Economics Bulletin

Volume 36, Issue 2

It is not structural breaks that earn average forecasts their fame

Dirk Ulbricht
institut für finanzdienstleistungen (iff)

Abstract

Structural change is a major challenge to the applied forecaster and a potential source of large forecast errors. Large forecasting competitions demonstrate the success of combined forecasts of simple linear models over forecasting devices that endogenously model structural change. Thereby, most studies look at the average performance over time, not at or around structural changes. However, is it really reliability in the presence of structural breaks that gives average forecasts an edge over their competitors? An analysis of real-time forecasts of UK inflation indicates that it is not their break performance that earns combined forecasts their fame.

The author likes to thank Kai Carstensen, Konstantin A. Kholodilin, an anonymous referee and the participants of the IWH-CIREQ Macroeconometric Workshop 2014 for their valuable comments.
Citation: Dirk Ulbricht, (2016) "It is not structural breaks that earn average forecasts their fame", *Economics Bulletin*, Volume 36, Issue 2, pages 1250-1259
Contact: Dirk Ulbricht - dirk.ulbricht@gmail.com.
Submitted: November 04, 2014. **Published:** June 22, 2016.

1 Introduction

Among others, Stock and Watson [1996a] show that most macroeconomic time series are subject to structural breaks. This has strong implications for the forecasting profession, as forecasting devices should incorporate past and anticipate future breaks.

However, the identification and locating of structural change is a highly demanding task, even if a sufficiently large number of observations is available in sample. This situation is aggrevated when it comes to real-time out-of-sample forecasting. If a structural break occurs at or very close to the forecast origin, the tests for structural breaks necessarily rely on a very limited information set. Furthermore, first vintages of data are frequently subject to considerable revisions, limiting the reliability of the most recent data. Finally, even if a break is correctly identified, it remains unclear how to incorporate this knowledge in the model. Therefore it is not surprising that while it is tempting to try to model breaks endogenously, thus implicitly predicting them as well, sucess has so far been limited, at best.

Considering structural breaks, it is inevitable to have a notion of a structural relationship that describes the data. Depending on the aim of the analysis and the variable to be described, there is a large variety of alternatives, ranging from large structural models, dynamic stochastic equilibrium models or single equations, such as the Phillips Curve, to simple dynamic processes. However, in the context of forecasting macroeconomic time series the latter, capturing an omnipresent feature of most series, their persistence, is the starting point of most approaches. Hence, it is the structure that is considered in the following.

In large forecast competitions, simple combinations of linear models not explicitly accounting for structural changes frequently turn out to be among the best forecasting devices, see e.g. Stock and Watson [2004]. Among others, the main argument put forward to explain their success is their reliability in the presence of structural changes of various types. Timmermann [2005] presents an overview of theoretical explanations for the success of combined forecasts.

However, while most studies considering combinations of linear models test for the presence of structural changes, very few actually locate and test for their effect on the accuracy of the forecasting devices. This paper tries to fill this void. In a real-time experiment, monthly UK inflation is iteratively predicted using combined forecasts of direct and indirect multivariate ordinary least squares estimate over a period of 17 years. Inflation is an attractive target variable, as it is not subject to revisions over the time period under consideration. Thus, the evaluation of forecast accuracy is not influenced by data revisions. Using the methodology of Bai and Perron [2003], one break is detected within this period, which can be associated with the Bank of England adopting an inflation targeting framework. The performance of the average forecasts at and after the break is contrasted with the individual linear and a benchmark non-linear forecasting device. While ranking second over the whole period, its performance is mediocre at and after the break. This indicates that it is not their performance at breaks that earn average forecasts their fame.

The following section presents the real-time data used in the experiment. Section 3 outlines the models used and the empirical approach. Section 4 presents the estimates of structural breaks in UK inflation. Section 5 compares the performance of the models over the whole period and at the break. The last section concludes.

2 The data

The data comprise of 16 monthly financial and economic time series. With the aim of being as close to the real forecasting exercise as possible, each iteration only makes use of the data as available at the time of forecast origin. This means that publication lags are explicitly considered even for the endogenous; the inflation rate is published in the month following the one that is reported, which implies that every forecast has to bridge at least one additional month. Furthermore, part of the variables is frequently revised, sometimes even years after the first publication date. The data used here are those printed in the Economic Trends and Financial Statistics of the Office of National Statistics (ONS, formally the Central Statistical Office) at the time the forecasts are made. These data are collected and described by Egginton et al. [2002]. The vintages go back as far as January 1980 and ends in June 1999. The set of real-time variables comprise industrial production (IP), the monetary aggregate M0, the retail sales index (RS), the total claimant count (U) as a measure of unemployment, and finally average earnings (AE). The variables are transformed to year on year percentage changes. Table 7 in the appendix gives an impression of the extent of the total revisions the data have been subject to up to today. It gives the mean of the absolute values of the revision, the extrema of these values and the ratio of the mean absolute revision to the mean absolute change of the respective variable. Especially in the case of IP the revisions are considerable.

The inflation rate (INFL), defined as the annual percentage change of the Retail Price Index (RPIX), is never revised. Over the period covered in this paper, it is the official measure of inflation and, after the Bank of England (BoE) adopted inflation targeting in 1993, the RPIX-inflation was the target. Furthermore, the data comprise financial and survey data. The exchange rate of the British Pound to the US Dollar (USD), the British Pound to the Deutsche Mark (DEM), the Treasury Bill rate of the United States of America deflated with the US consumer price inflation USINFL, (TBUS), and the price of Brent oil (OIL) are included as indicators of the comparative financial and trading situation of the British economy; the UK Treasury Bill rate deflated with INFL, (TB), the yields of the 10-year UK Government benchmark bonds (BD), and the monthly average of the Financial Times Stock Exchange Index (FTSE) reflect the domestic financial markets; Furthermore, three of the leading qualitative indicators are included: the Industrial Trends Survey of UK manufacturing of the Confederation of British Industry (CBI), the Business Climate Indicator (BS) and the Economic Sentiment Indicators (CS) of the Directorate General for Economic and Financial Affairs (DG ECFIN). A list of the variables, the data transformations, and some descriptive statistics are provided in Table 6 and Table 7 in the appendix.

3 The models and the experiment

The empirical experiment iteratively estimates models of the inflation rate on a monthly basis and makes forecast, where the forecast origins start in January 1982:12 and end in 1999:6 exclusively using data that would have been available to the forecaster at the time of forecast. In turn, 1, 3, 6 and 12-month-horizons forecasts are made (see Table 1).

The models analysed comprise direct and indirect linear models and – as a benchmark – a non-linear model. The indirect linear model is the VAR(p) defined as

Table 1: The forecast experiment

Information set	Forecast ahead			
	h = 1	h = 3	h = 6	h = 12
1948:6 ⟶ 1982:12	1983:1	1983:3	1983:6	1983:12
1948:6 ⟶ 1983:1	1983:2	1983:4	1983:7	1984:1
1948:6 ⟶ 1983:2	1983:3	1983:5	1983:8	1984:2
1948:6 ⟶ 1983:3	1983:4	1983:6	1983:9	1984:3
...				
1948:6 ⟶ 1999:6	1999:7	1999:9	1999:12	2000:6

h = forecast horizon

$$Y_t = \mu + A_1 Y_{t-1} + ... + A_p Y_{t-p} + \varepsilon_t \tag{1}$$

$Y_t = (y_{1t}, ..., y_{nt})$ denotes an $(n \times 1)$ vector of variables, A_p the $n \times n$ matrix of regression coefficients, p is the number of lags to be considered, μ is a vector of constants, and ε_t is the $n \times 1$ unobservable zero mean white noise vector process (serially uncorrelated or independent) with time invariant covariance matrix Σ. It is indirect, as, for $h \geq 2$, the predictions are made in a so called "plug-in" approach, i.e. recursively inserting the forecasts for period $t+1$ to obtain the forecast for $t+2$. First, the 1-step forecast based on information available at time t is composed in a direct way inserting data up to time t into the estimated model:

$$Y_{t+1|t} = \hat{\mu} + \hat{A}_1 Y_t + ... + \hat{A}_p Y_{t-p+1} \tag{2}$$

h+1-step forecasts in contrast are obtained using the chain-rule as

$$Y_{t+h|t} = \hat{\mu} + \hat{A}_1 Y_{t+h-1|t} + ... + \hat{A}_p Y_{t+h-p|t} \tag{3}$$

where $Y_{t+j|t} = Y_{t+j}$ for $j \leq 0$. In the experiment, the number p of the lags to be used in the single models is iteratively optimized using the Bayesian information criterion (BIC) allowing for a maximum of 12 lags.

Along the lines of Marcellino et al. [2006] the direct linear model is the simple OLS. Its minimum lag is set such that $p_{min} = h$, so that no chain-rule is required to compose the forecasts. The endogenous reduces to the target variable y_t^*, i.e. the inflation rate, resulting in estimating the model $y_{t+h}^* = \mu + bX_t + \varepsilon_t$, where $X_t = (y_{1t}, x_{1t}, ..., x_{nt})$ is a vector of regressors including exogenous variables and the endogenous. The forecasts are composed as $y_{t+h|t}^* = \hat{\mu} + \hat{b}X_t$. This implies, that for each forecast horizon and each recursion a new model has to be estimated. Here again, the number of additional lags is optimized using BIC.

For both the direct and the indirect approach, various variable combinations plus inflation are used as regressors. Inflation is included in each model. However, as the implementation of *all* possible combinations would result in the estimation of $2^{16} = 65536$ models for each of the 199 iterations, the maximum number of variables per model is restricted to 6, giving 1941 models for each of the two approaches each recursion.

A common method of dealing with structural breaks is to use only a fixed number of past observations of the data. The rationale is to facilitate the adaption of the model parameters to the new (post-break) situation giving more recent observations more weight. However, this obviously comes at the cost of artificially increasing estimation uncertainty. This rolling window approach is contrasted with the recursive window approach that uses all observations available up to the forecast origin. In the following, only the results for the rolling window including 50 observations are presented. However, the main findings remain unchanged for alternative window sizes of 40 and 60 observations.

As a benchmark, a model that explicitly accounts for coefficient changes, the time varying parameter (TVP) model is used. As in Stock and Watson [1996b], we will consider a random walk coefficient time-varying parameter model of the form

$$y_t = \beta_t y_{t-1} + e_t \tag{4}$$

$$\beta_t = \beta_{t-1} + \eta_t \tag{5}$$

where $e_t \sim i.i.d.N(0, \sigma_e^2)$ and $\eta_t \sim i.i.d.N(0, \sigma_\eta^2)$ and $E(\eta_t e_t) = 0 \forall t$ and where the coefficient β_t is allowed to change over time. We estimate the model by using the Kalman filter and setting diffuse priors for the hyperparameters.

4 The break point analysis

Following the approach of Hansen [2001] and Bai and Perron [2003], the analysis presented here focuses on the breaks in the structure of a linear AR(p)-model. It represents a simple and, in most cases, quite close description of the data. Furthermore, it has a proven track record as a forecasting model, and is most often hard to beat benchmark model in forecast comparisons (see e.g. Stock and Watson [2004]). BIC selects a lag length of six over the whole sample. Bai and Perron [1998] propose to first test for the presence of one structural break using the WDmax$F_T(M,q)$ and the UDmax$F_T(M,q)$ test statistics. If the null hypothesis of no structural break can be rejected, the number of breaks is determined employing the sequential $\sup F_T(l+1|l)$. Table 2 gives the empirical results for the UK inflation rate.

Table 2: Test results for the presence and number of breaks

$\sup F_T(1)$	UDmax	WDmax	$\sup F_T(2\|1)$	$\sup F_T(3\|2)$	$\sup F_T(4\|3)$
33.10	33.1	42.12	40.41	41.63	16.63
(22.62)	(22.80)	(24.34)	(22.62)	(24.64)	(26.54)

Asymptotic critical values at the 5 % significance level are given in parenthesis

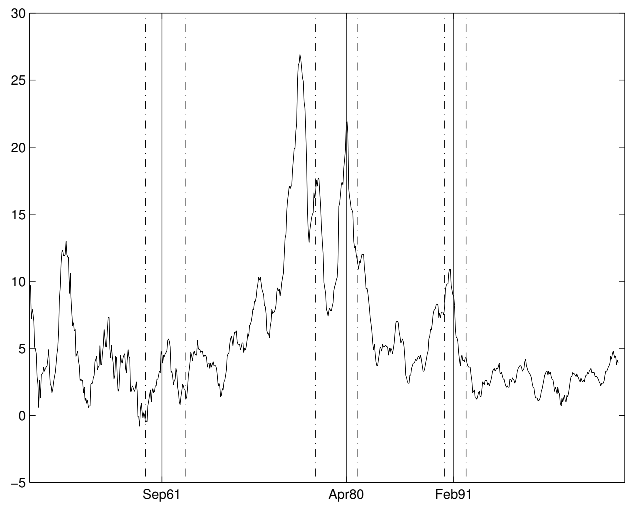
The $\sup F_T(1)$, the UDmax and the WDmax reject the null hypothesis at the 5 percent level, i.e., at least one structural break is present. The $\sup F_T(l+1|l)$ rejects up to $l = 3$, that is the estimated number of breaks is $m = 3$. The estimated break points employing the sequential procedure to find a global minimum are presented in Table 3.

Table 3: Breakdates and confidence intervalls

Estimators	$\hat{T}_1$	$\hat{T}_2$	$\hat{T}_3$
Break dates	1961:9	1980:4	1991:2
(95 % confidence intervall)	(1960:10; 1963:6)	(3/31/1977; 6/30/1981)	(1990:3; 1992:5)

The 5 percent confidence intervall dates are given in parenthesis. As the estimator allows for different variances across segments, the intervalls are not symmetric. The confidence band for the break point that lies in the period for which real-time data is available, the one at February 1991, is thereby especially tight spanning 26 month. A plot of the inflation rate, the breakpoint estimates, and 95 percent confidence intervalls is given in Figure 1.

Figure 1: UK inflation, breaks and 95 percent confidence bands



The sum of the estimated autoregressive coefficients and constants for the periods between the break dates are given in Table 4. The estimated structural changes and the development of the persistence correspond to the results found in the literature (see e.g. Benati et al., 2003). The last break in the early nineties is frequently associated with the Bank of England adopting an inflation targeting framework.

Table 4: Sum of estimated AR coefficients and constants between the breaks

Segment	I	II	III	IV
	(1948:6–1961:8)	(1961:9–1980:3)	(1980:4–1991:1)	(1991:1–2007:10)
Σ of AR coefs.	0.93	0.99	0.97	0.92
constant	0.29	0.13	0.20	0.24

The persistence rises sharply in the second period to approximately one and declines to 0.92 in the last segment after a moderate decline in the third segment. The estimated constant drops from 0.29 to 0.13 in the second segment and rises to a value of 0.24 in the last segment.

5 Results

Table 5 presents the RMSE for the average performance at and during the 11 month after the estimated break date. It shows of the naive combination scheme, averaging over all 7764 individual models of the direct and indirect OLS approach, using the rolling and the recursive estimation scheme. The ranking given in parenthesis is the outcome of the comparison of the respective model with all individual models, the TVP and the naive forecast[1].

For the 1-month-horizon forecasts the naive combination scheme ranks merely 285^{th}. For the three, six and 12-months horizon forecasts, it performs even worse, ranking 2212^{th}, 2672^{th}, and 3978^{th}, respectively. The performance of the TVP as well as the naive forecasting scheme is very poor for all horizons.

Table 5: RMSE and ranking at and 11 month after the break date considered

Horizon	1	3	6	12
NC^0	0.81	2.40	8.24	18.69
	(285)	(2,212)	(2,672)	(3,978)
TVP	1.53	4.56	12.87	21.61
	(3,888)	(5,010)	(4,572)	(4,476)
NA	1.51	4.46	12.77	22.44
	(3,784)	(4,939)	(4,549)	(4,558)

NC^0 = naive combination over all models, TVP = time varying parameter model, and the NA = naive foreast;

Table 9 gives the RMSE and the ranking of the models including the AR(p) and the combined forecasts using the direct and indirect, as well as the rolling and recursive approaches, seperately. The main result holds.

6 Conclusion

Following intuition and theoretical results - averaging over different forecasters should represent a hedge against structural change. However, in the case of a violent break in the UK inflation rate, this expectation is in sharp contrast with our findings. Using the methodology of Bai and Perron [2003], a structural break in the persistence of UK inflation is identified within the time window of the real-time experiment. A huge number of linear pseudo-out-of-sample forecasts is generated each recursion. While still a good forecasting model in the class of linear forecasting instruments, the naive combination of these forecasts underperforms at and shortly

[1]The naive forecast uses the last value observed as the forecast

after the structural break relative to its forecasting success over the whole window. Both its performance in terms of the RMSE as well as its ranking in comparison with the individual linear models employed deteriorate considerably. However, though simple averages have largely proven to be a good choice for many forecasting excercises, they might not be the best option. Recently, Tian and Anderson [2014] demonstrate that weighting schemes where average forecasts are based on different estimation windows are very useful to account for structural breaks. Furthermore, Barnett et al. [2014] show that some nonlinear models exhibit considerable performance to forecast UK data. Still, the results of this paper indicate that breaks in the persistence of the forecasted variable merit a closer look for the optimization of averages of linear forecasts.

References

J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998.

J. Bai and P. Perron. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22, 2003.

Alina Barnett, Haroon Mumtaz, and Konstantinos Theodoridis. Forecasting uk gdp growth and inflation under structural change. a comparison of models with time-varying parameters. *International Journal of Forecasting*, 30(1):129–143, 2014.

Luca Benati, George Kapetanios, et al. Structural breaks in inflation dynamics. In *Computing in Economics and Finance*, volume 169, pages 563–587, 2003.

D.M. Egginton, A. Pick, and S.P. Vahey. Keep it real!: a real-time UK macro data set. *Economics Letters*, 77(1):15–20, 2002.

B.E. Hansen. The New Econometrics of Structural Change: Dating Breaks in US Labor Productivity. *The Journal of Economic Perspectives*, 15(4):117–128, 2001.

M. Marcellino, J.H. Stock, and M.W. Watson. A Comparison of Direct and Iterated Multistep AR Methods for ForecastingMacroeconomic Time Series. *Journal of Econometrics*, 135 (1-2):499–526, 2006.

J.H. Stock and M. Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23:405–430, 2004.

J.H. Stock and M.W. Watson. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14(1):11–30, 1996a.

J.H. Stock and M.W. Watson. Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business & Economic Statistics*, 14(1):11–30, 1996b.

Jing Tian and Heather M Anderson. Forecast combinations under structural break uncertainty. *International Journal of Forecasting*, 30(1):161–175, 2014.

A. Timmermann. Forecast combinations. In Granger C. W. J. Elliott, G. and A. Timmermann, editors, *Handbook of Economic Forecasting*. North Holland, 2005.

Table 6: List of variables and transformations

Variable	Acronym	Transformation
Government bonds	BD	ln(t-t(-12))*100
Financial Times Stock Exchange index	FTSE	ln(t-t(-12))*100
US Dollar exchange rate	USD	ln(t-t(-12))*100
Deutschmark exchange rate	DEM	ln(t-t(-12))*100
Treasury Bills	TB	Level
Confederation of British Industry indicator	CBI	ln(t-t(-12))*100
Oil price	OIL	ln(t-t(-12))*100
Real US treasury bill rate	TBUS	Level
Economic sentiment indicator	CS	Level
Business climate indicator	BS	Level
Inflation rate	INFL	ln(t-t(-12))*100
Retail sales index	RS	ln(t-t(-12))*100
Industrial production	IP	ln(t-t(-12))*100
Average earnings	AE	ln(t-t(-12))*100
Monetary aggregate 0	M0	ln(t-t(-12))*100
Unemployment rate	U	ln(t-t(-12))*100

Table 8: The extent of the total revision of the real-time data

Series	Mean absolute revision	Minimum revision	Maximum revision	Mean absolute change	Ratio
Industrial production	1.17	-2.78	4.65	3.21	0.37
Average earnings	0.28	-1.03	1.46	7.50	0.04
Retail sales	0.66	-2.54	2.36	3.13	0.21
Unemployment	1.48	-4.34	5.27	16.20	0.09
Money	0.29	-0.94	3.34	4.80	0.06

Table 7: Descriptive statistics

	BD	FTSE	USD	DEM	TB	CBI	OIL	TBUS
Mean	-0.17	7.98	0.66	-2.16	1.48	17.82	23.41	1.29
Median	0.12	10.34	0.08	-0.58	2.55	16	19.06	1.47
Max	35.84	87.52	32.10	32.25	7.69	78	91.60	6.83
Min	-41.81	-89.05	-29.31	-40.39	-16.67	-30	2.23	-7.31
Std. Dev.	13.20	19.92	9.51	10.51	4.14	24.55	15.50	2.07
# of Obs.	599	536	599	599	525	393	455	609

	CS	BS	INFL	RS	IP	AE	M0	U
Mean	-8.88	-8.17	5.82	2.86	1.11	8.02	6.80	0.95
Median	-7	-6.25	4.2	3.26	1.20	7.26	6.14	-1.93
Max	12	28	26.9	12.72	20.42	29.82	16.04	57.96
Min	-32	-56	-0.8	-11.08	-12.62	2.06	-1.81	-41.41
Std. Dev.	8.92	14.68	4.87	2.94	3.91	4.82	3.26	19.62
# of Obs.	406	382	712	428	465	524	431	428

Table 9: RMSE of combined forecasts and benchmarks on basis of the squared errors at and 11 after the break month respectively at and 23 month after the break

RMSE over				breakdate + 11				breakdate+23
Horizon	1	3	6	12	1	3	6	12
$AR^{1,3}$	1.22	2.87	9.42	17.16	0.80	1.86	5.59	11.95
	(2,004)	(2,931)	(3,336)	(3,637)	(1,575)	(2,866)	(2,925)	(3,172)
$NC^{1,3}$	0.92	1.83	6.74	11.72	0.56	1.11	3.91	7.20
	(627)	(1,353)	(1,851)	(2,071)	(334)	(603)	(1,257)	(1,394)
$NC^{1,4}$	0.62	2.46	11.55	38.53	0.40	1.65	7.21	28.20
	(49)	(2,294)	(4,122)	(6,052)	(7)	(2,186)	(3,863)	(4,627)
$NC^{2,3}$	0.92	2.14	7.18	13.75	0.58	1.28	4.08	8.00
	(603)	(1,803)	(2,072)	(2,919)	(370)	(1,117)	(1,421)	(2,008)
$NC^{2,4}$	0.94	3.72	9.05	17.77	0.59	2.28	6.51	24.44
	(714)	(4,246)	(3,137)	(3,783)	(440)	(3,988)	(3,517)	(4,287)
NC^{0}	0.81	2.40	8.24	18.69	0.50	1.44	4.80	14.00
	(285)	(2,212)	(2,672)	(3,978)	(133)	(1,580)	(2,166)	(3,352)
TVP	1.53	4.56	12.87	21.61	0.94	2.58	7.10	14.84
	(3,888)	(5,010)	(4,572)	(4,476)	(2,950)	(4,447)	(3,795)	(3,401)
NA	1.51	4.46	12.77	22.44	0.92	2.50	6.96	15.05
	(3,784)	(4,939)	(4,549)	(4,558)	(2,752)	(4,361)	(3,728)	(3,418)

Rankings are given in parenthesis; 1=indirect OLS, 2=direct OLS, 3=recursive approach, 4=rolling window approach; 0=all models, that is 1+2+3+4, AR=autoregressive model, NC=naive combination, NA=naive forecast;