Forecasting U.K. Inflation: The Roles of Structural Breaks and Time Disaggregation∗

Jennifer L. Castle       David F. Hendry
Oxford University        Oxford University

November 24, 2006 (revised)

Abstract

Structural models’ inflation forecasts are often inferior to those of naïve devices. This chapter theoretically and empirically assesses this for U.K. annual and quarterly inflation, using the theoretical framework in Clements and Hendry (1998, 1999). Forecasts from equilibrium-correction mechanisms, built by automatic model selection, are compared to various robust devices. Forecast-error taxonomies for aggregated and time-disaggregated information reveal that the impacts of structural breaks are identical between these, helping to interpret the empirical findings. Forecast failures in structural models are driven by their deterministic terms, confirming location shifts as a pernicious cause thereof, and explaining the success of robust devices.

JEL classifications: C32, C53

Keywords: Inflation forecasting; Structural breaks; Robust forecasts; Time disaggregation; Forecast-error taxonomies

∗Castle (corresponding author): Department of Economics, Manor Road Building, Manor Road, Oxford, OX1 3UQ, UK (jennifer.castle@nuffield.ox.ac.uk); Hendry: Department of Economics, Manor Road Building, Manor Road, Oxford, OX1 3UQ, UK (david.hendry@nuffield.ox.ac.uk)
1 Introduction

Systematic mis-forecasting of economic events has led to extensive research on economic forecasting, culminating in a new theory of forecasting developed by Clements and Hendry (1998, 1999, 2002). This theory extends the previous assumptions that the econometric model coincides with the data-generating process (DGP) in a stationary world. The forecast-error taxonomy they developed allows for a mis-specified model with measurement error in the data, within a non-stationary world that is subject to structural breaks. That is more representative of the conditions in which many economic forecasts are made, including U.K. inflation. This chapter assesses their theory in light of empirical evidence, examining the forecast performance of structural models of inflation in comparison to alternative forecasting devices. Two forecast periods are examined: one with a break in the level of inflation (1990q4–1995q3), thereby capturing a period of structural change including exiting from the European Exchange Rate Mechanism (ERM) and the transition period to the new inflation targeting regime, and the other without a clear break, in which the inflation targeting regime operated by an independent central bank was well established (1998q3–2003q2).¹

The structural forecasts are based on single-equation and vector equilibrium-correction models of inflation built using the automatic model selection algorithm, PcGets. The inflation models perform well over the in-sample periods, incorporating most extant theories of inflation and, despite major changes in the sample and in data frequency, closely match the annual U.K. inflation model of Hendry (2001). We compare the forecasts from these structural models with various transformations which attempt to ‘robustify’ forecasts to structural breaks, and we consider differencing devices (Hendry and Clements, 2000; Hendry, 2006), rapid updating tools (Clements and Hendry, 2005), and forecast pooling (Hendry and Clements, 2004).

We focus on two key questions. First, we consider whether robust forecasting devices do prove useful in forecasting macroeconomic time series, and especially whether they outperform the dominant congruent in-sample model, both when there are structural breaks in the data and during relatively quiescent periods.² A forecast competition is undertaken in which 15

¹In October 1992, inflation targeting was introduced with a target for the Retail Price Index (RPIX) of between 1% and 4%, with a preference for the lower half of the range by the end of parliament. In June 1995, a more formal target of 2.5% or less was introduced. A month after operational independence was granted to the Bank of England in May 1997, the symmetric target of ±1% around a 2.5% target was established, along with a system of accountability. In December 2003, the target was changed to 2% based on the Harmonized Index of Consumer Prices (HICP), with a band of ±1%.

²The dominant congruent in-sample model refers to a model that is coherent with the available evidence, assessed by a range of mis-specification tests, and is dominant in the sense that it encompasses other model specifications, (see Hendry, 1995, p. 365). The dominant congruent in-sample model can be viewed as the ‘best’
models are used to compute 1-, 4-, and 8-step-ahead forecasts of quarterly inflation for both 5-year periods. The evidence suggests that alternative forecasting models do yield benefits over and above the congruent, dominant in-sample models. Secondly, we develop a forecast-error taxonomy for time-disaggregated information and use this to interpret the empirical results for forecasting annual inflation. Two models of annual inflation are developed, one being a lower-frequency model using annual-data analogues and the other a higher-frequency model using quarterly data. The taxonomy demonstrates that location shifts have the same impact regardless of the data frequency, but we find that increasing the information set by disaggregation does yield substantial improvements, providing empirical support for the theory of predictability in Clements and Hendry (2005).

The structure of the chapter is as follows. Section 2 presents two econometric models of quarterly inflation, including a single-equation equilibrium-correction model obtained using automatic model selection and a multivariate model based on various input prices and conditioning on excess demand. Section 3 outlines the forecasting models used to forecast quarterly inflation and discusses direct multi-step versus iterated forecasts and forecast evaluation criteria. Section 4 assesses the forecast performance of the econometric models and robust forecasting devices, ranking the models in an attempt to predict which models should forecast well. Section 5 outlines the theory of predictability and develops the forecast error taxonomy for time-disaggregated information, as well as examining the double-differenced forecast. Section 6 builds models of annual inflation using both annual and quarterly data, and Section 7 undertakes forecast comparisons for annual inflation. Section 8 concludes.

2 Quarterly Models of U.K. Inflation

The forecast exercise compares structural econometric models based on equilibrium-correction mechanisms (EqCMs) to a large suite of alternative rules, including robust forecasting devices designed to adapt to structural breaks, univariate time-series models, and pooled forecasts.\(^3\) A substantial literature motivates the choice of EqCM as the structural model for inflation; see, \textit{inter alia}, Rowlatt (1988), Juselius (1992), Metin (1995), De-Brouwer and Ericsson (1998), and Hendry (2001). There are a number of theories that have been postulated to explain the determinants of inflation, including the Phillips (1958, 1962) curve, formalized by Lipsey (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960); the natural rate hypothesis developed independently by Friedman (1968) and Phelps (1960).\(^3\) We exclude factor forecasts from the set examined due to the substantial data requirements for such models.

\(^3\)We exclude factor forecasts from the set examined due to the substantial data requirements for such models.
(1967); the rational expectations hypothesis of Lucas (1973) and Sargent and Wallace (1975); monetarism proposed by Friedman (1970, 1971); excess demand and supply pressures in all sectors of the economy, including the output gap (see, e.g., De-Brouwer and Ericsson, 1998; Hendry, 2001; Bowdler and Jansen, 2004); excess demand in the labor market or competition over the profit share captured by the mark-up (Dicks-Mireaux and Dow, 1959; Sargan, 1964, 1980; Nickell, 1990; Layard et al., 1991); and labor demand pressures through the NAIRU (see, e.g., Ball and Mankiw, 2002). Given the range of potential hypotheses, some form of model selection is essential. We propose the use of automatic general-to-specific model selection, using the PcGets algorithm of Hendry and Krolzig (2001). The algorithm is one of a large class of multi-path search procedures for selecting a parsimonious undominated representation of a general unrestricted model (GUM) using a general-to-specific (Gets) search strategy, and is detailed in the Appendix. This method allows all possible explanations of inflation to play a role if the data support their relevance. The resulting structural models support the conclusion of Hendry (2001) that there is no single cause of U.K. inflation.

2.1 Data

The data set consists of quarterly data for the U.K. over 1965q1–2003q2 and is derived from a number of sources detailed in the Appendix. Two forecast periods are examined (1990q4–1995q3 and 1998q3–2003q2), each consisting of 20 forecast observations (denoted by n). Figure 1 records quarterly and annual inflation based on the GDP deflator, highlighting the two forecast periods. The latter forecast period is relatively quiescent, whereas the earlier forecast period is one of structural change. Over this period, the mean and variance of inflation fell systematically. The smoothness of the annual differences highlights that structural shift more sharply than the movements in quarterly inflation. Table 1 records descriptive statistics for the two forecast periods: the 1990s exhibit a reduction towards lower and more stable inflation, implying that the forecast performance of econometric models should improve. However, structural changes throughout the in-sample period need to be modeled in order for the econometric models to perform well.

**FIGURE 1 ABOUT HERE**

4 All data are seasonally adjusted and lower case letters represent logarithms. The difference operator, $\Delta_i$, is defined as $(1 - L^i)$ where $L$ is the lag operator. Subscripts are omitted for first differences but are indicated for fourth differences, denoting the annual change.

5 The Bank of England targeted the RPIX over this period and not the GDP deflator. Hendry (2001) shows that the GDP deflator and consumer price index do not cointegrate, so caveats apply to interpreting the results for policy.
We use a progressive research strategy, building on previous empirical evidence to specify the initial GUM; see Hendry (1995, p. 550). The EqCMs of inflation are based on a mark-up model, with excess demand pressures causing short-run cyclical movements in inflation while the long-run price level is determined by sectoral price levels. We assume that the price level is $I(1)$ but subject to structural breaks which give the impression that the series is $I(2)$. The main series include producer prices ($ppi$), import prices ($imp$), housing rent ($rent$), unit labor costs scaled for the decline in average hours ($e^*$), oil prices ($oil$), national debt ($n$), external prices ($pw$), profit mark-up ($\pi^*$), output gap ($y^d$), excess demand for unemployment ($U^d$), growth rate of broad money ($\Delta m4$), short and long interest rates ($R_s, R_l$), short-long real interest rate spread ($s$), and real effective exchange rate ($e_r$). Two non-linear purchasing power parity (PPP) interaction terms are also included, given by $[\Delta e_t e_r^{t-j}]$ and $[\Delta e_t e_r^{2t-j}]$. These variables are motivated by Hendry (2000), who finds that devaluations are more inflationary if they occur when far from PPP equilibrium and may enter asymmetrically. We do not include further interaction effects due to the complexity of the GUM, but Castle and Hendry (2005) develop an automatic Gets selection algorithm within a non-linear framework that is designed to automatically generate interaction effects in the initial general model. However, theory, past evidence, and congruency tests support the inclusion of PPP interaction effects.\(^6\)

Castle (2006) provides details of the excess demand measures, based on Hendry (2001). $y^d_t$ is calculated as a Solow residual measure, in which deviations from a measure of potential capacity are used to estimate the gap. The measure of capacity allows for an improvement in productivity growth after 1980 and is given by

$$\text{cap}_t = \begin{cases} 
2.53 + 0.0026t + 0.36(k_t - wpop), & 1966q2–1980q4 \\
2.46 + 0.0033t + 0.36(k_t - wpop), & 1981q1–2003q2,
\end{cases} \quad (1)$$

where $k$ is the capital stock and $wpop$ is the working population. Excess demand for goods and services is given by $y^d_t = y^l_t - \text{cap}_t$, where $y^l_t$ is output per worker.

$U^d_t$ is derived from an EqCM where disequilibrium unemployment is based on steady-state growth, with unemployment rising when the real interest rate exceeds the real growth rate and vice versa. The excess demand for labor measure is given by

$$U^d_t = U_{r,t} - 0.73(R_{l,t} - \Delta p_t - \Delta y_t), \quad (2)$$

\(^6\)Evidence for non-linear inflation effects from the output gap have been examined in Turner (1995) and Clements and Sensier (2003), although the evidence for such effects is less robust. Evidence for threshold effects has been examined by Teräsvirta and Anderson (1992).
where $U_{r,t}$ is the unemployment rate.

Finally, the mark-up is calculated by developing an inflation model based on real input prices and undertaking a reduction, assuming long-run log-linear price homogeneity and that the adjustment speeds are the same in response to $c^*$, $ppi$, and $e_r$, resulting in

$$\pi^*_t = p_t - 0.7c^*_t - 0.1ppi_t - 0.2pw_{E,t} + 0.03,$$

where $pw_{E,t}$ denotes world prices in sterling. Figure 2 records $y^d_t$, $U^d_t$, and $\pi^*_t$ in panels a–c respectively.

**FIGURE 2 ABOUT HERE**

### 2.2 Single-Equation Equilibrium-Correction Models

The initial models of $\Delta p_t$ include three lags of $y^d_t$, $U^d_t$, $s$, $\pi^*$, $\Delta ppi$, $\Delta rent$, $\Delta imp$, $\Delta oil$, $\Delta p$, $\Delta c^*$, $\Delta m4$, $\Delta n$, $\Delta R_s$, $\Delta R_l$, $\Delta pw$, two lags of $\Delta e_{r,t-j}$ and $\Delta e_{r,t-j}^2$, and an intercept and trend. Three blip indicators are included: $D_{73.2,79.3}$ accounts for the substantial negative inflation outlier of $-2.2\%$ in 1973q2, which is assumed to be measurement error, and the VAT increase implemented by Mrs Thatcher; $D_{72.4,74.1}$ accounts for oil price shocks; and $D_{84.1,84.2}$ accounts for exchange rate fluctuations. We exclude contemporaneous covariates from the initial general models to facilitate *ex ante* forecasts. Their inclusion would require known values of the covariates over the forecast horizon, biasing the forecast errors of the partial model downwards in comparison to the system model. The use of the single-equation framework does require weak exogeneity in the regressors, and we relax this assumption for the VEqCM.

Direct estimation techniques are used to obtain multi-step forecasts within the single-equation framework. This implies that 16 EqCMs are estimated in total, using the PcGets ‘conservative’ strategy (approximately 1% per test), based on the same initial general model. As information in periods $t - 1$ to $t - h$ is excluded, residual autocorrelation is likely to bias the estimated coefficient standard errors. To overcome this, looser significance levels are used to select the model, and the autocorrelation mis-specification test is excluded from the test battery for the $h$-step forecasting models (where $h > 1$). The resulting specific models are estimated in PcGive to establish heteroskedasticity and autocorrelation consistent standard errors (HACSE) based on Andrews (1991). The two 1-step forecasting models for the later and earlier periods are reported in (4) and (5), respectively; the Appendix reports the 4- and 8-step forecasting models with HACSEs.$^7$

---

$^7$Coefficient standard errors are shown in parentheses. $R^2$ is the squared multiple correlation, $\hat{\sigma}$ is the residual
The two 1-step single-equation models are quite similar, demonstrating stability over the
period; both equations’ recursive graphics support this conclusion, and the models pass all
diagnostic tests. Both the mark-up and excess demands for goods and labor markets are key
driving forces, along with world inflation. The anomalous negative coefficient on producer
prices for the shorter sample period is likely to be driven by collinearity with world prices,
and may also be driving up the coefficient on world prices. The presence of the intercept in the
longer sample period is concerning, as it entails that ‘autonomous’ inflation exists when all
regressors take their equilibrium values of zero. The parsimony achieved by both models is due
to the selection strategy used: the PcGets conservative strategy is aimed at a 1% significance
level.8 Furthermore, the similarity between the models and the annual inflation model reported
in the Clements and Hendry chapter of this volume is striking, although there is no role for
monetary variables in the quarterly inflation models. The selected variables for the 4-step and
8-step-ahead forecasts differ from those for the 1-step-ahead forecasts. Most models retain
the key determinants of inflation identified by the 1-step models ($y^d, U^d, \Delta pw, \Delta e^s$, and $\pi^*$),

\[
\Delta p_t = 0.205y^d_{t-1} - 0.118U^d_{t-3} + 0.667\Delta pw_{t-2} + 0.131\Delta e^s_{t-1} - 0.125\pi^*_{t-1} \\
+ 0.044D_{t,73:2,79:3} + 0.024D_{72:4,74:1} + 0.016D_{84:1,84:2} + 0.006 \\
(0.032) \\
(0.026) \\
(0.121) \\
(0.042) \\
(0.005) \\
(0.005) \\
(0.004) \\
(0.002)
\]

\[
R^2 = 0.844 \hat{\sigma} = 0.634\% \text{ } SC = -9.850
\]

\[
F_{ar}(5,112) = 0.620 \quad F_{arch}(4,109) = 0.524 \quad F_{het}(16,100) = 1.349 \quad \chi^2_{nd}(2) = 0.846
\]

\[
F_{reset}(1,116) = 1.642 \quad F_{Chow}(20,117) = 0.607 \quad T = 1967q1 - 1998q2.
\]

\[
\Delta p_t = 0.311y^d_{t-1} - 0.154U^d_{t-1} + 1.263\Delta pw_{t-2} - 0.076\Delta ppi_{t-2} - 0.109\pi^*_{t-2} \\
+ 0.046D_{t,73:2,79:3} + 0.021D_{72:4,74:1} + 0.015D_{84:1,84:2} \\
(0.045) \\
(0.028) \\
(0.052) \\
(0.005) \\
(0.005) \\
(0.005) \\
(0.023) \\
(0.003)
\]

\[
\hat{\sigma} = 0.669\% \text{ } SC = -9.702
\]

\[
F_{ar}(5,82) = 1.168 \quad F_{arch}(4,79) = 1.066 \quad F_{het}(16,70) = 1.287 \quad \chi^2_{nd}(2) = 0.928
\]

\[
F_{reset}(1,86) = 0.267 \quad F_{Chow}(20,87) = 1.063 \quad T = 1967q1 - 1990q3.
\]

standard deviation, and $SC$ is the Schwarz criterion. The diagnostic tests are of the form $F_j(k,T-l)$, which
denotes an $F$-test against the alternative hypothesis $j$ for the following: $k$-order serial correlation ($F_{ar}$), $k$-order
autoregressive conditional heteroscedasticity ($F_{arch}$), heteroskedasticity ($F_{het}$), the RESET test ($F_{reset}$), and
parameter constancy over $k$ periods ($F_{Chow}$); finally, $\chi^2_{nd}(2)$ is a chi-square test for normality.

8Castle (2006) estimates an inflation model based on a similar general unrestricted model and finds a role
for an acceleration of inflation term and a purchasing-power parity interaction term using the liberal selection
strategy.
but some models retain other variables such as $\Delta oil$. The models selected to produce direct forecasts are based on different information sets (i.e., excluding information lagged $t-1, \ldots, t-3$ or $t-1, \ldots, t-7$ and hence including information lagged $t-4, \ldots, t-7$ or $t-8, \ldots, t-11$) such that the model selection algorithm identifies slightly different specifications. The longer the forecast horizon, the less timely information that is available when selecting the direct forecasting model and hence the poorer the in-sample model specification.

A caveat applies when using these models to forecast: the excess-demand measures are all calculated using the full sample. Hence, some information in the forecast period is used in the forecasts, and as such the forecasts are necessarily ex post. The computational intensity of estimating 16 sets of excess-demand pressures, exacerbated by end-of-sample problems in estimating the output gap, precluded computing ex ante forecasts. However, both the single-equation and vector EqCMs use the same information set, thereby avoiding any bias in the forecast competition from the use of forecast information included in the models.

The automatic model selection tool is designed to select the dominant congruent in-sample model. The strategy does not use model fit as a criterion, but it is well known that an in-sample encompassing model should fit best. The PcGets program is not designed to select forecasting models; as Clements and Hendry (1998, 1999) demonstrate, the best in-sample model need not be the best forecasting model in a non-stationary world in which models are mis-specified. Hence, the use of automatic model selection enables us to objectively compare models designed on in-sample criteria with alternative devices designed with forecasting as their explicit objective.

### 2.3 Vector Equilibrium-Correction Models

In order to develop system dynamic forecasts, we relax the weak exogeneity assumptions made in the single-equation analysis and build two VEqCMs using the techniques outlined in Johansen (1995). Five endogenous variables are modeled, reducing the information set compared to the single-equation analysis, but enabling us to avoid degrees of freedom problems associated with too general a VAR. Cointegration analysis is conducted on the unrestricted VAR, allowing for a partial system by conditioning on the output gap.\(^9\) As in the single-equation framework, estimates of the output gap are based on the full sample and as such, the forecasts are again ex post. We adopt the framework outlined in Harbo et al. (1998), and Castle (2006) provides full details of the models.

\(^9\)Weak exogeneity of the output gap is confirmed prior to conditioning.
The model developed over the period 1966q2–1998q2 is based on the unrestricted VAR for
\[
x_t = [p_t, c^*_t, ppi_t, pw_t, e_t]'\]
augmented by \(y^d_{t-1}\). A set of indicator variables is included to account for outliers, including an indicator that takes the value 1 in 1974q1 and 0 otherwise \((I_{74:1})\), and two blip dummies defined analogously to the single-equation models. The lag length of the unrestricted VAR is set to 3 based on tests of model reduction, and the VAR passes all diagnostics apart from autocorrelation at the 5% significance level, driven by world prices. Rigorous cointegration analysis leads us to conclude that there is one cointegrating vector, and various restrictions imply the cointegrating vector can be interpreted as a mark-up. The full system is not reported for brevity, but the inflation equation is reported in (6):

\[
\Delta p_t = 0.478 \Delta pw_{t-1} + 0.673 \Delta pw_{t-2} + 0.168 y^d_{t-1} + 0.144 \Delta c^*_{t-1} \\
-0.048 cv_{t-1} - 0.028 I_{74:1} + 0.043 D_{73:2,79:3} + 0.014 D_{73:1,75:1}
\]

\[
\hat{\sigma} = 0.715\% \quad F_{ar}(5, 107) = 5.780^{**} \quad F_{het}(31, 91) = 0.778
\]

\[
\chi^2_{nd}(2) = 0.685 \quad F_{arch}(4, 115) = 0.482 \quad T = 1966q2 - 1998q2,
\]

where \(cv_t = p_t - 0.87c^*_t - 0.09ppi_t - 0.05pw_{e,t} - 0.0012\).

The model for the period 1966q2–1990q3 is based on a slightly different unrestricted VAR given by
\[
y_t = [p_t, c^*_t, imp_t, pw_t, e_t]'.
\]
Selection of the initial VAR regressors was based on general-to-specific modelling of input prices and the results suggested replacing \(ppi\) with \(imp\) in the unrestricted VAR. As the two regressors are highly correlated, the resulting model is similar to the previous model. Again we condition on the output gap, confirmed by a test of weak exogeneity, and the same set of indicators is included. A lag length of 2 is selected using tests of model reduction, and the model passes all diagnostics apart from autocorrelation at the 5% significance level. We find evidence for one cointegrating vector that is similar to the previous cointegrating vector, and is interpreted as the mark-up, although a time trend is found to be significant, possibly due to
unmodeled changes in hours worked. The inflation equation is reported in (7):

\[
\Delta p_t = 1.017 \Delta pw_{t-1} + 0.241 yd_{t-1} + 0.163 \Delta c^*_t - 0.170 cv_{t-1} \\
- 0.032 I_{74:1} + 0.041 D_{73:2,79:3} \\
\hat{\sigma} = 0.749\% \quad F_{ar}(5, 81) = 6.133^{**} \quad F_{het}(19, 74) = 0.653 \\
\chi^2_{nd}(2) = 0.004 \quad F_{arch}(4, 86) = 0.255 \quad T = 1966q2 - 1990q3, \\
\text{where } cv_t = p_t - 0.88c^*_t - 0.12imp_t - 0.07 pw_{w.t} - 0.001t.
\]

Both models exhibit residual autocorrelation that is not removed by including extra lags, hence we report HACSE estimates (see Andrews, 1991). If the inflation equation is estimated independently there is no evidence of autocorrelation. However, the covariance matrix reflects collinearity between \( p, c^* \), and \( ppi (imp) \) in the residuals which is not fully modeled because of the restricted number of variables in the system. Ideally, a larger VEQCM should be built to account for the non-congruency, as the misspecification is a likely indication of omitted relevant variables given the reduced information set, but we proceed with the current models for the forecasting exercise.

The models are similar and both represent a reasonable fit, with equation standard errors just marginally larger than their single-equation counterparts. The low coefficient on the mark-up for the longer sample period is concerning; the coefficient on the mark-up for the shorter sample period corresponds to the single-equation models more closely. As in the single-equation models, the output gap and world inflation are important determinants, and as no role for excess demand for labor is allowed in the VAR framework, the growth of unit labor costs proxies labor market pressures. Observe that a trend is included in the cointegrating vector for the shorter sample period, but the trend is insignificant over the longer sample period. The cointegrating vector for the smaller sample is recorded in Figure 2d, which is scaled for a zero mean over the in-sample period, and the divergent behavior of the cointegrating vector over the out-of-sample period is evident.

3 Quarterly Forecasting Models

The two structural models developed in Section 2 (denoted EqCM and VEQCM for the single-equation and vector models, respectively) are compared to thirteen alternative forecasting
devices. Represent the VEqCMs in (6) and (7) as
\[
\Delta x_t = \gamma + \alpha (\beta' x_{t-1} - \mu) + \sum_{i=1}^{p} \Gamma_i \Delta x_{t-i} + \epsilon_t, \tag{8}
\]
where \( \epsilon_t \sim IN[0, \Omega] \), for \( t = 1, \ldots, T \).\(^{10}\) \( x_t \) is the vector of 5 variables, \( \beta' x_{t-1} - \mu \) is the cointegrating vector defined as the mark-up, and short-run dynamics are captured by \( \Delta x_{t-i} \).

We can use this simplified representation to define the alternative forecasting devices (the single-equation EqCM can also be represented in this framework with \( x \) as a scalar). Clements and Hendry (1999) outline the forecast errors and variances of (8) when a break occurs in \( \mu, \gamma, \) or \( \alpha \), along with the analogous forecast errors and variances for a variety of forecasting devices.

1. A range of forecasting devices attempt to robustify the forecasts to shifts in the parameters of deterministic terms. A differenced vector equilibrium-correction model (\( \Delta \text{VEqCM} \)) is computed as the first difference of the initial VAR with the rank restrictions from cointegration imposed:
\[
\Delta x_t = (I + \alpha \beta') \Delta x_{t-1} + \sum_{i=1}^{p} \Gamma_i \Delta^2 x_{t-i} + \nu_t, \tag{9}
\]
where \( \nu_t = \Delta \epsilon_t \). As the device differences the mean, a shift in \( \mu \) will imply the forecast will fail in the period following the break, but will then correct as \( \Delta \mu = 0 \) in subsequent periods. Hence, the differenced VEqCM will robustify forecasts to location shifts, although a cost is incurred in the form of increased forecast error variance; see Hendry (2006).

2. As the \( \Delta \text{VEqCM} \) contains \( I(-1) \) terms resulting from differencing the short-run dynamics of the model, the forecast error variance will be inflated by highly volatile double-differenced variables. The differenced VEqCM excluding \( I(-1) \) terms (denoted \( \Delta \text{VEqCM}_\beta \)) sets \( \Gamma_i = 0 \) \( \forall i \) to deliver
\[
\Delta x_t = (I + \alpha \beta') \Delta x_{t-1} + \tilde{\nu}_t, \tag{10}
\]
In practice, we include the difference of the output gap in \( \Delta \text{VEqCM}_\beta \), as this is less volatile than the second differences of the input prices.

3. We also compute a differenced single-equation equilibrium-correction model (\( \Delta \text{EqCM} \)) analogous to that of equation (9), where \( x_t \) is a scalar, and
\[\text{Observe that } E[\beta' x_t] = \mu \text{ is the mean of the cointegrating relation and } E[\Delta x_t] = \gamma \text{ is the unconditional growth rate of the system. } \alpha \text{ and } \beta \text{ are } n \times r \text{ matrices of rank } r.\]

\(^{10}\)
4. Next is a differenced single-equation equilibrium-correction model excluding double-differenced regressors ($\Delta \text{EqCM}_\beta$), which is based on the mark-up, output gap, and excess demand for labor.

5. A further robust forecasting device is a differenced device (DD) which differences the VAR, given by

$$\Delta x_t = \gamma + \eta_t,$$

which will be mis-specified unless the cointegrating rank is 0. The elimination of $\mu$ and $\alpha$ robustifies the model to changes in these parameters over the forecast horizon. In practice, a five-year rolling average is used rather than the in-sample inflation rate because of the regime changes over the in-sample period. To calculate $h$-step-ahead forecasts, direct projections are made based on a 20 period moving average up to $T - h$.

6. One of the more popular robust forecasting devices is the double-differenced device (DDD) which tracks past inflation,

$$\Delta x_t = \Delta x_{t-1} + \epsilon_t,$$

thereby eliminating $\mu$, $\gamma$, and $\alpha$, all the potentially damaging terms; see Hendry (2006). To compute $h$-step forecasts, direct projections are made:

$$\Delta \hat{x}_{T+h|T} = \Delta x_T.$$  

7. The smoothed difference model (SMD) attempts to capture a more rapid updating of the coefficients on the deterministic terms such that structural breaks are picked up faster. Assume a moving average of past actual growth rates given by $\tilde{\theta}_T = [1/(m + 1)] \sum_{i=0}^m \Delta x_{T-i}$, so that the forecast at $T + 1$ is $\tilde{x}_{T+1|T} = \tilde{x}_T$. Then

$$(m + 1)\tilde{\theta}_T = \sum_{i=0}^m \Delta x_{T-i} = \Delta x_T - \Delta x_{T-(m+1)} + (m + 1)\tilde{\theta}_{T-1},$$

so that

$$\tilde{\theta}_T = \tilde{\theta}_{T-1} + \frac{1}{m+1} \Delta \Delta x_{m+1},$$

reflecting aspects of Kalman filtering. Larger values of $m$ smooth intercept estimates but adapt more slowly. Setting $m = 3$, the SMD for quarterly data is given by

$$\Delta x_{T+i|T+i-1} = \frac{1}{4} \sum_{j=1}^4 \Delta x_{T+i-j} = \frac{1}{4} \Delta \Delta x_{T+i-1},$$

and direct multi-step forecasts are given by $\Delta \hat{x}_{T+h|T} = \frac{1}{4} \Delta \Delta x_T$. 

11
In comparison to the econometric models, we also examine a set of time-series models based on simple AR($p$) processes selected by the automatic procedure. The four issues we address using these models include: (i) the reduced information set for univariate models; (ii) direct versus iterated multi-step forecasts; (iii) the role of intercepts and trends; and (iv) the in-sample estimation period.

8. The first AR model computes direct forecasts and includes $p$ lags of the dependent variable (after data-based selection), an intercept, and a trend:

$$\Delta x_{t+h} = \mu + \sum_{i=0}^{p-1} \rho_i \Delta x_{t-i} + \delta t + \xi_t. \quad (15)$$

We label the model DAR$_\mu$ to denote a direct forecast from an AR model with deterministic terms.

9. The second AR model also computes direct forecasts but excludes the deterministic terms (DAR), again selecting the optimal lag length using PcGets:

$$\Delta x_{t+h} = \sum_{i=0}^{p-1} \rho_i \Delta x_{t-i} + \zeta_t. \quad (16)$$

10. The third AR model computes iterated multi-step forecasts based on an AR(1) model with intercept, estimated over the full-sample period (commencing in 1965q2). The forecasts are computed as

$$\Delta \hat{x}_{T+h|T} = \sum_{i=0}^{h-1} \rho^i \mu + \rho^h \Delta x_T. \quad (17)$$

We label the model IAR to represent an iterated AR forecast.

11. The final AR model computes iterated multi-step forecasts based on an AR(1) model with intercept, but estimated over a shorter sample period in an attempt to reduce the impact of structural breaks over the in-sample period. The models are estimated over the 20 observations prior to the forecast origin, and are denoted IAR$_{5yr}$.

12. The combination of forecasts is often found to outperform individual forecasts. Hence, we examine a pooled forecast (Pool) computed as an unweighted average. Hendry and Clements (2004) show that, in the presence of structural breaks, pooling may outperform any individual device, and averaging could dominate over estimating the weights for the forecast combination.
Finally, a trimmed pooled forecast is computed, which excludes from the unweighted average the worst forecast based on a mean-square forecast error criterion (Trim). Hence, this is an ex post forecast using out-of-sample information to determine which forecasts to trim. Theoretical analysis in Hendry and Reade (2004) shows that pooling can be improved substantively by eliminating badly performing models.

3.1 Direct Versus Iterated Forecasts

We compute 1-, 4-, and 8-step-ahead forecasts to coincide with the Bank of England’s decision-making process. Monetary policy decisions are based on 2-year-ahead forecasts, as the feed through effects of a change in policy are thought to take approximately two years.\(^{11}\)

In order to forecast more than 1-step-ahead in a multivariate framework, either an ‘iterated’ 1-step estimator or a direct \(h\)-step estimator can be used. The iterated 1-step forecast is most common, defined generally as

\[
\Delta \hat{x}_{T+h|T} = \hat{\Pi}^h \Delta x_T,
\]

where the average conditional error is

\[
E[(\Delta x_{T+h} - \Delta \hat{x}_{T+h|T}) | \Delta x_T] = (\Pi^h - E[\hat{\Pi}^h]) \Delta x_T,
\]

assuming that the estimators and the latest observations are approximately independent.

To compute 20 dynamic \(h\)-step forecasts for the VEqCM, the model is estimated over \(t = 1, \ldots, T\) and \(h\) dynamic forecasts are computed for \(T + 1, \ldots, T + h\). The model is then re-estimated over \(t = 1, \ldots, T+1\) and \(h\) dynamic forecasts are computed for \(T+2, \ldots, T+h+1\). This recursive process is repeated to obtain forecasts over the full forecast horizon. The iterated AR models are based on coefficient estimates over the in-sample period, i.e., no recursive estimation is undertaken.

The direct \(h\)-step estimator is non-recursive in that all information needed to derive an \(h\)-step forecast is available at time \(T\). The forecast is obtained by regressing \(\Delta x_T\) on the regressors lagged \(h\) periods. The estimator is given generally as

\[
\hat{\Pi}_h = \arg\min_{\Pi_h} \sum_{t=h}^{T} (\Delta x_t - \Pi_h \Delta x_{t-h}) (\Delta x_t - \Pi_h \Delta x_{t-h})'.
\]

Hence, in comparison to (18) and (19), the forecasts and average conditional errors are given

\(^{11}\)The Bank of England inflation fan charts are based on a 2-year forecast horizon; see Coyle (2001) for a discussion.
as

\[
\Delta \hat{x}_{T+h|T} = \tilde{\Pi}_h \Delta x_T, \quad \text{(21)}
\]

\[
E[(\Delta x_{T+h} - \Delta \hat{x}_{T+h|T}) | \Delta x_T] = (\Pi^h - E[\tilde{\Pi}_h]) \Delta x_T. \quad \text{(22)}
\]

The relative forecast accuracy of the two multi-step forecasts depends upon the accuracy of the estimators, \(\hat{\Pi}^h\) and \(\tilde{\Pi}_h\). Chevillon and Hendry (2005) find that the iterated 1-step forecasts are preferable when the model is well specified, for both stationary and \(I(1)\) processes. However, in the case of a mis-specified model for a non-stationary DGP, or if negative residual serial correlation or deterministic shocks are unaccounted for, direct multi-step estimation may lead to more accurate forecasts. The key factor is the size of the drift; see Chevillon (2006). As the drift increases, the benefits of the direct multi-step forecasts outweigh the iterated 1-step procedure. However, only the direct \(h\)-step forecast can be used in the single-equation framework. We also compare direct versus iterated AR forecasts.

3.2 Forecast Evaluation Criteria

Since ‘forecast-accuracy’ rankings depend on the measure used, we compute a range of statistics to evaluate the resulting forecasts. Define:

\[
e_{j,T+h|T} = \Delta x_{T+h} - \Delta \hat{x}_{j,T+h|T}, \quad \text{(23)}
\]

where \(\Delta \hat{x}_{j,T+h|T}\) denotes the \(h\)-step-ahead forecast for inflation from model \(j\) \((j = 1, \ldots, 15)\). \(n = 20\) forecasts are computed for each model \(j\) over each forecast period. The criteria calculated include:

1. Mean error (ME):
   \[
   \frac{1}{n} \sum_{i=1}^{n} e_{i,j,T+h|T}
   \]

2. Mean absolute error (MAE):
   \[
   \frac{1}{n} \sum_{i=1}^{n} |e_{i,j,T+h|T}|
   \]

3. Root mean square forecast error (RMSFE):
   \[
   \sqrt{\frac{1}{n} \sum_{i=1}^{n} e^{2}_{i,j,T+h|T}}
   \]

4. Determinant of the generalized forecast error second moment matrix (GFESM), reported as the square root (RGFESM) to maintain comparability with RMSFE.

For multi-step forecasts, the MSFE criterion precludes comparisons between different isomorphic representations of the same system because it is not invariant to non-singular, scale-preserving linear transformations (see Clements and Hendry 1993a,b). In order to obtain an
invariant measure, the covariances between different step-ahead forecast errors need to be taken into account. The determinant of the GFESM provides an invariant measure that accounts for such covariances by effectively measuring the volume of space around the forecast errors. The forecast error second moment matrix is obtained by stacking the forecast errors from all previous step-ahead forecasts:

\[
\Phi_{j,h} = E[E_j E_j'] \quad \text{where} \quad E_j' = (e_{j,T+1|T}, e_{j,T+2|T}, \ldots, e_{j,T+h|T}),
\]

and the statistic is calculated by obtaining the determinant of \( \Phi_{j,h} \). Observe that the diagonal of \( \Phi_{j,h} \) will contain the MSFEs.

In practice, we undertake a scaling to calculate % errors, given by \((100e_{j,T+i+k|T+i})/\hat{\sigma}\), where \(\hat{\sigma}\) is the equation standard error for the inflation equation in the VEqCM. A different \(\hat{\sigma}\) is used for the two forecast periods. The statistic is scaled by the inverse of the horizon to standardize the matrix size (e.g., \(|\Phi_{j,8}|^{\frac{1}{8}}\) for the 8-step measure). Furthermore, the square root of the statistic is reported, analogous to the RMSFE, and the 1-step RMSFE and 1-step RGFESM are equivalent; the one-to-one mapping is given by

\[
\text{RGFESM}_{1\text{-step}} = \hat{\sigma}^{-2}\text{[RMSFE}_{1\text{-step}}/100].
\]

The GFESM criterion is used to compare forecasts for the same horizon. Comparisons across horizons cannot be made because any non-unitary scaling can alter the inter-temporal comparisons due to the number of terms in the matrix. If the model parameters are known and the model is correctly specified, the generalized second moment of the 1-step forecast error determines the complete ranking, but under model mis-specification and parameter uncertainty, forecast uncertainty is not a monotonically increasing function of the forecast horizon, so rankings can alter as \(h\) increases (Clements and Hendry, 1998, Ch. 3 provide a discussion).

4 Quarterly Inflation Forecasts

To assess the forecast performance of the econometric models developed in Section 2, we undertake a forecast comparison exercise in which their partial \(ex \ ante\) forecasts are compared with the forecasts from alternative forecasting devices. Conventional forecasting theory from the seminal works of Box and Jenkins (1970), Klein (1971), and Granger and Newbold (1986) assumes that the best in-sample model is the best out-of-sample model, delivering the lowest MSFE matrix. Hence, the structural models developed should produce the best forecasts,
conditional on the strong assumption that they are the best in-sample models (slight evidence of non-congruency in the VEqCM does suggest that an alternative specification may be preferable). However, the empirical forecasts produced in this section show that there is no definitive ranking of forecasting models, and often models with little or no economic content can outperform well-specified econometric models. Such a finding is not new, and has provoked considerable debate historically; for some early example, see, *inter alia*, Nelson (1972), Naylor et al. (1972), and Cooper and Nelson (1975). A key difference here is that we have a well-tested theory as to why such an outcome can occur, namely, unmodeled and/or unanticipated location shifts.

4.1 Forecasting Results

Table 2 records the forecasting results for the 15 models examined over the 1-, 4-, and 8-step horizons. Figure 3 summarizes the information by recording the ME, MAE, and RMSFE for the 1- and 8-step forecasts in panels a–c and the determinant of the RGFESM statistic for the 4- and 8-step forecasts in panel d.

**TABLE 2 ABOUT HERE**

**FIGURE 3 ABOUT HERE**

The VEqCM and EqCM perform reasonably well, particularly in the latter period, with most of the realized outcomes lying within the $\pm 2\hat{\sigma}_f$ error bands. The models deliver much poorer forecasts over the earlier period. The 1- and 8-step forecasts are recorded in Figure 4; panels a and b record the VEqCM forecasts for the periods 1990q4–1995q3 and 1998q3–2003q2, respectively, and panels c and d record the same for the EqCM. Forecast error bars/bands are also recorded for the 1- and 8-step forecasts. All forecasts are recorded on the same axes for comparison. For the latter forecast period, both forecasts are biased upwards and the bias increases over the forecast horizon. There has been a structural shift towards lower and more stable inflation in the late 1990s, which may not be captured by models built over the entire in-sample period. Econometric models are not only susceptible to breaks over the forecast period, but shifts that have occurred previously that have not been fully modeled can also lead to large forecast errors. The bias is most evident over 2000–2001, when actual inflation was very low but the output gap suggests that excess demand pressures should be driving inflation upwards. Figure 2a records the estimated output gap used in the analysis. The figure indicates that there were excess demand pressures over 2000–2002. There are two plausible explanations:
first, the difficulty of estimating the output gap is most acute at the end of sample and a high uncertainty regarding estimates of the gap could be driving the forecast errors; secondly, the change in regime over the 1990s toward inflation targeting may have reduced the impact of the gap on inflation, implying that the coefficient on the gap is smaller over the forecast period. Both the single-equation and vector models do demonstrate a smaller coefficient on the gap for the models estimated up to 1998 compared to those estimated up to 1990.

The VEqCM and EqCM forecasts over the earlier sample period are substantially downward biased (except for the EqCM at longer horizons). This is striking given that inflation was falling over the forecast period. The models are over-estimating the decline, and the VEqCM is even forecasting a long period of deflation. Figure 2d records the cointegrating vector, interpreted as a mark-up, for the VEqCM model. The presence of the trend in the cointegrating vector (see equation (7)) is driving the forecast failure, emphasizing the problems of deterministic terms in forecasting. Differencing the cointegrating vector to remove the impact of the trend results in a substantial improvement in forecast performance. The mark-up variable for the single-equation models is recorded in Figure 2c, highlighting the differences due to the inclusion of a trend.

**FIGURE 4 ABOUT HERE**

The performance of the econometric models is poor in comparison to alternative models, ranked on either RMSFE or RGFESM criteria. While never the poorest models, they are consistently in the bottom third of the rankings. Table 3 records the rankings of each forecasting model based on RMSFE and RGFESM criteria. Theory would predict that the forecasts from the econometric models are better over quiescent periods where no breaks occur, and this is somewhat evident in Figure 4. However, while the measures of forecast accuracy are not comparable across forecast periods, we would expect the rankings of the econometric models to improve for the latter period. There is no evidence for this conjecture, suggesting that while the econometric models deliver better forecasts, the alternative forecasting models also deliver improved forecasts due to the reduction in mean and variance of inflation, such that the ranking of the models does not change dramatically. Mis-specification in the form of omitted variables or non-modeled in-sample breaks may be driving the poor performance of the econometric models over the latter forecast period.

**TABLE 3 ABOUT HERE**

To determine the impact of the increased information set used in the single-equation models,
we can compare the 1-step forecasts for the EqCM and VEqCM; the EqCM does marginally dominate the VEqCM. At longer horizons, the forecasts cannot be compared as the VEqCM computes dynamic forecasts whereas the EqCM computes direct forecasts. It is evident that at longer horizons direct estimation techniques are likely to yield poorer forecasts during stable periods: the increase in RMSFE over horizons for the EqCM in the latter period demonstrates this. Furthermore, at the 8-step horizon, the EqCM performs poorly on the RGFESM criterion, implying that there is little correlation between $h$-step forecasts due to the use of different model specifications at each forecast horizon. If very different models are selected at each horizon, the RGFESM will reflect the low correlation in forecast errors between horizons.

The differenced equilibrium-correction mechanisms tend to remove the bias in the forecasts (see Figure 3a), particularly if deterministic terms are driving the bias as in the VEqCM for the earlier forecast period. However, the forecast error variance can be large. Excluding the double differenced $I(-1)$ regressors results in a substantial improvement, both for the VEqCM and EqCM. This is most evident in the RGFESM criterion which reflects the covariances between forecast horizons because much of the forecast error variation between horizons is removed. The benefits of using the differenced equilibrium-correction mechanisms are most substantial over the earlier period when structural change was occurring, as predicted by theory. This is observed in the rankings in Table 3, where $\Delta$VEqCM$_\beta$ is consistently ranked higher over 1990q4–1995q3 than over 1998q3–2003q2 and on RGFESM criteria is ranked as the best forecast. Furthermore, these robust devices perform well at longer horizons, with the RMSFEs remaining fairly constant between the 1- and 8-step horizons. Figure 5 records the forecasts for the $\Delta$VEqCM$_\beta$ and $\Delta$EqCM$_\beta$ models. The $\Delta$VEqCM$_\beta$ forecasts are fairly constant as the double-differenced regressors are excluded. The $\Delta$EqCM$_\beta$ forecasts exhibit slightly more volatility due to the differenced output gap, excess demand for unemployment, and mark-up.

**FIGURE 5 ABOUT HERE**

Three other robust forecasting devices include the DD, DDD, and SMD. The DD provides excellent forecasts when the mean of inflation $[E(\Delta x_t) = \gamma]$ is unchanging, as in the latter forecast period. When there are shifts in $\gamma$, as in the earlier forecast period, the DD delivers poorer forecasts. The RGFESM ranking is very high as the correlation between $h$-step forecasts is near unity. If the forecasting rule was a constant (as opposed to a moving average up to $T-h$) the forecasts for $\Delta\hat{x}_{T+h|T+h-1}$ and $\Delta\hat{x}_{T+h|T+h-2}$ would be identical, resulting in
identical forecast errors such that the covariances between \( h \)-step forecasts would be unity, giving a very low RGFESM statistic. The DDD produces fairly poor forecasts over both periods. This is because the DDD is designed for step shifts as opposed to gradual trends like the decline in inflation over the earlier period. Also, quarterly inflation is fairly volatile and smoothed forecasts tend to outperform forecasts that match this volatility. The poor RGFESM reflects the low correlation between consecutive realizations given a stationary process (as \( \Delta \hat{x}_{T+h|T+h-1} = \Delta x_{T+h-1} \) and \( \Delta \hat{x}_{T+h|T+h-2} = \Delta x_{T+h-2} \)). The SMD is a ‘safe’ forecasting device like pooling. The forecasts do perform better in the stable period, mostly driven by the negligible bias.

Four alternative univariate time-series models are also considered. The results indicate that the inclusion of deterministic terms is detrimental to forecasting when breaks occur. The DAR is preferable to DAR\(\mu\) in the earlier forecast period (other than on RGFESM criteria for 8-steps), but in quiescent periods the forecasts are comparable, leading us to conclude that breaks in the deterministic terms may be driving the large forecast errors. Iterated forecasts are extremely dependent on parameter estimates of \( \mu \) and \( \rho \), and forecasts based on the full in-sample estimates are extremely poor (mostly ranked the worst forecasts). Both a substantial bias and variance drive the forecast failure. Estimating over a short sub-sample reduces the probability of breaks occurring in the in-sample period, and this has a dramatic effect on the forecast performance, most notably reducing the bias substantially; see Figure 3a. In the latter stable period, the IAR based on 5-year in-sample estimates performs extremely well, ranked first across all forecast horizons for both RMSFE and RGFESM. However, the forecasts are extremely reliant on correct coefficient estimates; in periods of change, the forecasts do not perform well as the model is not robust to such breaks. In periods of no change the iterated forecasts would seem to be preferable, whereas when there is structural change, direct forecasts may well outperform iterated forecasts. The RGFESM for the iterated forecasts is small. For longer horizons, the iterated forecasts converge to the long-run equilibrium, implying there is very little variation across forecast horizons resulting in high forecast error covariances.

Finally, we examine the performance of the pooled forecasts. Pooled forecasts perform well, both in quiescent periods but most notably in periods of change. The trimmed pooled forecast dominates the pooled forecast as the IAR forecast is an outlier which is excluded from the trimmed pooled forecast. In the more volatile period, the trimmed pooled forecast is the best on most criteria, and this is because the equilibrium-correction forecasts are biased downwards,
whereas all the univariate models are biased upwards. The trends in the VEqCM forecasts drive the downward bias, whereas the univariate time-series models are based on in-sample means that are higher than over the forecast period, resulting in upward biased forecasts. The pooled forecasts offset these two sets of forecasts, resulting in near unbiased forecasts. There are few costs associated with pooling when no one model dominates: even in the quiescent period, pooling is deemed preferable to using the econometric models.

To conclude, we find that robust forecasting devices do prove useful, both when there are breaks in the forecast period but also when the forecast period is stable with no evident breaks. The ‘best’ in-sample econometric models may not be the optimal forecasting models. Furthermore, the Clements and Hendry (1998, 1999) theory of forecasting provides explanations for all the empirical results obtained. In practice, computing a suite of forecasting models would be preferable, and is undertaken by many forecasting bodies such as the Bank of England.

4.2 Ranking of Forecasting Models

As well as assessing the models on various forecast criteria, we examine the models to see if there is any autocorrelation in which models perform best (and worst) over the forecast horizon, with a view to providing a more informed choice regarding which models to use. We examine which models deliver the closest forecast at each forecast observation to identify whether any patterns exist. Figure 6 records the number of individual forecast realizations in which the model is ranked best or worst based on MAE. The figure merges the results from both forecast periods and all three forecast horizons, resulting in a total of 120 observations. As the IAR dominates in terms of the worst forecast, we also calculate rankings excluding this model. The $\Delta\text{VEqCM}_\beta$, $\Delta\text{EqCM}_\beta$, pooled, and trimmed pooled forecast rarely deliver the worst forecast. No one model appears to consistently deliver the best forecast, again reinforcing the benefits of pooling.

FIGURE 6 ABOUT HERE

Table 4 provides more detail to Figure 6, ranking the best and worst forecasting models based on the absolute forecast error at every horizon. No model systematically outperforms the other models over the forecast period, and likewise, no model systematically has the largest absolute errors, other than the IAR. There is substantial fluctuation with regard to which models deliver the smallest absolute errors. Forecast accuracy appears to be rather volatile, with forecasting models being ranked both ‘best’ and ‘worst’ over the forecast period. Most
models tend to perform well on some occasions and poorly on others, and so we cannot draw conclusions regarding a systematic ranking, either during periods of change or during periods of stability.

TABLE 4 ABOUT HERE

Our initial hypothesis was that during quiescent periods, the econometric models would dominate and therefore these models alone should be used to forecast. However, during volatile periods in which breaks occur, the robust devices should outperform. This would lead to a forecasting rule based on using the econometric model until a break hits and then switch immediately to the robust devices. The evidence does not seem to support these conclusions: even in quiescent periods the robust devices perform well and there are benefits to pooling in either regime. Hence, a suite of forecasting models appears to be the most appropriate when forecasting economic data, given the caveat that the results are conditional on the specification of the structural model.

5 Forecasting Annual Inflation

Predictability is a necessary but not sufficient property for forecastability. A process \( y_t \) is defined as unpredictable with respect to an information set, \( I_{t-1} \), over \( T \) if

\[
D_{y_t}(y_t \mid I_{t-1}) = D_{y_t}(y_t) \forall t \in T.
\]  

(25)

Hence, if \( y_t \) is to be predictable, we require \( y_t = \phi(I_{t-1}, \nu_t) \), where \( \nu_t \) is unpredictable; see Clements and Hendry (2005). Predictability is relative to the information that is used. Predicting from a reduced information set, i.e., using \( J_{t-1} \subset I_{t-1} \) to predict \( y_t \), entails less accurate but unbiased predictions will be obtained providing \( J_{t-1} \) includes the \( \sigma \)-field of \( y_{t-k} \forall k \geq 1 \). Unpredictability is not invariant to the data frequency used, and so temporal disaggregation cannot lower the predictability of \( y_t \). This implies that as lower frequency data is a subset of higher frequency data, more accurate predictions should be obtained when forecasting annual inflation using quarterly data as opposed to annual data, although both forecasts should be unbiased. Increased measurement errors at the higher frequency could attenuate these implications. However, if there are location shifts the benefits of disaggregation are unclear as higher frequency data does not mitigate their effect. Section 5.1 develops a forecast error taxonomy for disaggregates to identify the impact of location shifts on time-disaggregated information,
and Section 5.2 assess the impact of time-disaggregation on the DDD. We use the following results to interpret the empirical evidence for annual inflation forecasts.

5.1 Disaggregating Forecasts Over Time

Let \( y_t (y_{\tau}) \) denote the variable of interest and \( z_t (z_{\tau}) \) denote a vector of \( n \) explanatory variables with elements \( z_{i,t} (z_{i,\tau}) \) where \( \tau = 2t \) for \( \tau = 1, \ldots, \Upsilon + H \) (e.g., \( \tau \) denotes semi-annual against annual frequency). It is essential to delineate what sort of variable \( y_{\tau} \) actually is, namely, a stock or price level, a flow or inflation measure. Given the context of our paper, we naturally take the last, in which case, all units need to be standardized at annualized rates of inflation for models to be comparable. Thus, we let \( y_t = (y_{\tau} + y_{\tau-1})/2 \), for even \( \tau \), be the time-aggregated variable. If the DGP is in terms of the time-aggregated measures, there can be no benefit from disaggregation, so we only consider the converse case.

We illustrate the analysis using as the DGP for the time-disaggregates an \( I(0) \) regression with constant parameters in-sample, and \( k \) unmodeled \( I(0) \) determinants \( z_{\tau-1} \):

\[
y_{\tau} = \mu + \rho y_{\tau-1} + \beta' z_{\tau-1} + \epsilon_{\tau} \quad \text{for } \tau = 1, \ldots, \Upsilon,
\]

where \( \epsilon_t \sim ID[0, \sigma_\epsilon^2] \), with a break at the forecast origin \( \Upsilon = 2T \) such that

\[
y_{\Upsilon+h} = \mu^* + \rho^* y_{\Upsilon+h-1} + (\beta^*)' z_{\Upsilon+h-1} + \epsilon_{\Upsilon+h} \quad \text{for } h = 1, \ldots, H,
\]

although the process stays \( I(0) \). Such a putative DGP reflects the prevalence of forecast failure in economics by its changing parameters, and allows an influence for both macroeconomic and policy variables. The aim of this section is to compare the forecast errors from the time disaggregates with directly forecasting the time aggregate. We construct a taxonomy of the time-disaggregate forecast errors using an estimated version of (26) when the forecast period is determined by (27), then combining with a two-step-ahead forecast to form the time-aggregate forecast. The analysis follows from the taxonomies in Clements and Hendry (1998), but considerably extended to allow for the time aggregation, although we only consider \( t = 1 \)-step-ahead forecasts when \((y_T : z_T)\) and \((y_{\Upsilon} : z_{\Upsilon})\) are known (forecast-origin uncertainty and multi-step-ahead forecasting would add further terms). Section 5.1.2 provides the corresponding taxonomy for the aggregate forecast errors, as well as a comparison between these. Because of parameter change, expected values also shift each period and so need explicit calculation for each point in time after the break.

---

12The empirical application compares annual and quarterly frequencies but the taxonomy implications hold for any even \( \tau \).
5.1.1 Post-Break Expected Values

First, taking expectations in (26) under in-sample stationarity,

\[ E[y_\tau] = \mu + \rho E[y_{\tau-1}] + \beta' E[z_{\tau-1}] = \mu + \rho \phi_y + \beta' \phi_z = \phi_y, \]

so that

\[ \phi_y = (1 - \rho)^{-1}(\mu + \beta' \phi_z), \tag{28} \]

and hence

\[ y_\tau - \phi_y = \rho(y_{\tau-1} - \phi_y) + \beta'(z_{\tau-1} - \phi_z) + \epsilon_\tau. \tag{29} \]

From (27), however,

\[ E[y_{\Upsilon+1}] = \mu^* + \rho^* E[y_{\Upsilon}] + (\beta^*)' E[z_{\Upsilon}] = \mu^* + \rho^* \phi_y + (\beta^*)' \phi_z, \tag{30} \]

so letting \( \phi^*_y = (1 - \rho^*)^{-1}(\mu^* + (\beta^*)' \phi^*_z) \) denote the post-break equilibrium mean, then from (30) we have

\[ y_{\Upsilon+1} - \phi^*_y = \rho^*(y_{\Upsilon} - \phi^*_y) + (\beta^*)'(z_{\Upsilon} - \phi^*_z) + \epsilon_{\Upsilon+1}. \tag{31} \]

Thus, the extent of the deviation of the outcome in the first forecast period, namely, \( y_{\Upsilon+1} - \phi^*_y \), depends on how far the forecast-origin values happened to deviate from their post-break equilibrium means. Conversely, (30) can be rewritten as

\[ E[y_{\Upsilon+1}] = \phi^*_y + \rho^*(\phi_y - \phi^*_y) + (\beta^*)'(\phi_z - \phi^*_z), \tag{32} \]

where the two final terms represent the deviation from the post-break equilibrium mean, \( \phi^*_y \), after one period.

Next, 2 periods later,

\[ y_{\Upsilon+2} = \phi^*_y + \rho^*(y_{\Upsilon+1} - \phi^*_y) + (\beta^*)'(z_{\Upsilon+1} - \phi^*_z) + \epsilon_{\Upsilon+2} \]
\[ = \phi^*_y + (\rho^*)^2(y_{\Upsilon} - \phi^*_y) + \rho^*(\beta^*)'(z_{\Upsilon} - \phi^*_z) + (\beta^*)'(z_{\Upsilon+1} - \phi^*_z) + \epsilon_{\Upsilon+2} + \rho^* \epsilon_{\Upsilon+1}. \tag{33} \]

Then, taking conditional expectations in (33), noting that \( E[z_{\Upsilon+1}] = \phi^*_z \),

\[ E[y_{\Upsilon+2} \mid y_{\Upsilon}, z_{\Upsilon}] = \phi^*_y + (\rho^*)^2(y_{\Upsilon} - \phi^*_y) + \rho^*(\beta^*)'(z_{\Upsilon} - \phi^*_z), \tag{34} \]

so that the unconditional expectation is

\[ E[y_{\Upsilon+2}] = \phi^*_y + (\rho^*)^2(\phi_y - \phi^*_y) + \rho^*(\beta^*)'(\phi_z - \phi^*_z) = \phi^*_y + \rho^*(E[y_{\Upsilon+1}] - \phi^*_y), \tag{35} \]

a pattern that is clear given (32).
5.1.2 Time-Aggregated Comparisons

Since the objective is to forecast the ‘annual’ (more generally, time-aggregated) outcome \( y_{T+1} \), we first derive its value from (31) and (33), namely,

\[
\begin{align*}
\bar{y}_{T+1} &= \phi_y^* + \rho^*(1 + \rho^*)(y_T - \phi_y^*) + (1 + \rho^*)(\beta^*)'(z_T - \phi_z^*) \\
&\quad - \frac{1}{2} \rho^*(1 + \rho^*)(y_{T-1} - \phi_y^*) \\
&\quad + \frac{1}{2} (\beta^*)'[(z_{T+1} - \phi_z^*) - (1 + \rho^*)(z_{T-1} - \phi_z^*)] \\
&\quad + \frac{1}{2} \epsilon_{T+2} + (1 + \rho^*)\epsilon_{T+1}.
\end{align*}
\]  

(36)

When only time-aggregated data are available, then the second and third lines must be omitted, so three important implications can be drawn from (36) when the post-break parameters are known. First, basing the time-aggregated forecast on two 1-step-ahead time-disaggregated forecasts from origins \( \Upsilon \) and then \( \Upsilon + 1 \) cannot worsen predictability, as then the only missing component is the innovation error \( \epsilon_{T+2} \). Secondly, using the combined 1- and 2-step-ahead time-disaggregated forecasts will worsen predictability relative to the updated forecast, as there will be an additional omitted term \( (z_{T+1}) \) and extra innovation errors \( (\epsilon_{T+1}) \). Thirdly, using only time-aggregated data will generally be worse still, as the further terms at \( \Upsilon - 1 \) will be omitted as well.

The next interesting case is when the in-sample parameters are known, but the break was not anticipated, so forecasts are based on

\[
\begin{align*}
\tilde{y}_{T+1|T} &= \phi_y + \rho(y_T - \phi_y) + \beta'(z_T - \phi_z), \\
\tilde{y}_{T+2|T} &= \phi_y + \rho^2(y_T - \phi_y) + \rho\beta'(z_T - \phi_z) + \beta'(z_{T+1} - \phi_z),
\end{align*}
\]

leading to

\[
\begin{align*}
\tilde{y}_{T+1|T} &= \phi_y + \rho(1 + \rho)(y_T - \phi_y) + (1 + \rho)\beta'(z_T - \phi_z) \\
&\quad - \frac{1}{2} \rho(1 + \rho)(y_{T-1} - \phi_y) \\
&\quad + \frac{1}{2} \beta'[(z_{T+1} - \phi_z) - (1 + \rho)(z_{T-1} - \phi_z)].
\end{align*}
\]  

(37)

When only time-aggregated data are available, the last two lines are unavailable to the investigator, who must perforce use just the first line. However, doing so will induce new parameters to reflect the omitted terms, and we denote that forecast by

\[
\hat{y}_{T+1|T} = \phi_y + \rho_p(1 + \rho_p)(y_T - \phi_y) + (1 + \rho_p)\beta'_p(z_T - \phi_z),
\]  

(38)
where the subscript $p$ in (38) denotes the probability limit of the estimated parameters, which thereby delivers the best in-sample fit.

We can now obtain two taxonomies of forecast errors, each of which decomposes the error between the ‘optimal forecast’ using in-sample information in (37) and then (38) from the realized outcome in (36), namely, $\tilde{y}_{T+1} - \tilde{y}_{T+1|T} = \tilde{u}_{T+1|T}$ and $\eta_{T+1} - \tilde{y}_{T+1|T} = \tilde{u}_{T+1|T}$, where the decomposition is into components deriving from both the break and time aggregation.\(^{13}\)

This leads to:

**Known-parameter forecast-error taxonomy from (37)**

\[
\tilde{u}_{T+1|T} = \begin{align*}
&[1 - \frac{1}{2}\rho^*(1 + \rho^*)](\phi_y^* - \phi_y) \\
&- \frac{1}{2}(2 + \rho^*)(\beta^*)'(\phi_z^* - \phi_z) \\
&+ [\rho^*(1 + \rho^*) - \rho(1 + \rho)](\bar{y}_T - \phi_y) \\
&+ \left[ (1 + \rho^*)(\beta^*)' - (1 + \rho)\beta' \right] (\bar{y}_T - \phi_z) \\
&+ \frac{1}{2}(\beta')' - \beta' \right] (\bar{y}_T + 1 - \phi_z) \\
&+ \frac{1}{2} [\epsilon_{Y+2} + (1 + \rho^*)\epsilon_{Y+1}] \\
\end{align*}
\]

(39) **equilibrium-mean shift** (ia) dynamic-slope change (ib) regressor-slope change (ic) aggregation mis-specification (iiia) lag regressor mis-specification (iib) regressor mis-specification (iiib) regressor-slope inconsistency (id) lag regressor mis-specification (iic) lag regressor mis-specification (iii) innovation error.

**Known-parameter forecast-error taxonomy from (38)**

\[
\tilde{u}_{T+1|T} = \begin{align*}
&[1 - \frac{1}{2}\rho^*(1 + \rho^*)](\phi_y^* - \phi_y) \\
&- \frac{1}{2}(2 + \rho^*)(\beta^*)'(\phi_z^* - \phi_z) \\
&+ [\rho^*(1 + \rho^*) - \rho(1 + \rho)](\bar{y}_T - \phi_y) \\
&+ \left[ (1 + \rho^*)(\beta^*)' - (1 + \rho)\beta' \right] (\bar{y}_T - \phi_z) \\
&+ \frac{1}{2}(\beta')' - \beta' \right] (\bar{y}_T + 1 - \phi_z) \\
&+ \frac{1}{2} [\epsilon_{Y+2} + (1 + \rho^*)\epsilon_{Y+1}] \\
\end{align*}
\]

(40) **equilibrium-mean shift** (Ia) dynamic-slope change (Ib) regressor-slope change (Id) dynamic-slope inconsistency (Ie) regressor-slope inconsistency (IIa) aggregation mis-specification (IIb) regressor mis-specification (IIc) lag regressor mis-specification (III) innovation error.

There are several surprising findings. First, the impacts of location shifts are identical between the two forecasts, (ia)≡(Ia) so the time-disaggregated information does not help offset that major problem. That component is in bold, as it is the only term which is non-zero on average. However, the break could almost certainly be detected much sooner using the time-disaggregated data. Secondly, the same applies to changes in all other parameters, (ib)≡(Ib) and (ic)≡(Ic), so again time disaggregation will not help the 2-period-ahead (i.e., one year) forecast. Thirdly, the impact of the innovation error is also identical despite time disaggregation, (iii)≡(III), which while it may appear counter-intuitive, makes sense on reflection as

\(^{13}\)We have separated out the equilibrium-mean shift in $z$ because time aggregation entails that some of its components are omitted, but it should be noted that when $z$ is correctly modeled, changes in $\phi_z$ do not induce forecast failure.
that is the component that is unpredictable on both information sets. Thus, all the differences
reside in the additional terms in (40), namely lines \((Id)\), \((Ie)\), and the differences between
\((iiia)-(iic)\) versus \((IIa)-(IIc)\); these are the information losses and parameter inconsistencies
deriving from time aggregation. Unfortunately, the combined effects of all these mistakes cannot
be signed, so one cannot prove that forecasts from the time-aggregated data will be worse.
Indeed, if terms like \((Ib)\), \((Id)\), etc. are combined to form, e.g.,
\[
[p^*(1 + p^*) - p_p(1 + p)](y_T - \phi_y),
\]
then situations in which the in-sample inconsistency lead to parameters that happened to be
closer to those after the break would favor the (mis-specified) time-aggregated forecasts.

5.1.3 Estimated Parameters
Allowing for parameters that are estimated from the in-sample data greatly complicates the
analysis but seems to add little insight, so we merely sketch the results. First, from (26), the
1-step-ahead model-based forecast becomes
\[
\hat{y}_{Y+1|Y} = \hat{\phi_y} + \hat{\beta}(z_Y - \hat{\phi}_z),
\]
so that from \(Y + 1\) onwards, the forecast errors \(\hat{e}_{Y+1|Y} = y_{Y+1} - \hat{y}_{Y+1|Y}\) are
\[
\hat{e}_{Y+1|Y} = \phi^*_y - \hat{\phi}_y + \rho^*(y_T - \phi^*_y) - \hat{\rho}(y_T - \hat{\phi}_y) + \beta^*(z_Y - \phi^*_z) - \beta^*(z_T - \hat{\phi}_z) + \epsilon_{Y+1}.
\]
Let \(\text{plim}_{T \to \infty} \hat{\rho} = \rho_p\), \(\text{plim}_{T \to \infty} \hat{\phi}_y = \phi_{y,p}\), etc.; also, \(\hat{\rho} = \rho_p + \Delta \rho, \hat{\phi}_y = \phi_{y,p} + \delta_y\) and similarly for other parameters \((\hat{\beta} = \beta_p + \Delta \beta\) and \(\hat{\phi}_z = \phi_{z,p} + \delta_z\), etc.). The forecast-error taxonomy
now follows by decomposing each term in (42) into its shift, mis-specification, and estimation
components as, e.g., in
\[
\phi^*_y - \hat{\phi}_y = (\phi^*_y - \phi_y) + (\phi_y - \phi_{y,p}) - \delta_y,
\]
where the objective is to isolate terms with zero and non-zero means, respectively. Thus, for
the autoregressive component,
\[
\rho^*(y_T - \phi^*_y) - \hat{\rho}(y_T - \hat{\phi}_y) = -(\rho^* - \rho)(\phi^*_y - \phi_y) - (\rho - \rho_p)(\phi^*_y - \phi_y) - \rho_p(\phi^*_y - \phi_y)
+ (\rho^* - \rho)(y_T - \phi_y) + (\rho - \rho_p)(y_T - \phi_y)
- \rho_p(\phi_y - \phi_{y,p})
- \Delta \rho(y_T - \phi_y) - \Delta \rho(\phi_y - \phi_{y,p})
+ \rho_p \delta_y + \Delta \rho \delta_y,
\]
and for the unmodeled variables,
\[
(\beta^*)'(z_T - \phi_z^*) - \beta'(z_T - \hat{\phi}_z) = -(\beta^* - \beta)'(\phi_z^* - \phi_z) - (\beta - \beta_p)'(\phi_z^* - \phi_z) - \beta'_p(\phi_z^* - \phi_z) \\
+ (\beta^* - \beta)'(z_T - \phi_z) + (\beta - \beta_p)'(z_T - \phi_z) - \beta'_p(\phi_z - \phi_{z,p}) \\
- \Delta'_\beta(z_T - \phi_z) - \Delta'_\beta(\phi_z - \phi_{z,p}) \\
+ \beta'_p \delta_z + \Delta'_\beta \delta_z.
\]

Collecting terms:
\[
\hat{\epsilon}_{T+1|Y} = (1 - \rho^*)(\phi_y^* - \phi_y) - (\beta^*)'(\phi_z^* - \phi_z) + (\rho - \rho)(y_T - \phi_y) + (\beta^* - \beta)'(z_T - \phi_z) + (1 - \rho)\beta'_p(\phi_z - \phi_{z,p}) \\
+ (\rho - \rho)(y_T - \phi_y) + (\beta - \beta_p)'(z_T - \phi_z) - \Delta_p(y_T - \phi_{y,p}) - \Delta'_\beta(z_T - \phi_{z,p}) \\
+ \Delta_p \delta_y + \Delta'_\beta \delta_z + \epsilon_{Y+1}.
\]

As can be seen, the non-estimated terms remain unchanged, and additional terms of $O_p(T^{-1/2})$ and $O_p(T^{-1})$ appear.

Next, to put the comparative forecasts on the same informational basis, we also need the 2-period-ahead forecast in order to forecast $y_{T+1} = (1/2)(y_{T+2} + y_{T+1})$. However, that 2-step-ahead model-based forecast requires some assumption about how $z_{T+1}$ would be forecast by the model proprietor, which might comprise a wide variety of approaches from extrapolation, a VAR, or an ‘off-line’ model. As we are merely illustrating the additional impact of parameter estimation, we use the random walk assumption, $z_{T+1} = z_T$, so that
\[
\hat{y}_{T+2|Y} = \hat{\phi}_y + \hat{\rho}^2(y_T - \hat{\phi}_y) + (1 + \hat{\rho})\beta'(z_T - \hat{\phi}_z),
\]
and hence the forecast errors $\hat{\epsilon}_{T+2|T} = y_{T+2} - \hat{y}_{T+2|T}$ are

$$
\hat{\epsilon}_{T+2|T} = \phi_y^* - \phi_y \\
+ (\rho^*)^2 (y_T - \phi_y^*) - \hat{\rho}^2 (y_T - \phi_y) \\
+ (1 + \rho^*) (\beta^* (z_T - \phi_z^*)) \langle z_T - \phi_z \rangle \\
+ \epsilon_{T+2} + \rho^* \epsilon_{T+1}.
$$

(45)

The same logic applies as for (43), but with the added complication of non-linear functions of estimated parameters; e.g.,

$$
\hat{\rho}^2 = (\rho_p + \Delta \rho)^2 = \rho_p^2 + 2\Delta \rho \rho_p + \delta^2 \approx \rho_p^2 + 2 \rho_p \Delta \rho
$$

(46)

and

$$
(1 + \hat{\rho}) \beta = (1 + \rho_p + \Delta \rho) (\beta_p + \Delta \beta) \approx (1 + \rho_p) \beta_p + (1 + \rho_p) \Delta \beta + \Delta \rho \beta_p,
$$

where second-order terms like $\Delta \rho \Delta \beta$ are neglected as $O_p(T^{-1})$.

A formal taxonomy of the overall error in forecasting $y_{T+1}$ can be built up by combining all the resulting terms. In the present situation, there are precisely the same number of parameters to be estimated in both aggregated and time-disaggregated models, namely, $n + 2$, so estimation does not involve any additional issues of parsimony. However, it is almost impossible to determine in general how the estimation variances compare between the two levels of time aggregation. At the two extremes, the time-disaggregated data could be infinitely more informative, if there were major within-year variations, but none between year; or add nothing if there was no within-year variation. Thus, we use the comparisons between (39) and (40) as the basis for interpreting our empirical findings.

A key feature of any taxonomy for an equilibrium-correction model like (26) is that the effects which impinge when forecasting immediately prior to a break for a point after the break also persist even when forecasting after the break has occurred; this is what induces systematic forecast failure following a location shift. Consequently, the results shown in (39) and (40) continue to apply when forecasting $y_{T+2}$ from $T+1$. There are procedures which can mitigate or help offset such persistent mis-forecasting, including updating parameter estimates and intercept corrections, but the former can, e.g., lose cointegration and the latter augments the forecast-error variance. It is this feature which distinguishes equilibrium-correction models (a large class which includes most widely used models) from robust forecasting devices, where behavior before a break is the same as an EqCM, but after is quite different. Hence, we now derive the taxonomy for a double-differenced device (DDD) when forecasting $y_{T+2}$ from $T+1$. 

28
5.2 DDD Taxonomy

Two periods after the break occurs at \( \Upsilon \):

\[
\begin{align*}
y_{\Upsilon+3} &= \phi_y^* + \rho^*(y_{\Upsilon+2} - \phi_y^*) + (\beta^*)'(z_{\Upsilon+2} - \phi_z^*) + \epsilon_{\Upsilon+3}, \\
y_{\Upsilon+4} &= \phi_y^* + \rho^*(y_{\Upsilon+3} - \phi_y^*) + (\beta^*)'(z_{\Upsilon+3} - \phi_z^*) + \epsilon_{\Upsilon+4}.
\end{align*}
\]

(47)

For \( \overline{y}_{T+2} = \frac{1}{2}(y_{T+4} + y_{T+3}) \), using (31) and (33):

\[
\begin{align*}
\overline{y}_{T+2} &= \phi_y^* + \rho^*(1 + \rho^*)(\overline{y}_{T+1} - \phi_y^*) + (1 + \rho^*)'(\beta^*)'(z_{T+1} - \phi_z^*) \\
&\quad - \frac{1}{2}\rho^*(1 + \rho^*)(y_{T+1} - \phi_y^*) \\
&\quad + \frac{1}{2}(\beta^*)'(z_{T+3} - \phi_z^*) - (1 + \rho^*)(z_{T+1} - \phi_z^*) \\
&\quad + \frac{1}{2}[\epsilon_{\Upsilon+4} + (1 + \rho^*)\epsilon_{T+3}].
\end{align*}
\]

(48)

The DDD forecast is given by

\[
\hat{y}_{T+2|T+1} = \overline{y}_{T+1},
\]

(49)

so from (36), the forecast error is \( \bar{u}_{T+2|T+1} = \overline{y}_{T+2} - \hat{y}_{T+2|T+1} \) (terms deriving from \( \overline{y}_{T+1} \) are shown in \( \cdot \)):

\[
\begin{align*}
\bar{u}_{T+2|T+1} &= \phi_y^* + \rho^*(1 + \rho^*)(\overline{y}_{T+1} - \phi_y^*) + (1 + \rho^*)'(\beta^*)'(z_{T+1} - \phi_z^*) \\
&\quad - \{(\phi_y^* + \rho^*(1 + \rho^*)(\overline{y}_{T} - \phi_y^*) + (1 + \rho^*)'(\beta^*)'(z_{T} - \phi_z^*)) \\
&\quad - \frac{1}{2}\rho^*(1 + \rho^*)(y_{T+1} - \phi_y^*) + (1 + \rho^*)(z_{T+1} - \phi_z^*) \\
&\quad + \frac{1}{2}(\beta^*)'(z_{T+3} - \phi_z^*) - (1 + \rho^*)(z_{T+1} - \phi_z^*) \\
&\quad - \frac{1}{2}(\beta^*)'(z_{T+1} - \phi_z^*) - (1 + \rho^*)(z_{T+1} - \phi_z^*)) \\
&\quad + \frac{1}{2}[\epsilon_{\Upsilon+4} + (1 + \rho^*)\epsilon_{T+3}] - \frac{1}{2}[\epsilon_{\Upsilon+2} + (1 + \rho^*)\epsilon_{T+3}].
\end{align*}
\]

(50)

The reason for the possible success of the DDD is immediately clear: the terms in \( \cdot \) contain precisely the same post-break parameters as those affecting \( \overline{y}_{T+2} \) and so cancel, yielding the forecast error,

\[
\begin{align*}
\bar{u}_{T+2|T+1} &= \rho^*(1 + \rho^*)\Delta\overline{y}_{T+1} + (1 + \rho^*)'(\beta^*)'(\Delta z_{T+1}) \\
&\quad - \frac{1}{2}\rho^*(1 + \rho^*)\Delta y_{T-1} \\
&\quad + \frac{1}{2}(\beta^*)'[\Delta_2 z_{T+1} - (1 + \rho^*)\Delta_2 z_{T-1}] \\
&\quad + \frac{1}{2}[\Delta_2 \epsilon_{T+2} + (1 + \rho^*)\Delta_2 \epsilon_{T+1}].
\end{align*}
\]

(51)
Since $E[\Delta y_{T+1}] = 0$ and $E[\Delta z_{T+1}] = E[\Delta_{2}z_{Y+1}] = E[\Delta_{2}z_{Y-1}] = 0$,

$$E[\pi_{T+2|T+1}] = 0,$$

so that the DDD forecast is unconditionally unbiased, regardless of whether the aggregated or disaggregated data are used. However, (50) also shows that the forecast-error variance is likely to be large, involving four innovation errors in the last line.

We can now compare the DDD to forecasting with the EqCM model for $y_{T+2}$ at $T + 1$ in (48):

$$y_{T+2} = \phi y + \rho(1 + \rho)(y_{T+1} - \phi_y) + (1 + \rho)(\beta'(z_{T+1} - \phi_z))$$

$$-\frac{1}{2}\rho(1 + \rho)(y_{T+1} - \phi_y)$$

$$+\frac{1}{2}(\beta')'(z_{T+3} - \phi_z) - (1 + \rho)(z_{T+1} - \phi_z)$$

$$+\frac{1}{2}[\epsilon_{T+4} + (1 + \rho)\epsilon_{T+3}].$$

When using aggregated data, only the first line will be available to base the forecast on, so the forecast is:

$$\bar{y}_{T+2|T+1} = \phi y + \rho(1 + \rho)(\bar{y}_{T+1} - \phi_y) + (1 + \rho)(\beta'z_{T+1} - \phi_z).$$

The forecast error will include both the effects of the unmodeled breaks and all the aggregation mis-specification:

$$\bar{y}_{T+2} - \bar{y}_{T+2|T+1} = (\phi^*_y - \phi_y)$$

$$+\rho^*(1 + \rho^*)(\bar{y}_{T+1} - \phi^*_y) - \rho(1 + \rho)(\bar{y}_{T+1} - \phi_y)$$

$$+(1 + \rho^*)(\beta'z_{T+1} - \phi^*_z) - (1 + \rho)(\beta'z_{T+1} - \phi_z)$$

$$-\frac{1}{2}\rho^*(1 + \rho^*)(y_{T+1} - \phi^*_y)$$

$$+\frac{1}{2}(\beta')'(z_{T+3} - \phi^*_z) - (1 + \rho^*)(z_{T+1} - \phi^*_z)$$

$$+\frac{1}{2}[\epsilon_{T+4} + (1 + \rho^*)\epsilon_{T+3}],$$

so that the forecasts will be biased when $\phi^*_y \neq \phi_y$. As noted in the previous section, the structural model does not mitigate the location shift, whereas the DDD forecast does after the location shift has occurred.

Finally, the earlier taxonomy showed that even if disaggregated information were available, the forecast error for the aggregated variable deriving from the break components would be the same, so the forecasts would again be biased.
6 Annual Forecasting Models

In order to test the theory we derive two models of annual inflation. We examine the case where \( \Delta_4 x_t = f_t(\mathcal{I}_{t-4}) + \nu_t \), for a model in which annual inflation is forecasted using a single-equation dynamic model based on \( \text{Gets} \) methodology. The first model we examine is the annual analogue of quarterly data, requiring information in fourth differences and lagged four periods only to emulate annual data. This is our lower frequency model, denoted AN.EqCM. The second model of annual inflation uses quarterly data, representing the higher frequency model, and is denoted QU.EqCM. Alternative forecasting devices are also compared, including a differenced EqCM of inflation using quarterly data excluding \( I(-1) \) double-differenced regressors (QU.\( \Delta \)EqCM\( _\beta \), analogous to (10)), a double-differenced device (DDD), a direct autoregressive model (DAR), an iterated autoregressive model (IAR), and a quarterly inflation model used to forecast 1-year-ahead inflation (QU.EqCM\( _{\text{in-sample}} \)). This forecasting rule is based on deriving a model of annual inflation using quarterly data and fixing the estimated coefficients to forecast 4-steps-ahead as opposed to 1-step-ahead. This is a variant of direct estimation techniques. If the DGP is given as

\[
\Delta_4 x_t = \gamma + \alpha(\beta' x_{t-1} - \mu) + \sum_{i=1}^{p} \Gamma_i \Delta x_{t-i} + \nu_t,
\]

we use the forecasting rule given by

\[
\Delta_4 x_{T+4|T} = \hat{\gamma} + \hat{\alpha}(\hat{\beta}' x_{T-1} - \hat{\mu}) + \sum_{i=0}^{p-1} \hat{\Gamma}_i \Delta x_{T-i},
\]

where the estimated coefficients are based on (55). While this is a mis-specified model, we can interpret the forecasting rule as

\[
\Delta_4 x_{T+4|T} = \Delta_4 x_T - \hat{\nu}_T,
\]

and so the forecasting rule is the DDD excluding the estimated error term. As well as protecting against breaks via the DDD component, all available information up to time \( T \) is used to develop the dominant, congruent in-sample model. However, to forecast 1-year-ahead we lose the quarterly information. If we develop an in-sample model based on annual data, relevant in-sample information is lost, but by using these coefficients lagged by one year we are making the implicit assumption that the exogenous variables have the same impact lagged one quarter as they do lagged one year. Finally, we compute a pooled forecast as the unweighted average of the seven forecasts for annual inflation.
Two forecast horizons are considered, 1- and 2-year-ahead forecasts. These correspond to the 4- and 8-step horizons based on quarterly data. As the models are single-equation models, we compute direct multi-step forecasts for the two horizons. The two forecast periods are identical to those for the quarterly models: 1990q4–1995q3 and 1998q3–2003q2 (Figure 1b records annual inflation). The models are evaluated on the criteria outlined in Section 3.2. Observe that there are only two consecutive $h$-step forecasts for annual inflation. Thus, the RGFESM statistic is based on a $(2 \times 2)$ matrix, $\Phi_{j,h}$, for the 2-year-ahead forecasts. The rankings for the 1-year-ahead forecasts are identical to the RMSFE criterion.

6.1 Models of Annual Inflation

6.1.1 Annual Analogue Model of Annual Inflation

The initial models of $\Delta_4p_t$ include either lags 4–7 or lags 8–11 of $y^d$, $U^d$, $\pi^*$, $s$, $\Delta_4 ppi$, $\Delta_4 rent$, $\Delta_4 imp$, $\Delta_4 oil$, $\Delta_4 e^*$, $\Delta_4 m4$, $\Delta_4 R_k$, $\Delta_4 R_{l}$, $\Delta_4 pw$, $\Delta_4 e_{e_{r,t-j}}$, and $\Delta_4 e^2_{e_{r,t-j}}$ for $j = 4, 5$ ($j = 8, 9$), and an intercept and trend. A set of indicator variables is also included. The mark-up ($\pi^*$) is calculated from an annual inflation model and is given by

$$\pi^*_t = (p_t - 0.58e_t^* - 0.34 pw_{L,t} - 0.07 ppi) + 4.12. \quad (58)$$

To derive a model of annual inflation in annual analogues using quarterly data, autocorrelation must be corrected for, as least squares will be inefficient. The first-order autocorrelation is of magnitude 0.7, with highly significant second- and third-order autocorrelation. As a rough guide, we can adjust the $t$-values that are selected in PcGets by a factor $\sqrt{(1 + \rho)/(1 - \rho)}$. Hence, we shall initially retain variables with a $t$-statistic greater than 4.76. To refine the selection using a more rigorous adjustment for autocorrelation, the model is then estimated in PcGive and tested down using HACSEs based on Andrews (1991).

The resulting model developed to forecast 1-year-ahead for the period 1998q3–2003q2 is reported in (59), with HACSEs reported in parentheses. The model fails autocorrelation, as expected, and the RESET test, which confirms model mis-specification. The model also fails the forecast Chow test, but out-of-sample fit is not a criterion on which to base in-sample model selection, as this would bias the forecasting results. The main determinants of inflation, including the output gap, excess demand for unemployment, mark-up, and world inflation are all significant and correctly signed. Input prices, including unit labor costs, rent, and oil prices,

---

14 This is implemented by adjusting the probabilities for the $t$-tests in the expert users strategy option in PcGets.

32
also play a role. There is a negative coefficient on lagged inflation due to omitting variables lagged $t - 1$ to $t - 3$.

$$\Delta_4 \hat{p}_t = \begin{equation}
-0.333 \Delta_4 \hat{p}_{t-5} + 0.801 y_{t-4}^d + 0.258 \Delta_4 c_{t-5}^e - 0.341 U_{t-6}^d \\
+ 0.092 \Delta_4 rent_{t-6} + 0.035 \Delta_4 oil_{t-4} + 0.640 \Delta_4 pw_{t-4} \\
-0.325 \pi_t^e - 0.058 I_{t4:1} - 0.033 I_{84:1} + 0.043 D_{79:3,80:2} + 0.026
\end{equation}
$$

$$R^2 = 0.951 \quad \hat{\sigma} = 1.202 \quad SC = -8.476$$

$$F_{ar}(5, 106) = 6.599^{**} \quad F_{arch}(4, 103) = 1.792 \quad F_{het}(19, 91) = 1.328 \quad \chi^2_{nd}(2) = 1.440$$

$$F_{reset}(1, 110) = 19.169^{**} \quad F_{Chow}(20, 111) = 2.469^{**} \quad T = 1967q4 - 1998q2.$$ 

The model to forecast 1-step-ahead over the earlier forecast period is given by

$$\Delta_4 \hat{p}_t = \begin{equation}
-0.254 \Delta_4 \hat{p}_{t-5} + 0.618 y_{t-4}^d + 0.201 \Delta_4 ppi_{t-4} - 0.564 U_{t-4}^d \\
+ 0.188 \Delta_4 \hat{r}_{s,t-6} + 0.039 \Delta_4 oil_{t-5} - 0.275 \Delta_4 imp_{t-4} + 0.894 \Delta_4 pw_{t-6} \\
-0.464 \pi_t^e + 0.041
\end{equation}
$$

$$R^2 = 0.916 \quad \hat{\sigma} = 1.592 \quad SC = -7.899$$

$$F_{ar}(5, 78) = 10.237^{**} \quad F_{arch}(4, 75) = 3.355^{**} \quad F_{het}(18, 64) = 3.242^{**} \quad \chi^2_{nd}(2) = 0.573$$

$$F_{reset}(1, 82) = 22.364^{**} \quad F_{Chow}(20, 83) = 0.710 \quad T = 1967q3 - 1990q3.$$ 

The model fails most diagnostic tests, as would be expected given the restricted information set in specifying the model. However, the model contains most explanations of inflation (although $imp$ has the incorrect sign, most likely due to collinearity with $ppi$), and no indicator variables are retained. The model is very similar to (59), although long-term interest rates are found to be significant. The 2-year-ahead forecasting models for both periods are reported in the Appendix.

### 6.2 Quarterly Model of Annual Inflation

Models of annual inflation using quarterly data are developed in order to compare the forecasts with the annual analogue models. The models are selected using the same methodology in which first differences are included in the GUM, reflecting the higher frequency data. Again, we only use lags dated $t-4, \ldots, t-7$ and $t-8, \ldots, t-11$ in order to forecast 1- and 2-year-ahead inflation. The 1-year-ahead models are reported in equations (61) and (62), with HACSEs in
tests of imposing the restriction of $\Delta p_{t-1}$ lagged one year. The models for both sample periods retained lagged dependent variables, and of annual inflation, the coefficients are fixed and the model is re-estimated on the variables mark-up. A set of indicator variables is also included. Having derived the quarterly model

The final models we examine are based on quarterly inflation. The initial models of $\Delta p_t$ include lags 1–4 of $\Delta p$, $y^d$, $U^d$, $\pi^*$, $s$, $\Delta ppi$, $\Delta rent$, $\Delta oil$, $\Delta c^*$, $\Delta m4$, $\Delta n$, $\Delta Rs$, $\Delta R_t$, $\Delta pw$, $\Delta e_{t-1}$, and $\Delta e_{t}^2$ for $j = 1, 2$, and an intercept and trend, where $\pi^*$ is the quarterly mark-up. A set of indicator variables is also included. Having derived the quarterly model of annual inflation, the coefficients are fixed and the model is re-estimated on the variables lagged one year. The models for both sample periods retained lagged dependent variables, and tests of imposing the restriction of $[\Delta p_{t-1} = \Delta p_{t-5} = \Delta p_{t-6} = 1]$ are accepted. The resulting

\[
\Delta_4 p_t = -0.736 \Delta p_{t-6} + 0.741 y_{t-4}^d + 0.679 \Delta c^*_{t-4} + 0.664 \Delta c^*_{t-5} - 0.455 U^d_{t-7} \\
+ 0.031 \Delta oil_{t-6} + 0.225 \Delta ppi_{t-5} + 1.623 \Delta pw_{t-4} - 0.352 \pi^*_{t-6} \\
- 0.033 I_{73:2} + 0.046 I_{79:3} + 0.033 \\
(0.204) \\
(0.095) \\
(0.110) \\
(0.067) \\
(0.012) \\
(0.332) \\
(0.072) \\
(0.005) \\
(0.004) \\
(0.004)
\]

\[R^2 = 0.919 \quad \hat{\sigma} = 1.540 \quad SC = -7.985\]

\[F_{ar}(5,109) = 7.397^{***} \quad F_{arch}(4,106) = 3.779^{**} \quad F_{het}(20,93) = 0.905 \quad \chi^2_{nd}(2) = 15.35^{**}\]

\[F_{reset}(1,113) = 20.59^{**} \quad F_{Chow}(20,114) = 1.004 \quad T = 1967q1 - 1998q2.\]

\[
\Delta_4 p_t = -0.835 \Delta p_{t-5} + 0.860 y_{t-4}^d + 0.727 \Delta c^*_{t-4} + 0.764 \Delta c^*_{t-5} - 0.531 U^d_{t-7} \\
- 0.266 \Delta imp_{t-6} + 0.359 \Delta ppi_{t-6} + 1.977 \Delta pw_{t-4} - 0.389 \pi^*_{t-6} \\
+ 0.050 I_{79:3} + 0.030 \\
(0.297) \\
(0.155) \\
(0.108) \\
(0.167) \\
(0.094) \\
(0.100) \\
(0.049) \\
(0.040) \\
(0.104) \\
(0.006) \\
(0.007)
\]

\[R^2 = 0.902 \quad \hat{\sigma} = 1.732 \quad SC = -7.707\]

\[F_{ar}(5,79) = 3.966^{***} \quad F_{arch}(4,76) = 2.613^{*} \quad F_{het}(19,64) = 1.682 \quad \chi^2_{nd}(2) = 7.247^{*}\]

\[F_{reset}(1,83) = 8.827^{**} \quad F_{Chow}(20,84) = 0.886 \quad T = 1967q1 - 1990q3.\]

Both models exhibit substantial mis-specification, evident in the failure of many diagnostic tests. The models are very similar, retaining the main determinants of inflation, including the output gap, excess demand for unemployment, mark-up, and world inflation. Similar to the annual analogue models, the lagged dependent variable is negative. The shorter estimation period retains import prices with a negative coefficient. Observe that the quarterly models have a poorer fit than the annual models over the same period, in contrast to our hypothesis regarding the models’ forecast performance.

6.3 Quarterly Inflation Model Used to Forecast 1-Year-Ahead Inflation

The final models we examine are based on quarterly inflation. The initial models of $\Delta_4 p_t$ include lags 1–4 of $\Delta p$, $y^d$, $U^d$, $\pi^*$, $s$, $\Delta ppi$, $\Delta rent$, $\Delta oil$, $\Delta c^*$, $\Delta m4$, $\Delta n$, $\Delta Rs$, $\Delta R_t$, $\Delta pw$, $\Delta e_{t-1}$, and $\Delta e_{t}^2$ for $j = 1, 2$, and an intercept and trend, where $\pi^*$ is the quarterly mark-up. A set of indicator variables is also included. Having derived the quarterly model of annual inflation, the coefficients are fixed and the model is re-estimated on the variables lagged one year. The models for both sample periods retained lagged dependent variables, and tests of imposing the restriction of $[\Delta p_{t-1} = \Delta p_{t-5} = \Delta p_{t-6} = 1]$ are accepted. The resulting
models for the 1-step-ahead forecasts are given in (63) and (64).

\[
\Delta p_t = 1.000\Delta_3 p_{t-4} + 0.204g^d_{t-4} + 0.131\Delta c^*_t - 0.118U^d_{t-6} \\
+ 0.667\Delta pu_{t-5} - 0.125\pi^*_t - 0.015D_{72:4,74:1} \\
+ 0.044D_{73:2,79:3} + 0.030D_{84:1,84:2} 
\]

(63)

\[
\hat{\sigma} = 3.235 \quad SC = -6.862
\]

\[
F_{ar}(5,121) = 53.58^{**} \quad F_{het}(18,107) = 20.34^{**} \quad \chi^2_{nd}(2) = 20.34^{**}
\]

\[
F_{Chow}(20,126) = 0.083 \quad T = 1967q1 - 1998q2.
\]

\[
\Delta p_t = 1.000\Delta_3 p_{t-4} + 0.324y^d_{t-4} - 0.076\Delta ppi^*_t - 0.150U^d_{t-4} \\
+ 1.247\Delta pu_{t-5} - 0.125\pi^*_t + 0.021D_{72:4,74:1} \\
+ 0.046D_{73:2,79:3} + 0.015D_{84:1,84:2} 
\]

(64)

\[
\hat{\sigma} = 3.710 \quad SC = -6.588
\]

\[
F_{ar}(5,86) = 43.76^{**} \quad F_{het}(18,72) = 7.639^{**} \quad \chi^2_{nd}(2) = 6.477^{*}
\]

\[
F_{Chow}(20,91) = 0.184 \quad T = 1968q1 - 1990q3.
\]

The models are a poor fit with very large standard errors and are clearly mis-specified, failing many diagnostics. However, the models need not be good in-sample models for them to perform well out-of-sample.

7 Annual Inflation Forecasts

Table 5 reports the 1- and 2-year-ahead ME, MAE, and RMSFE along with the 2-year-ahead RGFESM, for the annual inflation forecasts over the two periods examined. The forecast performance of the annual models is worse than that of the quarterly models, evident on MAE, RMSFE, and RGFESM criteria. The taxonomy demonstrates that gains from disaggregation must be due to information loss and parameter inconsistency derived from time disaggregation. These results suggest that there are gains to be made from moving to higher frequency data. Figure 7 records the 1- and 2-year-ahead forecasts for the annual and quarterly models for both forecast periods. The forecasts are on the same axes for comparison. The patterns
over the latter period are similar across models, with the 2-year-ahead forecasts consistently over-forecasting. For the earlier forecast period, the annual forecasts have a stronger downward trend, particularly for the 2-year forecasts, delivering poorer forecasts. The taxonomy demonstrates that location shifts and parameter changes have the same impact on the aggregated and disaggregated data. Thus, the differences in forecast performance driven by the trend component are a result of parameter estimation in the aggregated model. Hence, time disaggregation may or may not be beneficial and will also depend on the magnitude of measurement error. The optimal data frequency will depend on the availability and reliability of the data, but the impact of location shifts cannot be abated by moving to higher frequency data.

8 Conclusions

In this chapter, we have highlighted the difficulties in empirical forecasting in the face of structural breaks, also demonstrating that even in periods of relative stability it is difficult
to build econometric models that dominate other classes of forecasting rules. We find that deterministic terms can be problematic when forecasting. In particular, shifts in the mean of the cointegrating vector or the unconditional growth rate of the system will lead to forecast failure. However, it is often difficult to identify when breaks occur, and furthermore, any breaks in-sample must be fully modeled to avoid the model correcting to an incorrect mean. Highlighting the importance of deterministic terms in forecasting motivates the use of robust forecasting devices.

We find that most robust forecasting devices not only perform well in periods of structural change, but also in quiescent periods. This implies that the models can be used in any regime and not just employed when breaks occur. We find that univariate time-series models are not robust to breaks. They are dependent on the in-sample estimation period and the presence of deterministic terms, and as such, would not be recommended to include in a suite of forecasting models. Pooling is shown to be successful in many situations, and this could be seen as an ‘insurance policy’ forecast, avoiding selecting a particular forecasting method *ex ante*. The main benefits appear to be driven by forecast biases offsetting each other. Trimming is found to be useful if there are outliers among the forecasts.

We develop a forecast error taxonomy for time disaggregates to compare annual and quarterly inflation forecasts. The taxonomy delivers some surprising results: location shifts and parameter changes have the same impact regardless of whether the aggregated or disaggregated data are used. There is no clear benefit to using higher frequency data, as it depends on the information losses and parameter inconsistencies deriving from time aggregation. However, the empirical evidence does provide support for moving to higher frequency data, and in a real-time context, breaks can be picked up sooner. The taxonomy for the double-differenced device demonstrates that robust devices adapt rapidly to structural breaks, regardless of whether time-aggregated or disaggregated information is used.

The empirical example of U.K. quarterly and annual inflation demonstrates the success of adaptive forecasting devices outlined in the Clements and Hendry (1998, 1999) theory of forecasting. There has been a clear regime shift throughout the 1990s, in which inflation has been reduced to levels previously seen in the 1950s. This follows the exit from the ERM and the subsequent switch to inflation targeting. The move to Central Bank independence in 1997 is associated with low and stable inflation, and whether permanent or transitory, robust forecasting devices are picking up this behavior, whereas congruent in-sample models are still
correcting to the ‘old’ equilibrium. Hence, the application of the forecasting theory to inflation yields the results predicted by the theory, confirming its relevance when forecasting in the face of structural breaks.
Appendix

**TABLE 6 ABOUT HERE**

**Automatic Gets Selection**

Automatic general-to-specific model selection came to the fore when Hoover and Perez (1999) re-analyzed Lovell (1983) using an algorithm that searched many different paths. The experimental design aimed to select a few (0–5) regressors from a large set of 40 variables. They reversed Lovell’s conclusion that data mining was disastrous. Hendry and Krolzig (1999) improve on Hoover and Perez’s algorithm, developing the automatic model selection algorithm, PcGets. The econometric theory and methodology of the program are discussed in many publications, including Hendry and Krolzig (1999, 2001, 2004, 2004b, 2004c).

There are four stages in the PcGets algorithm, including estimation and testing of the GUM, the pre-search process, the multi-path search procedure, and finally post-search evaluation. Initially, the GUM is formulated based on theory and previous evidence, which is sufficiently general to nest the LDGP. A batch of mis-specification tests are performed on the GUM to ensure congruency, which means that the model matches the unknown LDGP in all measured aspects (see Bontemps and Mizon, 2003). If there is a failure of congruency—for example, a relevant variable may be omitted from the GUM—the search procedure would not commence and the econometrician would need to re-specify the GUM.

Pre-search reduction tests are undertaken at loose significance levels to remove highly insignificant variables. Then the multi-path search strategy commences from every feasible deletion, searching along each path using t-tests and F-tests until no more reductions can be made, checking the diagnostics at each reduction to ensure congruence. The algorithm will deliver a set of non-null terminal models that comprise the set of distinct, minimal, congruent reductions found along all search paths. When more than one such model is found, encompassing tests are used to select between the candidate congruent models. Each model is tested against the union of the models until a unique undominated congruent model is selected.\(^\text{15}\) Sub-sample reliability of the final model is evaluated by using overlapping sub-samples to formulate a reliability weighting, depending on whether a variable is significant in the two sub-samples and the full sample or not.

The properties of PcGets have been established by Krolzig and Hendry (2001) and Hendry

\(^{15}\)For the case in which a unique model does not emerge and the models are mutually encompassing and undominated, selection of the preferred model is made on the basis of information criteria.
and Krolzig (2003, 2005), who demonstrate that PcGets retains relevant variables close to the theoretical maximum given by the power of a single t-test on the relevant variable, and eliminates irrelevant variables at the chosen significance level, such that with \( n \) irrelevant variables and a significance level of \( \alpha \), \( n\alpha \) variables will be retained on average. Furthermore, extensive Monte Carlo evidence demonstrates that the equation standard error is close to that of the DGP, and while selected estimates are conditionally upward biased as the decision rule truncates the distribution to discard the mass near zero, coefficients have appropriate standard errors despite selection and can be bias corrected using simple formulae; see Hendry and Krolzig (2005). Finally, Campos et al. (2003) show PcGets model selection to be consistent.

**Forecasting Models**

EqCM model for quarterly inflation to forecast 4-steps-ahead over 1998q3–2003q2:

\[
\Delta p_t = 0.298 y_{t-4}^d - 0.120 U_{t-7}^d + 0.329 \Delta pw_{t-4} - 0.056 \pi_{t-4}^* + 0.086 \Delta ppi_{t-4} \\
+ 0.157 \Delta rent_{t-4} + 0.016 \Delta oil_{t-5} + 0.051 D_{73:2,79:3} + 0.025 D_{72:4,74:1} \\
+ 0.021 D_{84:1,84:2} + 0.007 \\
\]

\[
R^2 = 0.814 \quad \hat{\sigma} = 0.698 \quad SC = -9.597 
\]

\[
F_{ar}(5,110) = 0.884 \quad F_{arch}(4,107) = 0.990 \quad F_{het}(20,94) = 0.642 \quad \chi^2_{nd}(2) = 2.239 
\]

\[
F_{reset}(1,114) = 2.703 \quad F_{Chow}(20,115) = 0.990 \quad T = 1967q1 - 1998q2. 
\]

EqCM model for quarterly inflation to forecast 8-steps-ahead over 1998q3–2003q2:

\[
\Delta p_t = 0.568 y_{t-8}^d - 0.321 U_{t-11}^d + 0.345 \Delta pw_{t-11} - 0.083 \Delta ppi_{t-10} \\
+ 0.265 \Delta nt_{-8} + 0.218 \Delta mt_{-10} + 0.053 D_{73:2,79:3} + 0.014 \\
\]

\[
R^2 = 0.742 \quad \hat{\sigma} = 0.819 \quad SC = -9.358 
\]

\[
F_{ar}(5,106) = 1.921 \quad F_{arch}(4,103) = 0.648 \quad F_{het}(14,96) = 0.854 \quad \chi^2_{nd}(2) = 6.941^* 
\]

\[
F_{reset}(1,110) = 6.236^* \quad F_{Chow}(20,111) = 1.305 \quad T = 1968q1 - 1998q2. 
\]
EqCM model for quarterly inflation to forecast 4-steps-ahead over 1990q4–1995q3:

\[
\Delta p_t = 0.389 y_{t-5}^d - 0.254 U_{t-7}^d - 0.114 z_{t-4}^* + 0.158 R_{5, t-5} + 0.023 \Delta oil_{t-5}
\]
\[
+ 0.043 D_{73.279:3} + 0.016 D_{84.1, 84.2} + 0.017 (0.006) (0.005) (0.001)
\]
\[
R^2 = 0.787 \quad \hat{\sigma} = 0.768 \quad SC = -9.434
\]

\[
F_{ar}(5, 78) = 1.351 \quad F_{arch}(4, 75) = 0.761 \quad F_{het}(14, 68) = 1.766 \quad \chi^2_{nd}(2) = 1.363
\]

\[
F_{reset}(1, 82) = 0.948 \quad F_{Chow}(20, 83) = 1.122 \quad T = 1968q1 - 1990q3.
\]

EqCM model for quarterly inflation to forecast 8-steps-ahead over 1990q4–1995q3:

\[
\Delta p_t = 0.572 y_{t-8}^d - 0.327 U_{t-8}^d - 0.098 \Delta p_{10} + 0.050 D_{73.279:3} + 0.020 (0.005) \]
\[
R^2 = 0.670 \quad \hat{\sigma} = 0.939 \quad SC = -9.143
\]

\[
F_{ar}(5, 81) = 2.219 \quad F_{arch}(4, 78) = 0.372 \quad F_{het}(8, 77) = 1.478 \quad \chi^2_{nd}(2) = 5.466
\]

\[
F_{reset}(1, 85) = 0.624 \quad F_{Chow}(20, 86) = 1.398 \quad T = 1968q1 - 1990q3.
\]

Annual model of annual inflation to forecast 2-years-ahead over 1998q3–2003q2:

\[
\Delta_4 p_t = 1.782 y_{t-8}^d - 1.018 U_{t-8}^d + 0.253 \Delta_4 p_{10} + 0.511 \Delta_4 p_{10}
\]
\[
- 0.305 \Delta_4 m_{t-11} + 0.529 \Delta_4 c_{t-11} + 0.259 \Delta_4 rent_{t-8} + 0.189 \Delta_4 m_{t-8}
\]
\[
- 0.040 \Delta_4 oil_{t-9} + 0.035 (0.006)
\]
\[
R^2 = 0.828 \quad \hat{\sigma} = 2.240 \quad SC = -7.283
\]

\[
F_{ar}(5, 104) = 17.75** \quad F_{arch}(4, 101) = 6.853** \quad F_{het}(18, 90) = 2.977** \quad \chi^2_{nd}(2) = 6.593*
\]

\[
F_{reset}(1, 108) = 21.03** \quad F_{Chow}(20, 109) = 1.105 \quad T = 1968q4 - 1998q2.
\]

Annual model of annual inflation to forecast 2-years-ahead over 1990q4–1995q3:

\[
\Delta_4 p_t = 1.639 y_{t-8}^d + 0.518 y_{t-10}^d - 1.346 U_{t-8}^d + 0.593 \Delta_4 p_{10}
\]
\[
+ 0.257 \Delta_4 rent_{t-11} - 0.414 \Delta_4 c_{t-11} + 0.201 \Delta_4 m_{t-8}
\]
\[
+ 0.250 \Delta_4 m_{t-8} - 0.023 \Delta_4 oil_{t-8} + 0.050 (0.012)
\]
\[
R^2 = 0.812 \quad \hat{\sigma} = 2.354 \quad SC = -7.109
\]

\[
F_{ar}(5, 73) = 11.12** \quad F_{arch}(4, 70) = 3.265* \quad F_{het}(18, 59) = 1.795* \quad \chi^2_{nd}(2) = 1.490
\]

\[
F_{reset}(1, 77) = 5.804* \quad F_{Chow}(20, 78) = 1.319 \quad T = 1968q4 - 1990q3.
\]
Quarterly model of annual inflation to forecast 2-years-ahead over 1998q3–2003q2:

\[
\Delta_4 p_t = 1.032 y^d_{t-8} - 0.567 U^d_{t-8} - 0.139 \pi^*_t - 0.131 \Delta m_{t-10} \\
+ 0.826 \Delta rent_{t-8} - 0.811 s_t - 11 + 0.079 I_{79:3} - 0.057 I_{84:1} \\
\hat{\sigma} = 2.473 \quad SC = -7.152
\]

\[
F_{ar}(5,109) = 16.26^{**} \quad F_{arch}(4, 70) = 3.265^{**} \quad F_{het}(14, 99) = 5.958^{**} \quad \chi^2_{nd}(2) = 7.008^*
\]

Quarterly model of annual inflation to forecast 2-years-ahead over 1990q4–1995q3:

\[
\Delta_4 p_t = 1.540 y^d_{t-8} + 0.663 y^d_{t-10} - 1.040 U^d_{t-8} + 0.739 \Delta n_{t-8} - 0.188 \Delta ppi_{t-11} \\
- 0.624 s_{t-10} - 0.072 I_{73:2} - 0.047 I_{74:1} + 0.048 I_{79:3} + 0.048 \\
R^2 = 0.823 \quad \hat{\sigma} = 2.292 \quad SC = -7.172
\]

\[
F_{ar}(5,76) = 10.71^{**} = 16.26 \quad F_{arch}(4, 73) = 2.924^{**} \quad F_{het}(15, 65) = 4.374^{**} \quad \chi^2_{nd}(2) = 8.891^*
\]

\[
F_{reset}(1,80) = 15.84^{**} \quad F_{Chow}(20, 81) = 0.696 \quad T = 1968q1 - 1990q3.
\]
Acknowledgments

We are indebted to participants at the 2004 Journal of Applied Econometrics Annual Lecture, 2006 International Symposium of Forecasting, SITE conference on Forecasting at Stanford University, Forecasting in the Presence of Structural Breaks conference at Saint Louis University, and 2006 European Economic Association meeting, as well as an anonymous referee, for helpful comments and suggestions. Financial support from the Economic and Social Research Council under grant RES 051 270035 is gratefully acknowledged.
References


**Titles for Figures**

1. Quarterly and annual U.K. inflation (GDP deflator)

2. The output gap, excess demand for labour, mark-up, and cointegrating vector for the model estimated over 1966q2–1990q3

3. Mean errors, mean absolute errors, and root mean square forecast errors for the 1- and 8-step horizons and the square root of the determinant of GFESM for the 4- and 8-step horizons

4. 1- and 8-step forecasts for the VEqCM and EqCM models with $\pm 2\hat{\sigma}_f$ error bars/bands for 1990q4–1995q3 and 1998q3–2003q2

5. Annual inflation forecasts from the differenced equilibrium-correction models excluding double-differenced regressors (error bars/bands do not correct for the autocorrelation induced by differencing)

6. The number of observations in which the forecasting model produced the best and worst forecasts across all horizons and forecast periods

7. Annual inflation forecasts from the annual and quarterly single-equation EqCMs

8. Annual inflation forecasts from alternative forecasting devices for the period 1990q4–1995q3

9. Annual inflation forecasts from alternative forecasting devices for the period 1998q3–2003q2
Table 1: Descriptive statistics of annual and quarterly inflation

<table>
<thead>
<tr>
<th></th>
<th>$\Delta p$</th>
<th>$\Delta 4p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St. dev.</td>
</tr>
<tr>
<td>Full sample</td>
<td>0.017</td>
<td>0.015</td>
</tr>
<tr>
<td>1965q1–1990q3</td>
<td>0.022</td>
<td>0.016</td>
</tr>
<tr>
<td>(1): 1990q4–1995q3</td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>(2): 1998q3–2003q2</td>
<td>0.006</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Notes: (1) and (2) denote the two forecast periods. St. dev is the standard deviation of the realisations and Range denotes the maximum observation minus the minimum observation over the period (reported as a percentage).
Table 2: Summary of quarterly inflation forecasting results

<table>
<thead>
<tr>
<th></th>
<th>ME 1-step</th>
<th>ME 4-step</th>
<th>ME 8-step</th>
<th>MAE 1-step</th>
<th>MAE 4-step</th>
<th>MAE 8-step</th>
<th>RMSE 1-step</th>
<th>RMSE 4-step</th>
<th>RMSE 8-step</th>
<th>R^2 1-step</th>
<th>R^2 4-step</th>
<th>R^2 8-step</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 1998q3–2003q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆VEqCM</td>
<td>0.010</td>
<td>0.243</td>
<td>0.512</td>
<td>0.448</td>
<td>0.499</td>
<td>0.655</td>
<td>0.564</td>
<td>0.645</td>
<td>0.762</td>
<td>0.788</td>
<td>0.489</td>
<td>0.405</td>
</tr>
<tr>
<td>∆EqCM</td>
<td>-0.083</td>
<td>-0.380</td>
<td>-0.681</td>
<td>0.411</td>
<td>0.588</td>
<td>0.861</td>
<td>0.499</td>
<td>0.740</td>
<td>1.043</td>
<td>0.698</td>
<td>0.421</td>
<td>0.442</td>
</tr>
<tr>
<td>∆VEqCM</td>
<td>-0.035</td>
<td>0.062</td>
<td>-0.009</td>
<td>0.485</td>
<td>0.439</td>
<td>0.364</td>
<td>0.605</td>
<td>0.486</td>
<td>0.440</td>
<td>0.845</td>
<td>0.468</td>
<td>0.249</td>
</tr>
<tr>
<td>∆EqCM</td>
<td>0.017</td>
<td>0.021</td>
<td>0.025</td>
<td>0.335</td>
<td>0.341</td>
<td>0.336</td>
<td>0.409</td>
<td>0.411</td>
<td>0.408</td>
<td>0.572</td>
<td>0.124</td>
<td>0.031</td>
</tr>
<tr>
<td>∆VEqCM</td>
<td>-0.040</td>
<td>-0.186</td>
<td>-0.049</td>
<td>0.471</td>
<td>0.601</td>
<td>0.861</td>
<td>0.573</td>
<td>0.701</td>
<td>1.084</td>
<td>0.802</td>
<td>0.509</td>
<td>0.528</td>
</tr>
<tr>
<td>∆EqCM</td>
<td>-0.030</td>
<td>-0.189</td>
<td>-0.084</td>
<td>0.345</td>
<td>0.366</td>
<td>0.343</td>
<td>0.413</td>
<td>0.441</td>
<td>0.423</td>
<td>0.577</td>
<td>0.249</td>
<td>0.268</td>
</tr>
<tr>
<td>DD</td>
<td>-0.029</td>
<td>-0.034</td>
<td>-0.054</td>
<td>0.341</td>
<td>0.339</td>
<td><strong>0.320</strong></td>
<td>0.405</td>
<td>0.396</td>
<td>0.385</td>
<td>0.566</td>
<td>0.086</td>
<td>0.055</td>
</tr>
<tr>
<td>SMD</td>
<td>-0.025</td>
<td><strong>0.013</strong></td>
<td>0.039</td>
<td>0.467</td>
<td>0.480</td>
<td>0.528</td>
<td>0.553</td>
<td>0.485</td>
<td>0.639</td>
<td>0.773</td>
<td>0.612</td>
<td>0.622</td>
</tr>
<tr>
<td>DAR</td>
<td>0.097</td>
<td>0.160</td>
<td>0.145</td>
<td>0.338</td>
<td>0.386</td>
<td>0.397</td>
<td>0.414</td>
<td>0.435</td>
<td>0.460</td>
<td>0.582</td>
<td>0.264</td>
<td>0.206</td>
</tr>
<tr>
<td>IAR</td>
<td>0.238</td>
<td>0.080</td>
<td>0.124</td>
<td>0.340</td>
<td>0.343</td>
<td>0.400</td>
<td>0.416</td>
<td>0.434</td>
<td>0.462</td>
<td>0.582</td>
<td>0.340</td>
<td>0.281</td>
</tr>
<tr>
<td>IAR5</td>
<td>0.511</td>
<td>-1.144</td>
<td>-1.294</td>
<td>0.593</td>
<td>1.144</td>
<td>1.294</td>
<td>0.705</td>
<td>1.205</td>
<td>1.318</td>
<td>0.985</td>
<td>0.389</td>
<td>0.121</td>
</tr>
<tr>
<td>Pool</td>
<td>-0.059</td>
<td>-0.145</td>
<td>-0.185</td>
<td>0.344</td>
<td>0.354</td>
<td>0.389</td>
<td>0.415</td>
<td>0.424</td>
<td>0.474</td>
<td>0.581</td>
<td>0.217</td>
<td>0.161</td>
</tr>
<tr>
<td>Trim</td>
<td>-0.019</td>
<td>-0.061</td>
<td>-0.093</td>
<td>0.340</td>
<td>0.357</td>
<td>0.375</td>
<td>0.413</td>
<td>0.405</td>
<td>0.452</td>
<td>0.578</td>
<td>0.206</td>
<td>0.156</td>
</tr>
<tr>
<td>B. 1990q4–1995q3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆VEqCM</td>
<td>0.742</td>
<td>0.741</td>
<td>0.911</td>
<td>0.800</td>
<td>0.791</td>
<td>0.932</td>
<td>0.865</td>
<td>0.945</td>
<td>1.101</td>
<td>1.210</td>
<td>0.448</td>
<td>0.396</td>
</tr>
<tr>
<td>∆EqCM</td>
<td>0.583</td>
<td>-0.017</td>
<td><strong>0.122</strong></td>
<td>0.702</td>
<td>0.695</td>
<td>1.099</td>
<td>0.856</td>
<td>0.845</td>
<td>1.264</td>
<td>1.197</td>
<td>0.683</td>
<td>0.625</td>
</tr>
<tr>
<td>∆VEqCM</td>
<td>0.165</td>
<td>0.037</td>
<td>0.179</td>
<td>0.466</td>
<td>0.530</td>
<td>0.584</td>
<td>0.554</td>
<td>0.668</td>
<td>0.710</td>
<td>0.775</td>
<td>0.408</td>
<td>0.296</td>
</tr>
<tr>
<td>∆EqCM</td>
<td>0.087</td>
<td>0.068</td>
<td>0.068</td>
<td>0.444</td>
<td>0.525</td>
<td>0.522</td>
<td>0.520</td>
<td>0.619</td>
<td>0.617</td>
<td>0.727</td>
<td>0.148</td>
<td>0.016</td>
</tr>
<tr>
<td>∆VEqCM</td>
<td>0.090</td>
<td>0.061</td>
<td>0.189</td>
<td>0.593</td>
<td>0.720</td>
<td>0.541</td>
<td>0.731</td>
<td>0.898</td>
<td>0.679</td>
<td>1.022</td>
<td>0.730</td>
<td>0.621</td>
</tr>
<tr>
<td>∆EqCM</td>
<td>0.039</td>
<td>0.073</td>
<td>0.183</td>
<td>0.553</td>
<td>0.560</td>
<td>0.524</td>
<td>0.653</td>
<td>0.627</td>
<td>0.635</td>
<td>0.913</td>
<td>0.542</td>
<td>0.350</td>
</tr>
<tr>
<td>DD</td>
<td>-0.586</td>
<td>-0.649</td>
<td>-0.674</td>
<td>0.658</td>
<td>0.711</td>
<td>0.756</td>
<td>0.796</td>
<td>0.865</td>
<td>0.911</td>
<td>1.113</td>
<td>0.114</td>
<td>0.079</td>
</tr>
<tr>
<td>SMD</td>
<td>-0.190</td>
<td>-0.378</td>
<td>-0.647</td>
<td>0.485</td>
<td>0.543</td>
<td>0.723</td>
<td>0.639</td>
<td>0.673</td>
<td>0.920</td>
<td>0.894</td>
<td>0.414</td>
<td>0.336</td>
</tr>
<tr>
<td>DAR</td>
<td>-0.586</td>
<td>-1.064</td>
<td>-1.348</td>
<td>0.643</td>
<td>1.064</td>
<td>1.348</td>
<td>0.789</td>
<td>1.206</td>
<td>1.471</td>
<td>1.103</td>
<td>0.540</td>
<td>0.348</td>
</tr>
<tr>
<td>IAR</td>
<td>-0.114</td>
<td>-0.301</td>
<td>-0.376</td>
<td>0.418</td>
<td>0.646</td>
<td>0.570</td>
<td>0.596</td>
<td>0.754</td>
<td>0.773</td>
<td>0.834</td>
<td>0.526</td>
<td>0.496</td>
</tr>
<tr>
<td>Pool</td>
<td>-0.764</td>
<td>-1.450</td>
<td>-1.529</td>
<td>0.824</td>
<td>1.450</td>
<td>1.529</td>
<td>0.970</td>
<td>1.530</td>
<td>1.636</td>
<td>1.356</td>
<td>0.456</td>
<td>0.111</td>
</tr>
<tr>
<td>Trim</td>
<td>-0.088</td>
<td>-0.297</td>
<td>-0.329</td>
<td>0.482</td>
<td>0.463</td>
<td>0.478</td>
<td>0.480</td>
<td>0.593</td>
<td>0.607</td>
<td>0.671</td>
<td>0.309</td>
<td>0.190</td>
</tr>
</tbody>
</table>

Notes: Results for 1-, 4-, and 8-step-ahead forecasts over 1998q3–2003q2 and 1990q4–1995q3. Figures reported as percentages, with the best forecasts highlighted in **bold** and the worst forecasts highlighted in italics.
### Table 3: Forecast rankings based on RMSFE and RGFESM criteria

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-step RM</td>
<td>4-step RM GF</td>
<td>8-step RM GF</td>
<td>1-step RM</td>
<td>4-step RM GF</td>
<td>8-step RM GF</td>
</tr>
<tr>
<td>VEqCM</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>EqCM</td>
<td>10</td>
<td>14</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>δVEqCM</td>
<td>14</td>
<td>11</td>
<td>12</td>
<td>5</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>δEqCM</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>DD</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>DDD</td>
<td>11</td>
<td>10</td>
<td>15</td>
<td>11</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>SMD</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>DAR&lt;sub&gt;u&lt;/sub&gt;</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>DAR</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>IAR&lt;sub&gt;5yr&lt;/sub&gt;</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td>15</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>IAR</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Pool</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Trim</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: RM denotes RMSFE and GF denotes RGFESM criteria. Numbers reported are the model rankings, where 1 is the best forecasting model (in **bold**) and 15 is the worst forecasting model (in *italics*).
Table 4: Model rankings at each forecast observation based on MAE

<table>
<thead>
<tr>
<th></th>
<th>1-step</th>
<th>4-step</th>
<th>8-step</th>
<th>1-step</th>
<th>4-step</th>
<th>8-step</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990q4</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>DDD</td>
<td>IAR</td>
<td>EqCM</td>
</tr>
<tr>
<td>1991q1</td>
<td>DAR</td>
<td>DAR</td>
<td>DAR</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>∆VEqCM</td>
</tr>
<tr>
<td>1991q2</td>
<td>DAR</td>
<td>IAR</td>
<td>DAR</td>
<td>∆VEqCM</td>
<td>∆EqCM</td>
<td>∆VEqCM</td>
</tr>
<tr>
<td>1991q3</td>
<td>DAR</td>
<td>Trim</td>
<td>Pool</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
</tr>
<tr>
<td>1991q4</td>
<td>SMD</td>
<td>Pool</td>
<td>DAR</td>
<td>EqCM</td>
<td>EqCM</td>
<td>VEqCM</td>
</tr>
<tr>
<td>1992q1</td>
<td>Trim</td>
<td>∆VEqCM</td>
<td>DDD</td>
<td>IAR</td>
<td>∆EqCM</td>
<td>EqCM</td>
</tr>
<tr>
<td>1992q2</td>
<td>IAR</td>
<td>Pool</td>
<td>Trim</td>
<td>VEqCM</td>
<td>∆EqCM</td>
<td>VEqCM</td>
</tr>
<tr>
<td>1992q3</td>
<td>EqCM</td>
<td>VEqCM</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
<td>DDD</td>
</tr>
<tr>
<td>1992q4</td>
<td>EqCM</td>
<td>∆EqCM</td>
<td>DDD</td>
<td>DD</td>
<td>IAR</td>
<td>DAR</td>
</tr>
<tr>
<td>1993q1</td>
<td>IAR</td>
<td>DAR</td>
<td>Pool</td>
<td>EqCM</td>
<td>VEqCM</td>
<td>VEqCM</td>
</tr>
<tr>
<td>1993q2</td>
<td>Trim</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>EqCM</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1993q3</td>
<td>∆VEqCM</td>
<td>∆EqCM</td>
<td>Pool</td>
<td>EqCM</td>
<td>VEqCM</td>
<td>VEqCM</td>
</tr>
<tr>
<td>1993q4</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1994q1</td>
<td>DDD</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1994q2</td>
<td>EqCM</td>
<td>EqCM</td>
<td>∆EqCM</td>
<td>DD</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1994q3</td>
<td>DAR</td>
<td>VEqCM</td>
<td>DDD</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1994q4</td>
<td>∆VEqCM</td>
<td>DD</td>
<td>DD</td>
<td>VEqCM</td>
<td>EqCM</td>
<td>EqCM</td>
</tr>
<tr>
<td>1995q1</td>
<td>Trim</td>
<td>∆VEqCM</td>
<td>∆VEqCM</td>
<td>VEqCM</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1995q2</td>
<td>SMD</td>
<td>SMD</td>
<td>SMD</td>
<td>IAR</td>
<td>IAR</td>
<td>EqCM</td>
</tr>
<tr>
<td>1995q3</td>
<td>Pool</td>
<td>Trim</td>
<td>∆VEqCM</td>
<td>VEqCM</td>
<td>IAR</td>
<td>EqCM</td>
</tr>
<tr>
<td>1998q3</td>
<td>∆VEqCM</td>
<td>∆EqCM</td>
<td>SMD</td>
<td>EqCM</td>
<td>VEqCM</td>
<td>∆EqCM</td>
</tr>
<tr>
<td>1998q4</td>
<td>VEqCM</td>
<td>EqCM</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1999q1</td>
<td>DDD</td>
<td>DAR</td>
<td>DAR</td>
<td>IAR</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1999q2</td>
<td>IAR</td>
<td>DAR</td>
<td>∆EqCM</td>
<td>VEqCM</td>
<td>EqCM</td>
<td>IAR</td>
</tr>
<tr>
<td>1999q3</td>
<td>DAR</td>
<td>∆VEqCM</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>1999q4</td>
<td>DAR</td>
<td>DDD</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>2000q1</td>
<td>DD</td>
<td>∆VEqCM</td>
<td>∆EqCM</td>
<td>DDD</td>
<td>IAR</td>
<td>∆EqCM</td>
</tr>
<tr>
<td>2000q2</td>
<td>DAR</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
<td>∆EqCM</td>
</tr>
<tr>
<td>2000q3</td>
<td>SMD</td>
<td>DAR</td>
<td>DAR</td>
<td>VEqCM</td>
<td>∆EqCM</td>
<td>IAR</td>
</tr>
<tr>
<td>2000q4</td>
<td>DAR</td>
<td>DAR</td>
<td>DDD</td>
<td>VEqCM</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>2001q1</td>
<td>VEqCM</td>
<td>VEqCM</td>
<td>EqCM</td>
<td>DAR</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
</tr>
<tr>
<td>2001q2</td>
<td>Pool</td>
<td>Pool</td>
<td>DD</td>
<td>IAR</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>2001q3</td>
<td>DAR</td>
<td>SMD</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>EqCM</td>
<td>IAR</td>
</tr>
<tr>
<td>2001q4</td>
<td>VEqCM</td>
<td>VEqCM</td>
<td>∆EqCM</td>
<td>DDD</td>
<td>DAR</td>
<td>DDD</td>
</tr>
<tr>
<td>2002q1</td>
<td>IAR</td>
<td>VEqCM</td>
<td>VEqCM</td>
<td>∆EqCM</td>
<td>SMD</td>
<td>∆EqCM</td>
</tr>
<tr>
<td>2002q2</td>
<td>∆VEqCM</td>
<td>DAR</td>
<td>∆EqCM</td>
<td>IAR</td>
<td>IAR</td>
<td>IAR</td>
</tr>
<tr>
<td>2002q3</td>
<td>DAR</td>
<td>∆EqCM</td>
<td>∆EqCM</td>
<td>DDD</td>
<td>DDD</td>
<td>IAR</td>
</tr>
<tr>
<td>2002q4</td>
<td>Trim</td>
<td>Pool</td>
<td>Trim</td>
<td>IAR</td>
<td>IAR</td>
<td>EqCM</td>
</tr>
<tr>
<td>2003q1</td>
<td>SMD</td>
<td>DAR</td>
<td>Pool</td>
<td>∆VEqCM</td>
<td>IAR</td>
<td>∆EqCM</td>
</tr>
<tr>
<td>2003q2</td>
<td>∆EqCM</td>
<td>DD</td>
<td>DAR</td>
<td>IAR</td>
<td>IAR</td>
<td>EqCM</td>
</tr>
</tbody>
</table>

Notes: Best and worst forecast based on MAE at each forecast observation for each horizon.
Table 5: Forecast results for 1- and 2-year-ahead annual inflation

<table>
<thead>
<tr>
<th></th>
<th>ME 1-year</th>
<th>MAE 2-year</th>
<th>RMSFE 1-year</th>
<th>RMSFE 2-year</th>
<th>RGFESM 2-year</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 1998q3–2003q2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN.EqCM</td>
<td>0.109</td>
<td>-0.976</td>
<td>1.959</td>
<td>2.532</td>
<td>2.076</td>
</tr>
<tr>
<td>QU.EqCM</td>
<td>-0.377</td>
<td>-1.253</td>
<td>1.458</td>
<td>1.721</td>
<td>1.638</td>
</tr>
<tr>
<td>QU.ΔEqCMβ</td>
<td>-0.117</td>
<td>-0.263</td>
<td>0.708</td>
<td>0.697</td>
<td>0.923</td>
</tr>
<tr>
<td>DDD</td>
<td>0.122</td>
<td><strong>0.065</strong></td>
<td>0.791</td>
<td>1.003</td>
<td>0.929</td>
</tr>
<tr>
<td>DAR</td>
<td>0.403</td>
<td>0.985</td>
<td>0.858</td>
<td>1.176</td>
<td>1.000</td>
</tr>
<tr>
<td>IAR</td>
<td>-0.206</td>
<td>-0.224</td>
<td><strong>0.588</strong></td>
<td><strong>0.596</strong></td>
<td><strong>0.751</strong></td>
</tr>
<tr>
<td>QU.EqCMin-sample</td>
<td>0.658</td>
<td>0.771</td>
<td>0.767</td>
<td>1.065</td>
<td>0.930</td>
</tr>
<tr>
<td>Pool</td>
<td><strong>0.084</strong></td>
<td>-0.369</td>
<td>0.653</td>
<td>0.860</td>
<td>0.787</td>
</tr>
</tbody>
</table>

B. 1990q4–1995q3

<table>
<thead>
<tr>
<th></th>
<th>ME 1-year</th>
<th>MAE 2-year</th>
<th>RMSFE 1-year</th>
<th>RMSFE 2-year</th>
<th>RGFESM 2-year</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN.EqCM</td>
<td><strong>0.068</strong></td>
<td>3.219</td>
<td>2.202</td>
<td>3.972</td>
<td>2.773</td>
</tr>
<tr>
<td>QU.EqCM</td>
<td>0.459</td>
<td>2.036</td>
<td>1.400</td>
<td>2.088</td>
<td>1.771</td>
</tr>
<tr>
<td>QU.ΔEqCMβ</td>
<td>-0.237</td>
<td>0.399</td>
<td>1.636</td>
<td>1.706</td>
<td>1.889</td>
</tr>
<tr>
<td>DDD</td>
<td>-1.007</td>
<td>-2.081</td>
<td>1.566</td>
<td>2.154</td>
<td>1.899</td>
</tr>
<tr>
<td>DAR</td>
<td>-2.780</td>
<td>-1.792</td>
<td>2.804</td>
<td>1.982</td>
<td>3.198</td>
</tr>
<tr>
<td>IAR</td>
<td>-1.776</td>
<td>-2.857</td>
<td>1.924</td>
<td>2.858</td>
<td>2.243</td>
</tr>
<tr>
<td>QU.EqCMin-sample</td>
<td>-0.211</td>
<td>-1.455</td>
<td>1.335</td>
<td>1.823</td>
<td>1.592</td>
</tr>
<tr>
<td>Pool</td>
<td>-0.783</td>
<td><strong>-0.362</strong></td>
<td><strong>1.095</strong></td>
<td><strong>0.791</strong></td>
<td><strong>1.297</strong></td>
</tr>
</tbody>
</table>

Notes: Figures reported as percentages, with the best forecast performance highlighted in **bold** and the worst forecast performance highlighted in *italics.*
Table 6: Data appendix

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_t$</td>
<td>Gross domestic product: chained volume measures, seasonally adjusted [NS, ABMI]</td>
</tr>
<tr>
<td>$P_t$</td>
<td>Domestic product (expenditure) at market prices deflator: seasonally adjusted [NS, YBGB]</td>
</tr>
<tr>
<td>$G_{v_{at}}$</td>
<td>Gross value added at basic prices: chained volume measures, seasonally adjusted [NS, ABMM]</td>
</tr>
<tr>
<td>$M_{At}$</td>
<td>Nominal broad money stock (end period), £million. [NS, AUYN]</td>
</tr>
<tr>
<td>$R_{S,t}$</td>
<td>Three-month treasury bill rate [DS, UKGBILL3]</td>
</tr>
<tr>
<td>$R_{L,t}$</td>
<td>Yield on 20-year gilts [DS, UKGBOND]</td>
</tr>
<tr>
<td>$N_t$</td>
<td>Public sector net debt, £million [NS, BKQK]</td>
</tr>
<tr>
<td>$W_{\text{pop}_t}$</td>
<td>Population aged 16–59/64, '000s [NS, YBTF from 1992; Pre-1992: EPG, DEG, EG]</td>
</tr>
<tr>
<td>$\text{Emp}_t$</td>
<td>Total number in employment, aged 16+, '000s [NS, MGRZ from 1992; Pre-1992: EPG, DEG, EG]</td>
</tr>
<tr>
<td>$N_{H_t}$</td>
<td>Average actual weekly hours of work (all workers in main &amp; second job)</td>
</tr>
<tr>
<td>$PPI_{t_1}$</td>
<td>PPI manufacturing input—raw materials [DS, UKOPP029F]</td>
</tr>
<tr>
<td>$IMP_t$</td>
<td>Import price index [DS, UKIMPPRCF]</td>
</tr>
<tr>
<td>$RENT_t$</td>
<td>Actual rentals for housing + imputed rentals for housing, £million. (seasonally adjusted using X11) [NS, ADFT+ADFU]</td>
</tr>
<tr>
<td>$C_t$</td>
<td>Unit labour cost index for the whole economy [NS, LNNL]</td>
</tr>
<tr>
<td>$K_t$</td>
<td>Net capital stock for the whole economy excluding dwellings sector, £million [BoE]</td>
</tr>
<tr>
<td>$I_t$</td>
<td>Total gross fixed capital formation, constant price, £million [NS, NPQT]</td>
</tr>
<tr>
<td>$E_t$</td>
<td>£ Effective Exchange Rate index [DS, UKXTW..NF]</td>
</tr>
<tr>
<td>$H_{O_t}$</td>
<td>M0 wide monetary base (end period), £million: seasonally adjusted. [NS, AVAE]</td>
</tr>
<tr>
<td>$PW_t$</td>
<td>OECD consumer price index [DS, OCICP009F]</td>
</tr>
<tr>
<td>$OIL_t$</td>
<td>World market price of crude petroleum [DS, WD176AAZA]</td>
</tr>
<tr>
<td>$I_{date}$</td>
<td>Indicator: $= 1$ in the quarter indexed and 0 otherwise</td>
</tr>
<tr>
<td>$D_{date_{i}, date_{j}}$</td>
<td>Dummy variable: combines 2 indicators $= -1, +1$ in the indexed quarters</td>
</tr>
</tbody>
</table>

Figure 1

Quarterly inflation $\Delta \pi_t$

Annual inflation $\Delta_4 \pi_t$
Figure 2
Figure 4

VEqCM forecasts

1-step Forecasts
±2δ\_1-step
8-step Forecasts
±2δ\_8-step

Quarterly inflation

EqCM forecasts

1-step Forecasts
±2δ\_1-step
8-step Forecasts
±2δ\_8-step
Figure 5

\[ \Delta \text{EqCM}_\beta \text{ forecasts} \]

\[ \Delta \text{p} \]

\[ \text{8-step Forecasts} \]

\[ \text{1-step Forecasts} \]

\[ \pm 2 \sigma \]

\[ \text{8-step} \]

\[ \pm 2 \sigma \text{ 1-step} \]
Figure 7

Annual EqCM forecasts

Quarterly EqCM forecasts
Figure 8

∆EqCM forecasts

Quarterly EqCM(in–sample) forecasts

Direct AR forecasts

Iterated AR forecasts

Annual inflation


−0.1 0.0 0.1 0.2


−0.1 0.0 0.1 0.2


−0.1 0.0 0.1 0.2


−0.1 0.0 0.1 0.2
Figure 9

**ΔEqCM\(_{\beta}\) forecasts**

- Direct AR forecasts
- Iterated AR forecasts

**Quarterly EqCM (in-sample) forecasts**

- 1-year Forecasts
- 2-year Forecasts

**Annual inflation**

- 1998
- 1999
- 2000
- 2001
- 2002
- 2003

- −0.1
- 0.0
- 0.1
- 0.2

- ±2\(\sigma\)_1-year
- ±2\(\sigma\)_2-year

- 1-year Forecasts
- 2-year Forecasts