Modelling UK Inflation Uncertainty, 1958-2006

Gianna Boero, Jeremy Smith and Kenneth F. Wallis*

Department of Economics
University of Warwick
Coventry CV4 7AL, UK

Robert F. Engle Festschrift Conference, San Diego, 21 June 2008

Abstract Robert Engle's celebrated article that introduced the concept of autoregressive conditional heteroskedasticity (ARCH) included an application to UK inflation, 1958-77. This paper updates the estimation of his model and investigates its stability in the light of the well documented changes in policy towards inflation, 1958-2006. A simple autoregressive model with structural breaks in mean and variance, constant within subperiods (and with no unit roots), provides a preferred representation of the observed heteroskedasticity. Several measures of inflation forecast uncertainty are presented; these illustrate the difficulties presented by instability, not only for point forecasts but also, receiving increased attention nowadays, their uncertainty.

Keywords Inflation; autoregressive conditional heteroskedasticity (ARCH); structural breaks; forecast uncertainty; United Kingdom; monetary policy

*Corresponding author [K.F.Wallis@warwick.ac.uk]

1. INTRODUCTION

Introducing the autoregressive conditional heteroskedastic (ARCH) process in his celebrated article in *Econometrica* in July 1982, Robert Engle observed that the ARCH regression model "has a variety of characteristics which make it attractive for econometric applications" (p.989). He noted in particular that "econometric forecasters have found that their ability to predict the future varies from one period to another," citing the recognition by McNees (1979, p.52) that "the inherent uncertainty or randomness associated with different forecast periods seems to vary widely over time," and McNees's finding that "the 'large' and 'small' errors tend to cluster together" (p.49). McNees had examined the track record of the quarterly macroeconomic forecasts published by five forecasting groups in the United States over the 1970s. He found that, for inflation, the median one-year-ahead forecast persistently underpredicted the annual inflation rate from mid-1972 to mid-1975, with the absolute forecast error exceeding four percentage points for five successive quarters in this period; outside this period forecast errors were more moderate, and changed sign from time to time, though serial correlation remained. Engle's article presented an application of the ARCH regression model to inflation in the United Kingdom over the period 1958-77, which included the inflationary explosion of 1974-75, whose magnitude had likewise been unanticipated by UK forecasters (Wallis, 1989). In both countries this "Great Inflation" is now seen as an exceptional episode, and the transition to the "Great Moderation" has been much studied in recent years. How this has interacted with developments in the analysis of inflation volatility and the treatment of inflation forecast uncertainty is the subject of this paper.

The quarter-century since the publication of ARCH has seen widespread application in macroeconomics of the basic model and its various extensions – GARCH, GARCH-M, EGARCH … – not to mention the proliferation of applications in finance of these and related models under the heading of stochastic volatility, whose precursors predate ARCH (Shephard, 2007). There has also been substantial development in the measurement and reporting of inflation forecast uncertainty (Wallis, 2008). Since 1996 the National Institute of Economic and Social Research (NIESR) and the Bank of England have published not only point forecasts but also density forecasts of UK inflation, the latter in the form of the famous fan chart. Simultaneously in 1996 the Bank initiated its Survey of External Forecasters, analogous to the long-running US Survey of Professional Forecasters; based on the responses it publishes quarterly survey average density forecasts of inflation in its *Inflation Report*.

Finally the last quarter-century has seen substantial development of the econometrics of structural breaks and regime switches, perhaps driven by and certainly relevant to the macroeconomic experience of the period.

These methods have been applied in a range of models to document the decline in persistence and volatility of key macroeconomic aggregates in the United States, where the main break is usually located in the early 1980s. Interpretation has been less straightforward, however, especially with respect to inflation, since "it has proved hard to reach agreement on what monetary regimes were in place in the US and indeed whether there was ever any change at all (except briefly at the start of the 1980s with the experiment in the control of bank reserves)" (Meenagh, Minford, Nowell, Sofat and Srinivasan, 2007). Although the corresponding UK literature is smaller in volume, it has the advantage that the various changes in policy towards inflation are well documented, which Meenagh *et al*. and other authors have been able to exploit. Using models in this way accords with the earlier view of Nerlove (1965), while studying econometric models of the UK economy, that model building, in addition to the traditional purposes of forecasting and policy analysis, can be described as a way of writing economic history. The modelling approach and the traditional approach to economic history each have limitations, but a judicious blend of the two can be beneficial. At the same time there can be tensions between the *ex post* and *ex ante* uses of the model, as discussed below.

The rest of this paper is organised as follows. Section 2 contains a brief review of UK inflationary experience and the associated policy environment(s), 1958-2006, in the light of the literature alluded to in the previous paragraph. Section 3 returns to Engle's original ARCH regression model, and examines its behaviour over the extended period. Section 4 turns to a fuller investigation of the nature of the nonstationarity of inflation, preferring a model with structural breaks, stationary within subperiods. Section 5 considers a range of measures of inflation forecast uncertainty, from these models and other UK sources. Section 6 considers the association between uncertainty and the level of inflation, first mooted in Milton Friedman's Nobel lecture. Section 7 concludes.

2. UK INFLATION AND THE POLICY ENVIRONMENT

Measures of inflation based on the Retail Prices Index (RPI) are plotted in Figure 1, using quarterly data, 1958-2006. We believe that this is the price index used by Engle (1982), although the internationally more standard term, “consumer price index”, is used in his text; in common with most time-series econometricians, he defined inflation as the first difference of the log of the quarterly index. In 1975 mortgage interest payments were introduced into the RPI to represent owner-occupiers’ housing costs, replacing a rental equivalent approach, and a variant index excluding mortgage interest payments (RPIX) also came into use. This became the explicit target of the inflation targeting policy initiated in October 1992, since it removed a component of the all-items RPI that reflected movements in the policy instrument. In December 2003 the official target was changed to the Harmonised Index of Consumer Prices, constructed on principles harmonised across member countries of the European Union and promptly relabelled CPI in the UK, while the all-items RPI continues in use in a range of indexation applications, including index-linked gilts. Neither of these indices, nor their variants, is ever revised after first publication. For policy purposes, and hence also in public discussion and practical forecasting, inflation is defined in terms of the annual percentage increase in the relevant index. We denote the “econometric” and “policy” measures of inflation respectively as $\Delta_1 p_t$ and $\Delta_4 p_t$, where $\Delta_i = 1 - L^i$ with lag operator *L*, and *p* is the log of the quarterly index. The former, annualised (by multiplying by 4), is shown in the upper panel of Figure 1; the latter in the lower panel. It is seen that annual differencing removes the mild seasonality in the quarterly RPI, which is evident in the first-differenced series, and also much reduces short-term volatility.

Episodes of distinctly different inflationary experience are apparent in Figure 1, and their identification in the context of different modelling exercises and their association with different approaches to macroeconomic policy have been studied in the UK literature mentioned above. Haldane and Quah (1999) consider the Phillips curve from the start of the original Phillips sample, 1861, to 1998. For the post-war period, with a specification in terms of price inflation (unlike the original Phillips curve specification in terms of wage inflation), they find distinctly different “curves” pre- and post-1980: at first the curve is “practically vertical; after 1980, the Phillips curve is practically horizontal” (p.266). Benati (2004), however, questions Haldane and Quah’s use of frequency-domain procedures that focus on periodicities between five and eight years, and argues for a more "standard" business-cycle range of six quarters to eight years. With this alternative approach he obtains a further division of each episode, identifying "a period of extreme instability (the 1970s), a period of remarkable stability (the post-1992 period), and two periods 'in-between' (the Bretton Woods era and the period between 1980 and 1992)" (p.711). This division is consistent with his prior univariate analysis of RPI inflation, 1947:1-2003:2, which finds three breaks in the intercept, coefficients and innovation variance of a simple autoregression, with estimated dates 1972:3, 1981:2 and 1992:2 (although the date of the second break is much less precisely determined than the other two dates).

Nelson and Nikolov (2004) and Meenagh *et al.* (2007) consider a wide range of "real-time" policy statements and pronouncements to document the vicissitudes of UK macroeconomic policymaking since the late 1950s. Until 1997, when the Bank of England gained operational independence, monetary policy, like fiscal policy, was in the hands of elected politicians, and their speeches and articles are a rich research resource. This evidence, together with their simulation of an estimated New Keynesian model of aggregate demand and inflation behaviour, leads Nelson and Nikolov to conclude that "monetary policy neglect", namely the failure in the 1960s and 1970s to recognize the primacy of monetary policy in controlling inflation, is important in understanding the inflation of that period. Study of a yet wider range of policymaker statements leads Nelson (2007) to conclude that the current inflation targeting regime is the result not of changed policymaker objectives, but rather of an "overhaul of doctrine", in particular a changed view of the transmission mechanism, with the divide between the "old" and "modern" eras falling in 1979.

Meenagh *et al*. (2007) provide a finer division of policy episodes, identifying five subperiods: the Bretton Woods fixed exchange rate system, up to 1970:4; the incomes policy regime, 1971:1-1978:4; the money targeting regime, 1979:1-1985:4; exchange rate targeting, 1986:1-1992:3; and inflation targeting, since 1992:4. They follow their narrative analysis with statistical tests in a three-variable VAR model, finding general support for the existence of the breaks, although the estimated break dates are all later than those suggested by the narrative analysis. These reflect lags in the effect of policy on inflation and growth outcomes and, when policy regimes change, "there may well be a lag before agents' behaviour changes; this lag will be the longer when the regime change is not clearly communicated or its effects are not clearly understood" (p.6). Meenagh *et al*. suggest that this applies to the last two changes: the switch to exchange rate targeting in 1986, with a period of "shadowing the Deutsche Mark" preceding formal membership of the Exchange Rate Mechanism of the European Monetary System, was deliberately kept unannounced by the Treasury, while in 1992 inflation targeting was unfamiliar, with very little experience in other countries to draw on. Independent evidence on responses to later changes to the detail of the inflation targeting arrangements is presented in Section 5.

None of the research discussed above is cast in the framework of a regime switching model, of which a wide variety is available in the econometric literature. The brief account of five policy episodes in the previous paragraph makes it clear that there was no switching from one regime to another and back again; at each break point the old policy was replaced by something new. Likewise no regime switching models feature in the analysis presented below.

3. REESTIMATING THE ORIGINAL ARCH MODEL

The original ARCH regression model for UK inflation is (Engle, 1982, pp.1001-2)

$$\Delta_1 p_t = \beta_0 + \beta_1 \Delta_1 p_{t-1} + \beta_2 \Delta_1 p_{t-4} + \beta_3 \Delta_1 p_{t-5} + \beta_4 \left(p_{t-1} - w_{t-1} \right) + \varepsilon_t, \tag{1}$$

$$\varepsilon_t \left| \psi_{t-1} \sim N\left(0, h_t\right), \quad h_t = \alpha_0 + \alpha_1 \left(0.4\varepsilon_{t-1}^2 + 0.3\varepsilon_{t-2}^2 + 0.2\varepsilon_{t-3}^2 + 0.1\varepsilon_{t-4}^2 \right) \right. \tag{2}$$

where p is the log of quarterly RPI and ψ_{t-1} is the information set available at time $t-1$. The wage variable used by Engle (in logs) in the real wage "error correction" term, namely an index of manual wage rates, was subsequently discontinued, and for consistency in all our reestimations we use the average earnings index, also used by Haldane and Quah (1999). For the initial sample period, 1958:1-1977:2, we are able to reproduce Engle's qualitative findings, with small differences in the quantitative details due to these minor variations. In particular, with respect to the *h*-process, our maximum likelihood estimate of α_0 is, like his, not significantly different from zero, while our estimate of α_1, at 0.897, is slightly smaller than his (0.955). The turbulence of the period is illustrated in Figure 2, which plots the square root of the estimates of h_t over the sample period: these are the standard errors of one-quarter-ahead forecasts of annual inflation based on the model. The width of an interval forecast with nominal 50% coverage (the interquartile range) varies from a minimum of 2.75 percentage points to a maximum of 14 percentage points of annual inflation. Engle concludes that "this example illustrates the usefulness of the ARCH model … for obtaining more realistic forecast variances", although these were not subject to test in an out-of-sample exercise.

Reestimation over the extended sample period 1958:1-2006:4 produces the results shown in Table 1. These retain the main features of the original model – significant autoregressive coefficients, insignificant α_0, estimated α_1 close to 1 – except for the estimate of the error correction coefficient, β_4, which is virtually zero. Forward recursive estimation shows that this coefficient maintains its significance from the initial sample to samples ending in the mid-1980s, but then loses its significance as more recent observations are added to the sample. Figure 3(a) shows the conditional standard error of annualised inflation over the fully extended period. The revised estimates are seen to extend the peaks in the original sample period shown in Figure 2; there is then a further peak around the 1979-81 recession, after which the conditional standard error calms down.

Practical forecasters familiar with the track record of inflation projections over the past decade may be surprised by forecast standard errors as high as two percentage points of annual inflation shown in Figure 3(a). Their normal practice, however, is to work with an inflation measure defined as the percentage increase in prices on a year earlier, $\Delta_4 p$, whereas $\Delta_1 p$ is used in Engle's model and our various reestimates of it. The latter series exhibits more short-term volatility, as seen in Figure 1. Replacing $\Delta_1 p$ in the original ARCH regression model given above by $\Delta_4 p$ and reestimating over the extended sample gives the conditional standard error series shown in Figure 3(b). This has the same profile as the original specification, but reflects a much lower overall level of uncertainty surrounding the more popular measure of inflation.

Over the last decade the time series plotted in Figures 1 and 3 have a more homoskedastic, rather than heteroskedastic appearance, despite the significance of the estimate of α_1 over the full sample including this period. As a final reestimation exercise on the original ARCH model, with $\Delta_1 p$, we undertake backward recursive estimation. We begin with the sample period 1992:4-2006:4, the inflation targeting period, despite reservations about a learning period having been required before the full benefits of the new policy became apparent. We then consider sample periods starting earlier, one quarter at a time, until the complete sample period 1958:1-2006:4 is reached. Equivalently, we could begin with full sample estimation then sequentially remove the earliest observation. Either way, the resulting estimates of the coefficient α_1 and the p-values of the LM test (Engle, 1982, Section 8) are plotted in Figure 4 against the starting date of the sample; the end date is 2006:4 throughout. There is seen to be a clear change around 1980. To exhibit significant conditional heteroskedasticity it is necessary to include periods earlier than this in the sample; samples starting after 1980 offer no support for the existence of ARCH in this model. Similar results are obtained when the model is rewritten in terms of $\Delta_4 p$, except that the sample has to start in 1990 or later for the significant ARCH effect to have disappeared. These findings prompt more general questions about nonstationarity.

4. THE NONSTATIONARY BEHAVIOUR OF UK INFLATION

We undertake a fuller investigation of the nature of the nonstationarity of inflation, in the light of the coexistence in the literature of conflicting approaches. For example, Garratt, Lee, Pesaran and Shin (2003; 2006, Ch.9) present an eight-equation conditional vector error correction model of the UK economy, estimated over 1965:1-1999:4, in which RPI inflation, $\Delta_1 p$, is treated as an I(1) variable. This leads them to express the target in their monetary policy experiment as a desired constant reduction in the rate of inflation from that observed in the previous period, which does not correspond to the inflation target which is the current focus of policy in the UK, nor anywhere else. In contrast, Castle and Hendry (2007) present error correction equations for inflation (GDP deflator) for use in forecast comparisons, with the same sample starting date as Garratt *et al*., assuming that "the price level is I(1), but subject to structural breaks which give the impression that the series is I(2)" (p.5).

Standard unit root tests without structural breaks reveal some of the sources of potential ambiguity. Tests are performed recursively, beginning with a sample of 40 observations, 1958:1-1967:4, then extending the sample quarter-by-quarter to 2006:4. Results for the augmented Dickey-Fuller test are representative of those obtained across various other tests. For the quarterly inflation series $\Delta_1 p$, the results presented in Figure 5 demonstrate sensitivity to the treatment of seasonality. The upper panel gives the ADF statistic with the inclusion of a constant term, and shows that over the 1970s and 80s the null hypothesis of I(1) inflation would not be rejected. The addition of quarterly dummy variables, however, gives the results shown in the lower panel, which lead to the clear rejection of the unit root hypothesis as soon as the end-point of the sample gets clear of the 1975 peak in inflation, and thereafter. Such constant additive seasonality can alternatively be removed by annual differencing, which also reduces short-term volatility, as noted above in the discussion of Figure 1. For the $\Delta_4 p$ series, in the corresponding figure (not shown) the ADF statistic lies in the unit root non-rejection region over the whole period. Backward recursive estimation of the ADF test for the $\Delta_4 p$ series, however, shows that the unit root hypothesis would be rejected in samples with start dates in 1990 or later. These results represent a simple example of the impact of a deterministic component, and different ways of dealing with it, on inference about unit roots, and the sensitivity of such inference to the choice of sample period.

The impact of structural breaks on inference about unit roots over the full data period is assessed using the procedures of Zivot and Andrews (1992), allowing for an estimated break in mean under the alternative hypothesis. Once this is done, the ADF statistic, relative to Zivot and Andrews's critical values, implies rejection of the unit root hypothesis in all three cases: $\Delta_1 p$, with and without seasonal dummy variables, and $\Delta_4 p$. These results motivate further investigation of structural change, in models that are stationary within subperiods.

We apply the testing procedure developed by Andrews (1993), which treats the break dates as unknown. Confidence intervals for the estimated break dates are calculated by the method proposed by Bai (1997). For the $\Delta_1 p$ series, in an autoregressive model with seasonal dummy variables, namely

$$\Delta_1 p_t = \beta_0 + \beta_1 \Delta_1 p_{t-1} + \beta_2 \Delta_1 p_{t-4} + \sum_{j=1}^{3} \gamma_j Q_{jt} + \varepsilon_t \ , \qquad (3)$$

we find three significant breaks in β_0, but none in the remaining coefficients, at the following dates (95% confidence intervals in parentheses):

1972:3 (1970:3-1974:3)

1980:2 (1979:2-1981:2)

1990:4 (1987:4-1993:4).

These are similar dates to those of the more general breaks identified by Benati (2004), noted above, although in our case it is the date of the second break that is most precisely estimated. Likewise our three break dates are close to the dates of the first three breaks estimated in the three-variable VAR of Meenagh *et al*. (2007, Table 1). We have no counterpart to their fourth break, in 1993:4, associated with the introduction of inflation targeting a year earlier, although this date is the upper limit of the 95% confidence interval for our third break, which is the least precisely determined of the three.

The resulting equation with shifts in β_0 shows evidence of ARCH over the whole period, but results given in the final paragraph of Section 3 about its time dependence suggest separate testing in each of the four subperiods defined by the three break dates. In none of the subperiods is there evidence of ARCH. As an alternative representation of heteroskedasticity we consider breaks in the error variance. Following Sensier and van Dijk (2004) we again locate three significant breaks, at similar dates, namely 1974:2, 1981:3 and 1990:2. Estimates of the full model are presented in Table 2, and the implied subperiod means and standard deviations of inflation are shown as horizontal lines in Figures 1(a) and 3(a) respectively.

For the $\Delta_4 p$ series seasonal dummy variables are not required, but a moving average error is included, and the autoregression is slightly revised, giving the model

$$\Delta_4 p_t = \beta_0 + \beta_1 \Delta_4 p_{t-1} + \beta_2 \Delta_4 p_{t-2} + \varepsilon_t + \theta \varepsilon_{t-4} \ . \tag{4}$$

Again we find three significant breaks in β_0, the first and third of which are accompanied by shifts in β_1, the dates being as follows:

1975:3 (1974:2-1976:4)

1981:4 (1981:2-1982:2)

1988:3 (1987:2-1989:4).

As in the quarterly difference series, ARCH effects persist over the whole period, but there are no ARCH effects in any of the subperiods defined by these shifts in mean. With the same motivation as above we also find three significant breaks in variance in this case, namely 1974:2, 1980:2 and 1990:2, the first and last dates exactly coinciding with those estimated for the $\Delta_1 p$ series. This again provides an alternative representation of the observed heteroskedasticity, and the corresponding subperiod means and standard deviations are shown in Figures 1(b) and 3(b) respectively. (Note that regression residuals sum to zero over the full sample period, but not in each individual subperiod, because some coefficients do not vary between subperiods. Hence the plotted values in Figure 1 do not coincide with the subperiod means of the inflation data.)

The ARCH regression model and the alternative autoregressive model with intercept breaks in mean and variance are non-nested, and can be compared via an information criterion which takes account of the difference in the number of estimated parameters in each model. We find that the three measures in popular use, namely Akaike's information criterion, the Hannan-Quinn criterion and the Schwarz criterion, unambiguously select the breaks model, for both $\Delta_1 p$ and $\Delta_4 p$ versions.

A final note on outliers is perhaps in order, since several empirical researchers identify inflation outliers associated with the increase in Value Added Tax in 1979:3 and the introduction of Poll Tax in 1990:2, and deal with them accordingly. We simply report that none of the modelling exercises presented in this section is sensitive to changes in the treatment of these observations.

5. MEASURES OF INFLATION FORECAST UNCERTAINTY

Publication of the UK Government's short-term economic forecasts began on a regular basis in 1968. The 1975 Industry Act introduced a requirement for the Treasury to publish two forecasts each year, and to report their margins of error. The latter requirement was first met in December 1976, with the publication of a table of the mean absolute error (MAE) over the past ten years' forecasts of several variables, compiled in the early part of that period from internal, unpublished forecasts. Subsequently it became standard practice to include a column of MAEs in the forecast table – users could then easily form a forecast interval around the given point forecast, if they so wished – although in the 1980s and 1990s these were often accompanied by a warning that they had been computed over a period when the

UK economy was more volatile than expected in the future. This publication practice continues to the present day.

We consider the RPI inflation forecasts described as "fourth quarter to fourth quarter" forecasts, published each year in late November-early December in Treasury documents with various titles over the years – *Economic Progress Report*, *Autumn Statement*, *Financial Statement and Budget Report*, now *Pre-Budget Report*. For comparability with other measures reported as standard errors or standard deviations we multiply the reported forecast MAEs, which are rounded to the nearest quarter percentage point, by 1.253 $\left(=\sqrt{\pi/2}\right)$, since Melliss and Whittaker's (2000) review of Treasury forecasts found that "the evidence supports the hypothesis that errors were normally distributed". The resulting series is presented in Figure 6(a). The series ends in 2003, RPI having been replaced by CPI in the 2004 forecast; no MAE for CPI inflation forecasts has yet appeared. The peak of 5 percentage points occurs in 1979, when the point forecast for annual inflation was 14%; on this occasion, following the new Conservative government's policy changes, the accompanying text expressed the view that the published forecast MAEs were "likely to understate the true margins of error".

For comparative purposes over the same period we also plot comparable forecast standard errors for the two models estimated in Sections 3 and 4 – the ARCH model and the breaks model. In common with the practice of the Treasury and other forecasters we use the annual inflation $(\Delta_4 p_t)$ versions of these models. Similarly we regard the "year-ahead" forecast as a five-quarter-ahead forecast, since when forecasting the fourth quarter next year we first have to "nowcast" the fourth quarter this year, given that only third-quarter information is available when the forecast is constructed. The forecast standard errors take account of the estimated autoregressions in projecting five quarters ahead, but this is an "in-sample" or *ex post* calculation that assumes knowledge of the full-sample estimates at all intermediate points including, for the breaks model, the dates of the breaks; the contribution of parameter estimation error is also neglected. It is seen that the ARCH model's forecast standard error shows a much more exaggerated peak than that of Treasury forecasts in 1979, and is more volatile over the first half of the period shown, whereas the breaks model's forecast standard error is by definition constant over subperiods. Of course, in real-time *ex* *ante* forecasting the downward shift in forecast standard error could only be recognised with a lag, as discussed below.

From 1996 two additional lines appear in Figure 6(a), following developments noted in the Introduction. As late as 1994 the Treasury could assert that "it is the only major forecasting institution regularly to publish alongside its forecasts the average errors from past forecasts" (HM Treasury, 1994, p.11), but in 1996 density forecasts of inflation appeared on the scene. We consider the Bank of England's forecasts published around the same time as the Treasury forecasts, namely those appearing in the November issue of the quarterly *Inflation Report*. From the Bank's spreadsheets that underlie the fan charts of quarterly forecasts, originally up to two years ahead (nine quarters), later extended to three years, we take the uncertainty measure (standard deviation) of the five-quarter-ahead inflation forecast. This is labelled MPC in Figure 6(a), because the Bank's Monetary Policy Committee, once it was established, in 1997, assumed responsibility for the forecast.

In 1996 the Bank of England also initiated its quarterly Survey of External Forecasters, at first concerned only with inflation, later including other variables. The quarterly *Inflation Report* includes a summary of the results of the latest survey, conducted approximately three weeks before publication. The survey asks for both point forecasts and density forecasts, reported as histograms, and from the individual responses Boero, Smith and Wallis (2008) construct measures of uncertainty and disagreement. Questions 1 and 2 of each quarterly survey concern forecasts for the last quarter of the current year and the following year, respectively, and for comparable year-ahead forecasts we take the responses to question 2 in the November surveys. For these forecasts our SEF average individual uncertainty measure is plotted in Figure 6(a).

The general appearance of Figure 6(a) has few surprises for the careful reader of the preceding sections. The period shown divides into two subperiods, the first with high and variable levels of forecast uncertainty, the second with low and stable levels of forecast uncertainty, where the different estimates lie within a relatively small range. The recent fall in the Treasury forecast standard error may be overdramatised by rounding, whereas the fall in SEF uncertainty is associated by Boero, Smith and Wallis (2008) with the 1997 granting of operational independence to the Bank of England to pursue a monetary policy of inflation targeting. Their quarterly series show a reduction in uncertainty until the May 1999 Survey of External Forecasters, after which the general level is approximately constant. This reduction in uncertainty about future inflation is attributed to the increasing confidence in, and credibility of, the new monetary policy arrangements.

The forecast evaluation question, how reliable are these forecasts, applies to measures of uncertainty just as it does to measures of location, or point forecasts. Wallis (2004) presents an evaluation of the current-quarter and year-ahead density forecasts of inflation published by the MPC and NIESR. He finds that both overstated forecast uncertainty, with more inflation outcomes falling in the central area of the forecast densities, and fewer in the tails, than the densities had led one to expect. Current estimates of uncertainty are based on past forecast errors, and both groups had gone back too far into the past, into a different monetary policy regime with different inflation experience. Over 1997-2002 the MPC's year-ahead point forecast errors have mean zero and standard deviation 0.42, and the fan chart standard deviation gets closest to this, at 0.48, only at the end (2002:4) of the period considered. Mitchell (2005), for the NIESR forecasts, asks whether the overestimation of uncertainty could have been detected, in real time, had forecasters been alert to the possibility of a break in the variance. Statistical tests can detect breaks only with a lag, and in a forecast context we must also wait to observe the outcome before having information relevant to the possibility of a break in uncertainty at the forecast origin. In a "pseudo real time" recursive experiment it is concluded that tests such as those used in Section 4 could have detected at the end of 1996 that a break in year-ahead forecast uncertainty had occurred in 1993:4. This is exactly the date of the most recent break identified by Meenagh *et al*. (2007), and Mitchell's estimate is that it would not have been recognised by statistical testing until three years later; in the meantime forecasters might have been able to make judgmental adjustments.

As an aside we discuss a recent inflation point forecast evaluation study in which the same issue arises. Groen, Kapetanios and Price (2008) compare the inflation forecasts published in the Bank of England's *Inflation Report* with those available in pseudo real time from a suite of statistical forecasting models. All of the latter are subject to possible breaks in mean, so following a breaks test, the identified break dates are used to demean the series prior to model estimation, then the statistical forecasts are the remeaned projections from the models. It is found that in no case does a statistical model outperform the published forecasts. The authors attribute the Bank forecasters' success to their ability to apply judgment in anticipating the important break, namely the change of regime in 1997:3

following Bank independence. As in Mitchell's study, the *ex ante* recursively estimated shift is not detected until three years later.

For Treasury forecasts, which started earlier, we can compare the *ex ante* uncertainty measures in Figure 6(a) with the forecast root mean squared errors of year-ahead inflation forecasts reported by Melliss and Whittaker (2000). Over subperiods, dated by forecast origin, these *ex post* measures are: 1979-1984, 2.3%; 1985-1992, 1.7%; 1993-1996, 0.8%. These are below, often substantially so, the values plotted in Figure 6(a), with the exception of the 1990 and 1992 forecasts, again illustrating the difficulty of projecting from past to future in times of change.

In the absence of direct measures of uncertainty it is often suggested that a measure of disagreement among several competing point forecasts may serve as a useful proxy. How useful such a proxy might be can be checked when both measures are available, and there is a literature based on the US Survey of Professional Forecasters that investigates this question, going back to Zarnowitz and Lambros (1987). However recent research on the SPF data that brings the sample up to date and studies the robustness of previous findings to the choice of measures finds little support for the proposition that disagreement is a useful proxy for uncertainty (Rich and Tracy, 2006, for example). In the present context we provide a visual illustration of this lack of support by plotting in Figure 6(b) two measures of disagreement based on year-ahead point forecasts of UK inflation. Although the series are relatively short, we use the same scales in panels (a) and (b) of Figure 6 to make the comparison as direct as possible and the lack of a relation as clear as possible. The first series is based on the Treasury publication *Forecasts for the UK Economy*, monthly since October 1986, which is a summary of published material from a wide range of forecasting organisations. Forecasts for several variables are compiled, and their averages and ranges are also tabulated. We calculate and plot the sample standard deviation of year-ahead inflation forecasts in the November issue of the publication. The shorter series is our corresponding disagreement measure from the Bank of England Survey of External Forecasters (Boero, Smith and Wallis, 2008). Other than a slight downward drift, neither series shows any systematic pattern of variation, nor any correlation of interest with the uncertainty measures. We attribute the lower standard deviation in the SEF to the Bank's care in selecting a well-informed sample, whereas the Treasury publication is all-encompassing.

6. UNCERTAINTY AND THE LEVEL OF INFLATION

The suggestion by Friedman (1977) that the level and uncertainty of inflation are positively correlated has spawned a large literature, both theoretical and empirical. Simple evidence of such an association is provided by our breaks model where, using Benati's (2004) characterisation of the four subperiods as a period of high inflation and inflation variability, a period of low inflation and inflation variability, and two "in-between" periods, we note that the high and low periods for both measures coincide. Compare the horizontal lines in Figures 1(a) and 3(a) for the $\Delta_1 p$ model, and in Figures 1(b) and 3(b) for the $\Delta_4 p$ model. For the unconditional subperiod means and standard deviations of inflation over a shorter period (1965-2003), Meenagh *et al.*'s data (2007, Table 2) show a stronger association: when their five policy subperiods are ranked by mean inflation and by inflation standard deviation, the ranks exactly coincide. Of course the empirical literature contains analyses of much greater sophistication although, perhaps surprisingly, they are not subjected to tests of structural stability.

Two leading examples in the empirical literature, on which we draw, are the articles by Baillie, Chung and Tieslau (1996) and Grier and Perry (2000), in which various extensions of the GARCH-in-mean (GARCH-M) model are developed in order to formalise and further investigate Friedman's proposition. The first authors analyse inflation in ten countries, the second authors analyse inflation and GDP growth in the US, including subsample analyses. Of particular relevance for the present purpose is the inclusion of the conditional variance (or standard deviation) in the inflation equation and, simultaneously, lagged inflation in the conditional variance equation. Then, with a GARCH representation of conditional heteroskedasticity, the model is

$$\Delta_1 p_t = \beta_0 + \beta_1 \Delta_1 p_{t-1} + \beta_2 \Delta_1 p_{t-4} + \sum_{j=1}^{3} \gamma_j Q_{jt} + \delta_1 \sqrt{h_t} + \varepsilon_t \qquad (5)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} + \delta_2 \Delta_1 p_{t-1} \, . \qquad (6)$$

Full-sample estimation results show positive feedback effects between the conditional mean and the conditional variance, with a highly significant coefficient on lagged inflation in the variance equation (δ_2), and a marginally significant coefficient (p-value 0.063) on the conditional standard deviation in the mean equation (δ_1); all other coefficients are highly significant. However the model is not invariant over subperiods. If we simply split the sample at 1980, then the estimate of δ_2 retains its significance while the GARCH-M effect drops out from equation (5), which may be associated with the insignificant estimates of α_1 and α_2 in equation (6). All of these statements apply to each half-sample, however further division reveals the fragility of the significance of δ_2. As a final test we return to the breaks model of Section 4 and add the conditional standard deviation in mean and lagged inflation in variance effects. Equivalently, we allow the separate intercept terms in equations (5) and (6), β_0 and α_0, to shift at the dates estimated in Section 4; the coefficients α_1 and α_2 are pretested and set to zero. This model dominates the originally estimated model (5)-(6) on the three standard information criteria, yet has completely insignificant estimates of δ_1 and δ_2. More elaborate models are not able to take us much beyond Friedman's simple association between the first and second moments of inflation, as reflected in the shifts of our preferred model.

7. CONCLUSION

Robert Engle's concept of autoregressive conditional heteroskedasticity was a major breakthrough in the analysis of time series with time-varying volatility, recognised by the joint award of the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel in 2003. "The ARCH model and its extensions, developed mainly by Engle and his students, proved especially useful for modelling the volatility of asset returns, and the resulting volatility forecasts can be used to price financial derivatives and to assess changes over time in the risk of holding financial assets. Today, measures and forecasts of volatility are a core component of financial econometrics, and the ARCH model and its descendants are the workhorse tools for modelling volatility" (Stock and Watson, 2007, p.657). His initial application was in macroeconometrics, however, and reflected his location in the United Kingdom at the time. This paper returns to his study of UK inflation in the light of the well-documented changes in economic policy from his original sample period to the present time.

Investigation of the stability of the ARCH regression model of UK inflation shows that little support for the existence of the ARCH effect would be obtained in a sample period starting later than 1980; data from the earlier period of "monetary policy neglect" (Nelson and Nikolov, 2004) are necessary to support Engle's formulation. Fuller investigation of the nature of the nonstationarity of inflation finds that a simple autoregressive model with structural breaks in mean and variance, constant within subperiods (and with no unit roots), provides a preferred representation of the observed heteroskedasticity from an economic historian's point of view. As noted at the outset, however, the ARCH model has a strong forecasting motivation, and forecasters using the breaks model need to anticipate future breaks. Nevertheless the shifts also provide a simple characterisation of the association between the level and uncertainty of inflation suggested by Friedman (1977), which more elaborate models of possible feedbacks are unable to improve upon.

The United Kingdom can claim several firsts in the measurement and public discussion of the uncertainty surrounding economic forecasts by official agencies, and we present a range of measures of inflation forecast uncertainty, from the models considered here and from other UK sources. The few available evaluations of their accuracy indicate that the well-known problems of projecting from past to future in times of change apply equally well to measures of uncertainty as to point forecasts. While the paper reemphasises the importance of testing the structural stability of econometric relationships, it also acknowledges the difficulty of dealing with instability in a forecast context, for both the levels of variables of interest and, receiving more attention nowadays, their uncertainty.

REFERENCES

Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61, 501-533.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory,* 13, 315-352.

Baillie, R.T., Chung, C.F. and Tieslau, M.A. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics*, 11, 23-40.

Benati, L. (2004). Evolving post-world war II U.K. economic performance. *Journal of Money, Credit, and Banking*, 36, 691-717.

Boero, G., Smith, J. and Wallis, K.F. (2008). Uncertainty and disagreement in economic prediction: the Bank of England Survey of External Forecasters. *Economic Journal*, 118, forthcoming.

Castle, J.L. and Hendry, D.F. (2007). Forecasting UK inflation: the roles of structural breaks and time disaggregation. Discussion Paper No.309, Department of Economics, University of Oxford.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.

Friedman, M. (1977). Nobel lecture: inflation and unemployment. *Journal of Political Economy*, 85, 451-472.

Garratt, A., Lee, K., Pesaran, M.H. and Shin, Y. (2003). A long run structural macroeconometric model of the UK. *Economic Journal*, 113, 412-455.

_____ (2006). *Global and National Macroeconometric Modelling: A Long-Run Structural Approach*. Oxford: Oxford University Press.

Grier, K.B. and Perry, M.J. (2000). The effects of real and nominal uncertainty on inflation and output growth: some GARCH-M evidence. *Journal of Applied Econometrics*, 15, 45-58.

Groen, J.J.J., Kapetanios, G. and Price, S. (2008). Real time evaluation of *Inflation Report* and *Greenbook* forecasts for inflation and growth. Presented at the Royal Economic Society Annual Conference, University of Warwick, March 2008.

Haldane, A. and Quah, D. (1999). UK Phillips curves and monetary policy. *Journal of Monetary Economics*, 44, 259-278.

HM Treasury (1994). Economic forecasting in the Treasury. Government Economic Service Working Paper No.121. London: HM Treasury.

McNees, S.K. (1979). The forecasting record for the 1970s. *New England Economic Review*, September/October 1979, 33-53.

Meenagh, D., Minford, P., Nowell, E., Sofat, P. and Srinivasan, N. (2007). Are the facts of UK inflation persistence to be explained by nominal rigidity or changes in monetary regime? Cardiff Economics Working Paper E2007/21, Cardiff University.

Melliss, C. and Whittaker, R. (2000). The Treasury's forecasts of GDP and the RPI: how have they changed and what are the uncertainties? In *Econometric Modelling: Techniques and Applications* (S. Holly and M.R. Weale, eds), pp.38-68. Cambridge: Cambridge University Press.

Mitchell, J. (2005). The National Institute density forecasts of inflation. *National Institute Economic Review*, No.193, 60-69.

Nelson, E. (2007). An overhaul of doctrine: the underpinning of U.K. inflation targeting. Working Paper 2007-026C, Federal Reserve Bank of St. Louis.

Nelson, E. and Nikolov, K. (2004). Monetary policy and stagflation in the UK. *Journal of Money, Credit, and Banking*, 36, 293-318.

Nerlove, M. (1965). Two models of the British economy: a fragment of a critical survey. *International Economic Review*, 6, 127-181.

Rich, R. and Tracy, J. (2006). The relationship between expected inflation, disagreement and uncertainty: evidence from matched point and density forecasts. Staff Report No.253, Federal Reserve Bank of New York.

Sensier, M. and van Dijk, D (2004). Testing for volatility changes in U.S. macroeconomic time series. *Review of Economics and Statistics*, 86, 833-839.

Shephard, N. (2007). Stochastic volatility. In *New Palgrave Dictionary of Economics*, 2nd Edn (S. Durlauf and L. Blume, eds). MacMillan, forthcoming.

Stock, J.H. and Watson, M.W. (2007). *Introduction to Econometrics*, 2nd Edn. Boston: Pearson Education.

Wallis, K.F. (1989). Macroeconomic forecasting: a survey. *Economic Journal*, 99, 28-61.

_____ (2004). An assessment of Bank of England and National Institute inflation forecast uncertainties. *National Institute Economic Review*, No.189, 64-71.

_____ (2008). Forecast uncertainty, its representation and evaluation. In *Econometric Forecasting and High-Frequency Data Analysis* (R.S. Mariano and Y.K. Tse, eds), Vol.13 of the Lecture Notes Series of the Institute for Mathematical Sciences, National University of Singapore, pp.1-51. Singapore: World Scientific.

Zarnowitz, V. and Lambros, L.A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, 95, 591-621.

Zivot, E. and Andrews, D.W.K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics*, 10, 251-270.

Table 1. Estimation of the original ARCH model over 1958:1-2006:4

$$\Delta_1 p_t = \beta_0 + \beta_1 \Delta_1 p_{t-1} + \beta_2 \Delta_1 p_{t-4} + \beta_3 \Delta_1 p_{t-5} + \beta_4 \left(p_{t-1} - w_{t-1} \right) + \varepsilon_t ,$$

$$\varepsilon_t \left| \psi_{t-1} \sim N\left(0, h_t\right), \quad h_t = \alpha_0 + \alpha_1 \left(0.4\varepsilon_{t-1}^2 + 0.3\varepsilon_{t-2}^2 + 0.2\varepsilon_{t-3}^2 + 0.1\varepsilon_{t-4}^2 \right) \right.$$

	Coeff.	Std. Error	*z*-statistic	p-value
$\hat{\beta}_0$	0.014	0.0097	1.44	0.150
$\hat{\beta}_1$	0.391	0.0852	4.59	0.000
$\hat{\beta}_2$	0.659	0.0504	13.07	0.000
$\hat{\beta}_3$	–0.337	0.0646	–5.22	0.000
$\hat{\beta}_4$	0.002	0.0062	0.39	0.696
$\hat{\alpha}_0$	0.0002	8E–05	2.99	0.003
$\hat{\alpha}_1$	1.009	0.1564	6.45	0.000

Log likelihood	398.9	Akaike info criterion	–4.00
		Schwarz criterion	–3.88
		Hannan-Quinn crit’n	–3.95

Table 2. Estimation of the "breaks" model, 1958:1-2006:4

$$\Delta_1 p_t = \beta_0 + \beta_1 \Delta_1 p_{t-1} + \beta_2 \Delta_1 p_{t-4} + \sum_{j=1}^{3} \gamma_j Q_{jt} + \delta_1 D72{:}3 + \delta_2 D80{:}2 + \delta_3 D90{:}4 + \varepsilon_t$$

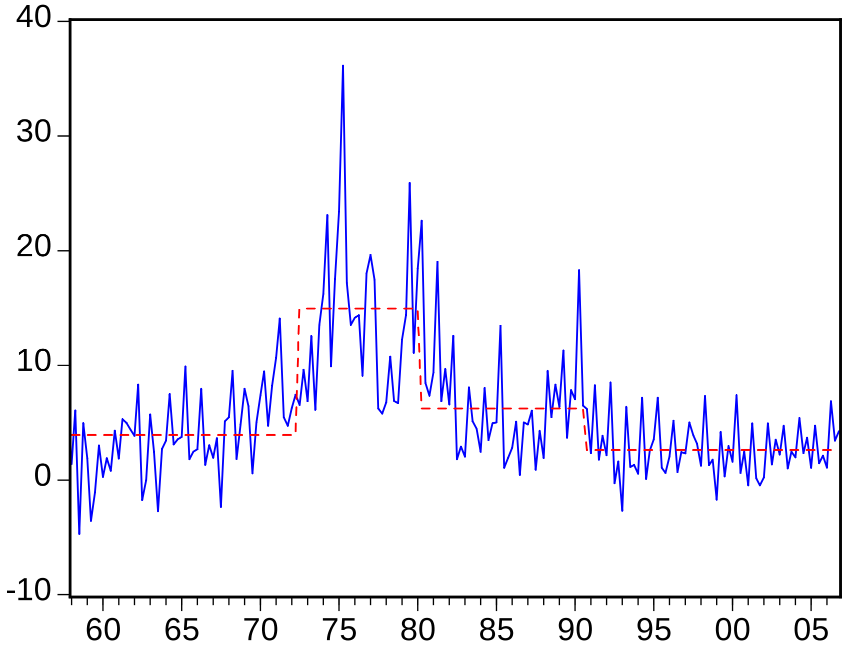
$$\varepsilon_t \left| \psi_{t-1} \sim N\left(0, h_t\right), \quad h_t = \alpha_0 + \alpha_1 D74{:}2 + \alpha_2 D81{:}3 + \alpha_3 D90{:}2\right.$$

	Coeff.	Std. Error	*z*-statistic	p-value
$\hat{\beta}_0$	0.024	0.007	3.62	0.000
$\hat{\gamma}_1$	–0.016	0.006	–2.73	0.006
$\hat{\gamma}_2$	0.030	0.006	4.92	0.000
$\hat{\gamma}_3$	–0.038	0.007	–5.30	0.000
$\hat{\beta}_1$	0.405	0.070	5.77	0.000
$\hat{\beta}_2$	0.138	0.074	1.88	0.061
$\hat{\delta}_1$	0.047	0.012	3.96	0.000
$\hat{\delta}_2$	–0.038	0.011	–3.50	0.001
$\hat{\delta}_3$	–0.015	0.005	–2.87	0.004
$\hat{\alpha}_0$	0.001	0.000	6.37	0.000
$\hat{\alpha}_1$	0.003	0.001	2.67	0.008
$\hat{\alpha}_2$	–0.003	0.001	–3.05	0.002
$\hat{\alpha}_3$	–0.001	0.000	–5.68	0.000

Log likelihood	449.5	Akaike info criterion	–4.45
		Schwarz criterion	–4.24
		Hannan-Quinn crit'n	–4.37

Figure 1. UK RPI inflation 1958:1-2006:4 (percentage points of annual inflation)

(a) $\Delta_1 p_t$



(b) $\Delta_4 p_t$

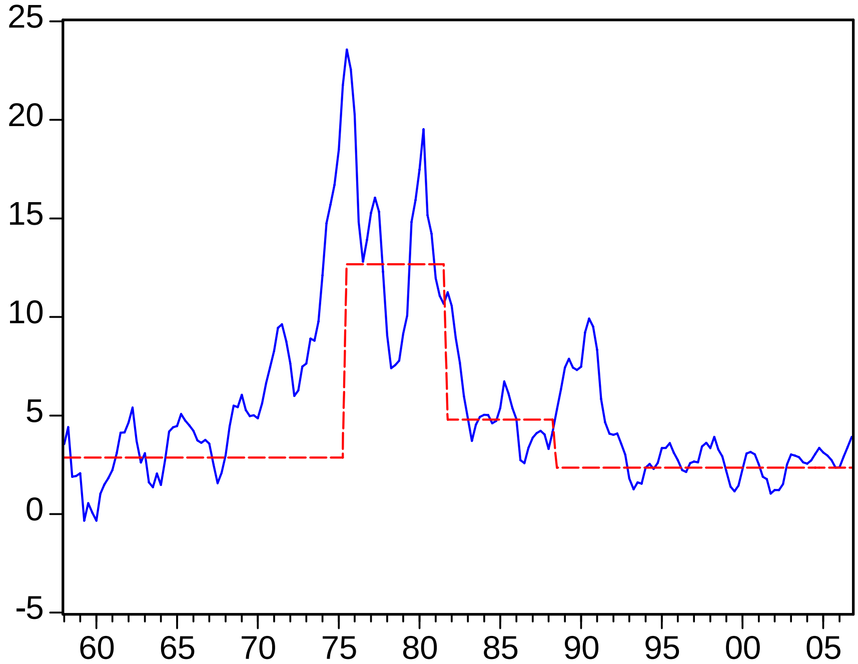


Figure 2. Conditional standard errors, 1958:1-1977:2, $\Delta_1 p_t$

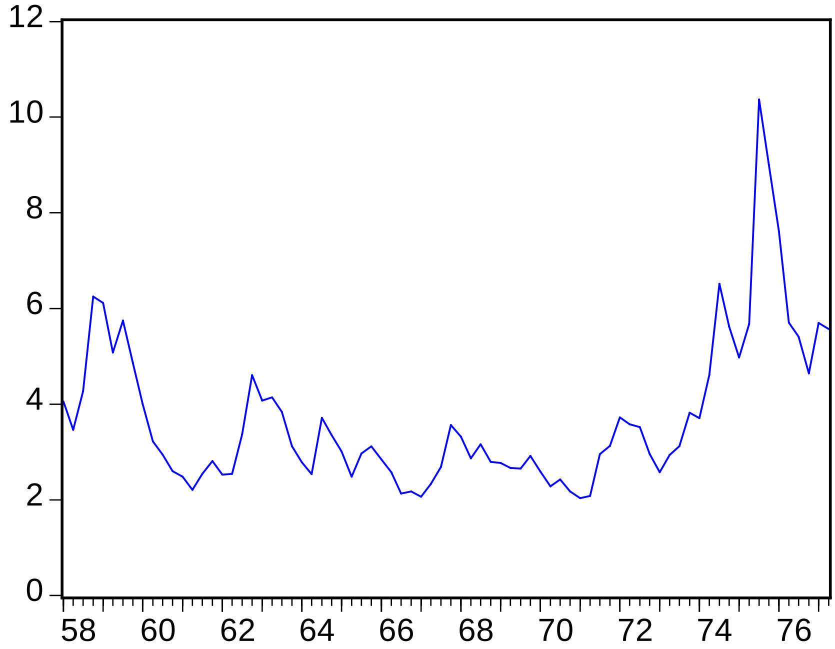
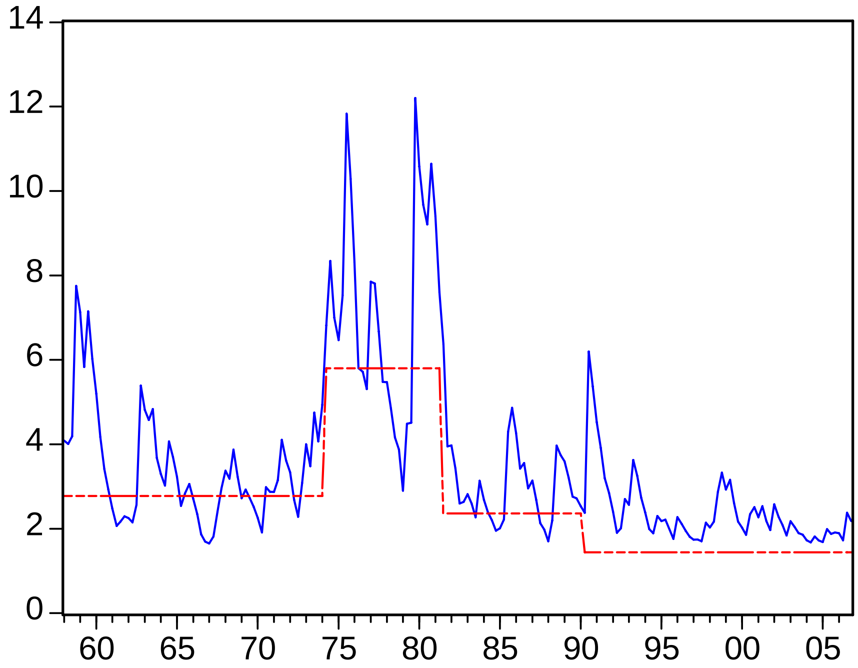


Figure 3. Conditional standard errors, 1958:1-2006:4

(a) $\Delta_1 p_t$



(b) $\Delta_4 p_t$

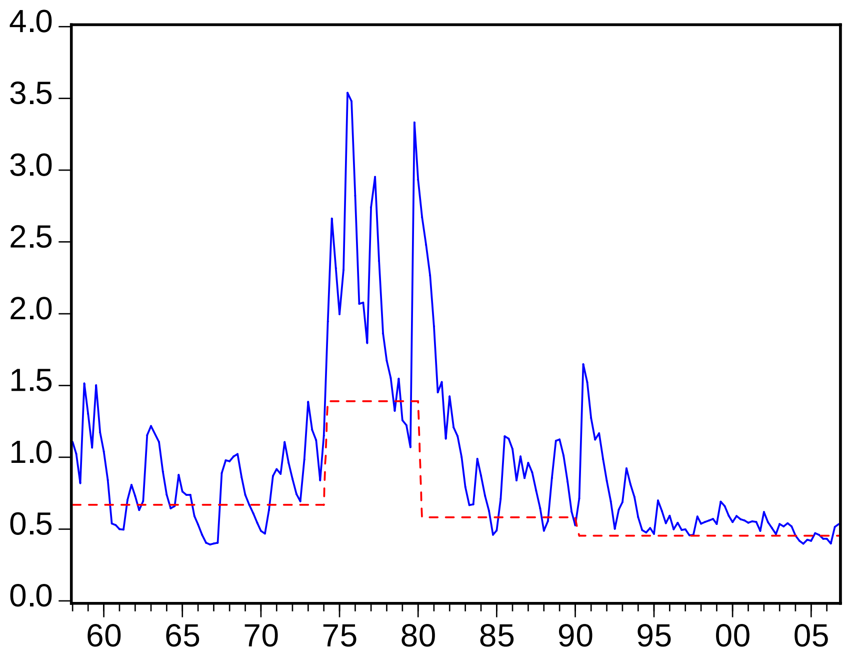


Figure 4. Backward recursive estimates of α_1 and the LM p-value, $\Delta_1 p_t$

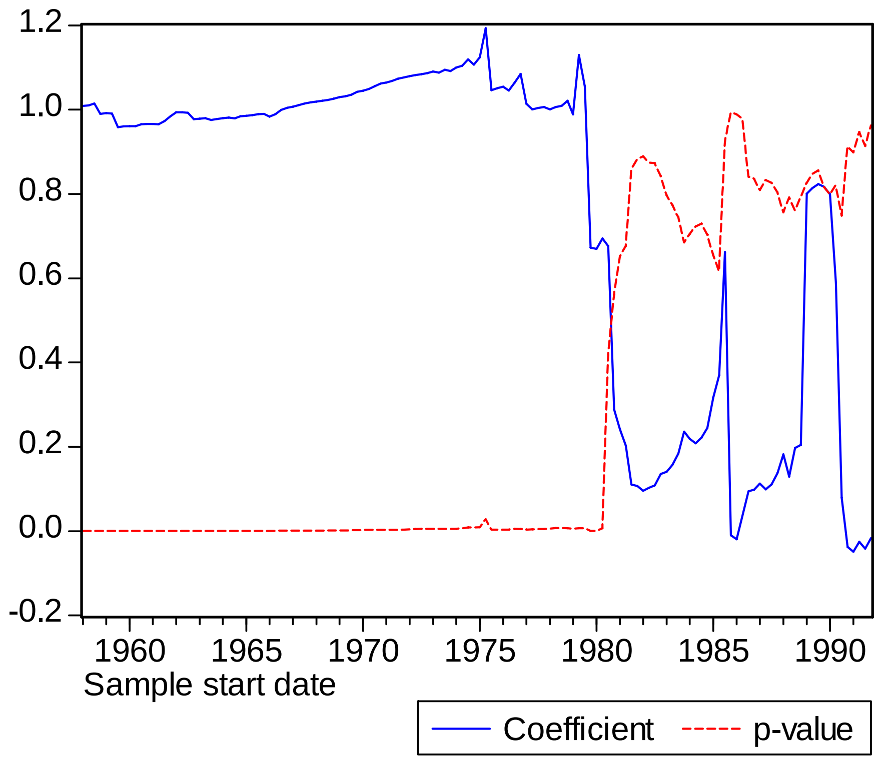
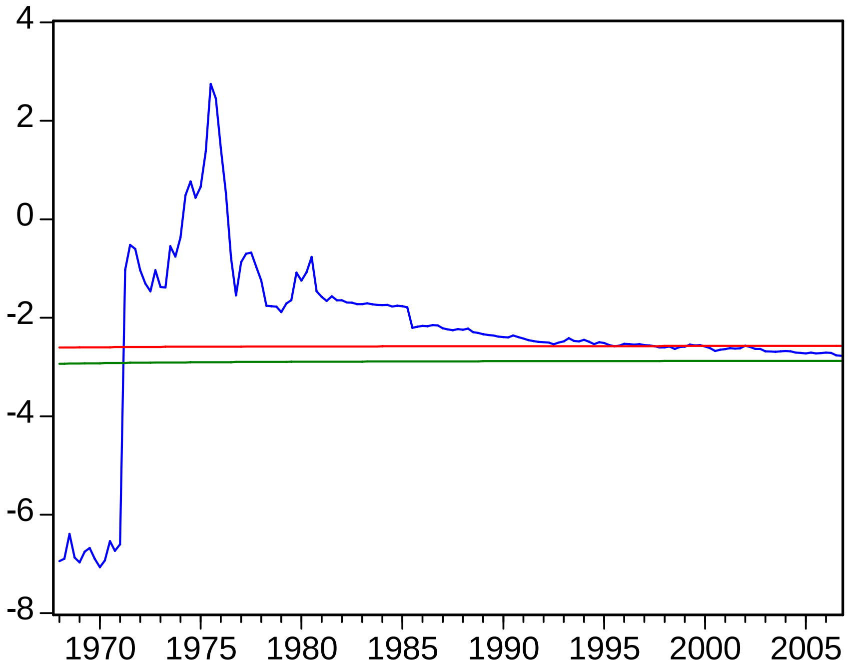


Figure 5. Recursive ADF tests for $\Delta_1 p_t$, with 5% and 10% critical values

(a) Constant only



(b) Constant and seasonal dummies

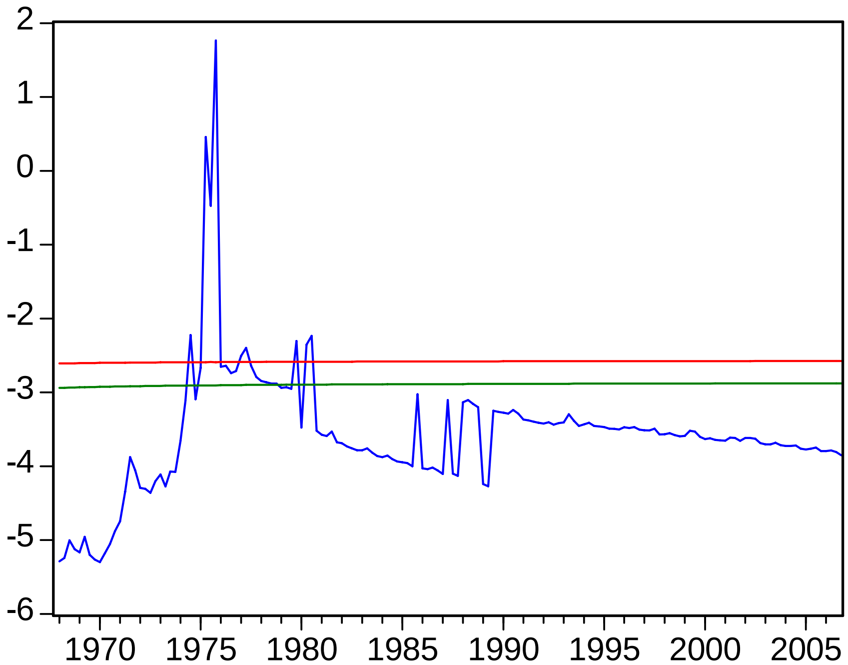
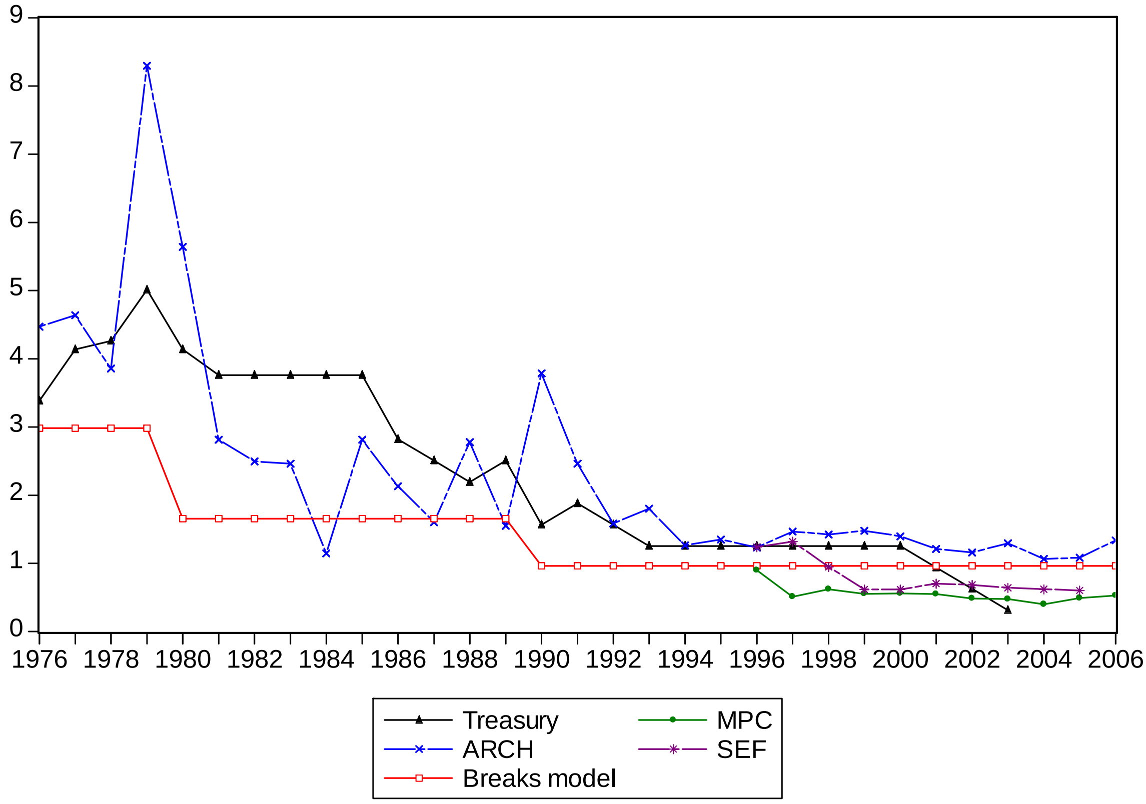


Figure 6. Measures of uncertainty and disagreement, year-ahead forecasts, 1976-2006

(a) Uncertainty



(b) Disagreement