Optimal Forecasts in the Presence of Structural Breaks By MH Peseran, A Pick and M Pranovich

Discussion presented at 7^{th} ECB Forecasting Workshop

Don Harding

La Trobe University

May 2012

Introduction

- Very ambitious paper. PPP derive optimal, in the minimum mean square error sense, forecasts in the presence of continuous and discrete structural breaks
- If correct the PPP paper will have a major impact on our understanding of structural breaks
- However, I have some issues with the paper and reach the conclusion that some of the findings of the paper may not be robust.
- Also have some issues with the assumptions about parameter values in the application but these empirical issues are of smaller order of significance.
- Accompanying note focuses on the technical issues. These slides focus intuition.

Plan of paper and discussion

The PPP paper is in four main parts

- Section 2: Optimal weights under various break processes when the date and size of the break are known by the econometrician.
- Section 3: Optimal weights under various break processes when the date and size of the break are unknown and must be estimated by the econometrician.
- Section 4: Monte Carlo evidence on forecasting performance
- Section 5: Application to the yield curve as a predictor of real economic activity.

I focus most of my discussion on section 2

Overview of issues

1. Treat $(\sigma_v, \sigma_\varepsilon)$ as if they are known. If $(\sigma_v, \sigma_\varepsilon)$ estimated that needs to be taken into account in defining the optimization problem
    1. Likely to be of great significance to practitioners as it is unlikely that available estimators for σ_v and σ_ε will be uncorrelated.
2. Optimization is undertaken subject to the equality constraint that the weights sum to one. Can aso treats the issue of whether the weights sum to one as being something that should be determined as part of the optimization problem.
3. Monte Carlos, Application etc. claims that a common value in the literature for the smoothing parameter γ is 0.95 to 0.98.
    1. An experienced (ExpWS) practitioner would not choose γ independent of the data.
    2. Where y_t corresponds to growth rates of GDP practitioners are likely to choose $\gamma = 0.5$

Optimal weights: Continuous breaks no regressors

The model is,

$$y_t = \beta_t + \sigma_\varepsilon \varepsilon_t \tag{1}$$

$$\beta_t = \beta_{t-1} + \sigma_v v_t \tag{2}$$

Use

$$\widehat{\beta}_{T+1} = \sum_{t=1}^{T} w_t y_t$$

The mean square error then is

$$Ee_{T+1}^2 = \sigma_v^2 \left(\boldsymbol{\iota}_T' - \mathbf{w}'\mathbf{H}\right)\left(\boldsymbol{\iota}_T - \mathbf{H}'\mathbf{w}\right) + \sigma_v^2 + \sigma_\varepsilon^2 \left(1 + \mathbf{w}'\mathbf{w}\right) \tag{3}$$

Normalizing the squared prediction error by σ_ε^2 and letting $\delta^2 = \frac{\sigma_v^2}{\sigma_\varepsilon^2}$

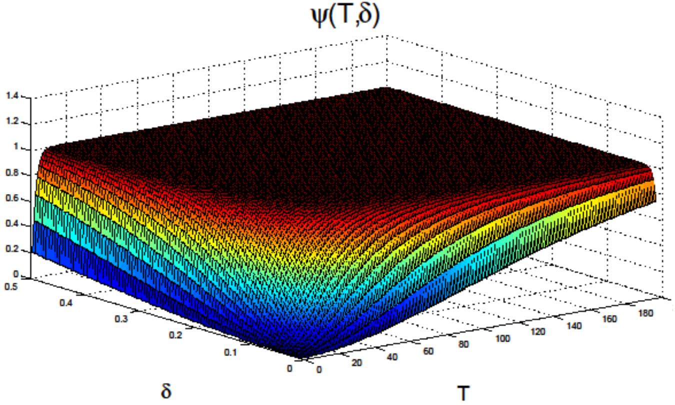
$$E\sigma_\varepsilon^{-2} e_{T+1}^2 = \delta^2 \left(\boldsymbol{\iota}_T' - \mathbf{w}'\mathbf{H}\right)\left(\boldsymbol{\iota}_T - \mathbf{H}'\mathbf{w}\right) + \delta^2 + \left(1 + \mathbf{w}'\mathbf{w}\right)$$

Intuition as to why optimal weights sum to less than one

- Mean square error is comprises of two parts $\delta^2 \left(\boldsymbol{\iota}_T' - \mathbf{w}'\mathbf{H}\right)\left(\boldsymbol{\iota}_T - \mathbf{H}'\mathbf{w}\right)$ and $(1 + \mathbf{w}'\mathbf{w})$ that are affected by the choice of w.
- $\delta^2 \left(\boldsymbol{\iota}_T' - \mathbf{w}'\mathbf{H}\right)\left(\boldsymbol{\iota}_T - \mathbf{H}'\mathbf{w}\right)$ is minimized by $\mathbf{w}' = (0, 0, ..., 0, 1)$
- $(1 + \mathbf{w}'\mathbf{w})$ is minimized by $\mathbf{w}' = 0$ (complete down weighting).
- The importance of the two components varies with T and δ^2.
- Thus, for mid range values of T the optimal weights sum to a number substantially less than one.
- As T becomes large the variance of the random walk component $\delta^2 \left(\boldsymbol{\iota}_T' - \mathbf{w}'\mathbf{H}\right)\left(\boldsymbol{\iota}_T - \mathbf{H}'\mathbf{w}\right)$ begins to dominate.
- For finite T the optimal never reach this vector $\mathbf{w}' = (0, 0, ..., 0, 1)$
- Hence, even for large T the optimal weights sum to a number that is less than one.

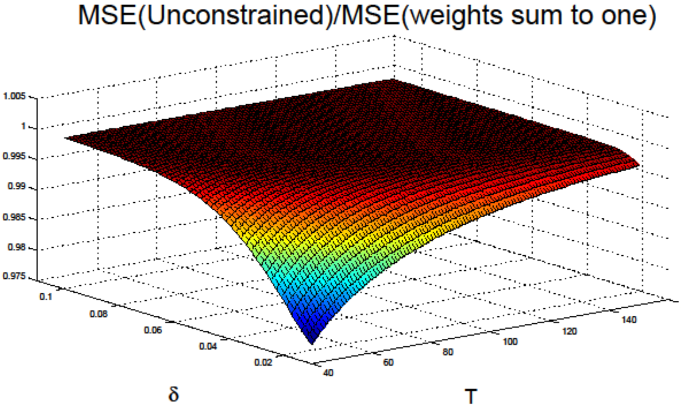
Evidence that optimal weights sum to less than one

Figure: $\psi\,(T, \delta)$



Evidence that optimal weights summing to less than one outperform the weights proposed in PPP

Figure: $E\left(\sigma_{\varepsilon}^{-2} e_{T+1}^{2} | \boldsymbol{\iota}_T' w \leq 1\right) / E\left(\sigma_{\varepsilon}^{-2} e_{T+1}^{2} | \boldsymbol{\iota}_T' w = 1\right)$



A single discrete break

Now consider the process

$$
\begin{aligned}
y_t &= \beta_{(1)} + \sigma_\varepsilon \varepsilon_t \qquad 1 \le t \le T_b && (4)\\
&= \beta_{(2)} + \sigma_\varepsilon \varepsilon_t \qquad T_b < t \le T+1 && (5)
\end{aligned}
$$

Following PPP define $\widehat{\beta}_T(w) = \sum_{t=1}^{T} w_t y_t$ and the prediction error $e_{T+1} = y_{T+1} - \widehat{\beta}_T(w)$

Allowing for the possibility that the weights might not sum to one the normalized mean square prediction error is

$$
\begin{aligned}
E\left[\sigma_\varepsilon^{-2} e_{T+1}^2(w)\right] &= 1 + \left(\sum_{t=1}^{T} w_t - 1\right)^2 \left(\frac{\beta_{(2)}}{\sigma_\varepsilon}\right)^2 + \lambda^2 \left(\sum_{t=1}^{T_b} w_t\right)^2 \\
&\quad + 2\frac{\beta_{(2)}}{\sigma_\varepsilon}\lambda \left(\sum_{t=1}^{T} w_t - 1\right)\left(\sum_{t=1}^{T_b} w_t\right) + \sum_{t=1}^{T} w_t^2 \qquad (6)
\end{aligned}
$$

A single discrete break: the weights compared

The weights then are

$$w_t = \frac{1}{T} \frac{1}{1 + Tb(1-b)\lambda^2} \text{ for } t \leq T_b$$

and

$$w_t = \frac{1}{T} \frac{1 + Tb\lambda^2}{1 + Tb(1-b)\lambda^2} \text{ for } t > T_b$$

A single discrete break: the minimum mean squared

The optimal normalized minimum mean square error is

$$E\left[\sigma_{\varepsilon}^{-2} e_{T+1}^{2}(w)\right] = 1 + \frac{b^2 + \frac{b}{T} + \frac{1}{T}(1-b)\left(\lambda^2 Tb + 1\right)^2}{\left[b + (1-b)\left(\lambda^2 Tb + 1\right)\right]^2}$$

Optimal window and post-break window

Here PPP following Peseran and Timmerman suggest truncating the window at $T_v < T_b$. There is a simple way to see that the optimal choice of v is zero (provided one estimates optimal weights) . Ie its optimal to use all of the data.

Setting $w_t = 0$ for $t < T_v$ can be put into the framework of the previous section by setting $b_v = b^* \frac{T}{T_v}$ and $\tau_v = T(1-v)+1$ then the minimum mean square prediction error conditional on v will be

$$E\left[\sigma_\varepsilon^{-2} e_{T+1}^2(w)\,|v\right] = 1 + \frac{\frac{b_v^2}{\tau_v} + \frac{b_v}{\tau_v^2} + (1-b_v)\left(\lambda^2 b_v + \frac{1}{\tau_v}\right)^2}{\tau_v\left[\frac{b_v}{\tau_v^2} + (1-b_v)\left(\lambda^2 b_v + \frac{1}{\tau_v}\right)\right]^2} \qquad (7)$$

Now τ_v increases as v decreases towards zero. The top line of (7) is decreasing in τ_v while the bottom line is decreasing in τ_v thus $E\left[\sigma_\varepsilon^{-2} e_{T+1}^2(w)\,|v\right]$ decreases as v decreases towards zero. Thus $E\left[\sigma_\varepsilon^{-2} e_{T+1}^2(w)\,|v\right]$ is minimized for $v = 0$.

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

There are two main issues. One is that as discussed in previous sections the weights may be suboptimal because they are forced to sum to one. A more difficult problem is that one cannot proceed to integrate out $\delta = \frac{\sigma_v}{\sigma_\varepsilon}$ as is done in this section. The reason for this is that the normalized mean square error is $\sigma_\varepsilon^{-2} e_{T+1}^2$ and if σ_ε must be estimated then $\sigma_\varepsilon^{-2} e_{T+1}^2$ will be random. In my view the correct approach would be to go back to the formulation of the mean square error and explicitly take account of the fact that σ_v and σ_ε must be estimated and the estimators will almost certainly be correlated.

Monte Carlo and application

Since the weights used in this section have been shown to be sub optimal using the fully optimal weights should improve the results for the proposed methods of dealing with breaks.

The values of $\gamma \in (0.8, 0.9, 0.95, 0.98)$ seem inappropriate if the y_t is the Monte Carlo relates to processes like the GDP growth rate where practitioners would typically use a value of about 0.5 for γ. There is also some good news here in that many of the issues that I raise about the optimal weights not summing to one are important for $\gamma \in (0.8, 0.9, 0.95, 0.98)$ but become unimportant for $\gamma = 0.5$.