Using all observations when forecasting under structural breaks*

Stanislav Anatolyev[†] Victor Kitov[‡]

New Economic School Moscow State University

December 2007

Abstract

We extend the idea of the trade-off window approach by Pesaran and Timmermann (2007, Journal of Econometrics 137, 134–161) of using observations preceding the last structural break to estimate model parameters for the purpose of forecasting. Our weighted least squares method utilizes information in all observations but with weights varying from one to another interval between breaks. This leads to a smaller mean squared prediction error which is illustrated by simulations. The proposed procedure is computationally simple having a convenient associated optimization program. We also describe and evaluate a cross-validation analog of the proposed method. (JEL codes: C12, C22, C32, C53)

*We thank the FEP co-editor Heikki Kauppi and anonymous referees of this and previous versions of the paper for useful comments and suggestions.

[†]New Economic School, Moscow, 117418 Russia; e-mail: sanatoly@nes.ru

[‡]Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, Russia; e-mail: vkitov@mail.ru

1 Introduction

A number of recent papers discuss forecasting when the estimated econometric relationship is subject to structural breaks (Pesaran and Timmermann, 2002, 2004; Clark and McCracken, 2005; Elliott, 2005), and, in particular, the question of how to exploit sample observations during estimation. Pesaran and Timmermann (2007) propose a method of a trade-off window (TOW) where, along with observations from the most recent stable period, the observations preceding the last break are partly employed in parameter estimation. These additional observations, albeit introducing bias, help reduce variance and hence may positively affect the forecast quality criterion. In this paper, we further develop and evaluate the idea of using pre-break observations in parameter estimation for further forecasting purposes.

We consider a linear model which is subject to m structural breaks at times $\tau_1, \cdots, \tau_m$. There are thus $m+1$ regimes in which the relationship is stable:

$$y_t = x_t'\beta_j + \varepsilon_t, \quad t = 1 + \tau_{j-1}, \cdots, \tau_j, \quad j = 1, \cdots, m+1,$$

where trivially $\tau_0 = 0$, $\tau_{m+1} = n$, and

$$E\left[\varepsilon_t | I_{t-1}\right] = 0,$$

where $I_{t-1} = \{x_t, x_{t-1}, x_{t-2,}, \cdots x_1, y_{t-1}, y_{t-2}, \cdots, y_1\}$. As noted in Pesaran and Timmermann (2005, section 4.5), a general estimator of β for the purpose of further forecasting has the form of weighted least squares:

$$\hat{\beta} = \left(\sum_{j=1}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} \omega_{j,t} x_t x_t' \right)^{-1} \sum_{j=1}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} \omega_{j,t} x_t y_t, \tag{1}$$

where, including the TOW method of Pesaran and Timmermann (2007),

$$\begin{aligned}
\omega_{j,t} &= \lambda^{n-t-1}(1-\lambda)/(1-\lambda^n) \quad \text{(discounted least squares)} \\
\omega_{j,t} &= 1 \quad \text{(expanding window)} \\
\omega_{j,t} &= 1\left[t \geq n - w + 1\right] \quad \text{(rolling window)} \\
\omega_{j,t} &= 1\left[t \geq \hat{\tau}_m + 1\right] \quad \text{(post-break window)} \\
\omega_{j,t} &= 1\left[t \geq \hat{\tau}_{\mathrm{W}}\right] \quad \text{(trade-off window)}
\end{aligned}$$

where λ is a discounting rate, w is a rolling window width, $\hat{\tau}_m$ is an estimate of the last break date, and $\hat{\tau}_{\mathrm{W}} \leq \hat{\tau}_m + 1$ is the first observation that falls into the optimal trade-off window. Our estimator is also a special case of the general form (1) with a flexible system of weights:

$$\omega_{j,t} = \hat{\alpha}_j, \quad j = 1, \cdots, m, \quad \omega_{m+1,t} = 1, \tag{2}$$

where each $\hat{\alpha}_j$ is a feasible version of a certain constant α_j^*, these m constants constituting a system of weights optimal in terms of the final prediction criterion, the mean squared prediction error (MSPE). We call an estimator (1) with the weights of the type (2) a *WLS estimator*, and the associated predictor – a *WLS predictor*.

Note that the TOW method places either zero or unit weights on observations preceding the last break, and in particular discriminates observations belonging to the last-but-one stable period. Evidently, neither such discrimination nor zero/unit weighting is optimal from the MSPE viewpoint. Our system of optimal weights α_j^* is a solution of a much more flexible (i.e. less constrained) optimization program. An additional convenience of the WLS method is the continuity of the optimization problem, in contrast to the TOW method where the optimization problem is discrete. On the other hand, the TOW optimization problem is one-dimensional, while the WLS problem is multidimensional, although computationally simple. The cross-validation version of the WLS method always requires prior estimation of break dates, while one of cross-validation versions of the TOW method does not.

One should also mention that there is a growing literature that alternatively employs Bayesian techniques in constructing forecasts in an unstable environment. See, for example, the studies by Clark and McCracken (2004), Koop and Potter (2007), and Pesaran, Pettenuzzo, and Timmermann (2007).

The paper is organized as follows. Section 2 presents the setup and WLS predictor, section 3 analyzes and compares the properties of the WLS and TOW predictors. Section 4 discusses the characterization of the optimal system of weights. Section 5 describes the related cross-validation method. Section 6 presents Monte–Carlo evidence. Finally, Section 7 concludes. The Appendix contains proofs of the theorems.

2 WLS predictor

In order to derive analytical results, we treat $\tau_1, \cdots, \tau_m$ as known. Because in practice the break dates are usually unknown, we study the case of unknown break dates in the simulation section. In that case the break dates are estimated together with regression parameters via minimization of the sum of squared residuals, following the approach of Bai and Perron (1998). Further, for the analytical results to be more transparent and clear-cut, we follow Pesaran and Timmermann (2007) and make a strict exogeneity assumption. Let $X = (x_1, \cdots, x_{n+1})$.

Assumption 1 (SE) *The errors ε_t and regressors x_t satisfy $E[\varepsilon_t|X] = 0$.*

We derive most of our results conditional on X. Sometimes we also consider special cases of conditional homoskedasticity and stationary regressor distribution.

Assumption 2 (CH) *The error exhibits conditional homoskedasticity, although regime-wise heteroskedasticity is allowed:*

$$E\left[\varepsilon_t^2|X\right] = \sigma_j^2, \quad t = 1+\tau_{j-1}, \cdots, \tau_j, \quad j = 1, \cdots, m+1.$$

Given the weights $(\alpha_1, \cdots, \alpha_m)$, the WLS estimator $\hat{\beta}(\alpha_1, \cdots, \alpha_m)$ is determined from the weighted least squares minimization problem:

$$\begin{aligned}\hat{\beta}(\alpha_1, \cdots, \alpha_m) &= \arg\min_b \sum_{j=1}^{m+1} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} (y_t - x_t' b)^2 \\ &= \left(\sum_{j=1}^{m+1} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \right)^{-1} \sum_{j=1}^{m+1} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t y_t,\end{aligned}$$

where $\alpha_{m+1} = 1$. Assume that there is no structural break at period $n+1$. The aim is to build a predictor $\hat{y}_{n+1}$ of y_{n+1} yielding the lowest conditional (given the set of explanatory variables) mean squared prediction error for prediction of y_{n+1} by explanatory variables x_{n+1}:

$$MSPE(\alpha_1, \cdots, \alpha_m) = E\left[(y_{n+1} - \hat{y}_{n+1}(\alpha_1, \cdots, \alpha_m))^2 |X\right],$$

where the forecast $\hat{y}_{n+1}$ has the form

$$\hat{y}_{n+1}(\alpha_1, \cdots, \alpha_m) = x'_{n+1}\hat{\beta}(\alpha_1, \cdots, \alpha_m).$$

Then the conditional MSPE is

$$MSPE(\alpha_1, \cdots, \alpha_m) = \sigma^2_{m+1} + x'_{n+1}(BB' + \Sigma)x_{n+1},$$

where B and Σ are the conditional bias and conditional variance of $\hat{\beta}(\alpha_1, \cdots, \alpha_m)$. The weights $\alpha_1, \cdots, \alpha_m$ need to be tuned so that $MSPE(\alpha_1, \cdots, \alpha_m)$ is minimal. This is a theme of section 4; for now, we take the system $(\alpha_1, \cdots, \alpha_m)$ as given.

3 Properties of WLS and TOW predictors

We have the following result on the bias and variance of $\hat{\beta}(\alpha_1, \cdots, \alpha_m)$.

Theorem 1 *Suppose assumption SE is satisfied. Then the conditional bias and conditional variance of $\hat{\beta}(\alpha_1, \cdots, \alpha_m)$ are*

$$\begin{aligned} B &= P^{-1}\sum_{j=1}^{m}\alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x'_t \left(\beta_j - \beta_{m+1}\right), \\ \Sigma &= P^{-1}\left(\sum_{j=1}^{m+1}\alpha_j^2 \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x'_t \sigma_t^2\right)P^{-1}, \end{aligned}$$

where

$$P = \sum_{j=1}^{m+1}\alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x'_t.$$

If assumption CH holds, then Σ simplifies further to

$$\Sigma = P^{-1}\left(\sum_{j=1}^{m+1}\alpha_j^2\sigma_j^2 \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x'_t\right)P^{-1}.$$

Using Theorem 1, we obtain the expression for the conditional MSPE:

$$MSPE(\alpha_1, \cdots, \alpha_m) = \sigma^2_{m+1} + x'_{n+1}(BB' + \Sigma)x_{n+1},$$

where $BB' + \Sigma$ is

$$P^{-1}\left\{\left(\sum_{j=1}^{m}\alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \left(\beta_j - \beta_{m+1}\right)\right)\left(\sum_{j=1}^{m}\alpha_j \left(\beta_j - \beta_{m+1}\right)' \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t'\right)\right.$$
$$\left. + \sum_{j=1}^{m+1}\alpha_j^2 \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \sigma_t^2\right\} P^{-1}.$$

Suppose that alternatively the trade-off window method of Pesaran and Timmermann (2007) is applied, where the first observation falling into the trade-off window is $\tau_{\mathrm{W}} \in [1+\tau_{\ell-1}, \tau_\ell]$ for some $\ell \in \{1, ..., m\}$. Let $\hat{\beta}_{\mathrm{W}}$ be a trade-off window estimate:

$$\hat{\beta}_{\mathrm{W}} = \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t x_t' + \sum_{j=1+\ell}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t'\right)^{-1} \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t y_t + \sum_{j=1+\ell}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t y_t\right).$$

We have the following result on the bias and variance of $\hat{\beta}_{\mathrm{W}}$.

Theorem 2 *Suppose assumption SE is satisfied. Then the conditional bias and conditional variance of $\hat{\beta}_W$ are*

$$\begin{aligned} B_W &= P_W^{-1}\left(\sum_{t=\tau_W}^{\tau_\ell} x_t x_t' \left(\beta_\ell - \beta_{m+1}\right) + \sum_{j=1+\ell}^{m} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \left(\beta_j - \beta_{m+1}\right)\right), \\ \Sigma_W &= P_W^{-1}\left(\sum_{t=\tau_W}^{\tau_\ell} x_t x_t' \sigma_t^2 + \sum_{j=1+\ell}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \sigma_t^2\right) P_W^{-1}, \end{aligned}$$

where

$$P_W = \sum_{t=\tau_W}^{\tau_\ell} x_t x_t' + \sum_{j=1+\ell}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t'.$$

If assumption CH holds, then Σ simplifies further to

$$\Sigma_W = P_W^{-1}\left(\sigma_\ell^2 \sum_{t=\tau_W}^{\tau_\ell} x_t x_t' + \sum_{j=1+\ell}^{m+1} \sigma_j^2 \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t'\right) P_W^{-1}.$$

Using Theorem 2, we obtain the expression for the conditional MSPE:

$$MSPE_{\mathrm{W}} = \sigma_{m+1}^2 + x_{n+1}' \left(B_{\mathrm{W}} B_{\mathrm{W}}' + \Sigma_{\mathrm{W}}\right) x_{n+1},$$

where $B_{\mathrm{W}} B_{\mathrm{W}}' + \Sigma_{\mathrm{W}}$ is

$$
\begin{aligned}
P_{\mathrm{W}}^{-1} & \left\{ \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t x_t' \left(\beta_\ell - \beta_{m+1}\right) + \sum_{j=1+\ell}^{m} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \left(\beta_j - \beta_{m+1}\right) \right) \right. \\
& \times \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} \left(\beta_\ell - \beta_{m+1}\right)' x_t x_t' + \sum_{j=1+\ell}^{m} \left(\beta_j - \beta_{m+1}\right)' \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \right) \\
& \left. + \sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t x_t' \sigma_t^2 + \sum_{j=1+\ell}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \sigma_t^2 \right\} P_{\mathrm{W}}^{-1}.
\end{aligned}
$$

It is useful to compare the expressions for the conditional MSPE provided by the WLS and TOW methods in a special case. Suppose for cleanness that x_t is scalar and CH holds with $\sigma_j^2 = \sigma^2$ for all $j = \ell, \cdots, m$. Set $\alpha_j = 0$ for all $j = 1, \cdots, \ell - 1$, $\alpha_\ell = \left(\sum_{t=1+\tau_{\ell-1}}^{\tau_\ell} x_t^2\right)^{-1} \sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t^2$, and $\alpha_j = 1$ for all $j = 1 + \ell, \cdots, m$ in order to equalize P and P_{W}. Then B and B_{W} are equal too, while

$$
\begin{aligned}
\Sigma &= \sigma^2 P^{-2} \left(\left(\sum_{t=1+\tau_{\ell-1}}^{\tau_\ell} x_t^2 \right)^{-1} \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t^2 \right)^2 + \sum_{j=1+\ell}^{m} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t^2 \right) \\
&\leq \sigma^2 P^{-2} \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t^2 + \sum_{j=1+\ell}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t^2 \right) = \Sigma_{\mathrm{W}}.
\end{aligned}
$$

It follows that with the chosen (yet non-optimal) system of weights the conditional prediction biases of both WLS and TOW methods coincide, while the conditional variance of the former is at most as large as that of the latter. The difference between respective conditional mean squared prediction errors will be further increased if the system of weights is tuned optimally; of course, the weights used above may be far from optimal.

4 Optimal weights

The optimal weights $\alpha_1^*, \cdots, \alpha_m^*$ are determined via minimization of $MSPE\left(\alpha_1, \cdots, \alpha_m\right).$ This leads to the following minimization program:

$$\begin{aligned}\min_{\alpha_1,\cdots,\alpha_m} x_{n+1}' \left(\sum_{j=1}^{m+1} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t'\right)^{-1} & \left\{ \left(\sum_{j=1}^{m} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \left(\beta_j - \beta_{m+1}\right)\right)\right. \\ & \times \left(\sum_{j=1}^{m} \alpha_j \left(\beta_j - \beta_{m+1}\right)' \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t'\right) \\ & \left. + \sum_{j=1}^{m+1} \alpha_j^2 \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \sigma_t^2 \right\} \left(\sum_{j=1}^{m+1} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t'\right)^{-1} x_{n+1}.\end{aligned}$$

Of course, all unknowns have to be replaced by empirical analogs. For example, each β_j is replaced with $\hat{\beta}_j$, the OLS estimate computed from observations in regime j, $j = 1, \cdots, m+1$. The problem above is a multivariate nonlinear minimization program that has to be solved typically using numerical methods.

In order to obtain some analytical characterization of the optimal weights, suppose that we use the asymptotic analogs of empirical moments

$$M = E\left[x_t x_t'\right], \quad V_j = E\left[x_t x_t' \sigma_t^2\right], \quad t = 1 + \tau_{j-1}, \cdots, \tau_j, \quad j = 1, \cdots, m+1,$$

under the assumption that M does not change when structural breaks take place (a natural condition under the strict exogeneity of x_t). Further, let

$$q_j = \frac{\tau_j - \tau_{j-1}}{n}, \quad t = 1 + \tau_{j-1}, \cdots, \tau_j, \quad j = 1, \cdots, m+1$$

denote the proportions of lengths of stable intervals to the total sample size. Then the program becomes approximately (assuming that asymptotically all $\tau_j \to \infty$ as $n \to \infty$ so that each q_j stays constant)

$$\begin{aligned}\min_{\alpha_1,\cdots,\alpha_m} x_{n+1}' & \left\{ n \sum_{j=1}^{m} \alpha_j q_j \left(\beta_j - \beta_{m+1}\right) \sum_{j=1}^{m} \alpha_j q_j \left(\beta_j - \beta_{m+1}\right)' \right. \\ & \left. + M^{-1} \left(\sum_{j=1}^{m+1} \alpha_j^2 q_j V_j\right) M^{-1} \right\} \left(\sum_{j=1}^{m+1} \alpha_j q_j\right)^{-2} x_{n+1}. \qquad (3)\end{aligned}$$

In particular, in the case of one structural break ($m = 1$),

$$\begin{aligned}
\alpha_1^* &\simeq \arg\min_{\alpha_1} x'_{n+1} \left\{ \frac{n\alpha_1^2 q_1^2 (\beta_1 - \beta_2)(\beta_1 - \beta_2)' + M^{-1} (\alpha_1^2 q_1 V_1 + q_2 V_2) M^{-1}}{(\alpha_1 q_1 + q_2)^2} \right\} x_{n+1} \\
&= \frac{x'_{n+1} M^{-1} V_2 M^{-1} x_{n+1}}{q_1 n x'_{n+1} (\beta_1 - \beta_2)(\beta_1 - \beta_2)' x_{n+1} + x'_{n+1} M^{-1} V_1 M^{-1} x_{n+1}}. \qquad (4)
\end{aligned}$$

It is worth noting that in cases of multiple structural breaks ($m > 1$) the first order conditions do not seem to lead to a nice closed form formula like (4).

Note that the optimal α_1^* *need not be less than unity*: if the pre-break error variance is much smaller than the post-break error variance while the regression parameters do not differ much across the intervals, it may be beneficial to weight pre-break data *higher* than the post-break data. Given though that when the breaks are small it is hard to identify their dates in practice, it is unlikely that the optimal α_1^* will end up exceeding unity.

Note also that in the single-break case the optimal α_1^* is necessarily positive. However, in the multiple-break case some of optimal weights may be *negative*: if in one of the intervals the error variance is much larger than it is in other intervals while the difference between regression parameters in this interval and those in the interval after the last break is larger than analogs for other intervals, the corresponding weight *may well turn out negative.*

5 Cross-validation

In developing their TOW method, Pesaran and Timmermann (2007) also suggest using instead of analytical formulas a cross-validation approach where the performance of candidate windows is evaluated on a validation subsample. Our WLS method may be also modified along these lines. In cross-validation, the observed data set is partitioned into two subsamples. The model is fit using the first subsample and a grid of possible weights, and its forecasting performance is evaluated on the second subsample. Then one selects the weights that yield most accurate forecasts according to the criterion of interest, MSPE in our case.

For each point of the grid, the following procedure is implemented. An initial subset of the first k observations ($\tau_m < k < n$) is used to find the optimal forecast $\hat{y}_{k+1}$ for y_{k+1}, and the forecast error $y_{k+1} - \hat{y}_{k+1}$ is calculated. Then observation $k+1$ is included into the subset used for model fitting, the optimal forecast $\hat{y}_{k+2}$ is calculated, etc. These steps are repeated until the end of sample is reached. The measure of forecast quality, which in our case is the average of squared forecast errors, is computed as

$$\frac{1}{n-k} \sum_{\tau=k+1}^{n} (y_\tau - \hat{y}_\tau)^2.$$

Finally, one selects the weights that minimize this measure, and uses this system of weights for forecasting purposes at $t = n$.

Because during this procedure the number of observations belonging to the final regime varies it is important to adapt $\alpha_1, \ldots, \alpha_m$ accordingly. Let $\gamma_1, \ldots, \gamma_m$ represent the relative "importance" of an observation in regimes $1, \ldots, m$, respectively, compared to its "importance" in regime $m+1$. There are $n - \tau_m$ observations in regime $m+1$, so the observations in regime i should have the weight of $\gamma_i(n - \tau_m)$. As the set of observations from regime i has the total weight of $\alpha_i(\tau_i - \tau_{i-1})$, we obtain the following relationship between α_i and γ_i:

$$\alpha_i = \frac{n - \tau_m}{\tau_i - \tau_{i-1}} \gamma_i.$$

During the cross-validation procedure, $\gamma_1, \ldots, \gamma_m$ take values on a grid, and for each interim forecast $\alpha_1, \ldots, \alpha_m$ are recomputed using the above formula.

Of course, the cross-validation method is more straightforward to implement than the optimization (analytical) approach, but at the same time it is much more computationally intensive in case there are multiple breaks; the number of points in the grid grows exponentially with m. Suppose γ_i can take one of N values for each of regimes $1, \ldots, m$, then the cross-validation procedure will have to compare forecasting performance of N^m alternative sets of weights.

6 Simulation evidence

In this section we present some Monte–Carlo evidence shedding light on the performance of the WLS method in comparison with the trade-off window method and primitive alternatives.

We consider first a single structural break case, then turn to a case of two breaks. The data generating processes (DGPs) are borrowed from Pesaran and Timmermann (2007). Note that the DGP design contradicts the assumption of strictly exogenous regressors, so the simulation exercise is closer to reality than the foregoing analytical results.

We consider a bivariate VAR(1) model

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} + A_t \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{pmatrix}$$

where

$$\begin{aligned} A_t &= \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}, \quad t \leq \tau_1, \\ A_t &= \begin{pmatrix} a_{11} + \Delta a_{11} & a_{12} + \Delta a_{12} \\ 0 & a_{22} + \Delta a_{22} \end{pmatrix}, \quad t > \tau_1, \end{aligned}$$

$$\begin{aligned} var\left(\varepsilon_{yt}\right) &= \sigma_y^2, \quad var\left(\varepsilon_{xt}\right) = \sigma_x^2, \quad t \leq \tau_1, \\ var\left(\varepsilon_{yt}\right) &= \left(\sigma_y + \Delta\sigma_y\right)^2, \quad var\left(\varepsilon_{xt}\right) = \left(\sigma_x + \Delta\sigma_x\right)^2, \quad t > \tau_1. \end{aligned}$$

The error terms ε_{yt} and ε_{xt} are conditionally uncorrelated and homoskedastic, but regimewise heteroskedasticity is allowed in some DGPs. We set $a_{11} = 0.9$, $a_{12} = 1$, $a_{22} = 0.9$, $\sigma_y = \sigma_x = 1$, and $\mu_y = \mu_x = 0$ for both regimes. Note that x_t Granger-causes y_t but not vice versa. The process starts from its pre-break stationary distribution. We set $n = 200$ and consider two cases: $\tau_1 = 50$ (so that $q_1 = \frac{1}{4}$, $q_2 = \frac{3}{4}$) and $\tau_1 = 150$ (so that $q_1 = \frac{3}{4}$, $q_2 = \frac{1}{4}$). The break dates, when not assumed known, are estimated using the Bai and Perron (1998) approach where all candidate points for a break date (except the first and the last 20 observations) are compared, and the break point that yields the lowest sum of squared residuals is selected. For simplicity, we do not estimate the number of breaks taking their true number as given. We consider the following set of after-the-break parameter changes:

DGP	Δa_{11}	Δa_{12}	Δa_{22}	$\Delta\sigma_y$	$\Delta\sigma_x$	Comments
1	-0.2	0	0	0	0	Small break in AR dynamics
2	-0.4	0	0	0	0	Large break in AR dynamics
3	0	0.5	0	0	0	Small break in marginal coefficient
4	0	1	0	0	0	Large break in marginal coefficient
5	-0.2	1	0	0	0	Break in dynamics and marginal coefficient
6	0	0	0	3	0	Increase in error term variance
7	0	0	0	-0.5	0	Decrease in error term variance
8	0	0	0	0	3	Increase in regressor variance
9	0	0	0	0	-0.5	Decrease in regressor variance
10	0	0	-0.2	0	0	Small break in regressor AR dynamics
11	0	0	-0.7	0	0	Large break in regressor AR dynamics

From 5,000 simulations, we estimate average MSPEs for six estimation methods: OLS applied to the whole data sample, OLS applied only to post-break data, the TOW and WLS methods, and cross-validation (CV) variations of both. The main (non-CV) variations are called optimization (Opt) ones, and the MSPE function in the Opt WLS method is minimized numerically with tolerance 10^{-5}. In the CV WLS method, γ runs on a grid of 100 elements: 0, 0.0125, 0.0250, $\cdots$, 1, 2, $\cdots$, 20. The Opt TOW and CV TOW procedures are described in Pesaran and Timmermann (2007).

The figures in Tables 1a and 1b are ratios of MSPEs to the natural benchmark – the MSPE when OLS is applied to all data. The variability across methods increases when the break timing is unknown, which can be expected as the break date uncertainty increases the resulting forecast error variance. The variability also goes up as the break gets more recent, which also has a natural explanation: the earlier the break the less pre-break data is available, so the smaller the difference is between various forecasts, as the methods differ only by different treatment of pre-break data. Opt WLS yields more accurate forecasts than Opt TOW in all experiments, although by a narrow margin. CV WLS produces less accurate forecasts than CV TOW only for DGPs 8 and 10 with known $\tau_1 = 50$, and for DGPs 2 and 4 with known $\tau_1 = 150$. Opt WLS yields smaller MSPE than use of only post-break data in all cases. In general, when breaks occur in regression parameters of the first equation (excluding the break in the variance in DGPs 1–5), all methods yield almost identical results. However, when the break occurs either in the variance of the first equation or in the second equation affecting the distribution of the regressor, use of pre-break data yields significantly better performance than use of only post-break data (DGPs 6–11). For such experiments smaller improvement is maintained by Opt variations than by CV methods. However, the CV methods do not improve forecasting performance compared to post-break estimation method for DGPs 1–5 when the break occurs in the regression parameters of the first equation. Taking into account that CV methods are much more computationally intensive than corresponding Opt variations and become almost computationally infeasible when the number of breaks is greater than two, preference can be given to Opt WLS as it is fast to compute and it dominates both the Opt TOW method and use of only post-break data.

Next we consider the same VAR model but which is subject to two structural breaks at $\tau_1 = 50$ and $\tau_2 = 150$ with $n = 150$ (so that $q_1 = q_2 = q_3 = \frac{1}{3}$). The variances of the error terms are normalized to unity and do not change, while the matrix A_t evolves in the following way:

$$\begin{aligned}
A_t &= \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}, \quad t \leq \tau_1, \\
A_t &= \begin{pmatrix} a_{11} + \Delta a_{11}^1 & a_{12} + \Delta a_{12}^1 \\ 0 & a_{22} + \Delta a_{22}^1 \end{pmatrix}, \quad \tau_1 < t \leq \tau_2, \\
A_t &= \begin{pmatrix} a_{11} + \Delta a_{11}^2 & a_{12} + \Delta a_{12}^2 \\ 0 & a_{22} + \Delta a_{22}^2 \end{pmatrix}, \quad t > \tau_2.
\end{aligned}$$

In all experiments, $a_{12} = 1$ and $a_{22} = 0.9$. The CV WLS method uses the grid for γ consisting of 30 elements: 0, $0.05, \cdots, 0.95$, 1, $2, \cdots, 10$. The TOW method uses only the data from the two last intervals. We consider the following set of after-the-break parameter changes:

DGP	a_{11}	Δa_{11}^1	Δa_{11}^2	Δa_{12}^1	Δa_{12}^2	Comments
1	0.9	−0.2	0	0	0	Mean reversion in AR dynamics
2	0.9	−0.2	−0.4	0	0	Decreasing trend in AR dynamics
3	0.3	0.2	0	0	0	Mean reversion in AR dynamics
4	0.3	0.2	0.4	0	0	Increasing trend in AR dynamics
5	0.9	0	0	1	0	Mean reversion in marginal coefficient
6	0.9	0	0	1	2	Increasing trend in marginal coefficient

The results are contained in Table 2 in the format of Table 1. The variability in forecasting performance increases further because some methods use data from only one regime (post-break), some from two (TOW) and some from the whole sample (WLS and full sample OLS). The results again demonstrate the dominance, although marginal, of WLS over TOW and use of post-break observations. CV WLS yields more accurate forecasts than use of post-break data and TOW in the majority of cases. As in the case of a single break we conclude that due to computational simplicity and accuracy of forecasts use of Opt WLS is most preferable.

7 Concluding remarks

We have further extended the idea of the trade-off window method by Pesaran and Timmermann (2007) of using pre-break observations to estimate model parameters for further forecasting. While Pesaran and Timmermann (2007) suggest exploiting partly those observations taking them and post-break observations with equal weights, we propose going the whole way in this direction and utilize all observations but with different weights. Theoretically, this is a more optimal thing to do in the sense that it leads to a smaller mean squared prediction error. Monte–Carlo simulations show that the proposed WLS method does beat the TOW method, although by a narrow margin. We also proposed a cross-validation analog of the WLS method which too is favorable.

Apart from the fuller account of information in the data, additional convenience of the WLS method is the continuity of the optimization problem for the optimal system of weights, in contrast to the TOW method where the optimization problem is discrete. Last, but not least, the WLS procedure is computationally simple and can be programmed easily.

References

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

Clark, T.E. and M.W. McCracken (2004). Improving forecast accuracy by combining recursive and rolling forecasts. Research Working Paper, Federal Reserve Bank of Kansas City.

Clark, T.E. and M.W. McCracken (2005). The power of tests of predictive ability in the presence of structural breaks. Journal of Econometrics 124, 1–31.

Elliott, G. (2005). Forecasting when there is a single break. Manuscript, University of California, San Diego.

Koop, G. and S.M. Potter (2007). Estimation and forecasting in models with multiple breaks. Review of Economic Studies 74, 763–789.

Pesaran, M.H., D. Pettenuzzo, and A. Timmermann (2007). Forecasting time series subject to multiple structural breaks. Review of Economic Studies 73, 1057–1084.

Pesaran, M.H. and A. Timmermann (2002). Market timing and return prediction under model instability. Journal of Empirical Finance 9, 495–510.

Pesaran, M.H. and A. Timmermann (2004). How costly is it to ignore breaks when forecasting the direction of a time series? International Journal of Forecasting 20, 411–425.

Pesaran, M.H. and A. Timmermann (2005). Real-time econometrics. Econometric Theory 21, 212–231.

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

A Appendix: proofs

Proof. [of Theorem 1] Expanding the expression for $\hat{\beta}(\alpha_1, \cdots, \alpha_m)$,

$$\begin{aligned}\hat{\beta}(\alpha_1, \cdots, \alpha_m) &= \beta_{m+1} + P^{-1} \sum_{j=1}^{m+1} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \left(\beta_j - \beta_{m+1}\right) \\ &\quad + P^{-1} \sum_{j=1}^{m+1} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t \varepsilon_t.\end{aligned}$$

Because the third term has zero expectation conditional on X, the bias term is

$$\begin{aligned}B &= E\left[\hat{\beta}(\alpha_1, \cdots, \alpha_m) | X\right] - \beta_{m+1} \\ &= P^{-1} \sum_{j=1}^{m} \alpha_j \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \left(\beta_j - \beta_{m+1}\right).\end{aligned}$$

The conditional variance is

$$\begin{aligned}\Sigma &= E\left[\left(\hat{\beta}(\alpha_1, \cdots, \alpha_m) - \beta_{m+1} - B\right)^2 | X\right] \\ &= P^{-1} \left(\sum_{j=1}^{m+1} \alpha_j^2 \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \sigma_t^2 \right) P^{-1}.\end{aligned}$$

Under the CH assumption, $\sigma_t^2 = \sigma_j^2$, $j = 1, \cdots, m+1$, and the expression for Σ simplifies.

■

Proof. [of Theorem 2] Expanding the expression for $\hat{\beta}_{\mathrm{W}}$,

$$\begin{aligned}\hat{\beta}_{\mathrm{W}} &= \beta_{m+1} + P_{\mathrm{W}}^{-1} \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t x_t' \left(\beta_\ell - \beta_{m+1}\right) + \sum_{j=1+\ell}^{m} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \left(\beta_j - \beta_{m+1}\right) \right) \\ &\quad + P_{\mathrm{W}}^{-1} \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t \varepsilon_t + \sum_{j=1+\ell}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t \varepsilon_t \right).\end{aligned}$$

Because the third term has mean zero conditional on x, the bias term is

$$\begin{aligned}B_{\mathrm{W}} &= E\left[\hat{\beta}_{\mathrm{W}} | X\right] - \beta_{m+1} \\ &= P_{\mathrm{W}}^{-1} \left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t x_t' \left(\beta_\ell - \beta_{m+1}\right) + \sum_{j=1+\ell}^{m} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \left(\beta_j - \beta_{m+1}\right) \right).\end{aligned}$$

The conditional variance is

$$\begin{aligned}\Sigma_{\mathrm{W}} &= E\left[\left(\hat{\beta}_{\mathrm{W}} - \beta_{m+1} - B\right)^2 |X\right] \\ &= P_{\mathrm{W}}^{-1}\left(\sum_{t=\tau_{\mathrm{W}}}^{\tau_\ell} x_t x_t' \sigma_t^2 + \sum_{j=1+\ell}^{m+1} \sum_{t=1+\tau_{j-1}}^{\tau_j} x_t x_t' \sigma_t^2\right) P_{\mathrm{W}}^{-1}.\end{aligned}$$

Under the CH assumption, $\sigma_t^2 = \sigma_j^2$, $j = 1, \cdots, m+1$, and the expression for Σ_{W} simplifies.

■

Table 1a: Single break case, MSPE ratios for $\tau_1 = 50, n = 200$

DGP	Post-break	Opt TOW	Opt WLS	CV TOW	CV WLS
Known break date					
1	0.69	0.70	0.69	0.69	0.69
2	0.52	0.55	0.52	0.53	0.52
3	0.92	0.92	0.92	0.92	0.92
4	0.70	0.71	0.70	0.71	0.71
5	0.68	0.69	0.68	0.69	0.69
6	1.01	1.00	1.00	1.00	1.00
7	1.00	1.00	0.99	1.00	1.00
8	1.00	1.00	1.00	1.00	1.01
9	1.01	1.01	1.00	1.01	1.01
10	1.01	1.01	1.00	1.00	1.01
11	1.01	1.01	1.00	1.00	1.00
Estimated break date					
1	0.67	0.68	0.67	0.68	0.67
2	0.53	0.56	0.53	0.53	0.53
3	0.91	0.92	0.91	0.92	0.92
4	0.71	0.72	0.71	0.72	0.71
5	0.69	0.70	0.69	0.70	0.70
6	1.24	1.18	1.18	1.08	1.07
7	1.02	1.02	1.02	1.01	1.01
8	1.14	1.11	1.10	1.05	1.04
9	1.18	1.16	1.14	1.06	1.06
10	1.14	1.11	1.10	1.05	1.04
11	1.14	1.12	1.10	1.05	1.04

Table 1b: Single break case, MSPE ratios for $\tau_1 = 150, n = 200$

DGP	Post-break	Opt TOW	Opt WLS	CV TOW	CV WLS
Known break date					
1	0.50	0.51	0.50	0.51	0.51
2	0.36	0.37	0.36	0.36	0.37
3	0.66	0.68	0.66	0.68	0.67
4	0.31	0.33	0.31	0.31	0.32
5	0.39	0.40	0.39	0.40	0.40
6	1.09	1.06	1.04	1.02	1.01
7	1.04	1.04	1.02	1.01	1.00
8	1.04	1.03	1.02	1.01	1.01
9	1.08	1.06	1.04	1.02	1.01
10	1.06	1.05	1.03	1.02	1.01
11	1.05	1.04	1.03	1.02	1.01
Estimated break date					
1	0.50	0.51	0.50	0.51	0.51
2	0.36	0.37	0.36	0.36	0.36
3	0.67	0.68	0.67	0.68	0.67
4	0.31	0.34	0.31	0.32	0.32
5	0.40	0.41	0.40	0.41	0.41
6	1.29	1.21	1.20	1.08	1.07
7	1.06	1.05	1.04	1.01	1.01
8	1.13	1.09	1.09	1.05	1.04
9	1.17	1.15	1.13	1.06	1.05
10	1.13	1.11	1.10	1.04	1.04
11	1.12	1.10	1.09	1.04	1.04

Table 2: Double break case, MSPE ratios for $\tau_1 = 50, \tau_2 = 100, n = 150$

DGP	Post-break	Opt TOW	Opt WLS	CV TOW	CV WLS
Known break date					
1	0.89	0.92	0.88	0.91	0.87
2	0.48	0.48	0.48	0.49	0.49
3	0.97	0.98	0.96	0.98	0.95
4	0.71	0.71	0.71	0.72	0.73
5	0.69	0.72	0.69	0.71	0.68
6	0.20	0.21	0.20	0.20	0.20
Estimated break date					
1	0.90	0.92	0.89	0.91	0.87
2	0.51	0.52	0.51	0.51	0.52
3	1.02	1.01	1.00	1.00	0.97
4	0.76	0.75	0.75	0.74	0.75
5	0.68	0.70	0.67	0.69	0.66
6	0.20	0.21	0.20	0.20	0.20