Comments on : "Forecasting UK GDP growth, inflation and interest rates under structural change: A comparison of models with time-varying parameters" by A. Barnett, H. Mumtaz and K. Theodoridis

Laurent Ferrara
Banque de France - International Macro Division

ECB Workshop *New directions for forecasting* / Frankfurt / 4-5 May 2012

Introduction

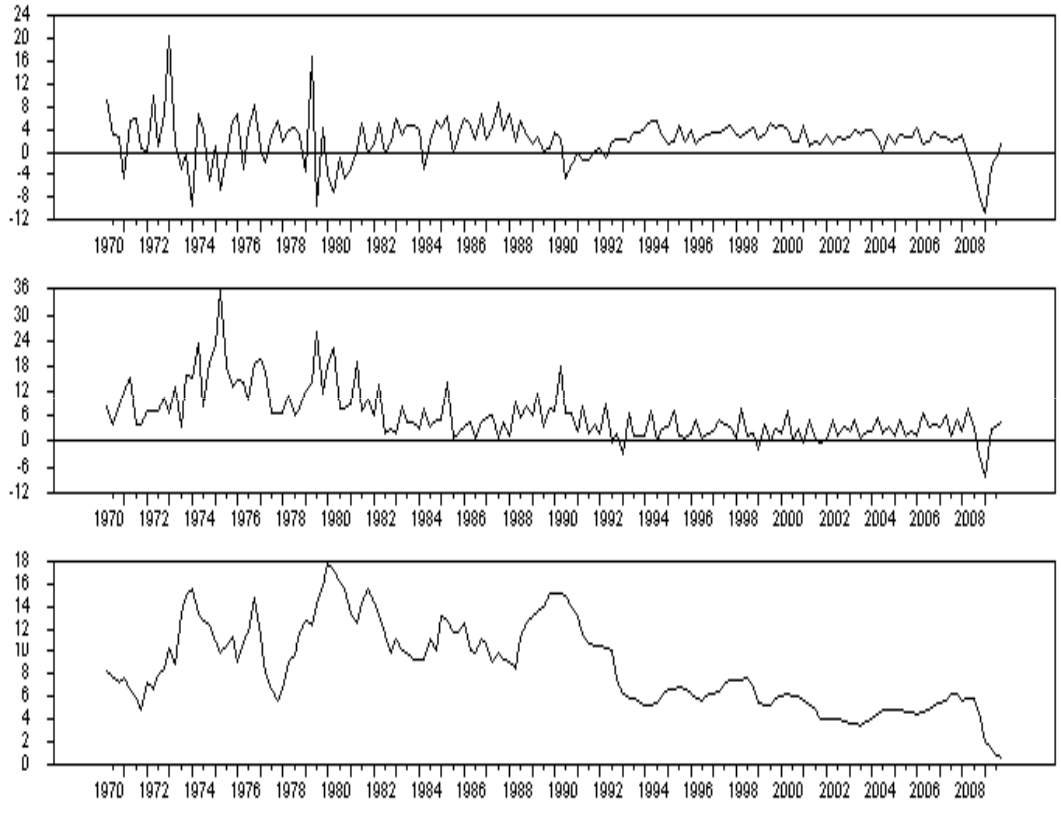
A forecasting competition of TV multivariate models to forecast UK GDP/Inflation/Interest rates.

Results:

1. Overall evidence of gains in RMSFE by introducing multivariate TV models by comparison with a AR(1) process
2. Uneasy to put forward a model that systematically outperforms, although adding information from larger data sets appears generally useful (FAVAR)
3. Interest rates are more difficult to predict than GDP and inflation, GDP is more difficult to predict than inflation

The main idea is to tackle the issue of **forecasting under structural changes** starting from a bunch of multivariate NL models.

Data 1970-2009



Discussion based on 3 points:

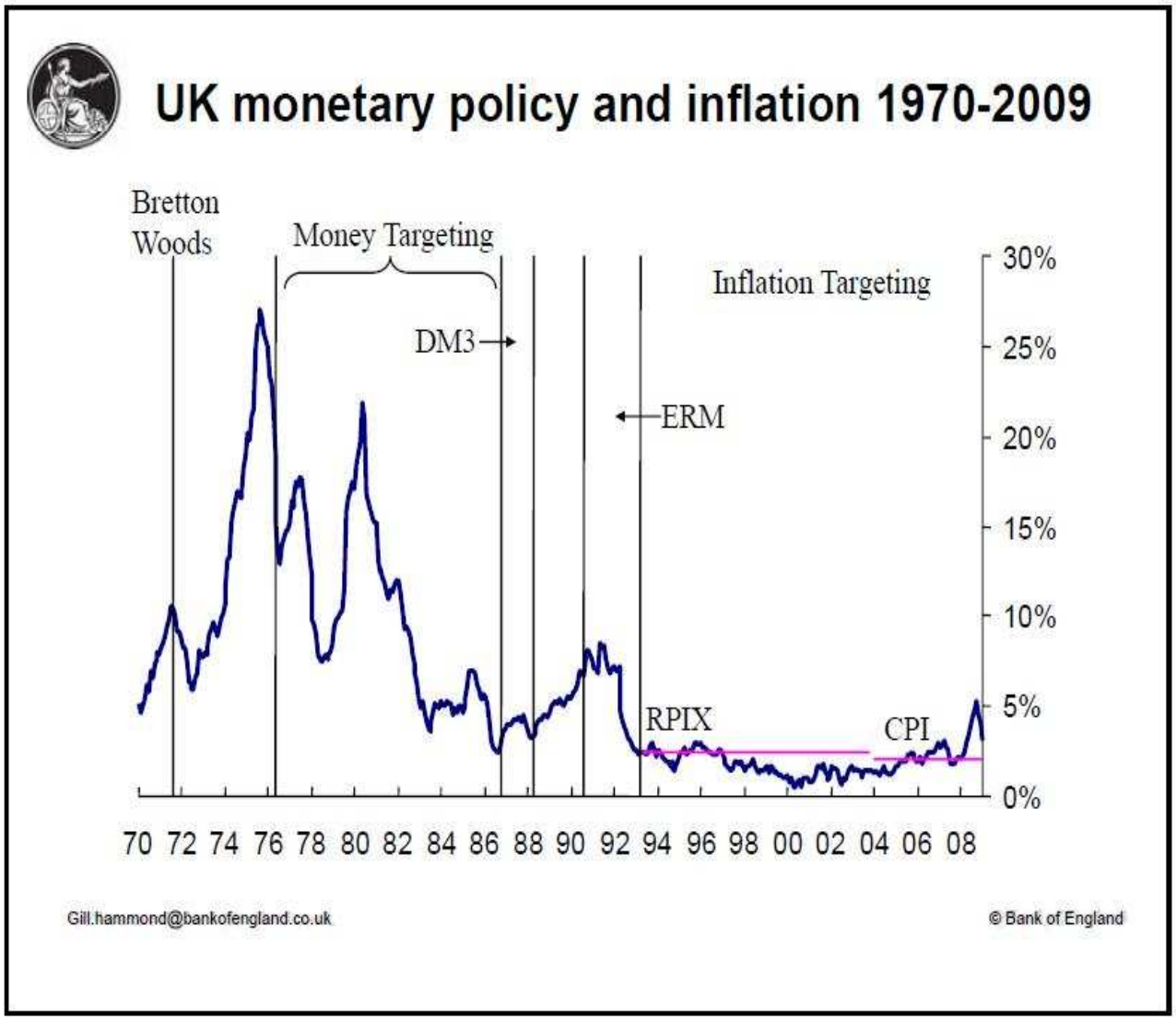
1. Some general comments of the forecasting competition
2. Linear vs Non-linear
3. Multivariate vs Univariate

General comments: Forecasting experiment

- Please precise clearly the dates.
  From my understanding: A learning sample from 1955q1 to 1975q4, an out-of-sample period from 1976q1 to 2007q4.
  Then a second out-of-sample period from 1993q1 to 2007q4.
  Why not providing quantitative measures during the 2008-09 recession?
  Rolling or recursive analysis?
- Direct or indirect multistep forecasting? ($h = 1, 4, 8, 12$)
  For $h > 1$ heavy numerical integration is needed when using NL models
  Direct forecasting reduces the misspecifcation bias (Chevillon, JoES 07, Marcellino, Stock and Watson, JoE 06)
- Poor benchmark : AR(1) does not fit the data
  A better linear specification should be implemented ($p > 1$, breaks in mean, in volatility ...)

General comments: Linear with breaks vs Non-linear

- Evidence of breaks in UK GDP : Bec, Bouabdallah, Ferrara (WP BdF, 10) find breaks in mean (1980q1) and in variance (1980q2)
- Running the Van Dijk and Sensier (2004) test leads to evidence of break in UK inflation and interest rate
- A fair comparison: Linear with breaks vs Non-linear. Which one outperforms for forecasting purposes?
- So why starting in 1993? Not related to the *Great Moderation* but to a change in monetary policy rule (introduction to IT with target at 2.5% for RPIX).
  In addition, this excludes all recessions from the sample


UK monetary policy and inflation 1970-2009
Bretton Woods
Money Targeting
Inflation Targeting
DM3
ERM
RPIX
CPI
30%
25%
20%
15%
10%
5%
0%
70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 00 02 04 06 08
Gill.hammond@bankofengland.co.uk
© Bank of England

General comments: Factor-Augmented models

- Very good forecasts, but for inflation only
- Surprising result: relative accuracy strongly increases with h

Table: Ratios of RMSFE for inflation

	1	4	8	12
TVP-FAVAR 76-07	0.89	0.77	0.70	0.68
TVP-FAVAR homo 76-07	0.95	0.78	0.68	0.65
TVP-FAVAR 93-07	0.76	0.55	0.44	0.41
TVP-FAVAR homo 93-07	0.81	0.49	0.38	0.35

- Counter-intuitive result! Any interpretation ?
- To a lesser extent, this phenomena is also visible pour some models, especially for inflation over the period 1992-2007: This reflects the very bad performance of the AR(1) benchmark model. Why not presenting directly RMSFE ?

General comments: Estimation in NL models

- Why put directly the estimation step into a Bayesian framework?
- For MS-VAR models: Evidence of good results using MLE (Krolzig's work). Especially if the model is constrained.
- For Threshold VAR models: LS estimation conditionally to the unknown threshold, using a grid-search procedure.

General comments: DM test

On the use of Diebold-Mariano in the NL framework, Costantini and Kunst (2011 IAS WP) show evidence of small-sample bias in favor of the null of linearity against a TAR alternative. They put forward 2 main arguments against the use of the DM test:

1. Classical testing with implausible null implies a sizeable small-sample bias if favor of this null
2. The original comparison based on RMSFE is a strong model selection on its own grounds, asymptotically to standard information criteria.

Linear vs Non-Linear: Motivation

- Overall, from the empirical literature: mixed evidence for out-of-sample NL forecasting, while in-sample NL analysis are fruitful (e.g.: business cycle analysis). See on this point Graham Elliot' speech tomorrow.
- Results depend on several factors:
  - estimation and forecasting periods
  - type of model implemented
  - macroeconomic variables considered
  - forecasting horizons
- Gains are not necessarily on the first moments.

Linear vs Non-Linear: Motivation

- In this paper, motivation on the use of NL models : NL models are used with the hope to account for breaks, rather than for changes in regimes (see MS-VAR with $p_{KK} = 1$).
- The number of series is small (3)
  +
  the number of models is small (6)
  =
  A careful investigation has to be done according to the usual sequence
  Test/Specification/Estimation/Validation/Forecasting
  (see Granger, Tjostheim and Teräsvirta, 2010)

Linear vs Non-Linear: NL Tests

Example taken from Bec, Bouabdallah and Ferrara (BdF WP, 10) for UK and US GDPs.
UK data from ONS over 1955Q1–2010Q1.
We implement the test of Carrasco, Hu and Ploberger (WP, 09)

Table: Carrasco, Hu and Ploberger linearity test

Series	T	SupTS	1% c.v.	5% c.v.	10% c.v.
US	252	11.96	4.75	3.51	2.96
UK	220	15.99	4.46	3.53	3.00

The null hypothesis of parameters stability is strongly rejected against a Markov-switching alternative

Linear vs Non-Linear: NL Test

Issues in testing for a bunch of NL models : no single test of a general NL pattern against the null of linearity.

Possibility: a general test put forward by Hamilton (2001, Econometrica) extended by Dahl and Gonzalez-Rivera (2003, JoE).

Lambda - Hamilton	GDPV	CPI	IRS
AR(1,1)	0.05	0.86	0.11
AR(2,2)	0.03	0.01	0.05
AR(3,3)	0.17	0.00	0.11
AR(4,4)	0.10	0.01	0.22
AR(5,5)	0.12	0.03	0.05
AR(6,6)	0.04	0.02	0.06
Lambda - DGR	**GDPV**	**CPI**	**IRS**
AR(1,1)	0.00	0.92	0.06
AR(2,2)	0.00	0.03	0.06
AR(3,3)	0.01	0.06	0.04
AR(4,4)	0.00	0.01	0.06
AR(5,5)	0.00	0.01	0.02
AR(6,6)	0.00	0.02	0.01
G - DGR	**GDPV**	**CPI**	**IRS**
AR(1,1)	0.00	0.72	0.97
AR(2,2)	0.01	0.00	0.13
AR(3,3)	0.01	0.00	0.03
AR(4,4)	0.01	0.01	0.06
AR(5,5)	0.01	0.01	0.00
AR(6,6)	0.02	0.00	0.00

Linear vs Non-Linear: Specification of TV-VAR

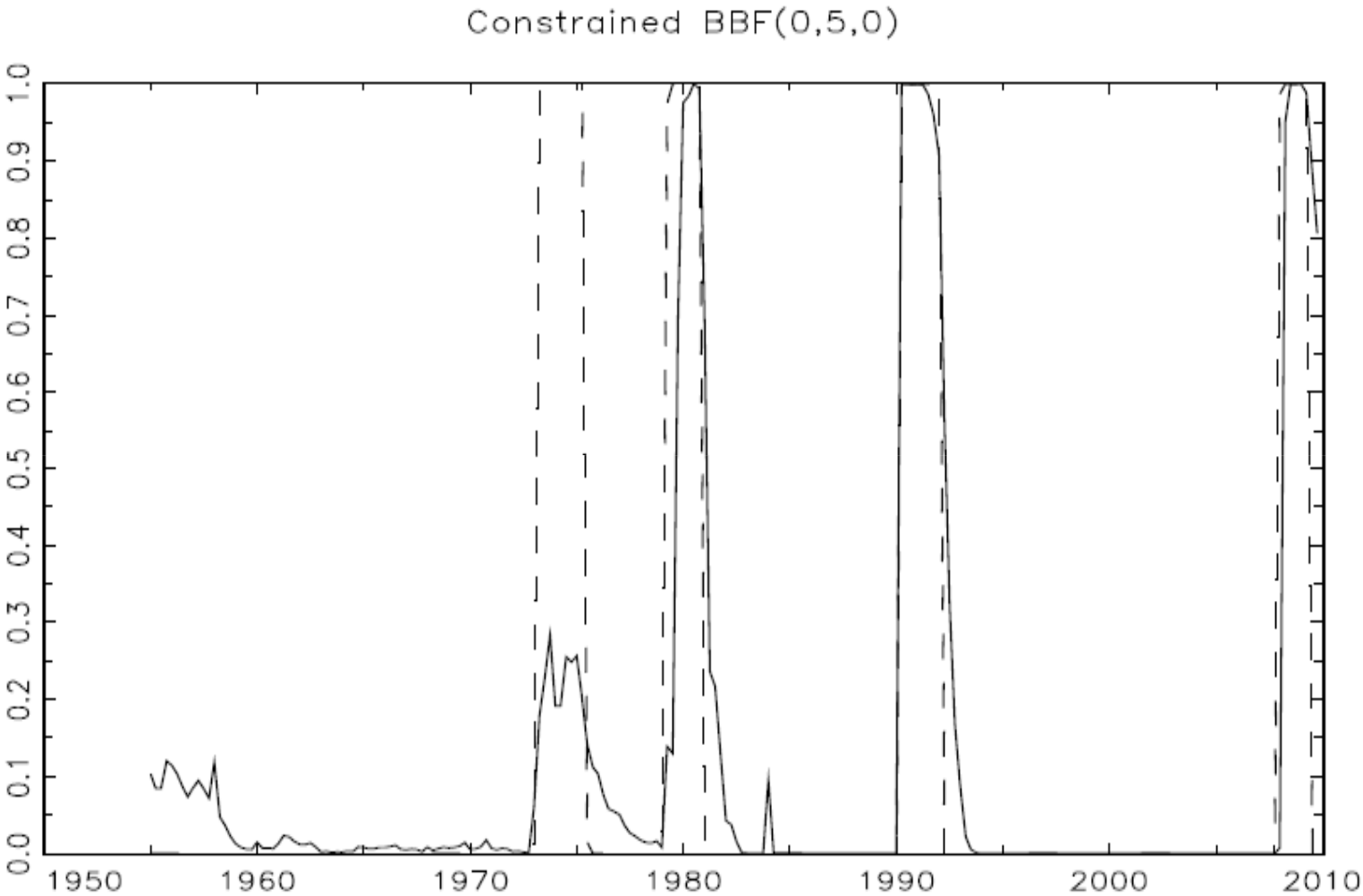
Several parameters have to be estimated for the TV-VAR specification leading to estimation issues:

$$Z_t = c_t + \sum_{j=1}^{p} B_t Z_{t-j} + \Omega_t^{1/2} \varepsilon_t \tag{1}$$

- If the aim is to account for structural changes: why not only allow c_t and keep B fixed?
- What are the results on B_t ?
- Is there really a change in its own dynamics and relationships between the 3 variables ?

Univ. vs Multi.: Where are the gains?

- In-sample analysis:
  Are the models able to reproduce the business cycle
  or
  to account for structural breaks ?
- Out-of-sample analysis:
  Is there really a gain in forecasting when going from univariate to multivariate?


Constrained BBF(0,5,0)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1950
1960
1970
1980
1990
2000
2010

Univ. vs Multi.: Results from Ferrara, Marcellino and Mogliani (BdF WP, forthcoming)

- Would non-linear univariate autoregressive models have been useful to predict the strong downturn during the 2008-2009 recession?
- **Large scale study**: broad international comparison (19 countries) for major macroeconomic and financial variables (23 variables) using a bunch of NL models.
- **Our main findings**:
  - On average, moderate evidence predictive ability of NL models (average RMSFE ratio is close but higher than 1), **and**
  - Absence of a systematic forecast gain for non-linear models (NL models provide a lower RMSE in around 40% of cases), **but**
  - Interesting predictive patterns for some models (TV-AR), countries (Japan, Mexico) and macroeconomic series (long and short interest rates, prices series).

Univ. vs Multi.: Results from Ferrara, Marcellino and Mogliani (BdF WP, forthcoming)

Ratios of RMSE for $h = 1$ over the period 2004-2009:

- ratio $= 0.91$ for GDP using TV-AR
- ratio $= 0.83$ for Inflation using ESTAR
- ratio $= 0.93$ for IR using LSTAR/SETAR

Though not directly comparable (different periods of estimation and evaluation, different benchmarks), univariate results appear competitive with the multivariate ones. Not really convinced at this stage by the multivariate approach for forecasting: an interesting topic for further research!