# Does modeling a structural break improve forecast accuracy?

Tom Boot
*University of Groningen*

Andreas Pick
*Erasmus University Rotterdam*
*Tinbergen Institute*
*De Nederlandsche Bank*

10th ECB Workshop on Forecasting Techniques 2018

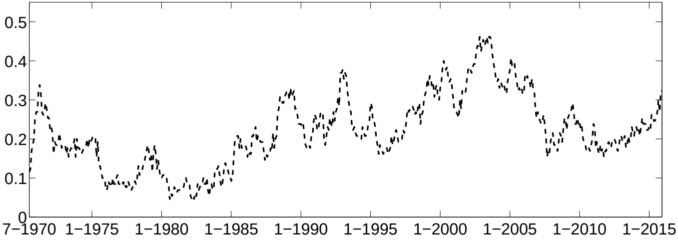
# INSTABILITY IN MACROECONOMIC SERIES

Stock and Watson (1996): strong evidence for structural breaks

# INSTABILITY IN MACROECONOMIC SERIES

Stock and Watson (1996): strong evidence for structural breaks

Figure: Fraction of series where a break is found




- 130 macroeconomic and financial time series (FRED-MD), 1960M01-2015M10
- AR(1) using a moving window of 120 observations
- Andrews (1993) (heteroskedasticity robust) sup-F test

# Forecasting under model instability

Clements and Hendry (1998) view structural breaks as a key reason for forecasting failure

Can explain lack of predictability in:

- Stock returns — Ang and Bekaert (2004); Rapach and Wohar (2006); Paye and Timmermann (2006)
- Interest rates and inflation — Rapach and Wohar (2005)
- Exchange rates — Rossi (2006)

# FORECASTING UNDER MODEL INSTABILITY

Clements and Hendry (1998) view structural breaks as a key reason for forecasting failure

Can explain lack of predictability in:

- Stock returns — Ang and Bekaert (2004); Rapach and Wohar (2006); Paye and Timmermann (2006)
- Interest rates and inflation — Rapach and Wohar (2005)
- Exchange rates — Rossi (2006)

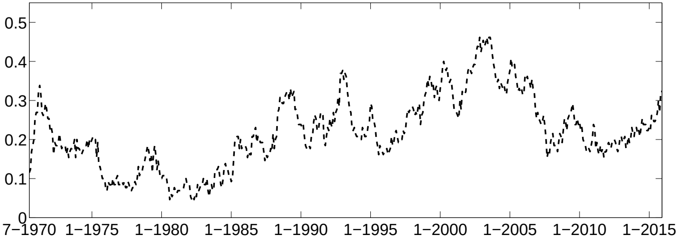
What should one do?

- 'Unbiasedness': post-break data
- Use some pre-break observations — Pesaran & Timmermann (2005, 2007)
- Optimally weight all observations — Pesaran, Pick & Pranovich (2013)

We need to know (1) whether there is a break and (2) the break date.

# Are breaks really that bad?

Figure: Fraction of series where a break is found




- Breaks in the parameters $\neq$ breaks in forecasts
- If breaks are ‘small’, break-models are not always better
  - Elliott and Müller (2007,2014): Large uncertainty around break date

# Testing for breaks from a forecasting perspective

We develop a break point test when forecasting under MSFE loss

1. Test for the break in the forecast, not in the parameters
2. Taking into account the full bias-variance trade-off:
   (a) shorter window, but no/smaller bias vs. longer window, but larger bias
   (b) break date uncertainty in break model

# TESTING FOR BREAKS FROM A FORECASTING PERSPECTIVE

We develop a break point test when forecasting under MSFE loss

1. Test for the break in the forecast, not in the parameters
2. Taking into account the full bias-variance trade-off:
   (a) shorter window, but no/smaller bias vs. longer window, but larger bias
   (b) break date uncertainty in break model

**Main empirical finding**
Breaks that are relevant for forecasting occur much less frequent than existing tests indicate

# Case 1: Known break date

Consider two forecasts:

1) Forecast conditional on a structural break at $T_b = T\tau_b$

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_1 \cdot \mathrm{I}[t < T_b] + \boldsymbol{x}_t'\boldsymbol{\beta}_2 \cdot \mathrm{I}[t \geq T_b] + \varepsilon_t$$

2) Forecast that ignores breaks

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

# Case 1: Known Break Date

Consider two forecasts:

1) Forecast conditional on a structural break at $T_b = T\tau_b$

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_1 \cdot \mathrm{I}[t < T_b] + \boldsymbol{x}_t' \boldsymbol{\beta}_2 \cdot \mathrm{I}[t \geq T_b] + \varepsilon_t$$

2) Forecast that ignores breaks

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

Expected mean squared forecast error

$$1) \quad \mathrm{E}\left[\left(\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_2 - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2 - \varepsilon_{T+1}\right)^2\right] = \frac{1}{T - T_b}\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1} + \sigma^2$$

with $\boldsymbol{V} = \mathrm{plim}_{T\to\infty}(T - T_b)\mathrm{Var}(\hat{\boldsymbol{\beta}}_2) = \mathrm{plim}_{T\to\infty}\, T\mathrm{Var}(\hat{\boldsymbol{\beta}}_F)$

# CASE 1: KNOWN BREAK DATE

Consider two forecasts:

1) Forecast conditional on a structural break at $T_b = T\tau_b$

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_1 \cdot \mathrm{I}[t < T_b] + \boldsymbol{x}_t'\boldsymbol{\beta}_2 \cdot \mathrm{I}[t \geq T_b] + \varepsilon_t$$

2) Forecast that ignores breaks

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

Expected mean squared forecast error

$$1) \quad \mathrm{E}\left[\left(\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_2 - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2 - \varepsilon_{T+1}\right)^2\right] = \frac{1}{T - T_b}\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1} + \sigma^2$$

$$2) \quad \mathrm{E}\left[\left(\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_F - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2 - \varepsilon_{T+1}\right)^2\right] = \frac{1}{T}\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1} + \sigma^2 + \left[\frac{T_b}{T}\boldsymbol{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2$$

with $\boldsymbol{V} = \mathrm{plim}_{T\to\infty}\,\mathrm{Var}(\hat{\boldsymbol{\beta}}_i)$

# When is the full sample forecast better?

$$\mathrm{E}\left[MSFE(\hat{\beta}_2)\right] = \frac{1}{T - T_b}\boldsymbol{x}_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1} + \sigma^2$$

$$\mathrm{E}\left[MSFE(\hat{\beta}_F)\right] = \frac{1}{T}\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1} + \sigma^2 + \left[\frac{T_b}{T}\boldsymbol{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2$$

Full sample forecast is more accurate if

$$\zeta = T\frac{\left[\boldsymbol{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2}{\boldsymbol{x}'_{T+1}\left(\frac{\boldsymbol{V}_1}{\tau_b} + \frac{\boldsymbol{V}_2}{1-\tau_b}\right)\boldsymbol{x}_{T+1}} \leq 1$$

# TEST STATISTIC FOR FORECASTING

Test for $H_0 : \zeta \leq 1$:

$$\hat{\zeta} = T \frac{(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)' \boldsymbol{x}_{T+1} \boldsymbol{x}'_{T+1} (\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)}{\boldsymbol{x}'_{T+1} \left( \frac{\hat{\boldsymbol{V}}_1}{\tau_b} + \frac{\hat{\boldsymbol{V}}_2}{1-\tau_b} \right) \boldsymbol{x}_{T+1}} \overset{H_0}{\sim} \chi^2(1, 1)$$

# Test statistic for forecasting

Test for $H_0 : \zeta \leq 1$:

$$\hat{\zeta} = T \frac{(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)' \boldsymbol{x}_{T+1} \boldsymbol{x}'_{T+1} (\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)}{\boldsymbol{x}'_{T+1} \left( \frac{\hat{\boldsymbol{V}}_1}{\tau_b} + \frac{\hat{\boldsymbol{V}}_2}{1-\tau_b} \right) \boldsymbol{x}_{T+1}} \overset{H_0}{\sim} \chi^2(1, 1)$$

Differs from the standard Wald statistic for $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$

$$\hat{W} = T(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)' \left[ \frac{\hat{\boldsymbol{V}}_1}{\tau_b} + \frac{\hat{\boldsymbol{V}}_2}{1 - \tau_b} \right]^{-1} (\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1) \overset{H_0}{\sim} \chi^2(\dim \boldsymbol{\beta})$$

We get

1. A test statistic weighted by $\boldsymbol{x}_{T+1}$
2. A non-central $\chi^2$ distribution

# Case II: Unknown break date

To account for uncertainty in the break date, consider 'local' breaks of $O(T^{-1/2})$.

$$\sup_{\tau} \hat{\zeta}(\tau) = T \frac{\left( \boldsymbol{x}'_{T+1} \hat{\boldsymbol{\beta}}_1(\tau) - \boldsymbol{x}'_{T+1} \hat{\boldsymbol{\beta}}_2(\tau) \right)^2}{\boldsymbol{x}'_{T+1} \left( \frac{\hat{\boldsymbol{V}}_1}{\tau} + \frac{\hat{\boldsymbol{V}}_2}{1-\tau} \right) \boldsymbol{x}_{T+1}}$$

# CASE II: UNKNOWN BREAK DATE

To account for uncertainty in the break date, consider 'local' breaks of $O(T^{-1/2})$.

$$\sup_{\tau} \hat{\zeta}(\tau) = T \frac{\left( \boldsymbol{x}'_{T+1} \hat{\boldsymbol{\beta}}_1(\tau) - \boldsymbol{x}'_{T+1} \hat{\boldsymbol{\beta}}_2(\tau) \right)^2}{\boldsymbol{x}'_{T+1} \left( \frac{\hat{\boldsymbol{V}}_1}{\tau} + \frac{\hat{\boldsymbol{V}}_2}{1-\tau} \right) \boldsymbol{x}_{T+1}}$$

$\hat{\zeta}(\tau)$ converges to a Gaussian process indexed by $\tau$

# Case II: Unknown Break Date

To account for uncertainty in the break date, consider 'local' breaks of $O(T^{-1/2})$.

$$\sup_{\tau} \hat{\zeta}(\tau) = T \frac{\left( \boldsymbol{x}'_{T+1} \hat{\boldsymbol{\beta}}_1(\tau) - \boldsymbol{x}'_{T+1} \hat{\boldsymbol{\beta}}_2(\tau) \right)^2}{\boldsymbol{x}'_{T+1} \left( \frac{\hat{\boldsymbol{V}}_1}{\tau} + \frac{\hat{\boldsymbol{V}}_2}{1-\tau} \right) \boldsymbol{x}_{T+1}}$$

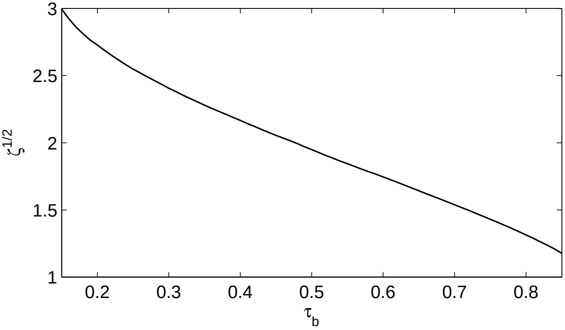
$\hat{\zeta}(\tau)$ converges to a Gaussian process indexed by $\tau$

For each break date $\tau_b$, there is a unique break size $\zeta(\tau_b)$ such that

$$\mathrm{E}\left[ MSFE_{Asy}(\hat{\boldsymbol{\beta}}_F) \right] = \mathrm{E}\left[ MSFE_{Asy}(\hat{\boldsymbol{\beta}}_2(\hat{\tau})) \right], \quad \hat{\tau} = \arg\max_{\tau} \hat{\zeta}(\tau)$$

When the break date is known $\zeta(\tau) = 1$.

# EQUAL MSFE UNDER LOCAL BREAKS OF UNKNOWN TIMING



# Weak Optimality

The test statistic, and therefore critical values, depend on the unknown break date

What if we substitute $\hat{\tau} \to \tau$?

# WEAK OPTIMALITY

The test statistic, and therefore critical values, depend on the unknown break date

What if we substitute $\hat{\tau} \to \tau$?

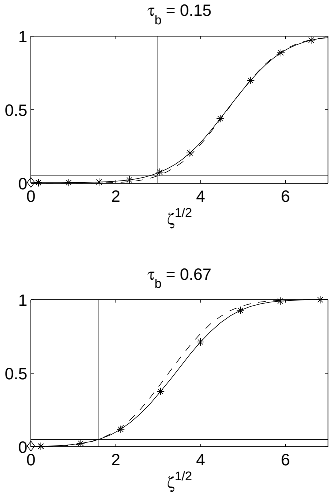
We show that the test is weakly optimal:

In the limit where the nominal size of the test $\alpha \to 0$

$$P_{H_a}\left(\sup_{\tau} \zeta(\tau) > b(\hat{\tau})\right) - P_{H_a}\Big(\zeta(\tau_b) > v(\tau_b)\Big) = 0$$

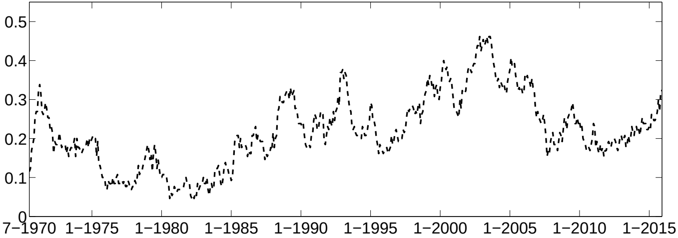
where $\hat{\tau} = \arg\max_{\tau} \zeta(\tau)$ and $P_{H_0}\left(\sup_{\tau} \zeta(\tau) > b(\tau_b)\right) = \alpha$ and $P_{H_0}\left(\zeta(\tau_b) > v(\tau_b)\right) = \alpha$

# ASYMPTOTIC POWER ($\alpha = 0.05$)



# Back to the data

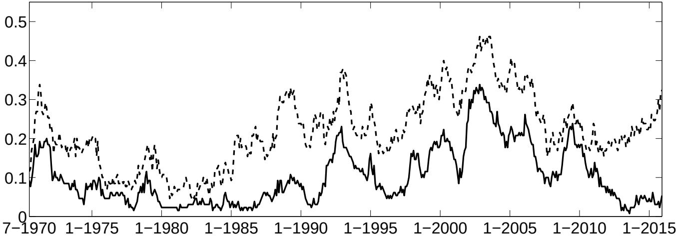
- 130 macroeconomic/financial time series between 1960:1 - 2015:10[1]
- Estimate AR(1) model on a moving window of 120 observations
- $y_t = \mu_1 I[t < T_b] + \mu_2 I[t \geq T_b] + \rho y_{t-1} + \varepsilon_t$
- Forecast period 1970:7 - 2015:10




[1] FRED-MD - https://research.stlouisfed.org/econ/mccracken/fred-databases/

# Back to the data

- 130 macroeconomic/financial time series between 1960:1 - 2015:10[2]
- Estimate AR(1) model on a moving window of 120 observations
- $y_t = \mu_1 I[t < T_b] + \mu_2 I[t \geq T_b] + \rho y_{t-1} + \varepsilon_t$
- Forecast period 1970:7 - 2015:10




[2] FRED-MD - https://research.stlouisfed.org/econ/mccracken/fred-databases/

# Relative MSFE

$$y_t = \mu_1 I[t < T_b] + \mu_2 I[t \geq T_b] + \rho y_{t-1} + \varepsilon_t$$
$$y_t = \mu + \rho y_{t-1} + \varepsilon_t$$

Decide between the post-break forecast and full-sample forecast using (1) test derived here / (2) standard sup-F test

| | Series | Relative MSFE | |
|---|---|---|---|
| | | AR(1) | AR(6) |
| All series | | 0.945 | 0.924 |
| OI | 17 | 0.970 | 0.945 |
| LM | 32 | 0.948 | 0.949 |
| CO | 10 | 0.978 | 0.955 |
| OrdInv | 11 | 0.953 | 0.920 |
| MC | 14 | 0.965 | 0.951 |
| IRER | 21 | 0.871 | 0.847 |
| P | 21 | 0.970 | 0.824 |
| S | 4 | 0.911 | 0.954 |

Excluding forecasts where both tests do not indicate a break

# CONCLUSIONS

- Existing structural break tests are inappropriate when forecasting
- We develop a nearly optimal test to find breaks that are important when forecasting
- The test has good finite sample performance
- Far fewer breaks that are important for forecasting
- **Shrinkage** estimators can be treated along the same lines: Optimal weights of Pesaran, Pick, and Pranovich (2013) can be writting in terms of our forecast Wald test statistic