \int_{\underline{b}}^{\bar{b}} \frac{1}{b(1-b)^2} I(b-a) db = -\frac{\bar{b}-\underline{b}}{(1-\bar{b})(1-\underline{b})}.$$

Combining these results, we obtain

$$w(a) \approx \begin{cases} 0 & \text{for } a < \underline{b} \\ \frac{-1}{T(\bar{b}-\underline{b})} \log\left(\frac{1-a}{1-\underline{b}}\right) + \frac{1}{T^2\phi^2(\bar{b}-\underline{b})} \times \\ \quad \times \left[ \frac{-(a-\underline{b})}{(1-a)(1-\underline{b})} + \log\left(\frac{\bar{b}}{a}\right) + \log\left(\frac{1-\bar{b}}{1-a}\right) \right] & \text{for } \underline{b} \leq a \leq \bar{b} \\ \frac{-1}{T(\bar{b}-\underline{b})} \log\left(\frac{1-\bar{b}}{1-\underline{b}}\right) - \frac{1}{T^2\phi^2(\bar{b}-\underline{b})} \frac{\bar{b}-\underline{b}}{(1-\bar{b})(1-\underline{b})} & \text{for } a > \bar{b} \end{cases}$$

and the discrete time version is

$$w_t \approx \begin{cases} 0 & \text{for } t < T\underline{b} \\ \frac{-1}{T(\bar{b}-\underline{b})} \log\left(\frac{1-t/T}{1-\underline{b}}\right) + \frac{1}{T^2\phi^2(\bar{b}-\underline{b})} \times \\ \quad \times \left[ \frac{-(t/T)-\underline{b}}{(1-t/T)(1-\underline{b})} + \log\left(\frac{\bar{b}(1-\bar{b})}{(t/T)(1-t/T)}\right) \right] & \text{for } T\underline{b} \leq t \leq T\bar{b} \\ \frac{-1}{T(\bar{b}-\underline{b})} \log\left(\frac{1-\bar{b}}{1-\underline{b}}\right) - \frac{1}{T^2\phi^2} \frac{1}{(1-\bar{b})(1-\underline{b})} & \text{for } t > T\bar{b} \end{cases} \quad (59)$$



In the case where  $\underline{b} = 0$ , and  $\bar{b}T = T - 1$ , or  $\bar{b} = 1 - 1/T$  we have for  $1 \leq t \leq T - 1$

$$w_t^* = \frac{-1}{T-1} \log(1 - t/T) - \frac{1}{T(T-1)\phi^2} \left[ \frac{t}{T-t} - \log\left(\frac{(T-1)}{t(T-t)}\right) \right], \quad (60)$$

and for the final date using the last part of (59) we obtain

$$w_T^* = \frac{\log(T)}{T-1} - \frac{1}{T\phi^2}. \quad (61)$$

The scaled version of these weights (that sum up to unity) are given by

$$w_t = \frac{w_t^*}{\sum_{s=1}^T w_s^*}, \text{ for } t = 1, 2, \dots, T.$$

In practice one could set  $\phi^2 = 1/2$  or 1.

## References

- Altissimo, Filippo, and Valentina Corradi (2003) ‘Strong rules for detecting the number of breaks in a time series.’ *Journal of Econometrics* 117, 207–244.
- Andrews, Donald W. K. (1993) ‘Tests for parameter instability and structural change point with unknown change point.’ *Econometrica* 61, 821–856.
- Andrews, Donald W. K., Inpyo Lee and Werner Ploberger (1996) ‘Optimal change-point tests for normal linear regression.’ *Journal of Econometrics* 70, 9–38.
- Bai, Jushan (1997) ‘Estimation of a change point in multiple regression models.’ *Review of Economics and Statistics* 79, 551–563.
- 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.
- Brown, Robert G. (1959) *Statistical Forecasting for Inventory Control*, McGraw-Hill, NY.
- Brown, R. L., J. Durbin, and J. M. Evans (1975) ‘Techniques for testing the constancy of regression relationships over time.’ *Journal of the Royal Statistical Society B* 37, 149–192.
- Clements, Michael P. and David F. Hendry (1999) *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.
- Clements, Michael P. and David F. Hendry (2006) ‘Forecasting with Breaks.’ in G. Elliott, C.W.J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Elsevier, 605–657.
- Estrella, Arturo, and Gikas A. Hardouvelis (1991) ‘The term structure as a predictor of real economic activity.’ *Journal of Finance* 46, 555–576.
- Estrella, Arturo, and Frederic S. Mishkin (1997) ‘The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank.’ *European Economic Review* 41, 1375–1401.
- Estrella, Arturo, Anthony P. Rodriguez, and Sebastian Schich (2003) ‘How stable is the predictive power of the yield curve? Evidence from Germany and the United States.’ *Review of Economics and Statistics* 85, 629–644.

- Giacomini, Raffaella, and Barbara Rossi (2006) ‘How stable is the forecasting performance of the yield curve for output growth?’ *Oxford Bulletin of Economic and Statistics* 68, 783–795.
- Giacomini, Raffaella, and Barbara Rossi (2009) ‘Detecting and predicting forecast breakdowns.’ *Review of Economic Studies* 76, 669–705.
- Hinkley, David V. (1970) ‘Inference about the change-point in a sequence of random variables.’ *Biometrika* 57, 1–17.
- Holt, Charles (1957) ‘Forecasting trends and seasonals by exponential weighted averages.’ *ONR Memorandum* 52/1957, Carnegie Mellon University.
- Hyndman, Rob J., Anna Koehler, J. Keith Ord, and Ralph D. Snyder (2008) *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer Verlag.
- Inoue, Atsushi, and Barbara Rossi (2011) ‘Identifying the sources of instabilities in macroeconomic fluctuations.’ *Review of Economics and Statistics* forthcoming.
- Koop, Gary, and Simon M. Potter (2007) ‘Estimation and forecasting in models with multiple structural breaks.’ *Review of Economic Studies* 74, 763–789.
- Pesaran, M. Hashem, Davide Pettenuzzo, and Allan Timmermann (2006) ‘Forecasting time series subject to multiple structural breaks.’ *Review of Economic Studies* 73, 1057–1084.
- Pesaran, M. Hashem, and Andreas Pick (2011) ‘Forecast combination across estimation windows.’ *Journal of Business and Economic Statistics* 29, 307–318.
- Pesaran, M. Hashem, Andreas Pick, and Allan Timmermann (2011) ‘Variable selection, estimation and inference for multi-period forecasting problems.’ *Journal of Econometrics* 164, 173–187.
- Pesaran, M. Hashem, and Allan Timmermann (2002) ‘Market timing and return predictability under model instability.’ *Journal of Empirical Finance* 9, 495–510.
- Pesaran, M. Hashem, and Allan Timmermann (2007) ‘Selection of estimation window in the presence of breaks.’ *Journal of Econometrics* 137, 134–161.
- Rossi, Barbara (2011) ‘Advances in forecasting under instability.’ Chapter prepared for the *Handbook of Economic Forecasting*, eds. G. Elliot and A. Timmermann.
- Schrumpf, Andreas, and Qingwei Wang (2010) ‘A reappraisal of the leading indicator properties of the yield curve under structural instability.’ *International Journal of Forecasting* 26, 836–857.
- Smith, L. V. and A. Galesi (2010), *GVAR Toolbox 1.0*, [www-cfap.jbs.cam.ac.uk/research/gvartoolbox/index.html](http://www-cfap.jbs.cam.ac.uk/research/gvartoolbox/index.html).
- Stock, James H., and Mark W. Watson (2003) ‘Forecasting output and inflation: The role of asset prices.’ *Journal of Economic Literature* 41, 788–829.