Optimal Forecasts in the Presence of Structural Breaks

Hashem Pesaran[1] Andreas Pick[2] Mikhail Pranovich[3]

[1]Cambridge University, CIMF

[2]Erasmus University Rotterdam, De Nederlandsche Bank

[3]Joint Vienna Institute

ECB Workshop on Forecasting Techniques | 4 and 5 May 2012

Introduction

- Parameter instability is an important source of forecast failure in macroeconomics and finance: Pesaran and Timmermann (2002), Pesaran, Pettenuzzo and Timmermann (2006), Koop and Potter (2007), Giacomini and Rossi (2009), Inoue and Rossi (2011) among others.

Introduction

- Parameter instability is an important source of forecast failure in macroeconomics and finance: Pesaran and Timmermann (2002), Pesaran, Pettenuzzo and Timmermann (2006), Koop and Potter (2007), Giacomini and Rossi (2009), Inoue and Rossi (2011) among others.
- Two approaches to modelling parameter instability:
  (1) Continuous break process and (2) discrete break process.

Under **discrete break** process a forecasters has a range of options:

- Estimate the break date; use post-break observations.

Under **discrete break** process a forecasters has a range of options:

- Estimate the break date; use post-break observations.
- Estimate the break date and size; use an optimal estimation window (Pesaran and Timmermann 2007, JoE)

Under **discrete break** process a forecasters has a range of options:

- Estimate the break date; use post-break observations.
- Estimate the break date and size; use an optimal estimation window (Pesaran and Timmermann 2007, JoE)
    - The window can be chosen to minimise the MSFE.

Under **discrete break** process a forecasters has a range of options:

- Estimate the break date; use post-break observations.
- Estimate the break date and size; use an optimal estimation window (Pesaran and Timmermann 2007, JoE)
  - The window can be chosen to minimise the MSFE.
  - But the procedure relies on estimating the break date and size.

Under **discrete break** process a forecasters has a range of options:

- Estimate the break date; use post-break observations.
- Estimate the break date and size; use an optimal estimation window (Pesaran and Timmermann 2007, JoE)
    - The window can be chosen to minimise the MSFE.
    - But the procedure relies on estimating the break date and size.
    - The optimal window can be difficult to derive for larger models.

Under **discrete break** process a forecasters has a range of options:

- Estimate the break date; use post-break observations.
- Estimate the break date and size; use an optimal estimation window (Pesaran and Timmermann 2007, JoE)
  - The window can be chosen to minimise the MSFE.
  - But the procedure relies on estimating the break date and size.
  - The optimal window can be difficult to derive for larger models.
- Estimate forecast for a range of estimation windows; Average the forecasts (**AveW**) (Pesaran and Pick 2011, JBES)

Under **discrete break** process a forecasters has a range of options:

- Estimate the break date; use post-break observations.
- Estimate the break date and size; use an optimal estimation window (Pesaran and Timmermann 2007, JoE)
    - The window can be chosen to minimise the MSFE.
    - But the procedure relies on estimating the break date and size.
    - The optimal window can be difficult to derive for larger models.
- Estimate forecast for a range of estimation windows; Average the forecasts (**AveW**) (Pesaran and Pick 2011, JBES)
    - No estimates of the break date and size are necessary; the method works for any model

Under **discrete break** process a forecasters has a range of options:

- Estimate the break date; use post-break observations.
- Estimate the break date and size; use an optimal estimation window (Pesaran and Timmermann 2007, JoE)
  - The window can be chosen to minimise the MSFE.
  - But the procedure relies on estimating the break date and size.
  - The optimal window can be difficult to derive for larger models.
- Estimate forecast for a range of estimation windows; Average the forecasts (**AveW**) (Pesaran and Pick 2011, JBES)
  - No estimates of the break date and size are necessary; the method works for any model
  - Good information about the break date and size can lead to better forecasts

Under **continuous break** process: Exponential smoothing forecasts

- No estimates of the break date and size are required

Under **continuous break** process: Exponential smoothing forecasts

- No estimates of the break date and size are required
- If available, they could improve the forecast

Under **continuous break** process: Exponential smoothing forecasts

- No estimates of the break date and size are required
- If available, they could improve the forecast
- Exponential smoothing forecasts are highly sensitive to the down-weighting parameter.

- We derive **optimal forecasts** under both break processes.

- We derive **optimal forecasts** under both break processes.
- We weigh each observation such that the MSFE is minimized.

- We derive **optimal forecasts** under both break processes.
- We weigh each observation such that the MSFE is minimized.
- Weighting observations encompasses all forecasting methods mentioned above.

- We derive **optimal forecasts** under both break processes.
- We weigh each observation such that the MSFE is minimized.
- Weighting observations encompasses all forecasting methods mentioned above.
- Under the continuous break process we recover the exponential smoothing forecast.

- We derive **optimal forecasts** under both break processes.
- We weigh each observation such that the MSFE is minimized.
- Weighting observations encompasses all forecasting methods mentioned above.
- Under the continuous break process we recover the exponential smoothing forecast.
- Under the discrete break forecast: for $k = 1$ and asymptotically for any k (conditional on the break size and date) the optimal weights follow a step function:
  constant weights within regimes but different weights across regimes.

- We derive **optimal forecasts** under both break processes.
- We weigh each observation such that the MSFE is minimized.
- Weighting observations encompasses all forecasting methods mentioned above.
- Under the continuous break process we recover the exponential smoothing forecast.
- Under the discrete break forecast: for $k = 1$ and asymptotically for any k (conditional on the break size and date) the optimal weights follow a step function:
  constant weights within regimes but different weights across regimes.
- For more regressors, the values of the weights depend on the regressors.

- We also extend the optimal weights to situations where the break date and size are not known.

- We also extend the optimal weights to situations where the break date and size are not known.
- Under a single break: continuous down-weighting and weights have similar shape to AveW and exponential smoothing weights.

- We also extend the optimal weights to situations where the break date and size are not known.
- Under a single break: continuous down-weighting and weights have similar shape to AveW and exponential smoothing weights.
- We provide Monte Carlo to assess the influence of uncertainty around the break date and size.

- We also extend the optimal weights to situations where the break date and size are not known.
- Under a single break: continuous down-weighting and weights have similar shape to AveW and exponential smoothing weights.
- We provide Monte Carlo to assess the influence of uncertainty around the break date and size.
- Finally, we provide an application to forecasting real GDP across 9 advanced and emerging economies using the yield curve.

Optimal weights under different break processes

Consider the linear regression model

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

$\boldsymbol{\beta}_t$ and σ_t^2, are subject to breaks.

The breaks can be **continuous**: $\boldsymbol{\beta}_t$ changes its value in every period. For example:

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

Alternatively, the breaks could be **discrete**: parameters change at distinct points in time, $T_{b,i}$, $i = 1, 2, \ldots, n$,

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

Additionally, σ_t may be subject to a similar break process.
The number of discrete breaks, n, is assumed to be small.
The break sizes, $\left\| \boldsymbol{\beta}_{(i)} - \boldsymbol{\beta}_{(i-1)} \right\|$, could be large relative to σ_t.

- The choice between specifications depends on the particular forecasting problem under consideration.

- The choice between specifications depends on the particular forecasting problem under consideration.
- We propose a general approach to achieve a minimum mean square forecast error: Weigh past observations by weights w_t in the estimation

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

subject to $\sum_{t=1}^{T} w_t = 1$.

- The choice between specifications depends on the particular forecasting problem under consideration.
- We propose a general approach to achieve a minimum mean square forecast error: Weigh past observations by weights w_t in the estimation

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

  subject to $\sum_{t=1}^{T} w_t = 1$.
- The weights $\mathbf{w} = (w_1, w_2, \ldots, w_T)'$ are chosen such that the MSFE of the one-step ahead forecast $\hat{y}_{T+1} = \hat{\beta}_T' \mathbf{x}_{T+1}$ is minimized.

Optimal weights in a model with a single, discrete break

- Consider the model

$$y_t = \beta_t + \sigma_\varepsilon \varepsilon_t, \tag{2}$$

and assume that β_t is subject to a single, discrete break at T_b, $1 < T_b < T$,

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

Optimal weights in a model with a single, discrete break

- Consider the model

$$y_t = \beta_t + \sigma_\varepsilon \varepsilon_t, \tag{2}$$

and assume that β_t is subject to a single, discrete break at T_b, $1 < T_b < T$,

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

- 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

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

- The MSFE scaled by the error variance is

$$E[\sigma_{\varepsilon}^{-2} e_{T+1}^{2}(\mathbf{w})] = 1 + \lambda^2 \left(\sum_{t=1}^{T_b} w_t \right)^2 + \sum_{t=1}^{T} w_t^2, \qquad (3)$$

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

- The MSFE scaled by the error variance is

$$E[\sigma_{\varepsilon}^{-2}e_{T+1}^2(\mathbf{w})] = 1 + \lambda^2 \left(\sum_{t=1}^{T_b} w_t\right)^2 + \sum_{t=1}^{T} w_t^2, \qquad (3)$$

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

- The optimal weights, obtained by minimizing (3) subject to $\sum_{t=1}^{T} w_t = 1$, yields

$$w_{(1)} = \frac{1}{T}\frac{1}{1 + Tb(1-b)\lambda^2}, \qquad (4)$$

and

$$w_{(2)} = \frac{1}{T}\frac{1 + Tb\lambda^2}{1 + Tb(1-b)\lambda^2}. \qquad (5)$$

where $b = T_b/T$.

Comparison to alternative forecasts

- Using post-break observations.
- Optimal estimation window (Pesaran and Timmermann 2007, JoE): Use window that results in minimum MSFE.
- Averaging over estimation windows (Pesaran and Pick 2011, JBES): Average over potential optimal windows to obtain robust forecast.
- Exponential smoothing (Holt 1957): Down-weighting for discrete break processes.

Exact relative MSFE for a single discrete, break for known b and λ

$T = 100$	b		0.95			0.9	
	λ	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($\nu_{\min} = 0.05$)		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

Forecasting with multiple breaks

Consider the linear regression model

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

where the parameter vector β_t is subject to n breaks at break points $b_i = \frac{T_{b,i}}{T}$, such that $b_1 < b_2 < \ldots < b_n$.
Again, initially assume that $n = 2$, such that the parameter vector is

$$\beta_t = \begin{cases} \beta_{(1)} & \text{for } 1 < t \leq T_{b,1} \\ \beta_{(2)} & \text{for } T_{b,1} < t \leq T_{b,2} \\ \beta_{(3)} & \text{for } T_{b,2} < t \leq T \end{cases}$$

Patterns of weights across regimes

- The pattern of the optimal weights depends on the pattern of the slope coefficients across the regimes.

Patterns of weights across regimes

- The pattern of the optimal weights depends on the pattern of the slope coefficients across the regimes.
- If the slopes are rising or falling monotonically ($\beta_{(1)} > \beta_{(2)} > \beta_{(3)}$ or $\beta_{(1)} < \beta_{(2)} < \beta_{(3)}$) the optimal weights also decay monotonically ($w_{(1)} < w_{(2)} < w_{(3)}$), which is in line with down-weighting of observations.

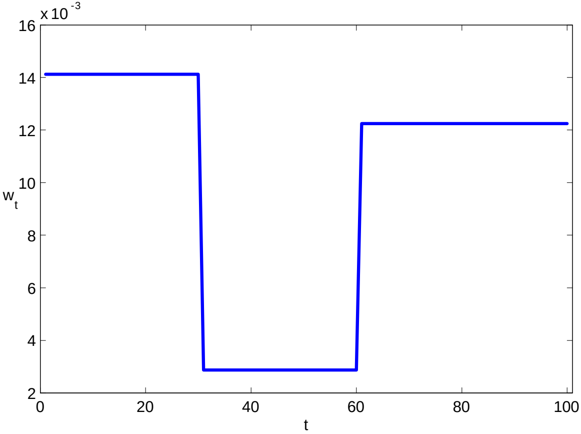
Patterns of weights across regimes

- The pattern of the optimal weights depends on the pattern of the slope coefficients across the regimes.
- If the slopes are rising or falling monotonically ($\beta_{(1)} > \beta_{(2)} > \beta_{(3)}$ or $\beta_{(1)} < \beta_{(2)} < \beta_{(3)}$) the optimal weights also decay monotonically ($w_{(1)} < w_{(2)} < w_{(3)}$), which is in line with down-weighting of observations.
- When this is not the case, it is possible for the middle regime to get less weight than the first and the third regimes.

Patterns of weights across regimes

- The pattern of the optimal weights depends on the pattern of the slope coefficients across the regimes.
- If the slopes are rising or falling monotonically ($\beta_{(1)} > \beta_{(2)} > \beta_{(3)}$ or $\beta_{(1)} < \beta_{(2)} < \beta_{(3)}$) the optimal weights also decay monotonically ($w_{(1)} < w_{(2)} < w_{(3)}$), which is in line with down-weighting of observations.
- When this is not the case, it is possible for the middle regime to get less weight than the first and the third regimes.
- Regimes with the same slope coefficients get the same weights.

Figure: Regime-specific slopes: $\beta_{(1)} = 3, \beta_{(2)} = 1$, and $\beta_{(3)} = 2.5$



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

- The time and the size of the break are unlikely to be known in practice.

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

- The time and the size of the break are unlikely to be known in practice.
- In particular, the size of the break is difficult to estimate unless a considerable number of post-break observations is available.

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

- The time and the size of the break are unlikely to be known in practice.
- In particular, the size of the break is difficult to estimate unless a considerable number of post-break observations is available.
- Need to obtain weights that allow for breaks but are robust to the time and size of break(s).

- For large values of T we get the following approximation

$$w(a,b,q^2,\phi^2) = \frac{1}{T}\left[\frac{1}{1-b} - \frac{1}{1-b}\mathrm{I}(b-a)\right] - \frac{1}{T^2\phi^2}\frac{1}{(1-b)^2} + \frac{1}{T^2\phi^2}\frac{1}{b(1-b)^2}\mathrm{I}(b-a) + O(T^{-3}).$$

where $\phi^2 = \lambda^2\hat{\omega}^2_{x,1}$ and $q^2 = \sigma^2_{(1)}/\sigma^2_{(2)}$.

- For large values of T we get the following approximation

$$\begin{aligned} w(a,b,q^2,\phi^2) &= \frac{1}{T}\left[\frac{1}{1-b} - \frac{1}{1-b}\mathrm{I}(b-a)\right] \\ &\quad -\frac{1}{T^2\phi^2}\frac{1}{(1-b)^2} + \frac{1}{T^2\phi^2}\frac{1}{b(1-b)^2}\mathrm{I}(b-a) + O(T^{-3}). \end{aligned}$$

  where $\phi^2 = \lambda^2\hat{\omega}^2_{x,1}$ and $q^2 = \sigma^2_{(1)}/\sigma^2_{(2)}$.
- The first order term does not depend on the break size or the ratio of error variances.

- For large values of T we get the following approximation

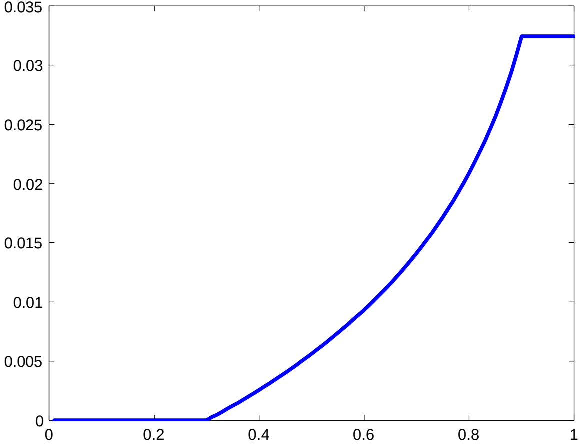
$$w(a,b,q^2,\phi^2) = \frac{1}{T}\left[\frac{1}{1-b} - \frac{1}{1-b}\mathrm{I}(b-a)\right] - \frac{1}{T^2\phi^2}\frac{1}{(1-b)^2} + \frac{1}{T^2\phi^2}\frac{1}{b(1-b)^2}\mathrm{I}(b-a) + O(T^{-3}).$$

  where $\phi^2 = \lambda^2\hat{\omega}^2_{x,1}$ and $q^2 = \sigma^2_{(1)}/\sigma^2_{(2)}$.
- The first order term does not depend on the break size or the ratio of error variances.
- The terms up to order T^{-2} are independent of q^2: a break in the variance is dominated by a break in the mean and slope parameters.

For large T robust weights can be obtained by integrating over $b \sim U(\underline{b}, \bar{b})$.

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

Figure: Approximate optimal weights for break in variance, $T = 100, \underline{b} = 0.3, \bar{b} = 0.9$



- If $\underline{b}$ and $\bar{b}$ are close to 0 and 1, we have

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

- If $\underline{b}$ and $\bar{b}$ are close to 0 and 1, we have

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

- A discrete time version is given by

$$w_t = \frac{-\log(1-t/T)}{T}, \text{ for } t = 1, 2, \ldots, T-1$$

- If $\underline{b}$ and $\bar{b}$ are close to 0 and 1, we have

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

- A discrete time version is given by

$$w_t = \frac{-\log(1-t/T)}{T}, \text{ for } t = 1, 2, \ldots, T-1$$

- The value of w_T is not defined, but can be computed using $t = T - 0.5$, namely

$$w_T = \frac{-\log(0.5/T)}{T}$$

- If $\underline{b}$ and $\bar{b}$ are close to 0 and 1, we have

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

- A discrete time version is given by

$$w_t = \frac{-\log(1-t/T)}{T}, \text{ for } t = 1, 2, \ldots, T-1$$

- The value of w_T is not defined, but can be computed using $t = T - 0.5$, namely

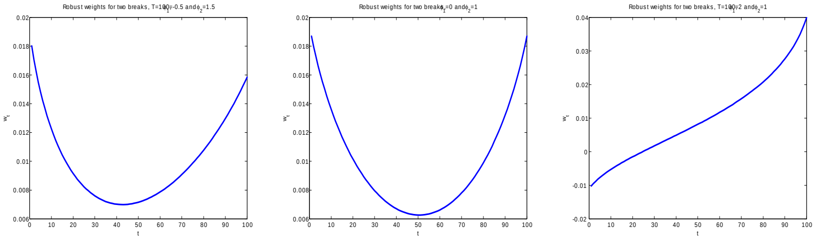
$$w_T = \frac{-\log(0.5/T)}{T}$$

- The resulting weights are

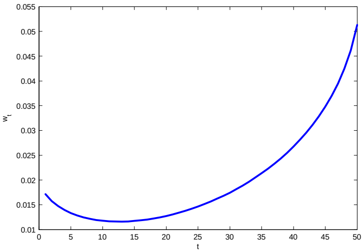
$$w_t^* = \frac{w_t}{\sum_{j=1}^{T} w_j}, \text{ for } t = 1, 2, \ldots, T$$

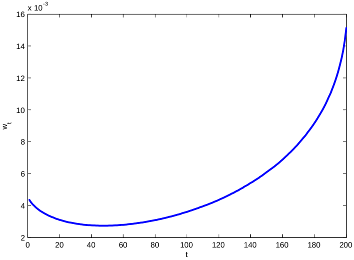
Robust weights for regression models with two breaks

Consider the case of two breaks. Numerical solutions by integrating over a grid for b_1 and b_2



The figure 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 $\phi_{(2)} = 1$.





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. The figure plots the weights when $\phi_{(1)}$ and $\phi_{(2)}$ are integrated with respect to a unform distribution in the range -2 to 2. The first graph shows the weights for $T = 50$ and the second for $T = 200$.

Monte Carlo experiments

- The first model considered is

$$y_t = \mu_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0,1) \quad t = 1, 2, \ldots, T, T+1$$

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

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

Monte Carlo experiments

- The first model considered is

$$y_t = \mu_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0,1) \quad t = 1, 2, \ldots, T, T+1$$

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

  $T = 50, 100, 200$, 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 has a discrete break

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

  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, λ and q are unknown and have to be estimated.

Monte Carlo experiments

- The first model considered is

$$y_t = \mu_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0,1) \quad t = 1, 2, \ldots, T, T+1$$

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

  $T = 50, 100, 200$, 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 has a discrete break

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

  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, λ and q are unknown and have to be estimated.
- We report the relative MSFE:

$$\mathrm{Re\,MSFE} = \frac{\mathrm{MSFE}_i}{\mathrm{MSFE}_{\text{equal w}}}, \; i = \mathit{OptW}, \mathit{PostB}, \mathit{RobustW}, \ldots$$

Relative MSFE: Monte Carlo results for continuous breaks
$T = 100$ and $q = 1$

γ	0.8	0.9	0.95	0.98
δ	4.472	9.487	19.494	49.497
opt.weight(cont.break; δ)	0.444	0.772	0.956	0.999
est.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
est.opt.weight(disc.break; $\hat{b}$, $\hat{\lambda}$)	0.510	0.856	1.085	1.121
rob.weights($\underline{b} = 0.75, \bar{b} = 0.98$)	0.508	0.781	0.963	1.029
rob.weights($\underline{b} = 0, \bar{b} = 1$)	0.620	0.829	0.958	1.007
rob.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

Relative MSFE: Monte Carlo results for a discrete break in drift
$T = 100$ and $q = 1$

b	0.95			0.9		
λ	0.5	1	2	0.5	1	2
opt.weight(disc.break; b, λ)	0.893	0.603	0.256	0.875	0.592	0.257
est.opt.weight(disc.break; $\hat{b}$, $\hat{\lambda}$)	1.022	0.826	0.320	1.014	0.737	0.263
rob.weights($\underline{b} = 0.75, \bar{b} = 0.98$)	0.934	0.796	0.648	0.901	0.705	0.480
rob.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
est.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

- For small breaks ($\lambda = 0.5$) the robust (optimal) weights that assume the break date to fall within a range perform best. Also both AveW and ExpS perform better than methods based on point estimates.
- For larger breaks, the methods using point estimates improve over other methods.
- For $\lambda = 1$ the optimal window forecast performs best, followed by the optimal weights forecast
- For $\lambda = 2$ the post break forecast performs best, closely followed by the optimal weights forecast

The yield curve as a predictor of real economic activity

- The slope of the yield curve has emerged as a valuable leading indicator of GDP growth (Stock and Watson 2003).
- Recent evidence suggests that the relationship between GDP growth and the yield curve changes over time (Estrella, Rodrigues and Schich 2003, Giacomini and Rossi 2006, Schrimpf and Wang 2010).
- The base line specification of our forecast regression is

$$y_{t,t+h} = \beta_0 + \beta_1 s_t + \varepsilon_t$$

where

$$y_{t,t+h} = 100 \ln(Y_{t+h}/Y_t),$$

Y_t is the level of real GDP, the slope of the yield curve, $s_t = i_t^L - i_t^S$, is the difference between the long term interest rate, i_t^L, and the short term interest rate, i_t^S.

- We evaluate the forecasts for various horizons, $h = 1, 2, 3, 4$.
- We use data on GDP and long and short term interest rates from the data set available with the GVAR toolbox (Smith and Galesi 2010).
- We use data for 9 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 with the first forecast for 1994Q1.

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

	Equally weighted ave.			
h	1	2	3	4
1994Q1–2009Q4 (Full forecast evaluations sample)				
equal weight(MSFE)	0.521	1.585	3.046	4.736
estim.opt.weight	1.056	1.035	1.053	1.036
rob.weight(1 break)	0.915	0.946	0.973	0.993
rob.weight(2 breaks)	0.953	0.963	0.974	0.980
post break	1.171	1.162	1.085	1.065
AveW	0.991	0.996	0.997	1.000
ExpS($\gamma = 0.95$)	0.909	0.951	0.978	0.999
ExpS($\gamma = 0.98$)	0.956	0.966	0.977	0.987

Conclusion

- Optimal weights for continuous and discrete break processes.
- Continuous break process: exponential smoothing.
- Discrete break process: new results.
- In practice, dates and sizes of breaks are unknown and estimates are highly unreliable: Robust weights
- Advantage of robust weights: Don't need choice of down-weighting factor.
- Possible to improve robust weights using some information on breaks, such as time interval over which breaks are likely.