Forecasting in the presence of recent structural change

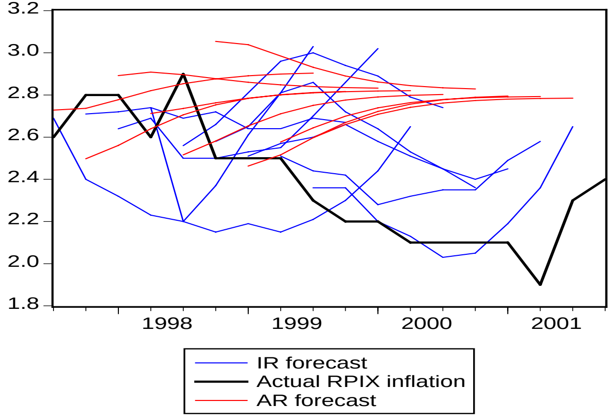
16th IWH-CIREQ Macroeconometric Workshop
Challenges for Forecasting
Structural Breaks, Revisions and Measurement Errors

George Kapetanios[2], Liudas Giraitis[2] and Simon Price[1,3,4]

[1]Bank of England, [2]Queen Mary, University of London, [3]City University and [4]CAMA, ANU

Disclaimer

Inflation Report and real-time AR forecasts of inflation



One thing after another

- Structural change is a major source of forecast error
- Usually assumed to have the form of breaks, characterized by abrupt parameter shifts
- Often appears as a location (mean) shift
- But change can take many forms: smooth, abrupt, stochastic, deterministic
- We don't know *a priori* what form
- Past focus: breaks. Two aspects received attention:
    1. How to detect a break? - Chow (1960), Andrews (1993), Bai and Perron (1998)
    2. How to modify forecasting strategy? - Pesaran-Timmermann (2007)
- But these issues apply more generally to structural change

Forecasting strategies for distant past breaks

Pesaran and Timmermann (2007)

1. Using basic model estimated over post-break data
2. Trading off the variance against the bias of the forecast by estimating the optimal size of the estimation window
3. Estimating optimal estimation window size by cross-validation
4. Combining forecasts from different estimation windows by using weights obtained through cross-validation as in 3
5. Simple average forecast combination with equal weights

Can we use these strategies immediately after we have identified a break?

Can we use these strategies immediately after we have identified a break?

No; due to lack of data

Recognising and dealing with **recent** change when it arrives in **real time**

Few observations available for either estimation or forecast evaluation: what should we do?

1. *Monitoring* for a break, ie, real-time break detection
    - Chu, Stinchcombe and White (1996) - asymptotic proper size under successive and repeated testing, although have low power
    - Problems mitigated by panel tests, Groen, Kapetanios and Price, forthcoming in JAE, but detection remains slow
2. How to modify forecasting strategy? - not discussed in the literature
    - Are breaks rare *or* recurring?
    - Detect a break and react *or* use robust methods?
    - Robust here means variations on discounting past data.

Eklund, Kapetanios and Price (2010) (BoE WP 406) discuss both issues in the case of recurrent breaks: find robust strategies best

Strategies robust to a recent break

- Time varying coefficient models specified in variety of ways
- Alternative: to consider β_t time dependent but deterministic - estimated nonparametrically (kernel based)
- Rolling regressions a pragmatic response
- Exponentially weighted moving averages is a generalisation with declining weights for older observations
- Pesaran and Timmermann forecast combination aggregates different estimation windows

Our previous results

- Systematic theoretical, experimental and empirical examination of strategies appropriate for real-life forecasting activities in the presence of breaks
- First examination of a monitoring-then-combining strategy
- Monitoring and combining works but has few benefits: is safe however
- In Monte Carlo evidence and real data rolling regressions and EWMA are not bad
- But forecast averaging à la Pesaran and Timmermann works well

What's new in the current paper

- Robust strategies need to select parameters that determine how much to discount past data - best to tune in a data dependent manner
- Good forecasts could be obtained by being nonparametrically adaptive to past structural change
- Turns out also to be a novel and simple way to accommodate trends of a completely generic nature - detrending unnecessary
- As persistence increases forecast performance improved by discounting more
- Eg, stationary iid series optimal discount zero: unit root just use last period

Robust strategies

- The model is

$$y_{T,t} = \beta_{T,t} + \epsilon_t$$

  allowing properties to depend on T
- Consider linear forecasts

$$\hat{y}_{t|t-1} \equiv \hat{y}_{t|t-1}(H) = \sum_{j=1}^{t-1} w_{tj;H} y_{t-j} = w_{t1;H} y_{t-1} + \cdots + w_{t,t-1;H} y_1$$

  ie standardized non-negative weights summing to unity
- Parametrised by a single tuning parameter H, controlling the rate at which past observations are downweighted
- Eg, set exponential weighted moving average

$$w_{tj} = w_{tj}(H) = H(1-H)^{t-j}$$

- In existing literature H set *a priori*; we make data-dependent

Cross-validation

- We use cross-validation for estimation problem: estimate H by minimising the RMSFE
- Objective function

$$Q_T(H) := \frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_{t|t-1} - y_t\right)^2; \qquad \hat{H} := \operatorname{argmin}_{H\in I_T} Q_T(H)$$

- Show that under reasonably mild assumptions including identifiability $\hat{H}$ has a limit and is consistent

Consider weights of this class

- For $t = 1, \cdots, T,\ T \geq 1$

$$w_{tj,H} = \frac{K(\frac{t-j}{H})}{\sum_{k=1}^{t} K(\frac{k}{H})}, \quad j = 1, \cdots, t, \qquad H \in I_T$$

- $K(x) \geq 0$, $x \geq 0$ continuous and smooth function such that weights sum to unity

Common examples

- Rolling window
- EWMA
- Triangular window

Consider these stochastic settings for β_t

b1	Stationary process	$\beta_t = \mu, \qquad y_t = \mu + u_t$
b2	Unit root	$\{\beta_t\} \in I(1)$
b3	Deterministic trend	$\beta_t = tg(t/T)$, where $g \in \mathcal{G}$
b4	Bounded unit root	$\beta_t = T^{-1/2}\tilde{\beta}_t$, $\{\tilde{\beta}_t\} \in I(1)$
b5	Bounded deterministic trend	$\beta_t = g(t/T)$, where $g \in \mathcal{G}$
b6	Break in the mean	$\beta_t = \begin{cases} \mu_1, & t = 1, \cdots, L \\ \mu_2, & t = L+1, \cdots, T \end{cases}$

$$T/2 \leq L \leq T$$
$$\mu_1 - \mu_2 \neq 0$$

$\mathcal{G}$ the set of smooth functions

Properties of method

- Thus consider
  - Stationary process (b1)
  - Strong persistence (b2 and b3)
  - Weakly persistent (b4, b5 and b6)
- Under this wide range of time series processes MSE minimisation is a well behaved method
- With these examples we show that the tuning parameter $\hat{H}$ is robustly adjusted to the unknown structure of the data
- The range of $\hat{H}$ may extend over the entire interval I_T

Extension to choice of subsample

- Select optimal subsample *via* a specific tuning parameter
- Two-parameter minimization
- $(\hat{H}, \hat{k})$ can be used to construct forecasts based on optimal subsample $[\hat{k}, \cdots, T]$ ('stability period') and an optimal tuning parameter $\hat{H} = \hat{H}(\hat{k})$ for it

$$Q_T(H, k) := \frac{1}{T-k} \sum_{t=k}^{T} (\hat{y}_{t|t-1} - y_t)^2$$

$$\{\hat{H}, \hat{k}\} := \operatorname{argmin}_{H \in I_T, k \in \{k_{\min}, \cdots, k_{\max}\}} Q_T(H, k)$$

Nonparametric extension

- The model is

$$y_{T,t} = \beta_{T,t} + \epsilon_t$$

  where β_t and $\sigma_{\epsilon,t}$ smooth deterministic functions of t, estimated nonparametrically using kernels

- Consider forecasts

$$\hat{y}_{t|t-1} = \sum_{j=1}^{t-1} w_{tj} y_{t-j}$$

- An attractive feature is that we do not impose monotonicity on the weights

Extension to dynamic weighting (AR)

$$\tilde{w}_{tj,H} = \begin{cases} w_{t-j}, & j = t-1, \cdots, t-p, \\ K(\frac{t-j}{H}), & j = 1, \cdots, t-p-1, \qquad H \in I_T, \end{cases}$$

- Allow the first p weights $w_1, \cdots, w_p$ $(p \geq 0)$ to vary freely and standardize weights: $w_{tj,H} = \frac{\tilde{w}_{tj,H}}{\sum_{j=1}^{t} \tilde{w}_{tj,H}}$
- Allows initial lags of y_t to enter freely into the forecast
- Q_T can be minimized jointly over $H, \tilde{w}_1, \cdots, \tilde{w}_p$, and, potentially, even p

Extension to regression models

$$y_{T,t} = \beta'_{T,t} x_t + u_t, \qquad t = 1, \cdots, T, \quad T \geq 1$$

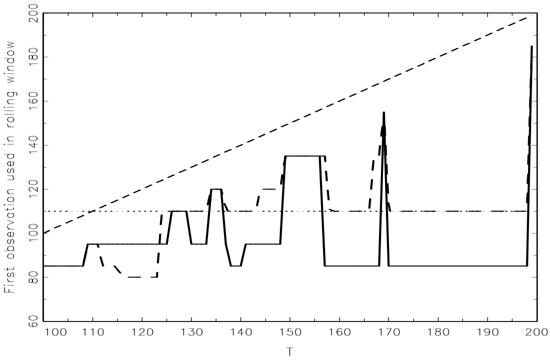
- x_t is a $K \times 1$ vector of predetermined (stochastic) variables, β_t's are $K \times 1$ vectors of parameters, u_t stationary dependent noise process that is independent of x_t
- Set $\beta_t = \left(E\left(x_t x_t'\right)\right)^{-1} E\left(x_t' y_t\right) = \left(\Sigma_t^{xx}\right)^{-1} \Sigma_t^{xy}$, where $\Sigma_t^{xx} = [\sigma_{ij,t}^{xx}]$, and $\Sigma_t^{xy} = [\sigma_{i,t}^{xy}]$ are corresponding covariance matrices
- Aim to estimate the expectations Σ_t^{xx} and Σ_t^{xy} over time by the robust methods outlined above
- $z_{ij,t} = x_{i,t} x_{j,t}$ and $z_{i,t} = x_{i,t} y_t$ are simple location models: $z_{ij,t} = \sigma_{ij,t}^{xx} + u_{ij,t}$, and $z_{i,t} = \sigma_{i,t}^{xy} + u_{i,t}$
- Regression reduced to estimation of a sequence of simple location models

Summary theoretical results

- Cut to the chase - why use a model for forecasting which was developed for some other purpose?
- Well behaved method of forecasting under structural change
- Also a very flexible way to allow for time variation in parameters
- Flexible way of dealing with different time series properties; eg
  - stationary series selects long window
  - a unit root process will select a short window
- Hence very powerful and flexible tool that requires no prior transformations
- Can be generalised to handle
  - dynamic weighting and allow the first p weights to vary freely
  - a regression model

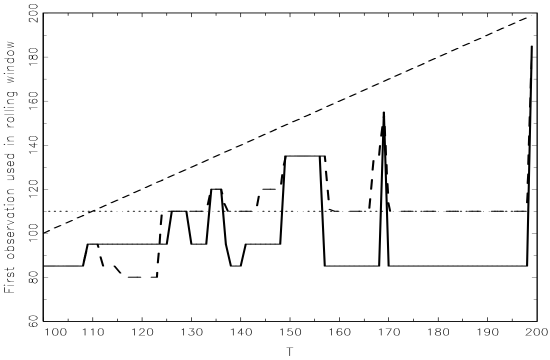
Example: window for a structural break

Solid line start window no structural change: long-dashed – – start window for break at observation 110: dotted ... the first post break observation: short-dashed - - - last observation in the window



Example: window for a structural break

Solid line start window no structural change: long-dashed – – start window for break at observation 110: dotted ... the first post break observation: short-dashed - - - last observation in the window

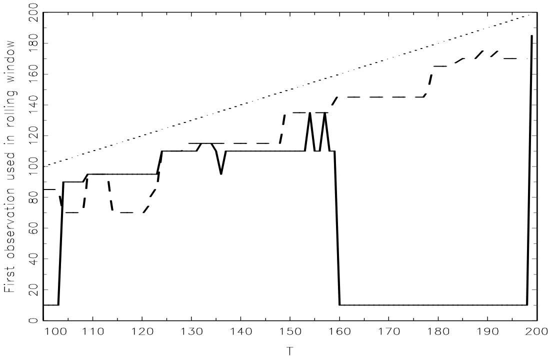


With structural break, window shorter

In this case picks up break accurately (window start 111)

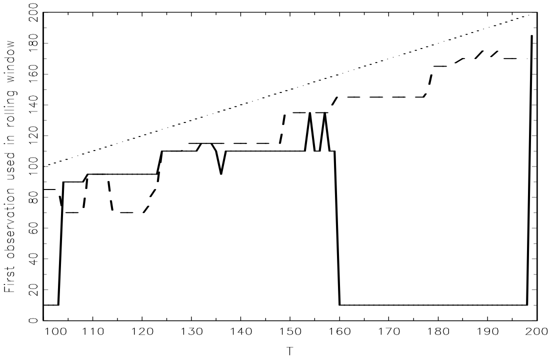
Example: window with a unit root

Solid line start window no structural change: dashed – – start point of the window: dotted ... last observation in the window



Example: window with a unit root

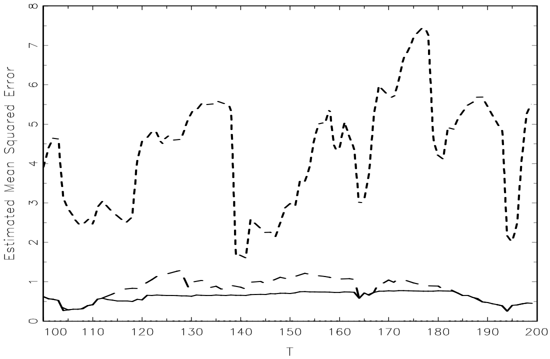
Solid line start window no structural change: dashed – – start point of the window: dotted ... last observation in the window



With unit root, window much smaller over most of the sample

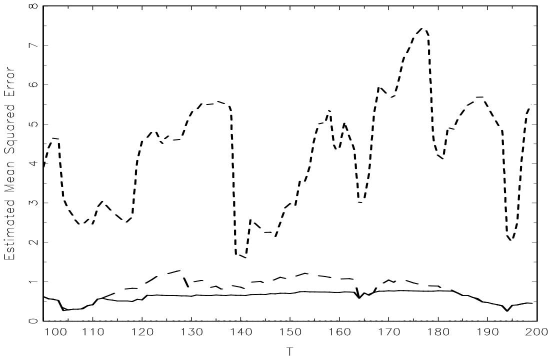
MSE

Solid line stationary: long-dashed – – for structural break: short-dashed - - - unit root



MSE

Solid line stationary: long-dashed – – for structural break: short-dashed - - - unit root



MSE ranked as expected (unit root > structural break > stationary)

Forecasts

- Now consider forecast performance - first Monte Carlo
- All forecasts considered explicitly weight past data in some way: deterministic weighting functions, Pesaran-Timmerman averaging, and nonparametric weights
- Forecasts start $T^0 = 100$, evaluated to $T = 200$

Forecast methods

- **Exponential** EWMA ρ fixed or cross-validated $(\hat{\rho})$
- **Rolling** Flat-weight window H fixed or cross-validated cross-validated $(\hat{H})$: also allowing for tuning subsample $\hat{k}$
- **Averaging** Pesaran-Timmerman averaging over all possible estimation periods
- **Polynomial** polynomial weights $w_{tj;H} = \frac{(t-j)^H}{\sum_{k-1}^{t-1} k^{-H}}$ $1 \leq j \leq t-1$:
- **Nonparametric**

Design of structural change

Model for generating the data:

$$y_t = \beta_t + \epsilon_t, \qquad t = 1, \ldots, T, \ldots$$

$\epsilon_t \sim niid(0, 1)$ and β_t is either a deterministic function of time or a normalised random walk.

1. $y_t = \epsilon_t$ *null case of no structural change*
2. $y_t = 0.05t + 5\epsilon_t$ *linear monotonic*
3. $y_t = 0.05t^{0.5+0.75\frac{t}{T}} + 5\epsilon_t$ *non-linear monotonic*
4. *break* $y_t = \begin{cases} \epsilon_t, & \text{if } t \leq \frac{11T}{20} \\ 1 + \epsilon_t, & \text{if } t > \frac{11T}{20} \end{cases}$
5. $y_t = 2\sin\left(\frac{2\pi t}{T}\right) + 3\epsilon_t$ *cyclical*
6. $y_t = 5\sin\left(\frac{2\pi t}{T}\right) + 3\epsilon_t$ *cyclical*
7. $y_t = (0.025t - 2.5)^2 + 5\epsilon_t$ *humped trends*
8. $y_t = (0.025t - 2.5)^2 + 3\epsilon_t$ *humped trends*
9. β_t is *bounded stochastic trend*, $y_t = \frac{2}{\sqrt{T}}\sum_{i=1}^{t} v_i + \epsilon_t$ where $v_t \sim niid(0,1)$
10. β_t is *bounded stochastic trend*, $y_t = \frac{2}{\sqrt{T}}\sum_{i=1}^{t} v_i + 0.05t + \epsilon_t$
11. β_t is *unit root*, $y_t = 2\sum_{i=1}^{t} v_i + \epsilon_t$

where $v_t \sim niid(0,1)$

Table 1: Monte Carlo results, T=200, 1-step ahead RRMSE against a full-sample benchmark

	Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	Ex11
Exponential rhohat	1.05	0.70	0.17	0.77	0.81	0.34	0.99	0.83	0.67	0.70	0.17
Rolling Hhat	1.13	0.75	0.20	0.83	0.87	0.37	1.04	0.88	0.76	0.73	0.33
Rolling H 20	1.05	0.67	0.21	0.76	0.77	0.34	0.94	0.80	0.67	0.67	0.29
Rolling H 30	1.03	0.67	0.27	0.76	0.78	0.38	0.94	0.82	0.69	0.66	0.36
Exponential rho 0.99	1.00	0.84	0.75	0.90	0.91	0.77	0.99	0.97	0.87	0.84	0.74
rho = 0.95	1.02	0.67	0.30	0.76	0.78	0.41	0.94	0.83	0.68	0.67	0.34
rho = 0.90	1.05	0.67	0.19	0.74	0.77	0.33	0.94	0.79	0.65	0.67	0.23
rho = 0.80	1.10	0.71	0.16	0.76	0.80	0.33	0.98	0.82	0.66	0.70	0.18
rho = 0.70	1.17	0.75	0.16	0.80	0.85	0.34	1.04	0.86	0.69	0.74	0.17
rho = 0.50	1.32	0.85	0.18	0.90	0.96	0.38	1.17	0.97	0.77	0.83	0.17
Averaging	1.01	0.75	0.64	0.84	0.86	0.63	0.99	0.97	0.80	0.75	0.61
Nonparametric	1.11	0.69	0.17	0.76	0.79	0.32	0.97	0.80	0.66	0.68	0.21
Polynomial ahat	1.01	0.77	0.44	0.94	0.92	0.56	1.01	0.95	0.79	0.77	0.35
Rolling Hhat khat	1.15	0.78	0.21	0.85	0.90	0.44	1.05	0.89	0.78	0.75	0.28

For each column, blue is best - red is worst

Summary of results Table 1

- Ex1 is the no-change baseline - downweighting worsens
- Down-weighting is good in almost all cases where there is structural change (RRMSE < 1)
- Tuning parameters can make a difference - EWMA fixed weights can perform poorly
- Tuned EWMA is never a poor performer
- Averaging not bad but towards worst end of methods
- Nonparametric methods offer a powerful alternative - in this case where there is change dominate the parametric tuned methods

Table 2: Monte Carlo results, T=200, $u_t \sim AR(0.7)$, 1-step ahead RRMSE against a full-sample benchmark

	Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	Ex11
Exponential rhohat	0.66	0.39	0.09	0.63	0.47	0.19	0.58	0.48	0.41	0.41	0.12
Rolling Hhat	1.02	0.66	0.13	0.86	0.62	0.28	0.79	0.67	0.56	0.57	0.14
Rolling H 20	1.03	0.65	0.20	1.01	0.77	0.33	0.89	0.77	0.69	0.65	0.33
Rolling H 30	1.02	0.66	0.26	1.03	0.79	0.37	0.92	0.80	0.72	0.66	0.41
Exponential rho 0.99	0.99	0.82	0.75	0.98	0.90	0.76	0.97	0.96	0.87	0.83	0.76
rho = 0.95	0.94	0.60	0.28	0.92	0.72	0.38	0.85	0.75	0.66	0.61	0.37
rho = 0.90	0.87	0.54	0.17	0.85	0.64	0.28	0.76	0.65	0.57	0.54	0.25
rho = 0.80	0.78	0.47	0.12	0.76	0.57	0.23	0.68	0.57	0.50	0.48	0.18
rho = 0.70	0.72	0.43	0.10	0.69	0.51	0.21	0.62	0.52	0.45	0.44	0.15
rho = 0.50	0.64	0.38	0.09	0.62	0.46	0.19	0.57	0.47	0.40	0.40	0.12
Averaging	1.00	0.73	0.63	0.99	0.85	0.62	0.97	0.95	0.81	0.74	0.65
Nonparametric	1.02	0.59	0.14	0.98	0.72	0.27	0.85	0.68	0.62	0.59	0.23
Polynomial ahat	0.71	0.47	0.18	0.74	0.56	0.35	0.64	0.53	0.55	0.45	0.21
Rolling Hhat khat	0.94	0.60	0.12	0.72	0.51	0.23	0.66	0.55	0.45	0.47	0.11

For each column, blue is best - red is worst

Summary of results Table 2

- Similar results
- Down-weighting also good in this more general case where structural change
- And even no change case often improved by downweighting
- The tuned EWMA does particularly well
- Tuned EWMA is never a poor performer

Table 3: Monte Carlo results, T=200, 2-step ahead RRMSE against a full-sample benchmark

	Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	Ex11
Exponential rhohat	1.05	0.69	0.17	0.76	0.80	0.34	0.98	0.82	0.66	0.70	0.19
Rolling Hhat	1.07	0.70	0.20	0.82	0.81	0.34	0.97	0.91	0.69	0.72	0.18
Rolling H 20	1.04	0.66	0.22	0.74	0.77	0.34	0.94	0.80	0.66	0.67	0.29
Rolling H 30	1.03	0.66	0.28	0.75	0.77	0.38	0.94	0.82	0.69	0.67	0.36
Exponential rho 0.99	1.00	0.84	0.76	0.89	0.91	0.77	0.99	0.98	0.87	0.84	0.74
rho = 0.95	1.02	0.67	0.31	0.75	0.78	0.41	0.94	0.83	0.68	0.68	0.35
rho = 0.90	1.05	0.66	0.20	0.73	0.77	0.33	0.94	0.79	0.64	0.68	0.24
rho = 0.80	1.10	0.70	0.17	0.75	0.80	0.32	0.98	0.81	0.65	0.71	0.20
rho = 0.70	1.17	0.74	0.17	0.79	0.84	0.34	1.04	0.85	0.68	0.75	0.19
rho = 0.50	1.32	0.83	0.18	0.89	0.95	0.38	1.17	0.95	0.76	0.84	0.19
Averaging	1.00	0.75	0.65	0.84	0.86	0.63	0.99	0.97	0.80	0.76	0.61
Nonparametric	1.10	0.68	0.17	0.75	0.78	0.32	0.96	0.80	0.65	0.69	0.22
Polynomial ahat	1.01	0.78	0.40	0.96	0.93	0.57	1.00	0.96	0.78	0.79	0.35
Rolling Hhat khat	1.06	0.71	0.20	0.81	0.82	0.35	0.95	0.90	0.67	0.72	0.15

For each column, blue is best - red is worst

Summary of results Table 3

- Once again, similar results
- Tuning is a good idea

Empirical exercise for the UK

- UK: 94 macro series, 1992Q1 to 2008Q2: *sub-periods* 1992Q1-1999Q4, 2000Q1-2008Q2
- All series transformed to stationarity for a fair comparison
- 33 series exhibited breaks (based on Bai-Perron mean shift in an AR)
- Unconditional mean often a good forecast - report location model
- Compare RMSFEs to the corresponding equal-weight / full-sample benchmark

Results reported

- Median RRMSE (relative to full sample benchmark) UK
- Summary statistics: RRMSE min, max, variance and skew
- Diebold Mariano 1: number of cases DM favours the benchmark (null: equality of robust method and benchmark against alternative: benchmark is better)
- Diebold Mariano 2: number of cases DM favours the robust method (null: equality of robust method and benchmark against alternative: robust method is better)

Table 4a: Location model, UK 200Q1 - 2008Q2

	Median	Min	Max	Var	Skew	DM1	DM2
Exponential rhohat	0.858	0.006	1.28	0.309	-1.233	2	21
Rolling Hhat	0.886	0.006	1.503	0.3	-1.207	2	19
Rolling H 20	0.887	0.005	1.518	0.309	-1.063	4	18
Rolling H 30	0.903	0.006	1.845	0.312	-0.821	6	19
Exponential rho 0.99	0.927	0.462	1.06	0.127	-1.703	3	26
rho = 0.95	0.858	0.007	1.252	0.27	-1.437	5	22
rho = 0.90	0.858	0.005	1.254	0.299	-1.233	6	20
rho = 0.80	0.884	0.005	1.273	0.327	-1.078	9	21
rho = 0.70	0.929	0.006	1.409	0.36	-0.907	12	20
rho = 0.50	1.047	0.007	1.755	0.438	-0.623	22	19
Averaging	0.883	0.069	1.203	0.235	-1.625	3	22
Nonparametric	0.926	0.034	1.699	0.351	-0.87	8	20
Polynomial ahat	0.863	0.011	1.263	0.266	-1.203	0	22
Rolling Hhat khat	0.86	0.005	1.292	0.292	-1.266	1	22

For each column, blue is best - red is worst

Table 4b: Location model, UK 1992Q1 - 1999Q4

	Median	Min	Max	Var	Skew	DM1	DM2
Exponential rhohat	0.80	0.01	1.45	0.37	-0.55	4	29
Rolling Hhat	0.90	0.01	1.59	0.30	-0.94	7	26
Rolling H 20	0.93	0.01	1.52	0.31	-0.96	12	22
Rolling H 30	0.90	0.01	1.33	0.27	-1.19	10	24
Exponential rho 0.99	0.95	0.70	1.02	0.08	-1.32	0	40
rho = 0.95	0.84	0.10	1.11	0.24	-1.15	2	37
rho = 0.90	0.81	0.01	1.22	0.30	-0.91	4	33
rho = 0.80	0.82	0.01	1.36	0.36	-0.67	7	31
rho = 0.70	0.84	0.01	1.47	0.40	-0.52	10	29
rho = 0.50	0.93	0.01	1.72	0.49	-0.32	13	27
Averaging	0.88	0.26	1.18	0.19	-1.32	2	33
Nonparametric	0.90	0.04	1.59	0.38	-0.77	7	21
Polynomial ahat	0.82	0.02	1.37	0.33	-0.82	0	26
Rolling Hhat khat	0.82	0.01	1.16	0.27	-1.20	2	30

For each column, blue is best - red is worst

Summary of results for location model, UK

- All these methods have value
- In the best cases do VERY well
- In the worst cases never a disaster
- Tuning parameters is a good idea
- All models significantly outperform the equal weight benchmark in at least 19% of cases and in the first half more often

Table 5a: Location model, US 200Q2 - 2008Q3

	Median	Min	Max	Var	Skew	DM1	DM2
Exponential rhohat	0.64	0.02	1.25	0.38	-0.18	0	37
Rolling Hhat	0.88	0.08	1.61	0.30	-0.76	14	24
Rolling H 20	0.87	0.08	1.64	0.32	-0.67	11	23
Rolling H 30	0.86	0.13	1.62	0.29	-0.84	6	30
Exponential rho 0.99	0.94	0.61	1.08	0.09	-1.30	9	42
rho = 0.95	0.80	0.20	1.35	0.27	-0.78	2	38
rho = 0.90	0.75	0.07	1.32	0.30	-0.67	1	39
rho = 0.80	0.69	0.04	1.20	0.32	-0.34	3	39
rho = 0.70	0.66	0.03	1.31	0.36	-0.09	4	39
rho = 0.50	0.66	0.02	1.56	0.42	0.18	4	38
Averaging	0.91	0.45	1.15	0.14	-1.18	11	40
Nonparametric	0.83	0.04	1.58	0.34	-0.52	2	22
Polynomial ahat	0.67	0.02	1.35	0.38	-0.22	2	38
Rolling Hhat khat	0.80	0.08	1.62	0.28	-0.66	2	40

For each column, blue is best - red is worst

Table 5b: Location model, US 1992Q2 - 2000Q1

	Median	Min	Max	Var	Skew	DM1	DM2
Exponential rhohat	0.65	0.01	1.29	0.39	-0.21	1	39
Rolling Hhat	0.90	0.14	1.92	0.36	0.31	6	30
Rolling H 20	1.00	0.12	2.74	0.51	1.22	10	27
Rolling H 30	0.96	0.16	2.06	0.38	0.67	10	30
Exponential rho 0.99	1.02	0.67	2.06	0.19	2.90	24	21
rho = 0.95	0.90	0.16	1.74	0.30	-0.08	11	30
rho = 0.90	0.82	0.07	1.79	0.38	0.05	5	34
rho = 0.80	0.76	0.03	1.66	0.41	-0.17	5	36
rho = 0.70	0.73	0.02	1.32	0.43	-0.20	4	37
rho = 0.50	0.73	0.01	1.59	0.50	0.15	5	39
Averaging	1.07	0.46	3.32	0.39	3.41	24	20
Nonparametric	1.11	0.05	4.81	0.92	2.18	16	30
Polynomial ahat	0.73	0.00	1.18	0.33	-0.56	3	37
Rolling Hhat khat	0.82	0.10	1.34	0.30	-0.69	2	30

For each column, blue is best - red is worst

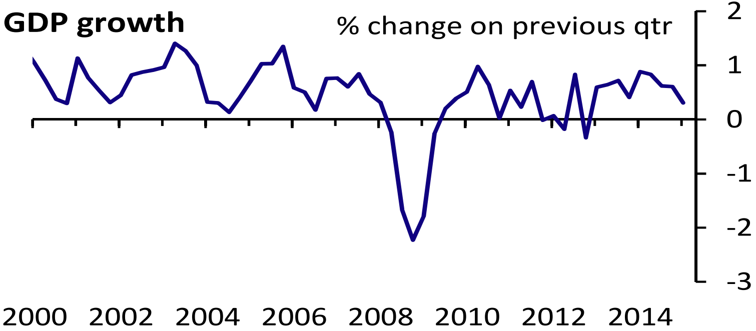
Summary of results for location model, US

- Similar results but more variable
- Some very large improvements

Update - using it in anger in forecasting growth

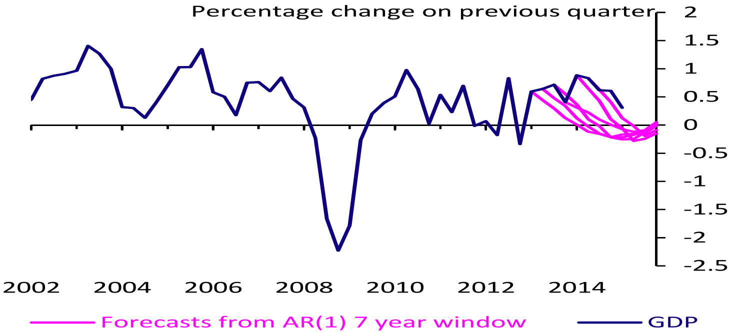
- Bank of England has a ‘suite’ of statistical forecasting models - perhaps better referred to as non-structural models
- Model combination using IC
- Initially handled mean-breaks by testing for breaks and demeaning, but not a good solution
- So adopted a pragmatic approach with a 7-year rolling estimation window
- Consequently the Suite outperformed judgemental and structural models in the recession
- Although also quickly gave random walk a large weight

GDP growth



- Rolling window quickly captures mean-shift

Forecasts from a 7-year window AR(1)



- Less good in the recovery though

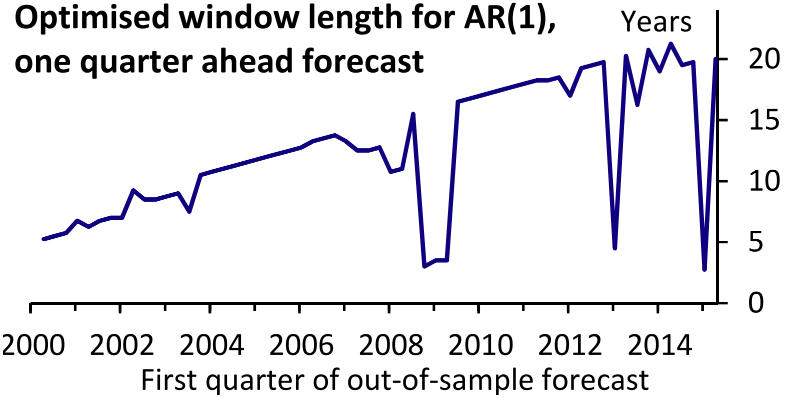
Bank Underground blog post

**Forecasting GDP in the presence of breaks:
when is the past a good guide to the future?**

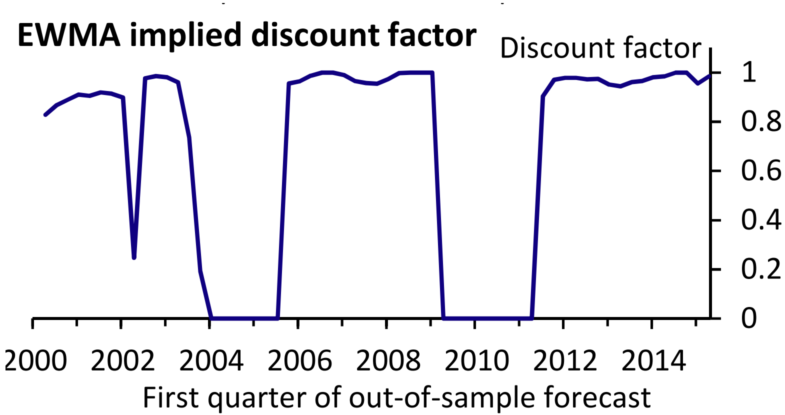
George Kapetanios, Simon Price and Sophie Stone

- We lengthened the window - was there a scientific rationale for that?

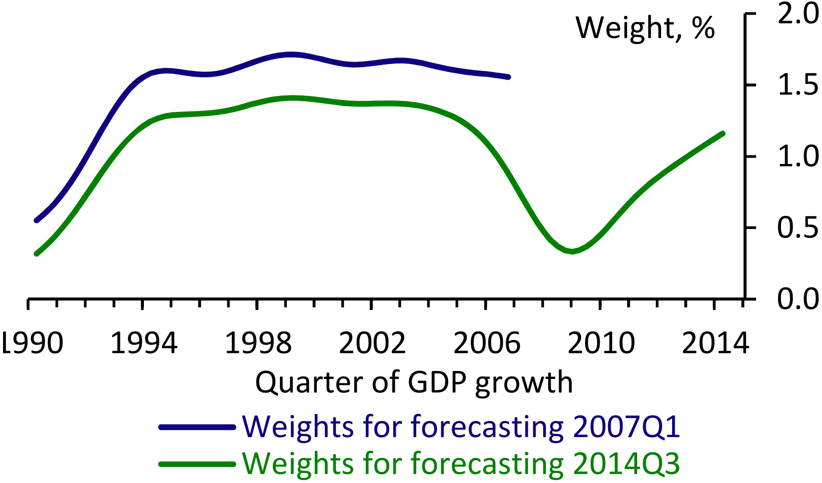
Optimal estimation window, AR(1)



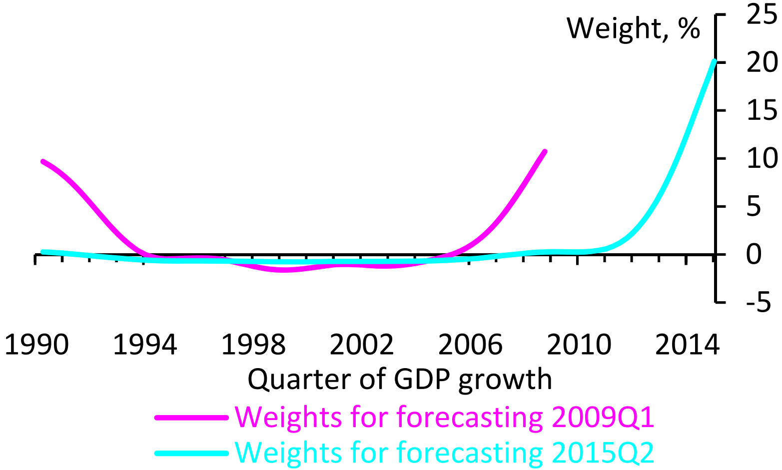
Optimal discount factor



Optimal non-parametric weights pre-crisis



Optimal non-parametric weights post-crisis



Out-of-sample RMSFEs for quarterly GDP growth, h quarters ahead

Forecast evaluation period:	2002-14			2014		
	h = 1	h = 4	h = 8	h = 1	h = 4	h = 8
AR(1) seven-year window	0.53	0.85	0.95	0.31	0.75	0.72
AR(1) optimised window	0.56	0.81	1.09	0.23	0.61	0.44
EWMA2	0.74	0.86	0.80	0.31	0.43	0.38
Non-parametric	0.70	0.87	0.84	0.27	0.46	0.56

- Direct forecasts in pseudo-real time
- Optimised using cross-validation window of ten quarters

Out-of-sample RMSFEs for quarterly GDP growth, h quarters ahead

Forecast evaluation period:	2002-14		
	h = 1	h = 4	h = 8
AR(1) seven-year window	0.53	0.85	0.95
AR(1) optimised window	0.56	0.81	1.09
EWMA2	0.74	0.86	0.80
Non-parametric	0.70	0.87	0.84

- Direct forecasts in pseudo-real time
- Optimised using cross-validation window of ten quarters

Out-of-sample RMSFEs for quarterly GDP growth, h quarters ahead

Forecast evaluation period:	2014		
	h = 1	h = 4	h = 8
AR(1) seven-year window	0.31	0.75	0.72
AR(1) optimised window	0.23	0.61	0.44
EWMA	0.31	0.43	0.38
Non-parametric	0.27	0.46	0.56

- Direct forecasts in pseudo-real time
- Optimised using cross-validation window of ten quarters

Conclusions

- Method tailored to task at hand - forecasting
- Downweighting of data robust - to time series properties and type of break
- It is possible to tune the parameter choice with cross-validation
- A novel approach to modelling time variation
- Can be completely agnostic about time series properties of the data - no need to transform series
- In practice, low-discount EWMA are good but cannot know this *ex ante* - tuning avoids bad discount choices - practically useful methods

Conclusions

- Method tailored to task at hand - forecasting
- Downweighting of data robust - to time series properties and type of break
- It is possible to tune the parameter choice with cross-validation
- A novel approach to modelling time variation
- Can be completely agnostic about time series properties of the data - no need to transform series
- In practice, low-discount EWMA are good but cannot know this *ex ante* - tuning avoids bad discount choices - practically useful methods
- You know what? It works!