Real-time Prediction with UK Monetary Aggregates in the Presence of Model Uncertainty[1]

Anthony Garratt
Birkbeck College, University of London

Gary Koop
University of Strathclyde

Emi Mise
University of Leicester

Shaun P. Vahey
Melbourne Business School, Reserve Bank of New Zealand and Norges Bank

27 July 2007

ABSTRACT: A popular account for the demise of the UK monetary targeting regime in the 1980s blames the weak predictive relationships between broad money and inflation and real output. In this paper, we investigate these relationships using a variety of monetary aggregates which were used as intermediate UK policy targets. We use both real-time and final vintage data and consider a large set of recursively estimated Vector Autoregressive (VAR) and Vector Error Correction models (VECM). These models differ in terms of lag length and the number of cointegrating relationships. Faced with this model uncertainty, we utilize Bayesian model averaging (BMA) and contrast it with a strategy of selecting a single best model. Using the real-time data available to UK policymakers at the time, we demonstrate that the in-sample predictive content of broad money fluctuates throughout the 1980s for both strategies. However, the strategy of choosing a single best model amplifies these fluctuations. Out-of-sample predictive evaluations rarely suggest that money matters for either inflation or real output, regardless of whether we select a single model or do BMA. Overall, we conclude that the money was a weak (and unreliable) predictor for these key macroeconomic variables. But the view that the predictive content of UK broad money diminished during the 1980s receives little support using either the real-time or final vintage data.

JEL Classification: C11, C32, C53, E51, E52.

Keywords: Money, Vector Error Correction Models, Model Uncertainty, Bayesian Model Averaging, Real Time Data.

[1]Contact: a.garratt@bbk.ac.uk. We thank participants at the 2006 CIRANO Data Revisions Workshop, the 2007 FRB Philadelphia "Real-Time Data Analysis and Methods in Economics" conference, and the 2007 CEF meetings. We are grateful to Todd Clark, Dean Croushore, Domenico Giannone, James Mitchell, Simon van Norden, Adrian Pagan, Barbara Rossi, Glenn Rudebusch, Yongcheol Shin, Tom Stark, Norman Swanson, and Allan Timmermann for helpful comments. We acknowledge financial support from the ESRC (Research Grant No RES-000-22-1342). The views in this paper reflect those of neither the Reserve Bank of New Zealand nor Norges Bank. Alistair Cunningham (Bank of England) kindly provided the GDP data. Gary Koop is a Fellow at the Rimini Centre for Economic Analysis.

1 Introduction

The demise of UK monetary targeting is generally argued to have taken place in 1985-86; see for example Cobham (2002, p. 61). A landmark speech by the Governor of the Bank of England in October 1986 indicated that the unpredictability of the relationships between broad monetary aggregates and inflation and economic growth undermined monetary targeting (Leigh-Pemberton, 1986). UK policymakers turned to exchange rate targeting for the remainder of the decade (see, Cobham, 2002, chapters 3 and 4, and Batini and Nelson, 2005, section 4). By the time of the Governor's 1986 speech, the most monetarist Government in the UK's post-WWII history had ceased to base policy on monetary aggregates.

In this paper, we investigate the predictability of UK inflation and output using the monetary aggregates $M0$, $M3$ and $M4$. We carry out a recursive analysis of whether the predictive content of money varies over time. Results are presented using both final vintage data and the real-time data which would have been available to a UK policymaker at each point in time. In terms of our set of models, we adopt the same Vector Error Correction Model (VECM) framework as Amato and Swanson (2001). By using Bayesian model averaging (BMA), we allow for model uncertainty with respect to the lag length and the number of cointegrating terms. We report probabilistic assessments of whether "money matters" by taking weighted averages across all models considered. The weights are the posterior model probabilities derived by approximate Bayesian methods based on the Schwarz Bayesian Information Criterion (BIC).

Our application uses a core set of variables comprising money, real output, prices, the short-term interest rate and the exchange rate. We find that, using both BMA and the best model in each period, the in-sample predictive power of money for output and inflation fluctuates in real time. These results are consistent with the US findings of Amato and Swanson (2001). However, the single best model selection strategy gives typically greater instability in the predictive relationships than BMA.

The view that the predictive content of broad money diminished during the 1980s receives little support for the revised data. Despite the emphasis given to it by policymakers in the 1980s, the broad monetary aggregate $M3$ displays little in-sample predictability for output growth, with either real-time or revised data. In contrast, using in-sample evidence, the probability that broad money predicts inflation exceeds 70 percent for most of the 1980s. This result requires the benefit of hindsight about data revisions however. With real-time data, the probability that money can predict inflation exhibits large fluctuations, which are mitigated considerably by BMA. Out-of-sample predictive evaluations rarely suggest that $M3$ matters for either inflation or real output, regardless of whether we select a single model or do BMA.

Although $M3$ was the monetary target preferred by UK policymakers for much of the 1980s, the government also published targets for $M0$ and $M4$ at times. We find that the narrower measure of money exhibits greater in-sample real-time predictability for inflation, and smaller fluctuations in predictability than with $M3$. We find little support for in-sample predictability of real output with $M0$. Nevertheless revised data suggest stronger support than is apparent with real-time data. Turning to the broader money measure, $M4$, we find strong support for predictability of inflation. The real-time fluctuations in the predictive power of $M4$ are much larger for real output growth than for inflation, and as with $M3$, these are mitigated considerably by BMA. Hence, we conclude that impact of data revisions on in-sample predictability is not restricted to a particular monetary aggregate. With the benefit of longer samples of real-time data for $M0$ and $M4$, the out-of-sample predictive performance improves slightly when we include money as an explanatory variable.

Our paper relates to the large literature which investigates whether money has predictive power for inflation and output. Numerous studies have assessed whether money predicts output and/or inflation, conditional on other macroeconomic variables. Nevertheless, the evidence on the extent of the marginal predictive content of money remains mixed. For example, (among others) Feldstein and Stock (1994), Stock and Watson (1989), Swanson (1998) and Armah and Swanson (2006) argued that US money matters for output; and Friedman and Kuttner (1992), and Roberds and Whiteman (1992) argued that it does not. Stock and Watson (1999, 2003) and Leeper and Roush (2002, 2004) apparently confirmed the earlier claim by Roberds and Whiteman (1992) that money has little predictive content for inflation; Bachmeier and Swanson (2004) claimed the evidence is stronger. These studies used substantially revised US data (which we refer to as final vintage data). Amato and Swanson (2001) argue that using the evidence available to US policymakers in real time, the evidence is weaker for the money-output relationship. Corradi, Fernandez and Swanson (2007) extend the classical econometric methodology of Amato and Swanson (2001) but also find a weak predictive relationship between money and output.

With the exception of Roberds and Whiteman (1992), the existing US literature does not allow formally for model uncertainty.[2] Implementation of the classical approach traditionally adopted in the literature requires the researcher to select one preferred model from a broad set of models by a sequential testing procedure. The process requires decisions about the number of cointegrating relationships, the sign and size of the long-run parameters, the number of lags and any restrictions on the short-run dynamics. After selecting the preferred specification at each point in time, the econometrician unconcerned about model uncertainty discards the other models regardless of the probability that those specifications are appropriate. In real-time analysis, the researcher typically selects a single "best" model in each period, and ignores completely models that may have been preferred in previous periods. Any probabilistic statements about the objects of economic importance, such as the marginal predictive content of money, are conditioned on the researcher's best model specification.

Whether model uncertainty compounds the real-time difficulties of assessing the predictive properties of money has not previously been studied. Egginton, Pick and Vahey (2003), Faust, Rogers and Wright (2006), Garratt and Vahey (2006) and Garratt, Koop and Vahey (2007) have shown that initial measurements to UK macroeconomic variables

[2]UK studies adopting a Bayesian limited information approach to modeling money demand include Lubrano, Pierse and Richard (1986), and Steel and Richard (1991). The US studies by Amato and Swanson (2001) and Corradi, Fernandez and Swanson (2007) discuss model robustness.

have at times been subject to large revisions. The phenomenon was particularly severe during the late 1980s. But these papers do not discuss the predictability of money for output or inflation.

In summary, the contributions of the present paper include the use of BMA to investigate a key macroeconomic issue (i.e. the predictive power of money), the use of real-time as well as final vintage data, the investigation of both in-sample and out-of-sample predictability of money, and the use of a new and little-studied UK data set.

The remainder of the paper is organized as follows. Section 2 provides a summary of the UK's monetary targeting experience. Section 3 discusses the econometric methods. Section 4 describes the UK data. Section 5 presents some results and section 6 concludes.

2 The UK's Monetary Targeting Experience

Although UK monetary targeting is often perceived as a 1980s phenomenon, attention to the behavior of monetary aggregates was a feature of UK macroeconomic policy in the 1960s and 1970s; see Bank of England (1978). The monetary regime became formalized as a target for $M3$ in July 1973, revised in late 1976 to refer to $\pounds M3$ which excludes private-sector foreign currency deposits.[3]

Uncertainty surrounds the end date for monetary targeting. The announcement by the Chancellor of the Exchequer, Nigel Lawson, to suspend the target for $\pounds M3$ in October 1985 clarifies the policymaker's dissatisfaction with monetary targeting (see Cobham, 2002, p. 46). However, the target for $\pounds M3$ is revived in the 1986 budget the following spring. The remarks in October 1986 by the Governor of the Bank of England (Leigh-Pemberton, 1986, p. 507) question again the predictability for output and inflation of broad money. Since the Governor draws attention to the discussions with Chancellor about the future of monetary targeting, some central bank watchers argue that the attention shifts towards exchange rate issues before October; see, for example Cobham (2002, chapters 3 and 4), and Nelson and Batini (2005, section 4). Policymakers set target or monitoring ranges for the growth of the monetary aggregates $M0$ and $M4$ from 1987; see Cobham (2002, p. 51, table 3A.5). This year also saw $\pounds M3$ relabeled $M3$ and the series previously known as $M3$ became $M3c$. (In our empirical analysis, we refer to the appropriate broad money aggregate as $M3$ regardless of whether the real-time label was $M3$ or $\pounds M3$.)

Leigh-Pemberton (1986) identifies financial innovation as distorting the underlying relationships between broad money and other macro aggregates. He also argues that these unanticipated events are responsible for the forecast failures for broad money. For example, the Government had missed its published forecast ranges for $\pounds M3$ growth for three successive years between 1984/5 and 1986/7; see Leigh-Pemberton (1986, p. 500, chart 3).[4] Many financial innovations stem from the financial deregulation taking place throughout the 1980s, including the Big Bang of 1986 which opens the London Stock

[3] Cobham (2002, p. 27) has noted informal narrow money ($M1$) monitoring between 1974 and 1976.

[4] Cobham (2002, p. 28) has argued that difficulties in meeting the broad money target led to temporary targets for $M1$ and $PSL2$. The latter aggregate referred to the components of $\pounds M3$ plus selected liquid assets and was subsequently relabeled $M5$.

Market to international competition (Bank of England, 1986, p. 71-73). Cobham (2002, p. 38, table 3.3) classifies the financial innovations by main area of impact: banks, building societies, money market, and capital markets. The first two categories have probably the largest direct impact on measures of money. Examples involving banks include: the move to automatic teller machines (from the late 1970s), the abolition of fixed reserve requirements for banks (1981), and the introduction of debit cards (1987).

The most notable disturbances to broad money come from periodic reclassification of the monetary sector. Narrow measures of money, such as $M0$, are largely unaffected.[5] Topping and Bishop (1989) describe the impacts of the inclusion of trustee savings banks in the monetary sector during 1981, and the conversion of the Abbey National building society into a limited liability bank in 1989. Both cause substantial shocks to the level of broad money. They estimate that the trustee savings banks adds approximately 10 percent to the £$M3$ stock. Subsequently, the conversion by Abbey National cause the Bank of England to suspend publication of $M3$. Cobham (2002, p. 39) notes that the merging of the bank and building society sectors and the increased provision of retail financial services are facilitated by developments in the money and capital markets, which become more competitive through the 1970s and 1980s. As a result of these many reforms, private and public companies hold increasing amounts of assets (and liabilities), including broad money.

To accompany these financial innovations, policymakers introduce many microeconomic reforms including: industrial relations laws; privatization of public companies; changes in social security benefits; and personal and corporate tax changes.[6] Nigel Lawson argues that the financial innovations are perceived as part of the government's policy of supply-side reforms which raises trend economic growth, although the size of this response is subject to uncertainty (Lawson, 1992, p. 804-5).

The wave of economic reforms, which starts in the 1970s and affects the UK economy over the subsequent decades, also extends to the provision of UK statistics. Egginton, Pick and Vahey (2003), Garratt and Vahey (2006) and Garratt, Koop and Vahey (2007) show that the Central Statistical Office (CSO) make substantial revisions to preliminary measurement of economic growth during the 1980s. Subsequently, policymakers at both HM Treasury and the Bank of England partly blame data inaccuracies for the late 1980s' inflationary boom; see Hibberd (1990) and Leigh-Pemberton (1990). In 1989, Nigel Lawson, Chancellor of the Exchequer 1983-1989, took responsibility for the CSO (later renamed the Office for National Statistics, ONS) and a number of major reforms that were introduced between 1989 and 1993, are described in detail by Wroe (1993).[7] Patterson and Heravi (1991) and Patterson (2002) show that pre-1990 UK data errors are

[5] Target ranges for narrow money (but not $M4$) were set from 1984; see Cobham (2002, p. 51, table 3A.4).

[6] Although many of the more notable microeconomic reforms take place in the 1980s, the process is ongoing. See (among others) Blanchflower and Freeman (1994) and Card and Freeman (2004) for discussions.

[7] As a result of these and more minor recent reforms (see Robinson, 2005), the magnitude of data revisions and volatility moderates after the late 1980s (see Garratt and Vahey, 2006, and Garratt, Koop and Vahey, 2007).

very persistent, and contribute to the uncertainty around the order of integration of UK national account variables.[8]

In the empirical analysis that follows, we assess whether the predictive content of broad monetary fluctuated through the period of monetary targeting using both UK real-time and final vintage data. Our econometric methodology is motivated by the considerable model uncertainty generated by the many financial innovations, the microeconomic reforms, and the persistent data inaccuracies throughout our sample period. In the subsequent section, we show how we incorporate uncertainty over the number of cointegrating relationships and lag structure in our empirical analysis by utilizing Bayesian methods.

3 Econometric Methods

3.1 Bayesian Model Averaging

Bayesian methods use the rules of conditional probability to make inferences about unknowns (for example, parameters, models) given knowns (for example, data). For instance, if $Data$ is the data and there are q competing models, $M_1, .., M_q$, then the posterior model probability, $\Pr(M_i|Data)$ where $i = 1, 2, ..., q$, summarizes the information about which model generated the data. If z is an unknown feature of interest common across all models (for example, a data point to be forecast, an impulse response or, as in our case, the probability that money has predictive content for output), then the Bayesian is interested in $\Pr(z|Data)$. The rules of conditional probability imply:

$$\Pr(z|Data) = \sum_{i=1}^{q} \Pr(z|Data, M_i) \Pr(M_i|Data). \tag{1}$$

Thus, overall inference about z involves taking a weighted average across all models, with weights being the posterior model probabilities. This is Bayesian model averaging (BMA). In this paper, we use approximate Bayesian methods to evaluate the terms in (1).

For each model, note that BMA requires the evaluation of $\Pr(M_i|Data)$ (that is, the probability that model M_i generated the data) and $\Pr(z|Data, M_i)$ (which summarizes what is known about our feature of interest in a particular model). We will discuss each of these in turn. Using Bayes rule, the posterior model probability can be written as:

$$\Pr(M_i|Data) \propto \Pr(Data|M_i) \Pr(M_i), \tag{2}$$

where $\Pr(Data|M_i)$ is referred to as the marginal likelihood and $\Pr(M_i)$ the prior weight attached to this model—the prior model probability. Both of these quantities require prior information. Given the controversy attached to prior elicitation, $\Pr(M_i)$ is often simply set to the noninformative choice where, *a priori*, each model receives equal weight. We will adopt this choice in our empirical work. Similarly, the Bayesian literature has proposed many benchmark or reference prior approximations to $\Pr(Data|M_i)$ which do not require

[8] Siklos (1996, 2006) discusses persistence in US data revisions.

the researcher to subjectively elicit a prior (see, e.g., Fernandez, Ley and Steel, 2001). Here we use the Schwarz or Bayesian Information Criterion (BIC). Formally, Schwarz (1978) presents an asymptotic approximation to the marginal likelihood of the form:

$$\ln \Pr(Data|M_i) \approx l - \frac{K \ln(T)}{2}. \tag{3}$$

where l denotes the log of the likelihood function evaluated at the maximum likelihood estimate (MLE), K denotes the number of parameters in the model and T is sample size. The previous equation is proportional to the BIC commonly used for model selection. Hence, it selects the same model as BIC. The exponential of the previous equation provides weights proportional to the posterior model probabilities used in BMA. This means that we do not have to elicit an informative prior and is familiar to non-Bayesians. It yields results which are closely related to those obtained using many of the benchmark priors used by Bayesians (see Fernandez, Ley and Steel, 2001).

With regards to $\Pr(z|Data, M_i)$, we avoid the use of subjective prior information and use the standard noninformative prior. Thus, the posterior is proportional to the likelihood function and MLEs are used as point estimates. Two of our features of interest, z, are the probability that money has no predictive content for (i) output, and (ii) inflation.

3.2 The Models

The models we examine are VECMs (and without error correction terms these become Vector Autoregressions (VARs)).[9] Let $\boldsymbol{x}_t$ be an $n \times 1$ vector of the variables of interest. The VECM can be written as:

$$\Delta \boldsymbol{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \boldsymbol{x}_{t-1} + \boldsymbol{d}_t \boldsymbol{\mu} + \boldsymbol{\Gamma}(L) \Delta \boldsymbol{x}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{4}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices with $0 \leq r \leq n$ being the number of cointegrating relationships. $\boldsymbol{\Gamma}(L)$ is a matrix polynomial of degree p in the lag operator and $\boldsymbol{d}_t$ is the deterministic term. In models of this type there is considerable uncertainty regarding the correct multivariate empirical representation of the data. In particular there can be uncertainty over the lag order and the number of cointegrating vectors (i.e. the rank of $\boldsymbol{\beta}$). Hence this framework defines a set of models which differ in the number of cointegrating relationships (r) and lag length (p). Note that the VAR in differences occurs when $r = 0$. If $r = n$ then $\boldsymbol{\alpha} = I_n$ and all the series do not have unit roots (i.e. this usually means they are stationary to begin with). In this version of the paper, we simply set $\boldsymbol{d}_t$ so as to imply an intercept in (4) and an intercept in the cointegrating relationship.

The next step is to calculate the "feature of interest" in every model. In all that follows we consider a VECM (or VAR) which contains the five ($n = 5$) quarterly variables used in this study; real output (y_t), the price level (p_t), a short-term nominal interest rate (i_t), exchange rate (e_t), and as a fifth variable a monetary aggregate referred to as

[9]These were the types of model used for forecasting and policy analysis both in the academic literature and at the Bank of England, see for example Nickell (1985), Hall and Henry (1986), Hall, Henry and Wilcox (1989), Hylleberg and Mizon (1989).

m_t. In our subsequent analysis, we use three different measures of money: $M0$, $M3$, and $M4$. (See section 4 for a detailed description of the data.) Hence, $\boldsymbol{x}_t = (y_t, p_t, i_t, e_t, m_t)'$ (where lower case denotes the natural logarithm of the variables). When cointegration does not occur, the key equation in the VAR (which we refer to generically as M_{var}), is:

$$\begin{aligned}\Delta y_t = a_0 + \sum_{i=1}^{p} a_{1i}\Delta y_{t-i} + \sum_{i=1}^{p} a_{2i}\Delta p_{t-i} + \sum_{i=1}^{p} a_{3i}\Delta i_{t-i} \\ + \sum_{i=1}^{p} a_{4i}\Delta e_{t-i} + \sum_{i=1}^{p} a_{5i}\Delta m_{t-i} + \epsilon_t\end{aligned} \tag{5}$$

and money has no predictive content for output if $a_{51} = \ldots = a_{5p} = 0$. From a Bayesian viewpoint, we want to calculate $p\left(a_{51} = \cdots = a_{5p} = 0 | Data, M_{var}\right)$.

Using the same type of logic relating to BICs described above (that is, BICs can be used to create approximations to Bayesian posterior model probabilities), we calculate the BICs for M_{var} (the unrestricted VAR) and the restricted VAR (that is, the VAR with $a_{51} = \cdots = a_{5p} = 0$). Call these BIC_U and BIC_R, respectively. Some basic manipulations of the results noted in the previous section says that:

$$\Pr\left(a_{51} = \cdots = a_{5p} = 0 | Data, M_{var}\right) = \frac{\exp\left(BIC_R\right)}{\exp\left(BIC_R\right) + \exp\left(BIC_U\right)}. \tag{6}$$

This is the "probability that money has no predictive content for output" for one model, M_{var}.

Note that we also consider the probability that money has no predictive content for inflation, in which case the equation of interest would be the price equation, and the "probability that money has no predictive content for prices" can be obtained as described in the preceding paragraph.

When cointegration does occur, the analogous VECM case adds the additional Granger causality restriction on the error correction term (see the discussion in Amato and Swanson, 2001, after their equation 2). The equation for output in any of the VECMs (one of which we refer to generically as M_{vec}) takes the form:

$$\begin{aligned}\Delta y_t = b_0 + \sum_{i=1}^{p} b_{1i}\Delta y_{t-i} + \sum_{i=1}^{p} b_{2i}\Delta p_{t-i} + \sum_{i=1}^{p} b_{3i}\Delta i_{t-i} \\ + \sum_{i=1}^{p} b_{4i}\Delta e_{t-i} + \sum_{i=1}^{p} b_{5i}\Delta m_{t-i} + \sum_{i=1}^{r} \alpha_i \xi_{r,t-1} + \varepsilon_t\end{aligned} \tag{7}$$

where the $\xi_{i,t}$ $(i = 1, \ldots, r)$ are the error correction variables constructed using the maximum likelihood approach of Johansen (1988, 1991). The restricted VECM would impose $b_{51} = \cdots = b_{5p} = 0$ and $\alpha_1 = \cdots = \alpha_r = 0$[10] and the probability that money has no predictive content for output is:

$$\Pr\left(b_{51} = \cdots = b_{5p} = 0 \text{ and } \alpha_1 = \cdots = \alpha_r = 0 | Data, M_{vec}\right). \tag{8}$$

[10]Imposing the additional restriction with respect to the loading coefficient for the ECM terms enables an assessment of whether money is long-run forcing for output and inflation (see Granger and Lin, 1995).

For any VECM, this probability can be calculated using BICs analogously to (6).

To summarize, for every single (unrestricted) model, $M_1, .., M_q$, we can calculate the probability that money has no predictive power for output (or inflation) using (6) or (8). The probability that money has predictive power for output is one minus this. Our goal is to do BMA and calculate an overall "money has predictive power for output" which averages over all the models. This can be done using the BICs for the unrestricted models as described in the previous sub-section. Hence, our econometric methodology allows for the model uncertainty apparent in any assessment of the predictive content of money. We stress that, although we adopt a Bayesian approach, it is an approximate one which uses data-based quantities which are familiar to the classical econometrician. That is, within each model we use MLEs. When we average across models we use weights which are proportional to (the exponential of) the familiar BIC. By using this methodology, we are able to assess the predictive content of money for various macro variables (and other objects of interest) using evidence from all the models considered.

4 The Data

In our recursive analysis of the predictive content of money, we consider two distinct data sets. The first uses final vintage data. This is the set of measurements available in 2006Q1. The second uses the vintage of real-time data which would have been available to a policymaker at each vintage date. For example, the real-time data set for the vintage date 1989Q4 contains the historical observations which were available to forecasters at the end of 1989Q4. In our application, we work with a "publication lag" of two quarters—a vintage dated time t includes time series observations up to date $t-2$. The differences between these observations at given vintage date and the final vintage measurements are often substantial; see Garratt and Vahey (2006).[11] The real-time data set comprises one vintage per quarter, starting with the 1979Q2 vintage and ending with 2006Q1. For each vintage, the time series observations start with the 1965Q4. All real-time measurements were published initially by the CSO and its successor, the ONS, in *Economic Trends* and *Economic Trends: Annual Supplement*. Since *Economic Trends* is published monthly, there could be as many as three measurements published for each quarter. We standardized the time interval between vintages by selecting the last monthly release for each quarter.

By analyzing successive vintages of data in our real-time data set, we mimic the common practice followed by applied econometricians in real-time. This approach is common in the real-time literature; see for example Amato and Swanson (2001). In contrast, others including Howrey (1978), Koenig et al (2003) and Garratt, Koop and Vahey (2007) consider measurements that have been revised the same number of times within the regression model. Alternatively, Aruoba (2005) and Croushore (2005) compare preliminary measurements with those taken just before a "benchmark" revision. Croushore and Evans

[11] We refer to a set of measurements published at a given vintage date as a 'vintage'. See, for example, Diebold and Rudebusch (1991) and Croushore and Stark (2001).

(2006) and Corradi, Fernandez and Swanson (2007) discuss the econometric issues related to the choice of revision definitions.

The (seasonally-adjusted) real-time GDP(E) data were taken from the Bank of England's on-line real-time database.[12] Garratt and Vahey (2006) describe and compare the various sources of UK real-time data, including the Bank of England's data set. This paper is also this source for prices. That is we use the real-time implicit price deflator data extended to the 2006Q1 vintage.

For the monetary aggregates, we compiled real-time data for the measures of money for which policymakers set targets for approximately 10 years during our sample.[13] The CSO and ONS did not report all three series between 1977Q4 and 2006Q1. As a result, our analyses of $M0$, $M3$ and $M4$ use different evaluation periods. $M0$ was initially labeled the "wide monetary base" when it was introduced for the 1981Q2 vintage.[14] The monetary aggregate (usually) referred to as £$M3$ was introduced for the 1977Q4 vintage and was phased out in the vintage dated 1989Q4. The broad money measure $M4$ was introduced for the vintage date 1987Q3. With the exception of $M3$ which ends in 1989Q4, the last vintage date is 2006Q1. Revisions to the UK monetary aggregates occurred mainly as a result of changes to the seasonal adjustment by the CSO and ONS.[15] (Unfortunately, real-time seasonally unadjusted figures were not published in *Economic Trends*.) The money series have also been affected by periodic re-classifications. As discussed in detail in section 2, many of these resulted from financial innovation. Our (vintages of) real-time are the exact measurements released by the ONS at the respective vintage dates. The one exception is that post-1981 $M3$ data are scaled to remove the impact of the trustee savings banks conversion, the effects of this on the monetary aggregates was common knowledge at that time (see Topping and Bishop, 1989).

Since the interest rate and the exchange rate are unaffected by revisions, the real-time measurements are exactly the same as those in the final vintage. The interest rate is the 90 day Treasury bill average discount rate; the exchange rate is the sterling effective exchange rate.[16]

In our empirical work, we take the natural logarithms of the raw data published by the ONS and CSO. All quarterly growth rates are defined as the first difference of the log variables. For the interest rate, we use $r_t = 0.25 \times \log(1 + (R_t/100))$.

Figures 1a through 1e plot the final vintage (2006Q1) measurements. Figures 1a shows that the UK experienced stronger economic growth and less volatility in the 1980s, relative to the 1970s. The inflationary pressures build from 1986, with inflation reaching roughly 10 percent in the first half of 1990 as shown in Figure 1b. The strong economic growth and associated build up of inflation is sometimes referred to as the "Lawson Boom". Figure

[12] The Bank of England provided the real GDP data up to the latest vintage on request. Vintages to 2002 can be downloaded from the Bank of England website.

[13] The $M0$ and $M4$ series were originally collected and described by Garratt and Vahey (2006).

[14] Since the ONS does not publish estimates for $M0$ before 1969Q2, we use notes and coins prior to this date.

[15] Mills (1987) analyzes the revision properties of pre-1987 $M3$ data; see also Topping and Bishop (1989).

[16] Both series are published by the ONS in *Financial Statistics*.

1c shows that $M3$ growth begins to ratchet up from mid 1983, and fluctuates in a band between 14 and 26 percent from 1986. Money growth gives a preliminary indication of the late 1980s inflationary boom. Nelson and Batini (2005, p. 69-70) argue that this monetary base expansion reflects a change in policy priorities. In particular, they argue that Chancellor Nigel Lawson preferred to target exchange rates rather than monetary aggregates. Cobham (2002, p. 58) notes the importance of Deutschemark shadowing prior to the UK's entry in to the Exchange Rate Mechanism. Figure 1d shows that about the time of the Governor's 1986 speech—which many commentators think indicated the demise of formal UK monetary targeting—the effective exchange rate dips sharply, and then recovers until 1989. Like the exchange rate, the interest rate plotted in Figure 1e display considerable volatility throughout the period. During the 1980s, interest rate volatility drops somewhat as the Government moves away from monetary targeting. But at the end of the Lawson Boom, interest rates rise sharply back to the levels seen in the 1970s and remain high during the slump of the early 1990s.

We emphasize that the Figures 1a through 1e show the final vintage data. In real-time a policymaker would not observe these measurements.

5 Empirical Results

Our empirical analysis assesses the predictive content of money for real output and inflation, both in and out of sample, taking into account model uncertainty, as outlined in Section 3. We begin by focussing on the $M3$ system (i.e. the set of VARs and VECMs using $M3$ as the monetary aggregate), before discussing the $M0$ and $M4$ systems. For a given system, the models are defined by the cointegrating rank, r, and the lag length, p. We consider $r = 0, \ldots, 4$ and $p = 1, \ldots, 8$ (see text below motivating this choice). Thus, we consider $5 \times 8 = 40$ models, $q = 40$ for each monetary aggregate. In every period of our recursive exercise, we estimate all of the models using the real-time data set and then repeat the exercise using final vintage data. We present results using the model averaging strategy described in Section 3. We also present results for the model with the highest BIC in each time period—a model selection strategy. We stress that, in our recursive exercise, the best model can vary over time.

Before beginning our Bayesian analysis, it is worth mentioning that even a classical econometric analysis using the entire sample reveals strong evidence of model uncertainty with these data. For instance, a preliminary analysis using the $M3$ system (with final vintage time series observations for 1965Q4 to 1989Q2) illustrates the considerable uncertainty encountered by an empiricist in the selection of p and r. Using likelihood ratio (LR) tests, the Akaike Information Criteria (AIC) and the BIC to select the lag length, the lags are 3, 1 or 0, respectively. This lag choice has implications for the cointegrating rank. For example, if we choose $p = 3$, the trace and maximum eigenvalue tests indicate a rank of 0, whilst if we choose $p = 1$, the trace test indicates a rank of 1 and the maximum eigenvalue test a rank of 2. Many issues of this sort have been discussed by others (e.g. Doornik, Hendry and Nielsen, 1998).

It is also worth digressing to discuss the relationship between our approach and several classical econometric approaches which have recently been discussed in the literature (e.g., among many others, Corradi and Swanson, 2006, Inoue and Kilian, 2005, and Inoue and Rossi, 2005). The Bayesian uses posterior model probabilities (i.e. $\Pr(M_i|Data)$) to compare models or do Bayesian model averaging. This holds regardless of whether we average over a huge set of possibly non-nested models (BMA) or calculate the probability that a restriction holds (e.g., the probability that money has in-sample predictive power for output). In the classical econometric literature recently (see citations at beginning of this paragraph), there has been concern about the properties of classical hypothesis testing procedures with repeated tests (as in a recursive testing exercise). In particular, there has been concern about getting the correct critical values for such tests. In our Bayesian approach, such considerations are not relevant. The Bayesian approach does not involve critical values. For instance, in our in-sample results, we simply calculate the probability that money causes output at each point in time. Since we do not carry out classical econometric hypothesis tests, no problems arise from their sequential use. Similarly, with our recursive forecasting exercise, we simply derive the predictive density at each point in time and then calculate various functions of the resulting densities. Issues relating to differential power of in-sample versus out-of-sample power of classical hypothesis test procedures are not applicable.

We present our empirical results in five sections. Since the $M3$ aggregate is of particular macroeconomic significance, in the first three sections we restrict our attention to this measure of money. The first section examines the in-sample behavior of the various models. The second section contains in-sample results, recursively estimated from 1965Q4 to t, where $t = 1978\text{Q}4, \ldots, 1989\text{Q}2$ (43 recursions), where we evaluate the probability that money has predictive power for inflation and output. The third section examines out-of-sample prediction, where the results are recursively generated using data from $1965Q4$ through to t, where $t = 1981\text{Q}1, \ldots, 1989\text{Q}2$ (34 recursions). Remember that $M3$ was phased out in the 1989Q4 vintage–the last recursion uses data up to 1989Q2 (given the publication lag). The fourth section describes the in-sample and out-of-sample analysis for the other monetary aggregates. The in-sample analysis for $M0$ uses $t = 1981\text{Q}1, \ldots, 2005\text{Q}3$ (a total of 99 recursions), and for $M4$, $t = 1987\text{Q}1, \ldots, 2005\text{Q}3$ (75 recursions). The out-of-sample analysis for $M0$ and $M4$ uses $t = 1987\text{Q}1, \ldots, 2003\text{Q}3$ (a total of 67 recursions). The fifth section briefly describes some extensions we have done, but for the sake of brevity have not included in the paper.

5.1 Model Comparison with $M3$ System

Before discussing the predictive power of money, we begin by summarizing evidence on which models are supported by the data. In particular, we present the probabilities attached to each model in our recursive BMA exercise based on the real-time data using $M3$ as the monetary aggregate. For each time period, the same three models almost always receive the vast majority of the posterior probability. These three preferred models all have the same lag length, $p = 1$, and hence, we focus on the uncertainty about the number of cointegrating relationships.

Figure 2 plots the probability of $r = 1, 2$ and 3 cointegrating relationships.[17] Since the sample includes many financial innovations, microeconomic reforms, and persistent data inaccuracies, we do not attempt an economic interpretation of the number of cointegrating relationships.

The overall impression one gets from looking at Figure 2 is that the number of cointegrating vectors varies over time and we cannot ignore model uncertainty. It is rare for a single model to dominate (for example, it is rare for one value of r to get more than 95% of the posterior model probability), and the model with highest probability does tend to vary over time. There is almost no evidence for three cointegrating relationships. Models with two cointegrating relationships tend to receive increasing probability with time. However, prior to 1982, there is more evidence for $r = 1$. During the critical period in which monetary targeting was abandoned, the $r = 2$ and $r = 1$ models are roughly equally likely, with several switches in the identity of the preferred model.

5.2 In-sample Prediction with $M3$ System

In this section, we examine the in-sample ability of the $M3$ monetary aggregate to predict inflation and output in our recursive exercise. Figure 3 plots the probability that money can predict output; and Figure 4 shows the corresponding plot for predicting inflation. All probabilities are calculated using the approach described at the end of Section 3. Each figure contains three lines. Two of these use the real-time data. The first of these lines uses BMA and the other uses the single model with highest probability—the best model. The third line uses the final vintage data, but we plot only BMA results.[18] Hence, two of the lines use real-time data, the third line has the advantage of hindsight. The dates shown on the x-axis refer to the last observation for each vintage; these differ from the vintage date by the publication lag. So, for example, the 1986Q4 vintage has 1986Q2 as the last time series observation.

Figure 3 shows that $M3$ has no predictive power for real output, regardless of the model selection/averaging strategy or the type of data. Remember that $M3$, the preferred monetary target of UK policymakers in this period, was phased out in 1989 where Figure 3 ends.

Figure 4 presents the probability that money has predictive power for inflation. The lines in this figure are very erratic. For instance, using real-time data and selecting a single model, the probability that money can predict inflation swings rapidly from near zero to near one several times in the 1980s. A researcher using a traditional approach which selects a single model could conclude that money mattered in some periods, but then next quarter did not matter all. These switches between times when money matters and when it does not occur with embarrassing frequency. On the other hand, model averaging yields a much less volatile pattern. It is worth noting that for the last vintage in 1986, when the Governor claimed predictability was causing difficulties with the monetary targeting regime, the probability that money could predict inflation, using BMA and real-time

[17] The probability of $r = 0$ is approximately zero for all money definitions.

[18] Results for the best model using final vintage data exhibit a similar pattern to BMA results, but are more erratic.

data, does drop to around 0.4 (plotted as 1986Q2). If we use the strategy of selecting a single model, this probability jumps from roughly one for 1986Q1, to less than 0.05 for 1986Q2, and back to nearly one again the following quarter. The BMA approach indicates some volatility too, but to a much lower degree. With final vintage data, the BMA approach reveals no marked deterioration in the predictive power of money during the mid 1980s. Instead, following a fall to around 0.7 in the early 1980s, this probability rises back to approximately one by mid 1984. We conclude that for $M3$, data revisions played a substantial part in the periodically weak evidence that money causes inflation.[19]

With regards to the question of what caused the UK to drop monetary targeting, note that Figures 3 and 4 show little evidence that the in-sample predictive power of $M3$ deteriorated during the mid 1980s based on final vintage data and using BMA to calculate the probabilities. It appears that $M3$ never had much predictive content for real output. $M3$ does appear (at least at times) to have more predictive power for inflation, but there is no marked deterioration in the $M3$-inflation relationship apart from a brief fall in the early 1980s. When using real-time data, however, the predictive power of money for inflation does fluctuate a lot.

It has been noted that using revised data could distort the predictive ability of broad money since data revisions might be aimed at strengthening the link between macroeconomic variables (see Amato and Swanson, 2001, and Diebold and Rudebusch, 1991). There were minor changes to UK $M3$ during the 1980s (for example, due to the status of institutions recording money, and various privatisations). However, these small changes (typically smaller than never 0.2% of the level) do not coincide with upward movement in the probabilities reported in figures 3 and 4. (In section 5.4 we also show that data revisions enhance predictive ability for the narrower $M0$ measure of money.)

5.3 Out-of-Sample Prediction with $M3$ Systems

Amato and Swanson (2001) argue that out-of-sample forecast performance should be used to judge the predictive content of money in real-time. Although this approach is appealing in principle, small sample problems can make inference based on out-of-sample performance difficult in practice; see Clark and McCracken, 2007b, and the references therein.

To illustrate the practical issues involved in real-time evaluation of out-of-sample prediction, consider a monetary policymaker evaluating the forecasting performance of our many models in 1986Q4. Publication lags for real-time data mean that the policymaker has time series observations up to 1986Q2. The lags between monetary policy implementation and other macroeconomic variables imply that monetary policymakers are typically concerned with predictions between one and two years ahead. If the out-of-sample horizon of interest is 8 quarters from the last available observation, i.e. 6 quarters ahead from the vintage date, the forecast of interest is for 1988Q2. Preliminary real-time outturns for this observation will only be released in the 1988Q4 vintage. Any changes in monetary

[19]The papers by Orphanides (2001, 2003), Rudebusch (2001), and Bernanke and Boivin (2003) debate the contribution of data revisions to US monetary policy from the perspective of policy rules.

policy at that date will have impacts roughly one to two years later—and by then, the UK business cycle has entered a different phase (see Figure 1a). A further complication is that the initial realizations of macroeconomic variables may not be reliable for forecast evaluations. Amato and Swanson (2001) argue that real-time forecasters should evaluate models using a number of vintages of outturns in real-time out-of-sample prediction evaluation, although this introduces further lags in any real-time evaluation process to allow for data revisions.

With these issues in mind, and given the short sample of UK data available with the $M3$ definition of money—out-of-sample we have just 34 recursions (1981Q1 to 1989Q2)—we limit our formal out-of-sample prediction analysis to using the final vintage for outturns. To be precise, we evaluate the performance of our predictions (regardless of whether they are produced using real-time or final vintage data) by comparing them to the "actual" outcome. We use final vintage data for this "actual" outcome. We emphasize that these measures of predictive performance are not timely indicators of model performance—only a forecaster with the 2006Q1 vintage (and the real-time data set) could reproduce the tables reported in this section. Nevertheless, the ex post analysis provides insight into the out-of-sample performance of our models.

As in the previous section, we discuss the forecasting performance of different models, methodologies and data sets. We compare the forecasting performance of the $M3$ system with money to the same system without money using both model averaging and model selection procedures. Separate sets of comparisons are made for real-time and final vintage data. Note, however, that given the broad similarity of the results for model averaging and selection procedures, we only report BMA results in Tables 1 to 3. Where appropriate, we briefly comment on the results obtained by selecting a single model in the text.

Technical details about how the forecasting is done are provided in Appendix A. Suffice it to note here that, if Δp_{t+h} is our variable of interest (i.e. inflation in this instance or output growth, h periods in the future), then we forecast using information available at time t (denoted as $Data_t$), where BMA provides us with a predictive density $\Pr(\Delta p_{t+h}|Data_t)$ which averages over all the models. This is what the BMA results below are based on. Our model selection strategy provides us with a predictive density $\Pr(\Delta p_{t+h}|Data_t, M_{Best})$ where M_{Best} is the model with the highest value for BIC. All of the features of interest in the tables below are functions of these predictive densities (i.e. point forecasts are the means of these densities, etc.).

Table 1 presents results relating to the predictability of inflation and output growth for the $M3$ system. The upper panels (a) and (b) present results for inflation using real-time and final data respectively; the bottom panels (c) and (d) results for output growth using real-time use final vintage data respectively. We begin by discussing results relating to inflation.

Let us first look at the point forecasts. The rows labeled "RMSE" are the root mean squared forecast errors (where the forecast error is the actual realization minus the mean of the predictive distribution). All results are presented relative to the RMSE produced by BMA using the models without money. A number less than one indicates an improved forecast performance relative to this case (i.e. including money helps improve forecast performance). The general picture presented is that including money does not improve forecasting performance. These conclusions hold regardless of whether we do BMA or select the best model. Thus, our results are robust to this important aspect of statistical methodology.

We can also take our Bayesian point forecasts and use any of the classical statistics for evaluating predictive performance. One of the most popular of these is the Diebold-Mariano (DM) statistic (see Diebold and Mariano, 1995). The tables contain results using the DM with the loss function defined as the difference in squared forecast errors of each model relative to the benchmark real time BMA model with no money. Note we do not report the p-values in this instance as our models are nested and therefore the test statistic no longer has a standard normal distribution (see Clark and McCracken, 2007a). But even without critical values, we note that the DM-stats are almost always negative (indicating excluding money is actually improving forecast performance).

The previous discussion related to point forecasts. Point forecasts use only the mean of the predictive distribution. But it is desirable to go beyond the mean and look at other features of the predictive density. Accordingly, the remaining information in the tables does this. The row labelled $\Pr(d_{t+h} > 0)$ calculates a Bayesian variant of the DM statistic. Details are given in Appendix B. The basic idea is that d_{t+h} is a measure of the difference in forecasting performance between models with and without money. It is (apart from an unimportant normalization) the DM statistic. From a Bayesian point of view, this is a random variable (since it depends on the forecast errors which are random variables). If $d_{t+h} > 0$ then models with money are forecasting better than models without money. Hence, if $\Pr(d_{t+h} > 0) > 0.5$ we are finding evidence in favor of of money having predictive power for output or inflation.

The values of the Bayesian variant of the DM statistic reported in Table 1 suggest that including $M3$ when forecasting makes very little difference to the forecast performance of the system. With the notable exception of the eight quarter ahead ($h = 8$) horizon for inflation (arguably the horizon and variable in our results which a policymaker would find most interesting), where the probabilities of 0.4 and 0.45 suggest money should be excluded, the statistic is either 0.5 or marginally above, reaching a maximum of 0.52. In general, our results suggest the systems with and without money have roughly equal forecast performance.

Using the BMA predictive density, $\Pr(\Delta p_{t+h}|Data_t)$, we can calculate predictive probabilities such as $\Pr(\Delta p_{t+h} < a|Data_t)$ for any value of a. Following Egginton, Pick and Vahey (2002), we assume that the inflation rate prevailing when Nigel Lawson started as Chancellor in July 1983 is the threshold of interest.[20] Hence, for (GDP deflator) inflation we set $a = 5$ percent. We define a "correct forecast" as one where $\Pr(\Delta p_{t+h} < a|Data_t) > 0.5$ and the observed revision is less than a. The proportion of correct forecasts is referred to as the "hit rate" and is presented in the tables.

An examination of Table 1 indicates that inclusion of money does improve some of the hit rates with real-time data, often by a substantial amount, at shorter horizons. For instance, the hit rate at $h = 4$ with real-time data is 53% when money is excluded.

[20] Nigel Lawson suggested that '(t)he acid test of monetary policy is its record in reducing inflation ... The inflation rate is judge and jury' (Lawson, 1992, p.481).

However, when money is included the hit rate rises to 65%. In contrast, with final vintage data, the inclusion of money causes no change in the hit rates for $h = 8$ and only a small change for $h = 4$. But at horizon $h = 1$ the inclusion of money actually causes the hit rate to deteriorate from 68% to 56%. In summary, we are finding some (weak) evidence that the inclusion of money has a bigger role in getting the shape and dispersion of the predictive distribution correct than in getting its location correct. That is, the RMSE results provided little evidence that including money improves point forecasts, but including money does seem to improve hit rates with real-time data (but not with final vintage data). Finally, it is worth mentioning that the hit rates are basically the same for BMA and the best single model.

Finally, the rows labeled "PT" present a classical econometric measure of absolute forecast performance from Pesaran and Timmermann (1992). This is the directional market timing statistic and the hypothesis test based on this statistic uses the same information as the Kuipers score (i.e. it measures the proportion of growth rates greater than the threshold a that were correctly forecast minus the proportion of below mean growth rates that were incorrectly forecast). Under the null hypothesis that the forecasts and realizations are independently distributed the PT statistic has a standard normal distribution. Only two of these are greater than the 10 percent critical value of 1.64.[21] Thus, once again we are finding almost no evidence that money can help forecast inflation.

Let us now turn to output growth. In panels (c) and (d) of Table 1, we report the same set of evaluation statistics for central and probability forecasts for output growth. We consider the probability event $\Pr(\Delta y_{t+h} < a|Data_t)$ where we set $a = 2.3\%$, the average annualized growth rate for the 1980Q1 to 2005Q3 period. The general conclusion is broadly similar to the inflation case: the RMSE does not improve when money is included. The final vintage RMSEs are typically marginally lower than the real-time data equivalents (the exception is for $h = 8$, with and without money). As with inflation, there is a strong similarity of the RMSE results for the statistical strategies of BMA and selecting the single best models.

The same conclusion, that the inclusion of money makes relatively small differences, can be drawn from the BMA predictive probabilities for output growth. Depending on the horizon we see a slight worsening through to a slight improvement. For example, including money with the real time data worsens an already poor hit rate from 41% to 35% at $h = 1$, stays the same at 68% for $h = 4$ and improves from 38% to 47% for $h = 8$. Using final vintage data and excluding money, the hit rates improve in two cases, but not for $h = 8$. Including money has mixed impacts on final vintage hit rates: both $h = 1$ and $h = 8$ have higher hit rates with money, but the $h = 4$ case deteriorates. The PT statistics indicate insignificance typically. Selecting the best model for each recursion produces results very similar to the BMA strategy.

The overall impression from our out-of-sample prediction analysis with the $M3$ monetary aggregate is that the $M3$ system with no money provides a benchmark that is difficult to beat. For our short sample, there is little evidence that including money in the system

[21]Note that some of the PT statistics are undefined. This is due to the denominator of the PT statistic being zero which arises due to the small sample size we have for $M3$.

makes substantial differences to out-of-sample prediction, either with real-time or final vintage data.

We stress that the out-of-sample results presented in this section are not timely. UK policymakers in the 1980s could not have used final vintage data to evaluate model performance in this way. So the relatively weak out-of-sample performance of the models with money might not have been apparent in real time. And (at times) the in-sample evidence for money predicting inflation is strong with real-time data, but prone to fluctuations. Recall that UK inflation reached double-digit values later in the decade (see Figure 1b), and that the decision to end monetary targeting was made in the mid 1980s. Subsequently, less emphasis was given to the $M3$ monetary aggregates despite the high risk of inflation indicated by the system with money. To shed more light on this, we provide plots of net contribution of $M3$ to the assessment of inflationary pressures and above-trend economic growth. Figure 5 plots the difference in the real-time BMA probability of inflation greater than five percent, with and without money at $h = 8$.[22] Although the differential fluctuates through the recursions, the net contribution of $M3$ is to increase the proportion of the predictive density above the five percent threshold—the inclusion of money gives a much stronger indication of inflationary pressures. For example, the 1986Q4 vintage (plotted as 1986Q2) raises the posterior probability of the high inflation event by nearly 40 percentage points, and the difference fluctuates between 30 and 10 thereafter. With final vintage data, the fluctuations are generally smaller about a slightly higher mean, but the message is similar: the exclusion of the monetary aggregates implies a much more benign inflation outlook with revised data.

Turning to Figure 6, which plots the difference of output growth greater than 2.3 percent with and without money, we see adding monetary aggregates to our systems causes the probability of above-trend economic growth to fall. The volatility is somewhat lower than for the inflation event shown in Figure 5; and the use of real-time data, rather than final vintage data, adds to the volatility. Although the plots for the two data sets differ at times, the general story is that real-time data leads to a similar—but slightly noisier—assessment of the net contribution of M3 to the probability of strong economic growth.

We conclude from Figures 5 and 6 that real-time forecasts made in the mid 1980s conditional on excluding money indicate stronger economic growth and weaker inflation than forecasts conditional on including $M3$.

5.4 $M0$ and $M4$ Results

Since the $M3$ aggregate preferred by UK policymakers for much of the 1980s was phased out in 1989, the evaluation of our systems in the previous section is based on small samples. The government also published targets for $M0$ and $M4$ at times during the 1980s, and for these monetary aggregates we have real-time data up to the vintage 2006Q1.

We focus initially on the in-sample evidence of whether money has predictive power

[22] The dates shown refer to the last observation in each vintage. Other horizons give broadly similar plots for Figures 5 and 6.

for output or inflation.[23] Figure 7, which is in the same format as Figures 3 and 4, shows that the case for believing that $M0$ causes economic growth is weak throughout the 1980s, regardless of whether we use BMA or select the best model. The real-time probability only rises above 0.1 in the early 1990s. Note however, that using final vintage data yields substantially more evidence that money has predictive power for real output. For instance, suppose we were to adopt a rule of thumb where probabilities greater than 0.5 lead us to conclude that money has predictive power for output. Using real-time data we would almost never conclude that money has predictive power for output. However, using final vintage data, we would conclude money does have predictive power for output for the recursions ending 1994Q2 through to 1998Q1, although more recent observations suggest weaker predictability.

Figure 8 shows that there is generally stronger predictability of $M0$ for inflation, particularly from 1983 onwards, regardless of whether we use real-time or final-vintage data, or we adopt BMA or select a single model. However, there are times when the real-time data indicate declines in predictability for both modelling strategies. For example, the probability of money predicting inflation using BMA falls approximately 30 percentage points in 1983Q2, and 17 percentage points in 1986Q2. After 1987, the probability of prediction for inflation is nearly one.

Figure 9, which uses $M4$, shows stronger evidence of predictability for real output growth than for either $M0$ or $M3$, but only after 1990 and rarely is the evidence extremely strong. All three of the lines in this figure indicate that the probability of predictability generally rises through time, reaching 0.65 or more by the end of the evaluation period. However, there are important and interesting differences between the three lines. Using real-time data, the strategy of selecting a single best model yields a much more volatile line. The differences between BMA results using real-time and final-vintage data are not so pronounced. However it is interesting to note that, in contrast to $M0$, the real-time evidence that $M4$ predicts output is slightly but consistently higher than with final vintage data.

Figure 8 shows the in-sample probability that $M4$ predicts inflation. Using real-time data, the late 1980s do seem to be a volatile time. As with many other cases, BMA results are much smoother than those produced using the strategy of selecting a single model. However, as of 1990 the probability that money can predict inflation reaches 0.8 or more, and settles down to approximately one by the mid 1990s, comfortably before Bank of England independence in 1997. Using final vintage data, the probability that money can predict inflation is almost one throughout the entire time period.

A story consistent with all three monetary aggregates is that there is little evidence in-sample that money can predict real output growth in the 1980s. After the early 1990s (at least for monetary aggregates where real-time data is available) there is more evidence of predictability, although this is rarely very strong using either real-time or final-vintage data. For $M4$ the evidence using real-time is obscured by fluctuations in probabilities, which are mitigated to some extent by BMA. The evidence that money predicts inflation

[23]Other empirical results, such in-sample evidence on the number of cointegrating relationships, etc. for the $M0$ and $M4$ systems can be obtained from the authors on request.

in-sample is stronger through the evaluation period. Although both $M0$ and (from 1987) $M4$ posterior probabilities display large fluctuations in real-time, the problem is less severe than with $M3$.

Turning to the out-of-sample evidence, Table 2 uses VARs and VECMs where $M0$ is the money variable ("M0 System"). Table 3 uses $M4$ ("M4 System"). In each case, the upper panels (a) and (b) refer to inflation; and the lower panels (c) and (d) to real output growth. Remember that these systems are estimated with much more data than the $M3$ system since this monetary aggregate was discontinued in 1989. The evaluation period in both cases begins in 1987Q1 (the first vintage for M4 being 1987Q3) and ends in 2003Q3—we require final vintage data to evaluate eight periods ahead and our last observation is for 2005Q3, given by the 2006Q1 vintage.

Looking first at the point forecasts, the general picture presented is that including $M0$ does not improve forecasting performance in real-time. In contrast, with the M4 System the RMSEs are typically less than one and decrease with forecast horizon. For the M0 system the RMSE ratio favours excluding money (is greater than one) for inflation and at horizon $h = 8$ for output growth, but including it for output growth (is less than one) for horizons 1 and 4. Whereas for the M4 System the RMSE ratios favours including money for inflation and for output growth at horizon $h = 8$.

The Bayesian variant of the DM is always greater than 0.5, for both M0 and M4, thus providing evidence that the inclusion of money does improve forecasting performance. However the evidence is weak for output growth, where values are 0.52 at most and slightly stronger for inflation where values are typically 0.55 to 0.56. Note also that these conclusions hold regardless of whether we do BMA or select the best model.

An examination of Table 3 for the $M4$ systems indicates that inclusion of money improves many of the hit rates in real-time, often by a substantial amount. This is most noticeable with long horizons ($h = 8$), where the PT statistics indicate statistical significance. For the $M0$ systems shown in Table 2, there is little improvement in the hit rates in real time from the inclusion of the monetary aggregate. These results hold with both the BMA and best model selection strategy.

For both $M0$ and $M4$ systems, the support for the inclusion of money is typically marginally stronger with final vintage data than with real-time data, but any improvement in central forecasts and hit rates remains modest. This characterization holds regardless of the model selection strategy.

To summarize our analysis of the $M0$ and $M4$ systems, we find that narrow money exhibits greater in-sample predictability for inflation, and slightly smaller real-time fluctuations in predictability than for $M3$ systems. We find little support for in-sample predictability of economic growth with $M0$ in real time, but more support with revised data. For broad money, $M4$, we find find strong support for predictability of inflation. The real-time fluctuations are much larger for real output growth than for inflation, and as with $M3$, these are mitigated considerably by BMA. With the benefit of longer samples of real-time data, the out-of-sample performance of the predictive densities for $M4$ systems, and to a lesser extent for $M0$ systems, improves marginally when money is included.

Remember that nearly all of the results presented in this section would not have been available for real-time analysis by UK policymakers in the 1980s. They would have been restricted to in-sample statistics for the $M3$ measure of money. (An exception is the in-sample results for narrow money, presented in Figures 7 and 8, which show results broadly consistent with our $M3$ analysis.) Nevertheless, our $M0$ and $M4$ results reveal many similarities with those for $M3$. It appears that sample size limitations did not distort our analysis of the predictive content of money using the policymakers' preferred aggregate.

5.5 Extensions

Like Amato and Swanson (2001), our analysis of whether money matters focuses on a theoretically-coherent set of endogenous variables (in our case, for an open economy). Our formal approach to model uncertainty involves Bayesian model averaging over a fairly wide set of models. We find little in-sample evidence that money has predictive power for output but stronger support for an in-sample predictive relationship with inflation. The in-sample evidence tends to be very erratic with real time data.

The reader may wonder whether these results are robust to the inclusion of other (perhaps, less theoretically appealing) variables. Unfortunately, there are many candidate variables that could be considered. As an example, we have repeated our econometric methods adding the oil price as an extra exogenous variable to our VECMs. Since oil supply disturbances were often accommodated by UK monetary policy during the 1970s and 1980s, empirical assessments of predictive relationships tend to be sensitive to the inclusion of oil price inflation as an exogenous variable. In general, our findings based on longer samples (for $M0$ and $M4$) are very robust. That is, the corresponding plots to Figures 7, 8, 9 and 10 look very similar. In particular, the evidence of in-sample predictive relationships from monetary aggregates is extremely erratic with real-time data. For the shorter sample with the $M3$ system, we still find that money has little predictive power for output growth (i.e., a figure similar to Figure 3 holds). But for inflation, the inclusion of the oil price yields much weaker evidence of predictability both with real-time and final vintage data. To disentangle the impact of the oil shocks from that of $M3$ on inflation requires longer runs of data than are available.

Mindful of the contribution of the oil shocks to our results, we also carried out a rolling (instead of recursive) forecasting exercise using a window of 40 observations for the $M0$ and $M4$ systems. In these cases, we generally find slightly stronger evidence that money has predictive power for output or inflation. Corradi, Fernandez, and Swanson (2007) find that excluding 1970s' data works well for the US. But we find that truncating the sample produces quite erratic results, masking the contribution of data revisions.

6 Conclusions

This paper investigates whether money has predictive power for output or inflation in the UK. We carry out a recursive analysis to investigate whether predictability has changed over time. We use real-time data which allows us to examine whether prediction would have been possible in real time (that is, using the data which would have been available at the time the prediction was made) and/or retrospectively (using final vintage data). We consider a large set of Vector Autoregressive (VAR) and Vector Error Correction models (VECM) which differ in terms of lag length and the number of cointegrating terms. Faced with this model uncertainty, we use Bayesian model averaging (BMA) and contrast it to a strategy of selecting a single best model.

Our empirical results can be divided into in-sample and out-of-sample components. With regards to in-sample results, using the real-time data available to UK policymakers, we find that the predictive content of $M3$ fluctuates throughout the sample. However, the strategy of choosing a single best model amplifies these fluctuations relative to BMA. With BMA and final vintage data, the in-sample predictive content of broad money did not diminish substantially during the 1980s. The $M3$ monetary aggregate displays little predictability for output at any point and the probability that broad money predicts inflation rises though the mid 1980s. But this result requires the benefit of hindsight about data revisions. With real-time data, the probability that money can predict inflation in-sample is very erratic during the period of formal monetary targeting.

Our out-of-sample analysis suggests little support for the hypothesis that $M3$ matters for inflation or output either with real-time or final vintage data. With the benefit of longer samples, the evidence of predictability is marginally stronger with $M0$ and, in particular, $M4$.

UK policymakers have argued that the unpredictability of the relationships between broad monetary aggregates and inflation and economic growth undermined the UK's monetary targeting experiment. The results in this paper indicate that they were right to worry about this issue. With the hindsight of revised data, our evidence provides only weak support of a predictive relationship between money and inflation.

Table 1: Evaluation of BMA Out of Sample Central and Probability Forecasts with $M3$, 1981Q1-1989Q2

	Without Money			With Money		
	h=1	*h=4*	*h=8*	*h=1*	*h=4*	*h=8*
(a) Inflation Real Time						
RMSE	1.00	1.00	1.00	1.19	1.40	1.50
DM	·	·	·	-1.93	-4.52	-4.10
$\Pr(d_{t+h} > 0)$	·	·	·	0.51	0.50	0.45
Hit Rate, $\Pr(\Delta p_{t+h} < 5.0\%)$	0.68	0.53	0.65	0.74	0.65	0.65
PT, $\Pr(\Delta p_{t+h} < 5.0\%)$	1.38	-1.60	·	2.33	·	·
(b) Inflation Final Vintage						
RMSE	1.00	1.00	1.00	1.14	1.43	1.78
DM	·	·	·	-2.84	-2.77	-6.34
$\Pr(d_{t+h} > 0)$	·	·	·	0.51	0.50	0.40
Hit Rate, $\Pr(\Delta p_{t+h} < 5.0\%)$	0.68	0.62	0.65	0.56	0.65	0.65
PT, $\Pr(\Delta p_{t+h} < 5.0\%)$	1.27	-0.76	·	-0.79	·	·
(c) Output Growth Real Time						
RMSE	1.00	1.00	1.00	1.06	1.09	1.12
DM	·	·	·	-0.63	-2.24	-0.90
$\Pr(d_{t+h} > 0)$	·	·	·	0.51	0.51	0.51
Hit Rate, $\Pr(\Delta y_{t+h} < 2.3\%)$	0.41	0.68	0.38	0.35	0.68	0.47
PT, $\Pr(\Delta y_{t+h} < 2.3\%)$	-1.69	1.81	-1.42	-1.97	1.71	0.44
(d) Output Growth Final Vintage						
RMSE	1.00	1.00	1.00	1.08	1.10	1.25
DM	·	·	·	-2.77	-1.61	-2.10
$\Pr(d_{t+h} > 0)$	·	·	·	0.51	0.52	0.50
Hit Rate, $\Pr(\Delta y_{t+h} < 2.3\%)$	0.47	0.71	0.38	0.53	0.65	0.47
PT, $\Pr(\Delta y_{t+h} < 2.3\%)$	-0.72	2.14	-1.20	-0.16	1.50	1.24

Notes: RMSE denotes Root Mean Square Error, defined as a ratio relative to the benchmark no money case. DM denotes the Diebold-Mariano (1995) statistic, where the loss function, d_t, is defined using the difference in squared forecast errors of the with and without money models. The probability $\Pr(d_{t+h} > 0)$, is a bootstrapped test statistic described in Appendix B and the text, computed using 5000 replications. The Hit Rate defines the proportion of correctly forecast events, where we assume that the event can be correctly forecast if the associated probability forecast exceeds 0.5. PT is the Pesaran Timmerman (1992) (PT) statistic described in the text.

Table 2: Evaluation of BMA Out of Sample Central and Probability Forecasts with $M0$, 1987Q1-2003Q3

	Without Money			With Money		
	h=1	*h=4*	*h=8*	*h=1*	*h=4*	*h=8*
(a) Inflation Real Time						
RMSE	1.00	1.00	1.00	1.03	1.08	1.12
DM	·	·	·	-0.29	-0.89	-1.48
$\Pr(d_{t+h} > 0)$	·	·	·	0.55	0.55	0.53
Hit Rate, $\Pr(\Delta p_{t+h} < 5.0\%)$	0.72	0.66	0.45	0.78	0.69	0.51
PT, $\Pr(\Delta p_{t+h} < 5.0\%)$	2.40	3.26	2.34	4.18	3.84	2.69
(b) Inflation Final Vintage						
RMSE	1.00	1.00	1.00	1.00	1.04	1.16
DM	·	·	·	-0.03	-0.55	-1.49
$\Pr(d_{t+h} > 0)$	·	·	·	0.56	0.57	0.52
Hit Rate, $\Pr(\Delta p_{t+h} < 5.0\%)$	0.75	0.55	0.34	0.81	0.67	0.48
PT, $\Pr(\Delta p_{t+h} < 5.0\%)$	4.04	2.73	1.71	4.95	3.40	2.01
(c) Output Growth Real Time						
RMSE	1.00	1.00	1.00	0.94	0.96	1.18
DM	·	·	·	0.80	0.34	-2.15
$\Pr(d_{t+h} > 0)$	·	·	·	0.52	0.52	0.52
Hit Rate, $\Pr(\Delta y_{t+h} < 2.3\%)$	0.54	0.46	0.46	0.51	0.57	0.54
PT, $\Pr(\Delta y_{t+h} < 2.3\%)$	-0.23	-1.15	-0.61	-0.35	0.98	0.64
(d) Output Growth Final Vintage						
RMSE	1.00	1.00	1.00	0.93	0.98	1.26
DM	·	·	·	0.78	0.44	-1.80
$\Pr(d_{t+h} > 0)$	·	·	·	0.52	0.52	0.52
Hit Rate, $\Pr(\Delta y_{t+h} < 2.3\%)$	0.49	0.52	0.49	0.55	0.52	0.61
PT, $\Pr(\Delta y_{t+h} < 2.3\%)$	-0.87	0.11	-0.06	0.53	0.08	1.92

Notes: See Notes to Table 1.

Table 3: Evaluation of BMA Out of Sample Central and Probability Forecasts with $M4$, 1987Q1-2003Q3

	Without Money			With Money		
	h=1	*h=4*	*h=8*	*h=1*	*h=4*	*h=8*
(a) Inflation Real Time						
RMSE	1.00	1.00	1.00	0.93	0.79	0.81
DM	·	·	·	2.16	2.38	1.41
$\Pr(d_{t+h} > 0)$	·	·	·	0.54	0.56	0.55
Hit Rate, $\Pr(\Delta p_{t+h} < 5.0\%)$	0.72	0.66	0.45	0.81	0.88	0.81
PT, $\Pr(\Delta p_{t+h} < 5.0\%)$	2.40	3.26	2.34	3.98	5.67	4.30
(b) Inflation Final Vintage						
RMSE	1.00	1.00	1.00	0.93	0.90	0.97
DM	·	·	·	0.79	0.71	0.13
$\Pr(d_{t+h} > 0)$	·	·	·	0.55	0.56	0.55
Hit Rate, $\Pr(\Delta p_{t+h} < 5.0\%)$	0.75	0.55	0.34	0.87	0.88	0.78
PT, $\Pr(\Delta p_{t+h} < 5.0\%)$	4.04	2.73	1.71	5.48	5.67	3.63
(c) Output Growth Real Time						
RMSE	1.00	1.00	1.00	1.00	1.12	0.92
DM				0.10	-2.14	0.58
$\Pr(d_{t+h} > 0)$	·	·	·	0.51	0.51	0.51
Hit Rate, $\Pr(\Delta y_{t+h} < 2.3\%)$	0.54	0.46	0.46	0.51	0.60	0.60
PT, $\Pr(\Delta y_{t+h} < 2.3\%)$	-0.23	-1.15	-0.61	-0.62	1.75	2.12
(d) Output Growth Final Vintage						
RMSE	1.00	1.00	1.00	1.00	1.02	0.85
DM	·	·	·	-0.08	-0.58	0.95
$\Pr(d_{t+h} > 0)$	·	·	·	0.51	0.51	0.51
Hit Rate, $\Pr(\Delta y_{t+h} < 2.3\%)$	0.49	0.52	0.49	0.51	0.51	0.63
PT, $\Pr(\Delta y_{t+h} < 2.3\%)$	-0.89	0.11	-0.06	-0.43	-0.36	2.74

Notes: See notes to Table 1.

References

Amato, J. and Swanson, N.R. (2001). 'The real-time predictive content of money for output', *Journal of Monetary Economics*, 48, pp. 3-24.

Armah, N.A. and Swanson, N.R. (2006). 'Predictive inference under model misspecification with an application to assessing the marginal predictive content of money for output', mimeo, Rutgers University.

Aruoba, B. (2005). 'Data revisions are not well behaved', mimeo, CIRANO Workshop on Data Revisions', October 2005, Montreal.

Bachmeier L.J. and Swanson, N.R. (2004). 'Predicting inflation: does the quantity theory help?', *Economic Inquiry*, 43, pp. 570-585

Bank of England (1978). *Quarterly Bulletin*, March 1978.

Bank of England (1986). *Quarterly Bulletin*, March 1986.

Batini, N. and Nelson, E. (2005). 'The UK's rocky road to stability', FRB St Louis Working Paper 2005-020, March.

Bernanke, B. and Boivin, J. (2003). 'Monetary policy in a data-rich environment', *Journal of Monetary Economics*, 50, pp. 525-546.

Blanchflower, D.G. and Freeman, R. (1994). 'Did the Thatcher reforms change British labour market performance?', in Barrell, R. (Ed.) *The UK labour market: Comparative aspects and institutional developments', Cambridge University Press.*

Card, D. and Freeman, R. (2004). 'What have two decades of British economic reform delivered?', in Card, D., Blundell, R. and Freeman, R. (Eds.) *Seeking a Premier Economy: the Economic Effects of British Economic Reforms, 1980-2000*, National Bureau of Economic Research.

Castle, J. and Ellis, C. (2002). 'Building a real-time database for GDP(E)', *Bank of England Quarterly Bulletin*, February, pp. 42-49.

Clark, T.E. and McCracken, M.W. (2007a).'Tests of Equal Predictive Ability with Real Time Data', Paper Presented at the 'Real-Time Data Analysis and Methods in Economics', held at the Federal Reserve Bank of Philadelphia.

Clark, T.E. and McCracken, M.W. (2007b).'The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence', *Journal of Money, Credit, and Banking*, forthcoming.

Cobham, D. (2002). *The making of monetary policy in the UK, 1975-2000*, Wiley and Sons, Chichester, UK.

Corradi, V. and Swanson, N. (2006). 'Predictive density evaluation', in *Handbook of Economic Forecasting*, eds. G. Elliott, C.W.J. Granger and A. Timmermann, North-Holland Elsevier.

Corradi, V., Fernandez, A. and Swanson, N.R. (2007).'Information in the revision process of real-time data', Paper Presented at the 'Real-Time Data Analysis and Methods in Economics Conference', held at the Federal Reserve Bank of Philadelphia.

Croushore, D. (2005). 'An evaluation of inflation forecasts from surveys using real-time data', mimeo, CIRANO Workshop on Data Revisions, October 2005, Montreal.

Croushore, D. and Evans, C.L. (2006). 'Data revisions and the identification of monetary policy shocks', *Journal of Monetary Economics*, 53, pp. 1135-1160.

Croushore, D. and Stark, T. (2001). 'A real-time data set for macroeconomists', *Journal of Econometrics*, 105, pp. 111-130.

Diebold, F.X. and R.S. Mariano (1995). "Comparing predictive accuracy", *Journal of Business Economics and Statistics*, 13, 134-144.

Diebold, F. X., and Rudebusch, G. D. (1991). 'Forecasting output with the composite leading index: a real-time analysis', *Journal of the American Statistical Association*, 86, pp. 603-610.

Doornik, J.A., Hendry, D.F. and Nielsen, B. (1998). 'Inference in cointegrating models: UK M1 revisited', *Journal of Economic Surveys*, 12, 5, pp. 533-572.

Egginton, D., Pick, A. and Vahey, S. (2002). 'Keep it real! A real-time UK macro data set', *Economics Letters*, 77, pp. 15-20.

Faust, J., Rogers, J. and Wright, J. (2005). 'News and noise in G7 GDP announcements', *Journal of Money, Credit and Banking*, 37, pp. 403-420.

Feldstein, M. and Stock, J.H. (1994). 'The use of a monetary aggregate to target nominal GDP', in Mankiw, N.G. (Ed.), *Monetary Policy*, Chicago University Press, Chicago, pp. 7–62.

Fernandez, C., Ley, E. and Steel, M. (2001). 'Benchmark priors for Bayesian model averaging', *Journal of Econometrics*, 100, pp. 381-427.

Friedman, B.M. and Kuttner, K.N. (1992). 'Money, income, prices and interest rates', *American Economic Review*, 82, pp. 472-492.

Garratt, A. and Vahey, S.P. (2006). 'UK Real-time data characteristics', *Economic Journal*, 116, F119-F135, February.

Garratt, A., Koop, G. and Vahey, S.P. (2007). 'Forecasting substantial data revisions in the presence of model uncertainty', *Economic Journal* forthcoming and RBNZ Discussion Paper 2006/02.

Granger, C.W.J. and Lin, J.L. (1995). 'Causality in the long-run', *Econometric Theory*, 11, pp. 530-536.

Hall, S.G. and Henry, S G B (1986), 'A dynamic econometric model of the UK with rational expectations', *Journal of Economic Dynamics and Control*, 10(1-2), pp. 219-23.

Hall, S.G., Henry S.G.B. and Wilcox, J.B. (1989), 'The Long-Run Determination of the UK Monetary Aggregates', *Bank of England Discussion Paper*, No. 41.

Howrey, E.P. (1978). 'The use of preliminary data in econometric forecasting', *Review of Economics and Statistics*, 60, pp. 193-200.

Hibberd, J. (1990). 'Official statistics in the Late 1980s', *Treasury Bulletin*, Summer, pp. 2-13.

Hylleberg, S. and Mizon, G.E.(1989), 'Cointegration and Error Correction Mechanisms', *Economic Journal*, 99, No.395, pp. 113-125.

Inoue, A. and Kilian, L. (2005). 'In-sample or out-of-sample tests of predictability: Which one should We Use?' *Econometric Reviews*, 23, 371-402.

Inoue, A. and Rossi, B. (2005). 'Recursive predictability tests for real-time data' *Journal of Business and Economic Statistics*, 23, pp. 336-345.

Koenig, E., Dolmas, S. and Piger, J. (2003). 'The use and abuse of 'real-time' data in economic forecasting', *Review of Economics and Statistics*, 85, pp. 618-628.

Koop, G. (2003). *Bayesian Econometrics*, John Wiley and Sons, Chichester.

Johansen, S. (1988). 'Statistical analysis of cointegration vectors', *Journal of Economic Dynamics and Control*, 12, pp. 231-254.

Johansen, S. (1991). 'Estimation and hypothesis testing of cointegrating vectors in gaussian vector autoregressive models', *Econometrica*, 59, pp. 1551-1580.

Lawson, N. (1992). *The View From No. 11: Memoirs of a Tory Radical*, Bantam Press, London.

Leeper, E.M. and Roush, J.E.(2003). 'Putting 'M' back into monetary policy', *Journal of Money, Credit, and Banking*, 35, 2, pp. 1217-1256.

Leigh-Pemberton, R. (1986). 'Financial change and broad money', *Bank of England Quarterly Bulletin*, December.

Leigh-Pemberton, R. (1990). 'Monetary policy in the second half of the 1980s', *Bank of England Quarterly Bulletin*, November.

Lubrano, M., Pierse, R.G., and Richard, J.-F. (1986). 'Stability of a UK money demand equation: a Bayesian approach to testing exogeneity', *The Review of Economic Studies*, 53, 4, pp. 603-634.

Mankiw, N.G., Runkle, D. and Shapiro, M. (1984). 'Are preliminary announcements of the money stock rational forecasts', *Journal of Monetary Economics*, 14, pp. 15-27.

Mills, T.C. (1987). 'Uncertainty in the UK monetary aggregates: modelling data revisions in economic time series', *Manchester School of Economic and Social Studies*, 4, pp. 337-352.

Nickell, S.J. (1985) 'Error correction, partial adjustment and all that: an expository note', *Oxford Bulletin of Economics and Statistics*, 47(2), pp. 119-29.

Office for National Statistics. *Economic Trends*, various issues, The Stationery Office, London.

Office for National Statistics. *Economic Trends Annual Supplement*, various issues, The Stationery Office, London.

Orphanides, A. (2001). 'Monetary policy rules based on real-time data', *American Economic Review*, 91, pp. 964-985.

Orphanides, A. (2003). 'Monetary policy evaluation with noisy information', *Journal of Monetary Economics*, 50, pp. 605-631.

Patterson, K. (2002). 'The data measurement process for UK GNP: stochastic trends, long memory, and unit roots', *Journal of Forecasting*, 21, 4, pp. 245-64.

Patterson, K. and Heravi, M. (1991). 'Data revisions and the expenditure components of GDP', *Economic Journal*, 101, pp. 887-901.

Pesaran, M. and Timmermann, A. (1992). 'A simple nonparametric test of predictive performance', *Journal of Business and Economic Statistics*, 10, pp. 461-465.

Roberds, W. and Whiteman, C.H. (1992). 'Monetary aggregates as monetary targets: A Statistical Investigation', *Journal of Money, Credit and Banking*, 24, 2, pp. 141-161.

Robinson, H. (2005). 'Revisions to quarterly GDP growth and its production (output), expenditure and income components', *Economic Trends,* 625, pp. 34-49.

Rudebusch, G.D. (2001). 'Is the Fed too timid? Monetary policy in an uncertain world', *Review of Economics and Statistics*, 83, pp. 203-217.

Schwarz, G. (1978). 'Estimating the dimension of a model', *Annals of Statistics*, 6, pp. 461-464.

Siklos, P.L. (1996). 'An empirical analysis of revisions in US macroeconomic aggregates', *Applied Financial Economics*, 6, pp. 59-70.

Siklos, P.L. (2006). 'What can we learn from comprehensive data revisions for forecasting inflation? Some US evidence', CIRANO conference paper, June.

Steel, M.F.J. and Richard, J.-F. (1991). 'Bayesian multivariate exogeneity analysis: An application to a UK money demand equation', *Journal of Econometrics*, 49, pp. 239-274.

Stock, J.H. and Watson, M.W. (1989). 'Interpreting the evidence on money-income causality', *Journal of Econometrics*, 40, pp. 161–182.

Stock, J.H. and Watson, M.W. (1999). 'Forecasting inflation', *Journal of Monetary Economics*, 44, pp. 293-335.

Stock, J.H. and Watson, M.W. (2003). 'Forecasting output and inflation: the role of asset prices', *Journal of Economic Literature*, 41, pp. 788-829.

Swanson, N.R. (1998). 'Money and output viewed through a rolling window', *Journal of Monetary Economics*, 41, pp. 455–473.

Topping, S.L. and Bishop, S.L. (1989). 'Breaks in monetary series', Bank of England Discussion Paper, Technical Series, No. 23.

Wroe, D. (1993). 'Improving macro-economic statistics', *Economic Trends*, 471, pp. 191-199.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley and Sons.

Appendix A: Predictive Densities for VARs and VECMs

Let the VAR be written as:

$$Y = XB + U \tag{A.1}$$

where Y is a $T \times n$ matrix of observations on the n variables in the VAR. X is an appropriately defined matrix of lags of the dependent variables, deterministic terms, etc.. B are the VAR coefficients and U is the error matrix, characterized by error covariance matrix Σ.

Based on these T observations, Zellner (1971, pages 233-236) derives the predictive distribution (using a common noninformative prior) for out-of-sample observations, W generated according to the same model:

$$W = ZB + V, \tag{A.2}$$

where B is the same B as in (A.1), V has the same distribution as U, etc. (see Zellner, 1971, chapter 8 for details of definitions, etc.). Crucially, Z is assumed to be known. In this setup, the predictive distribution is multivariate Student-t (see page 235 of Zellner, 1971). Analytical results for predictive means, variances and probabilities such as $\Pr(\Delta p_{t+h} < 5.0\%)$ can be directly obtained using the properties of the multivariate Student-t distribution. For other predictive features of interest, predictive simulation, involving simulating from this multivariate Student-t can be done in a straightforward manner.

The previous material assumed Z is known. How can we handle Z in our case? In the case of one period ahead prediction, $h = 1$, then Z is known. That is, in (A.1), if Y contains information available at time t, then X will contain information dated $t - 1$ or earlier. Hence, in (A.2) if W is a $t + 1$ quantity to be forecast, then Z will contain information dated t or earlier. But what about the case of h period ahead prediction where $h > 1$? Then Z is not known. But, following common practice, we can simply estimate a different VAR for each h. For $h = 1$ work with a standard VAR as described above. For $h > 1$, still work with a VAR defined as in (A.1), except let Y contain information at time t, but let X only contain information through period $t - h$ (i.e. let X contain lags of explanatory variables lagged at least h periods). All these predictive densities will be multivariate Student-t and, hence, their properties can be evaluated (either analytically or simply by simulating from the multivariate Student-t predictive density).

The preceding describes how we derive h step ahead predictive densities for VAR models. The VECM can be written as in (A.1) if we include in X the error correction terms (in addition to all the VAR explanatory variables). We replace the unknown cointegrating vectors which now appear in X by their MLEs. If we do this, analytical results for predictive densities can be obtained exactly as for the VAR. Note that this is an approximate Bayesian strategy and, thus, the resulting predictive densities will not fully reflect parameter uncertainty. We justify this approximate approach through a need to keep the computational burden manageable. Remember that we have 80 models (i.e. 40 VARs and VECMs, each of which has a variant with money and a variant without money) and six different data combinations (e.g. we have real-time and final vintage versions of our variables and use three different monetary aggregates). For each of the six different data combinations we have to do a recursive prediction exercise (involving up to 99 forecast periods). Furthermore, we have to do all this for $h = 1, 4$ and 8. In total, our empirical results involve posterior and predictive results for 315,360 VARs or VECMs. Thus, it is important to make modelling choices which yield analytical posterior and predictive results. If we had to use posterior simulation, the computational burden would have been overwhelming.

Appendix B: A Bayesian Variant of the Diebold-Mariano Statistic

To develop a Bayesian interpretation of various classical ways of assessing predictive accuracy, consider the approach of Diebold and Mariano (1995). This involves comparing the predictive performance of two models (call them models 1 and 2). Their approach is based upon the forecast errors, e_{1t} and e_{2t} for $t = 1, .., T$ for the two models. They let $g(e_{it})$ for $i = 1, 2$ be the loss associated with each forecast and suppose interest centers on the difference between the losses of the two models:

$$d_t = g(e_{1t}) - g(e_{2t}).$$

Diebold and Mariano derive a test of the null hypothesis of equal accuracy of the two forecasts. The null hypothesis is $E(d_t) = 0$. A test statistic they use is:

$$DM = \frac{\overline{d}}{\sqrt{var}},$$

where $\overline{d} = \frac{\sum d_t}{T}$ and var is an estimate of the variance of $\overline{d}$ used to normalize the test statistic so that it is asymptotically $N(0, 1)$. From a Bayesian point of view, we will simply take d_t as an interesting feature of interest; as a feature useful for providing evidence on whether model 1 or model 2 is forecasting better. We will ignore var (as it is merely a normalizing constant relevant for deriving classical asymptotic theory).

As a digression, it is worth noting that Diebold and Mariano's method assumes there are two models. We are dealing with many more than that. However, this is simple to deal with in one of two ways. First of all, we can simply say model 1 is "the best model with money included" and model 2 is "the best model with money excluded" and then we do have two models, conventionally defined. However, when doing BMA it is valid to interpret "the BMA average of all models with money included" as a single model and "the BMA average of all models with money excluded" as a single model.

Before beginning a discussion of a Bayesian analogue to the DM statistic, we stress that the DM statistic depends on the forecast errors and, as used by Diebold and Mariano, is based on a point forecast. Our Bayesian methods provide us with point predictions and, thus, we can simply use the DM test in exactly the same way as they do. Our statistical methods are then a combination of Bayesian methods (to produce the predictions) and classic methods (to evaluate the quality of the predictions).

Our fully Bayesian procedure goes beyond point forecasts and treats e_{it} as a random variable. That is, if y_t is the actual realized value of the dependent variable and y_{it}^* is the random variable which has the predictive density under model i, then:

$$e_{it} = y_t - y_{it}^*$$

is a random variable and d_t will also be a random variable. Remember that y_{it}^* has a t-distribution and, hence, if we treat y_t as a fixed variable (as the Bayesian would), e_{it} will also have a t-distribution. But d_t is a nonlinear function of t-distributed random variables and, hence, will not have a distribution of convenient form. Nevertheless, by using predictive simulation methods (i.e. drawing y_{it}^{*}s from the t-distributed predictive density and then transforming these draws as appropriate to produce draws of d_t) we can obtain the density of d_t which we label:

$$\Pr(d_t).$$

Remember that model 1 is better than model 2 (in terms of the loss function $g()$) if $d_t > 0$. So we can calculate:

$$\Pr(d_t > 0)$$

which will directly answer questions like "what is the probability that the model with money is forecasting better at time t?". If we average this over time:

$$\frac{\sum \Pr(d_t > 0)}{T}$$

we can shed light on the issue "are models with money predicting better overall than models without money?". To make the notation more compact, in the text we label this Bayesian metric as $\Pr(d_{t+h} > 0)$.

In our empirical work, we use a quadratic loss function.

Figure 1a: Output Growth (annualized percent, final vintage)

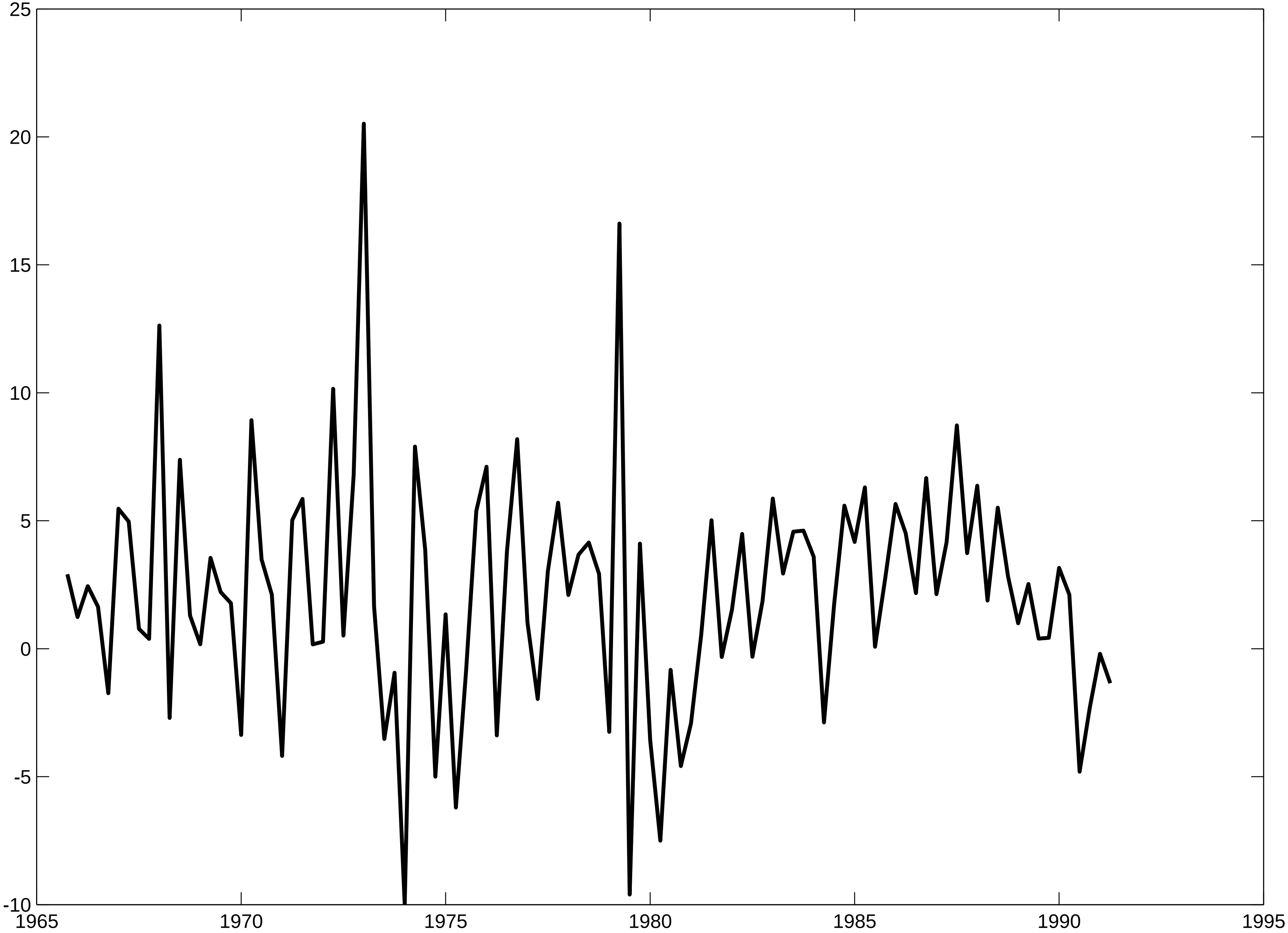

Figure 1b: Inflation (annualized percent, final vintage)

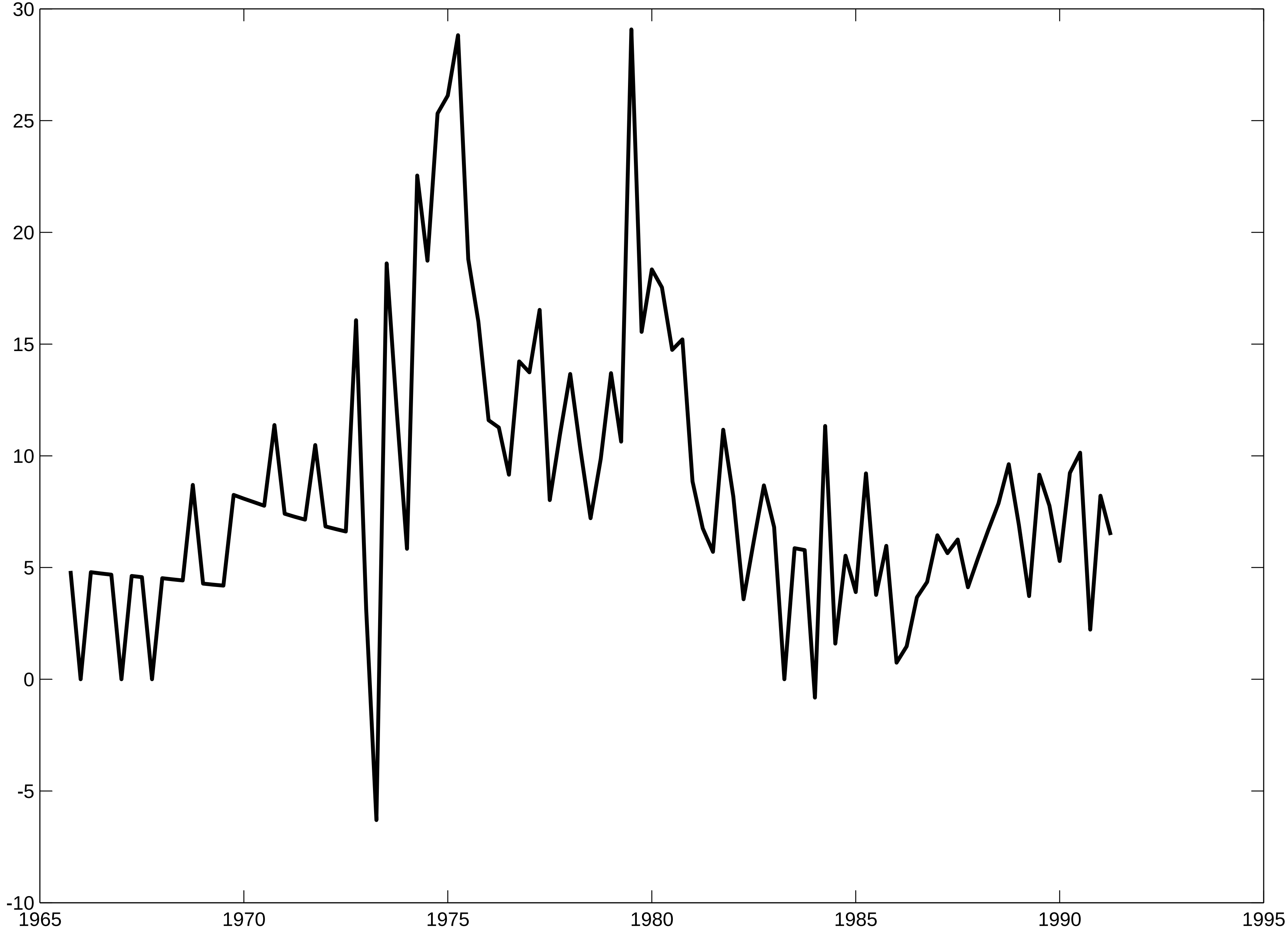

Figure 1c: M3 Growth (annualized percent, final vintage)

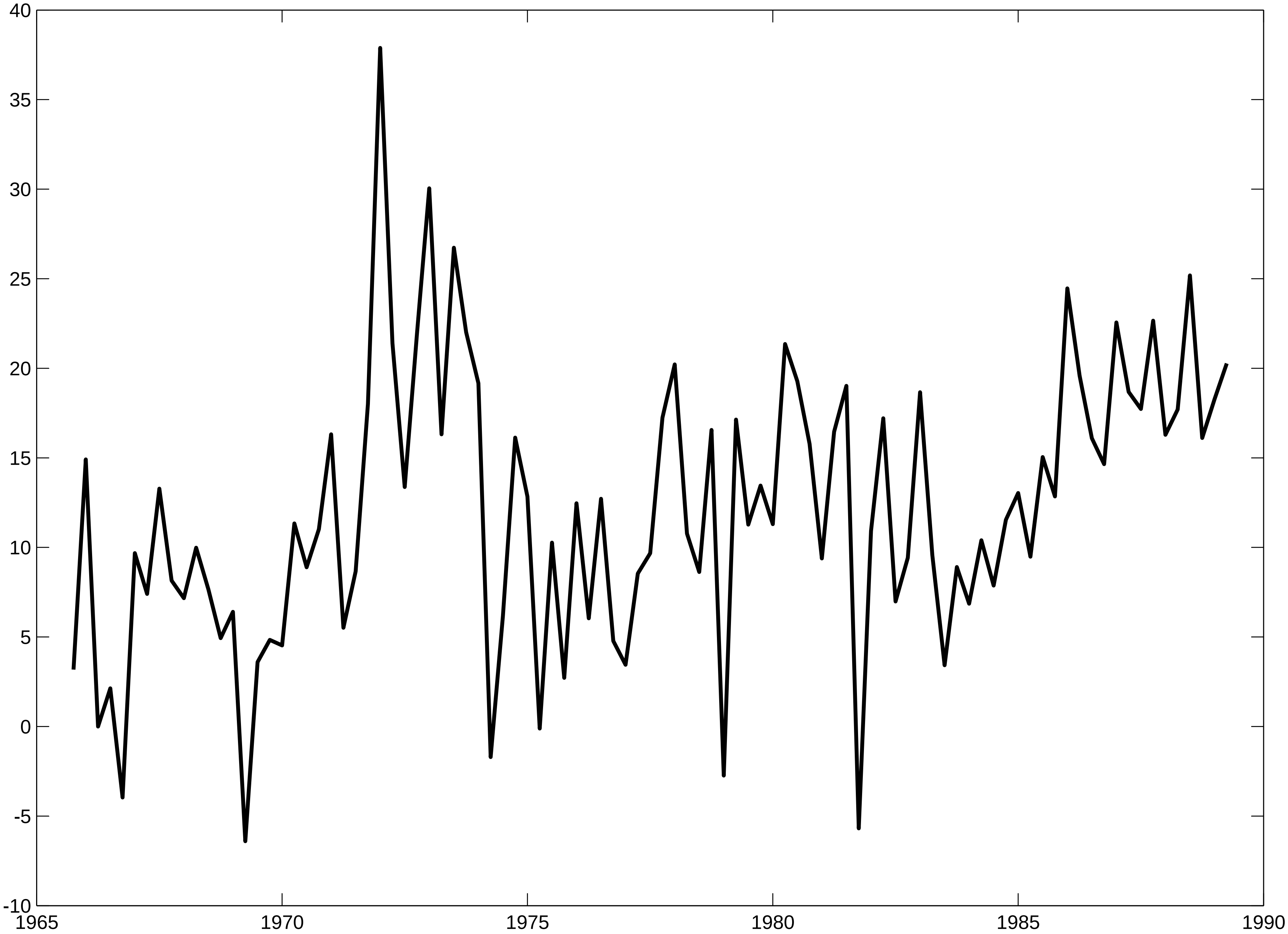

Figure 1d: Change in Effective Exchange Rate (annualized percent)

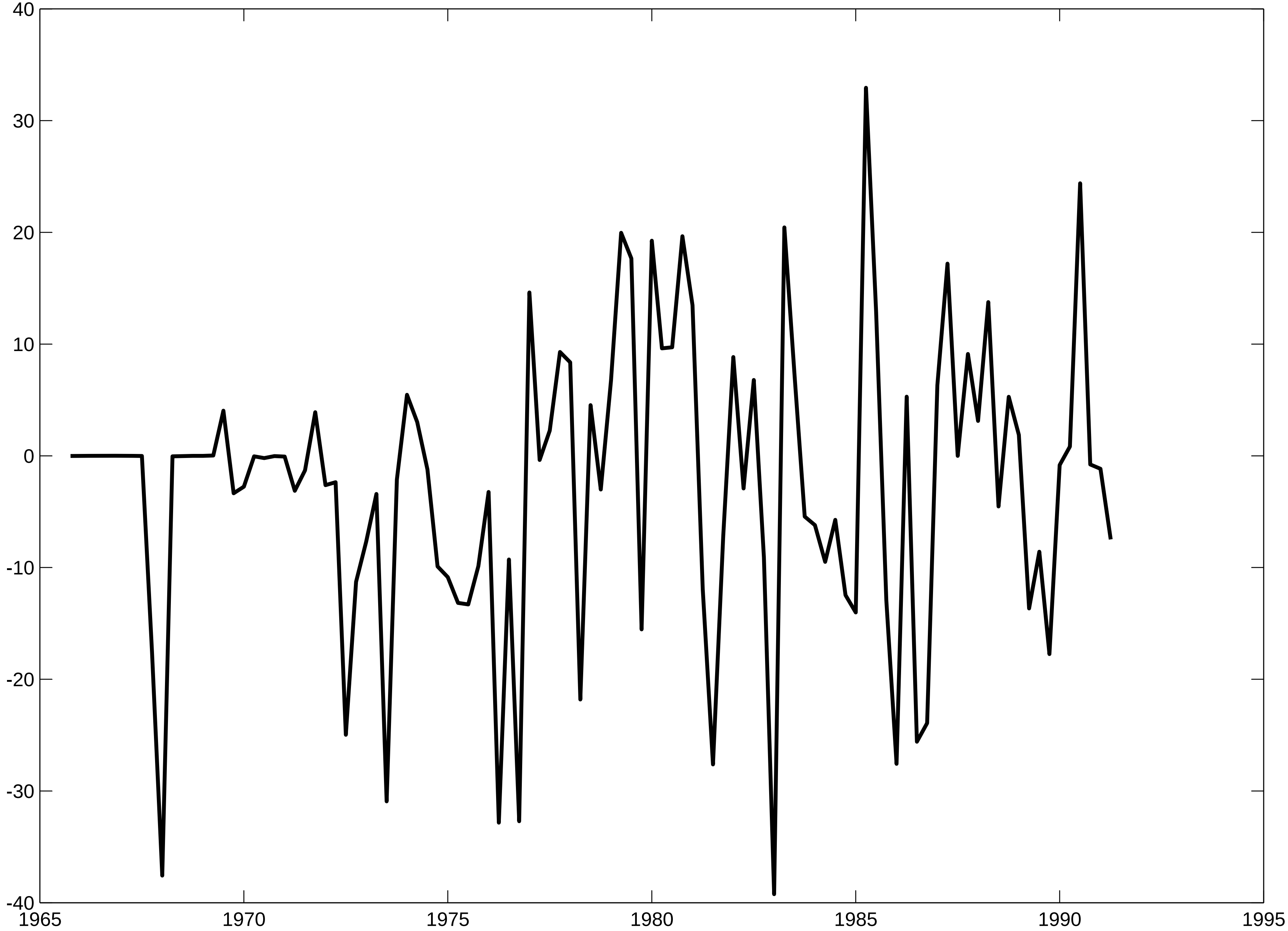

Figure 1e: Interest Rate (percent)

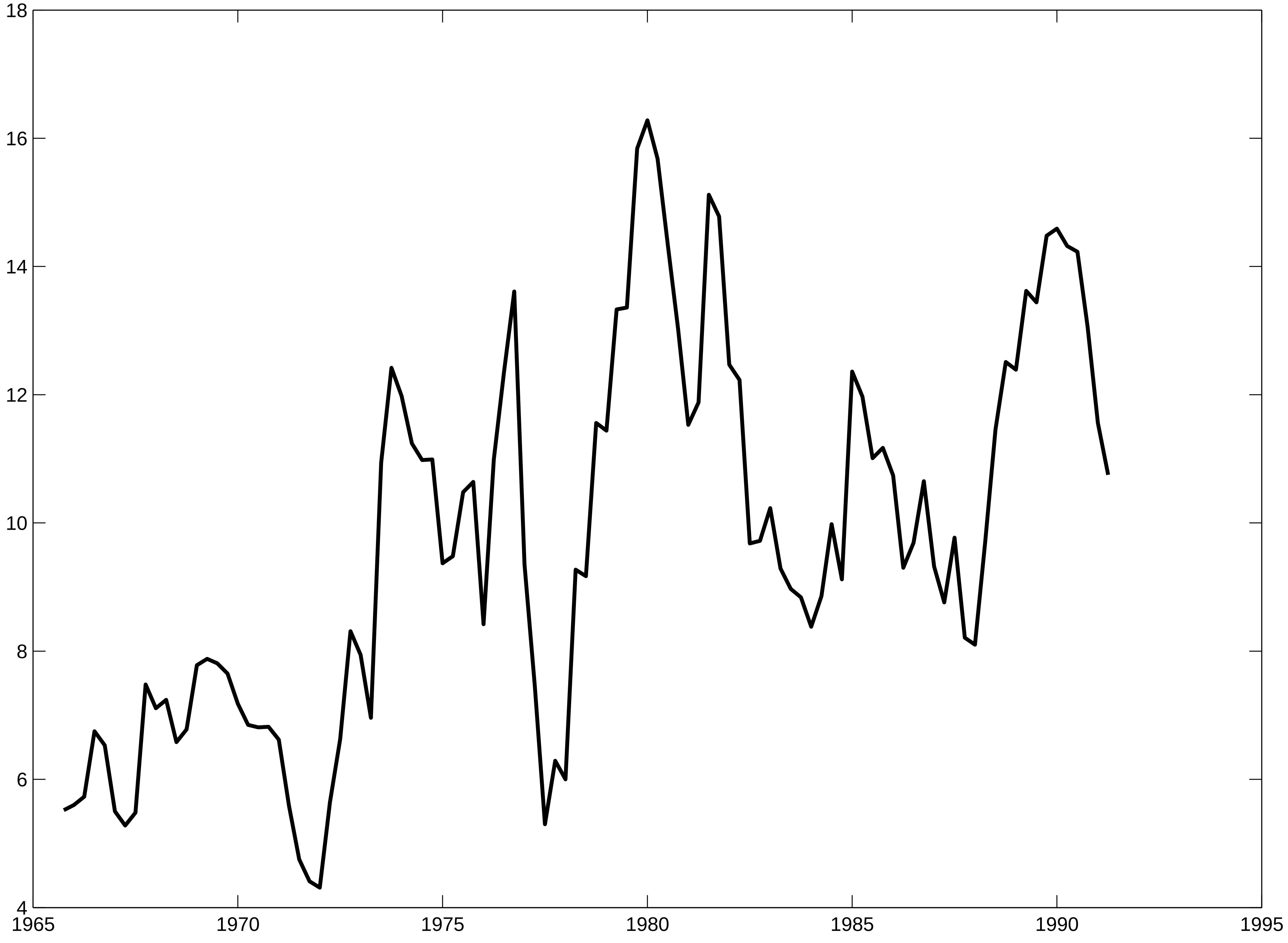

Figure 2: Probability of Various Models with M3

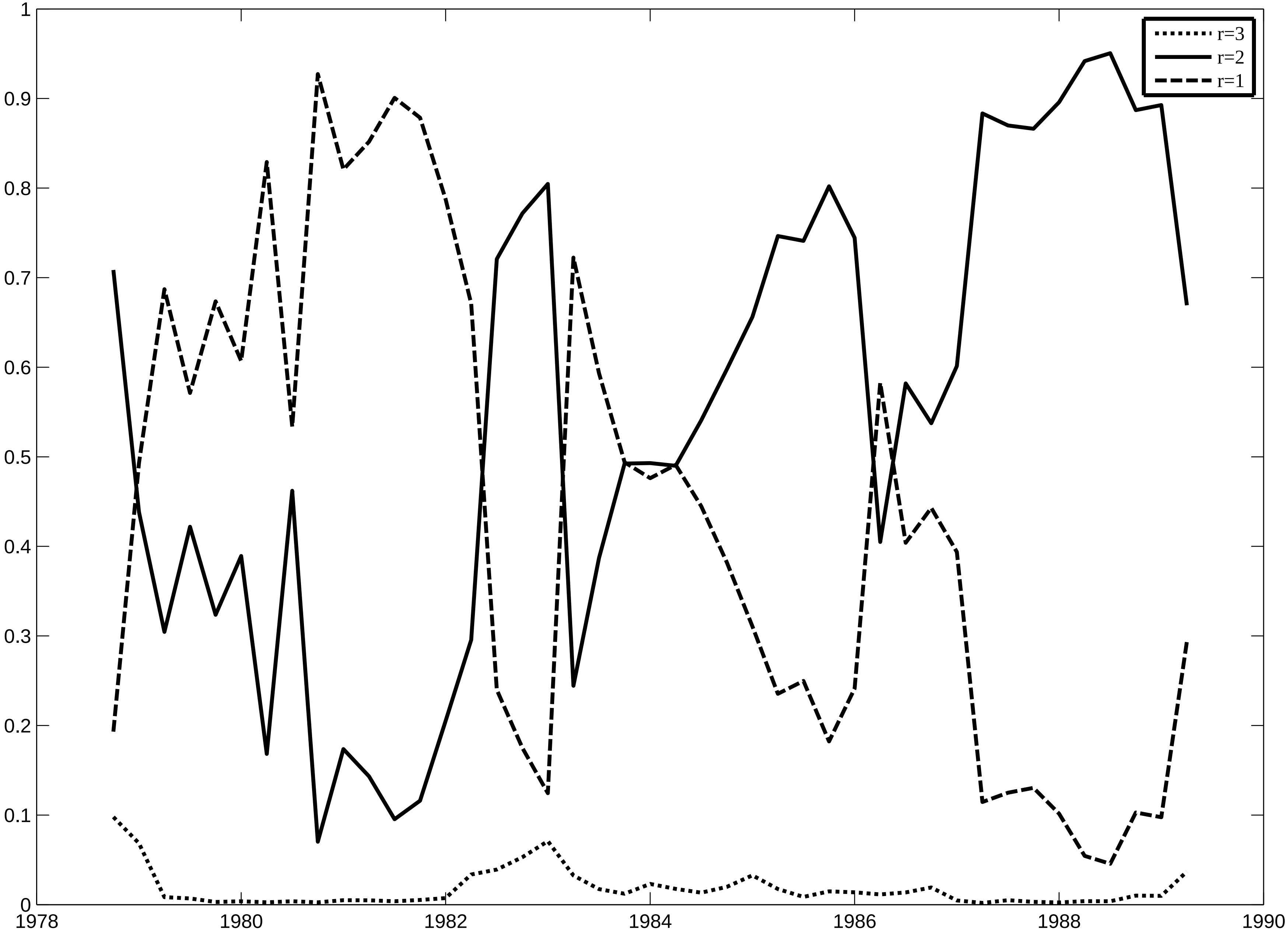

Figure 3: Probability M3 Predicts Output Growth

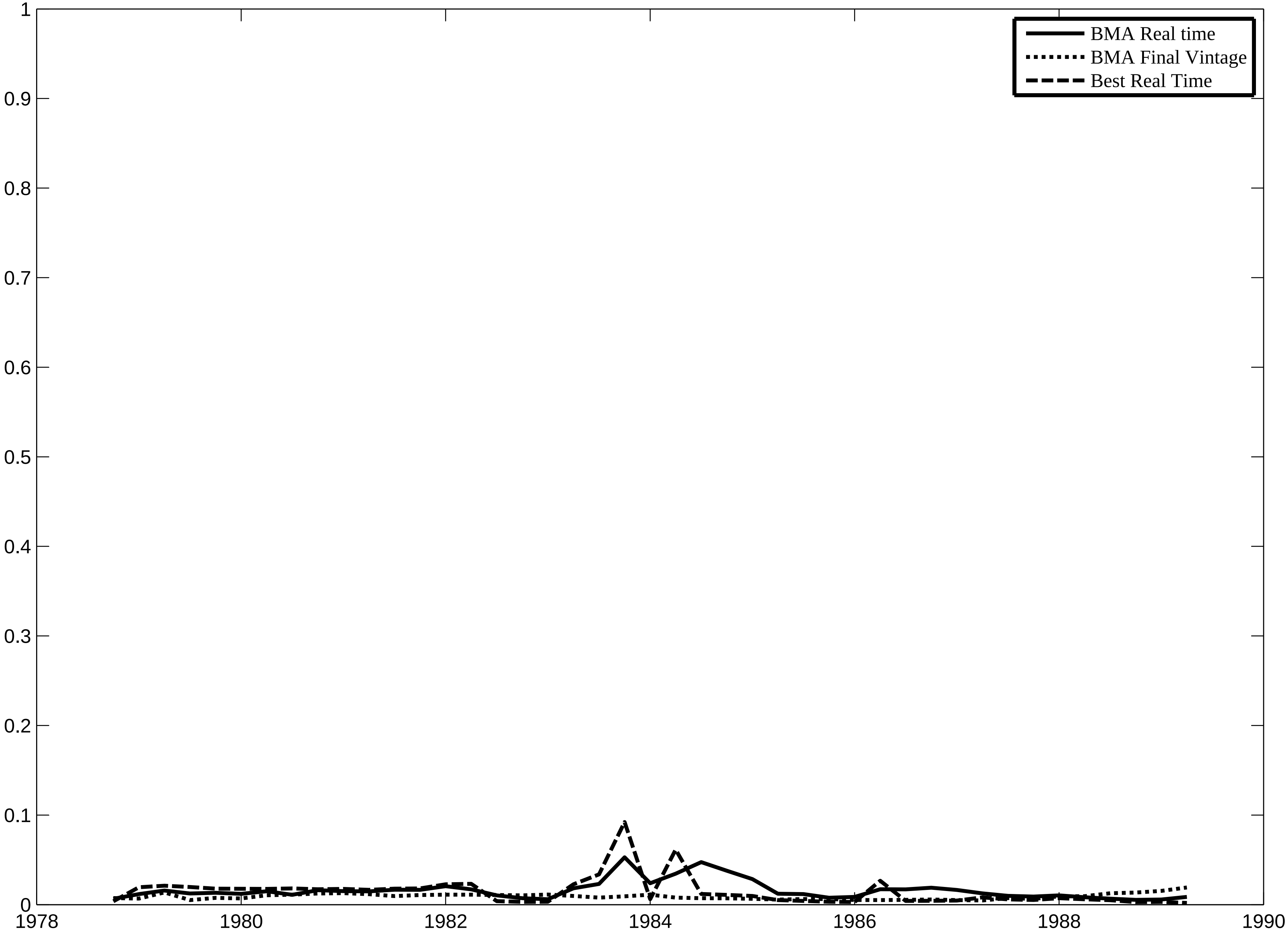

Figure 4: Probability M3 Predicts Inflation

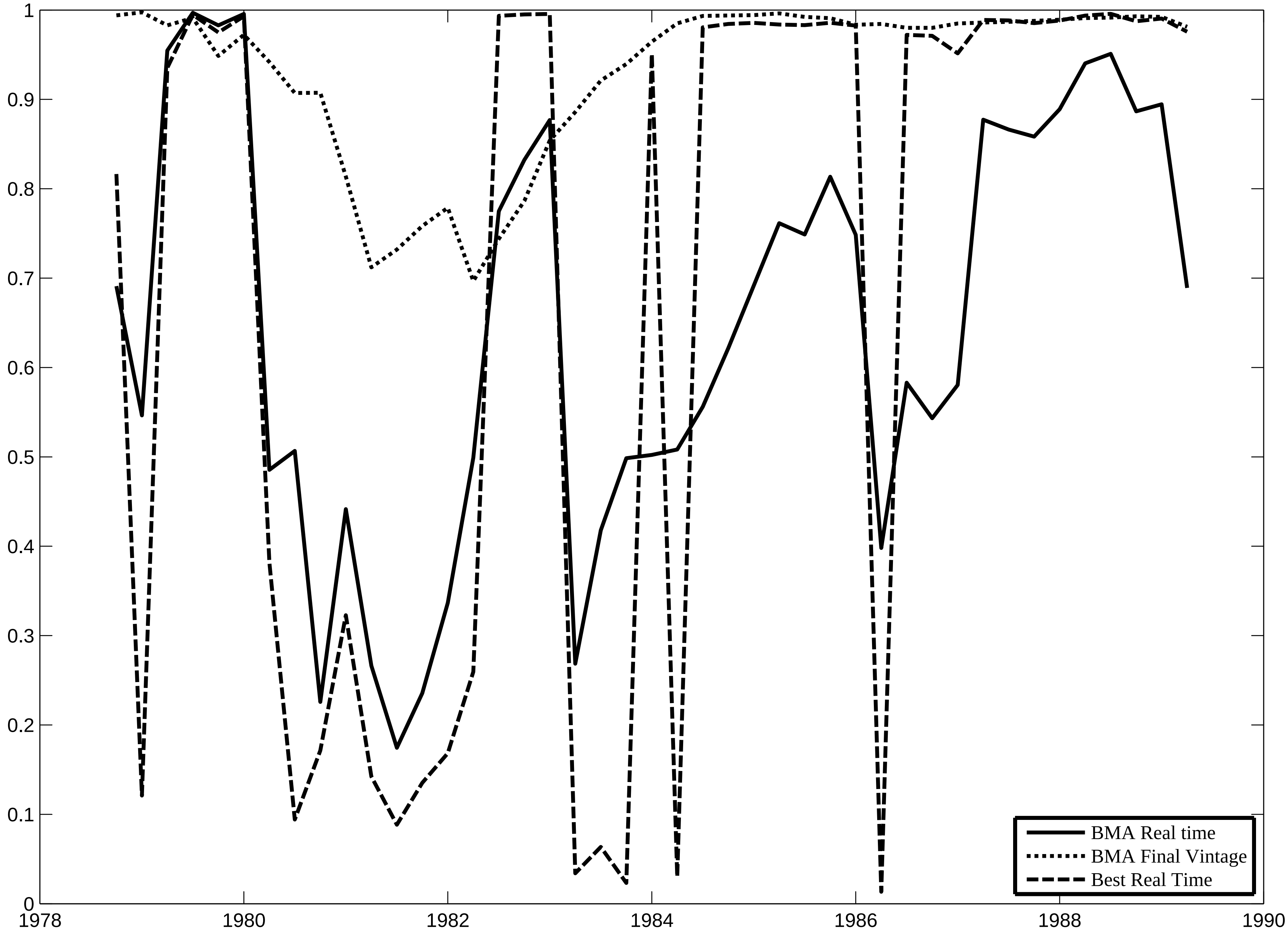

Figure 5: Probability of Inflation > 5 percent, h=8, Net Contribution of M3

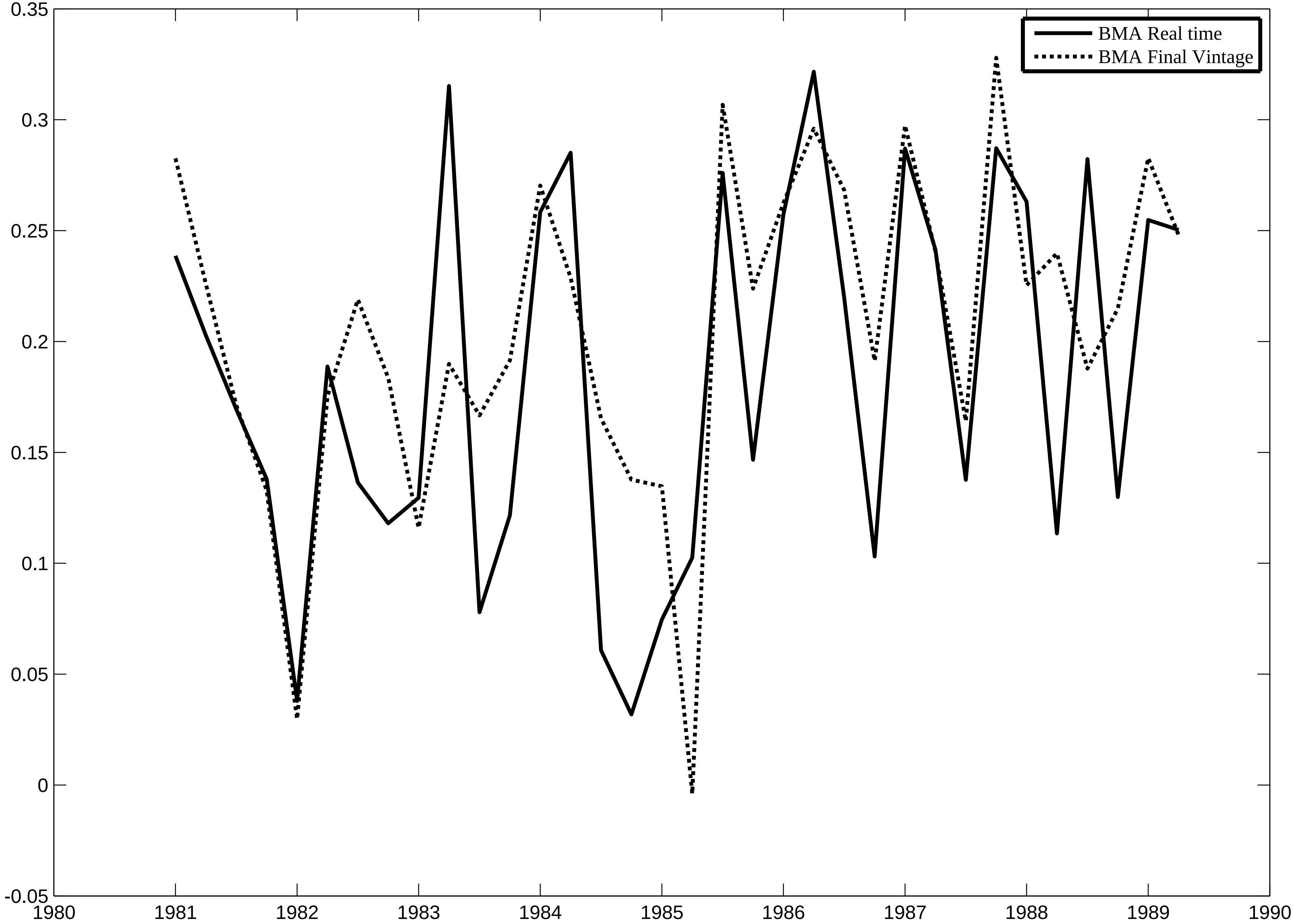

Figure 6: Probability of Output Growth > 2.3 percent, h=8, Net Contribution of M3

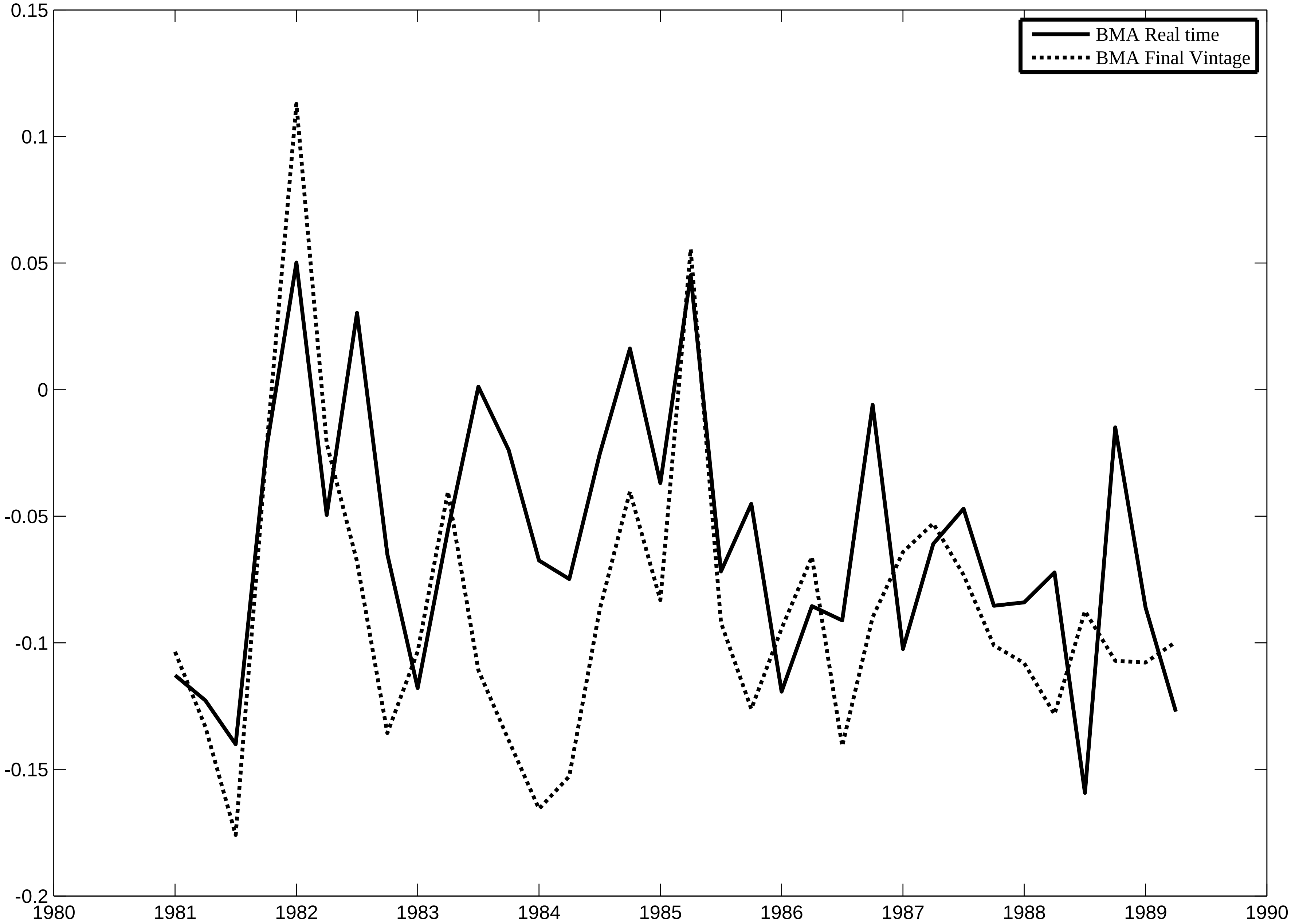

Figure 7: Probability M0 Predicts Output Growth

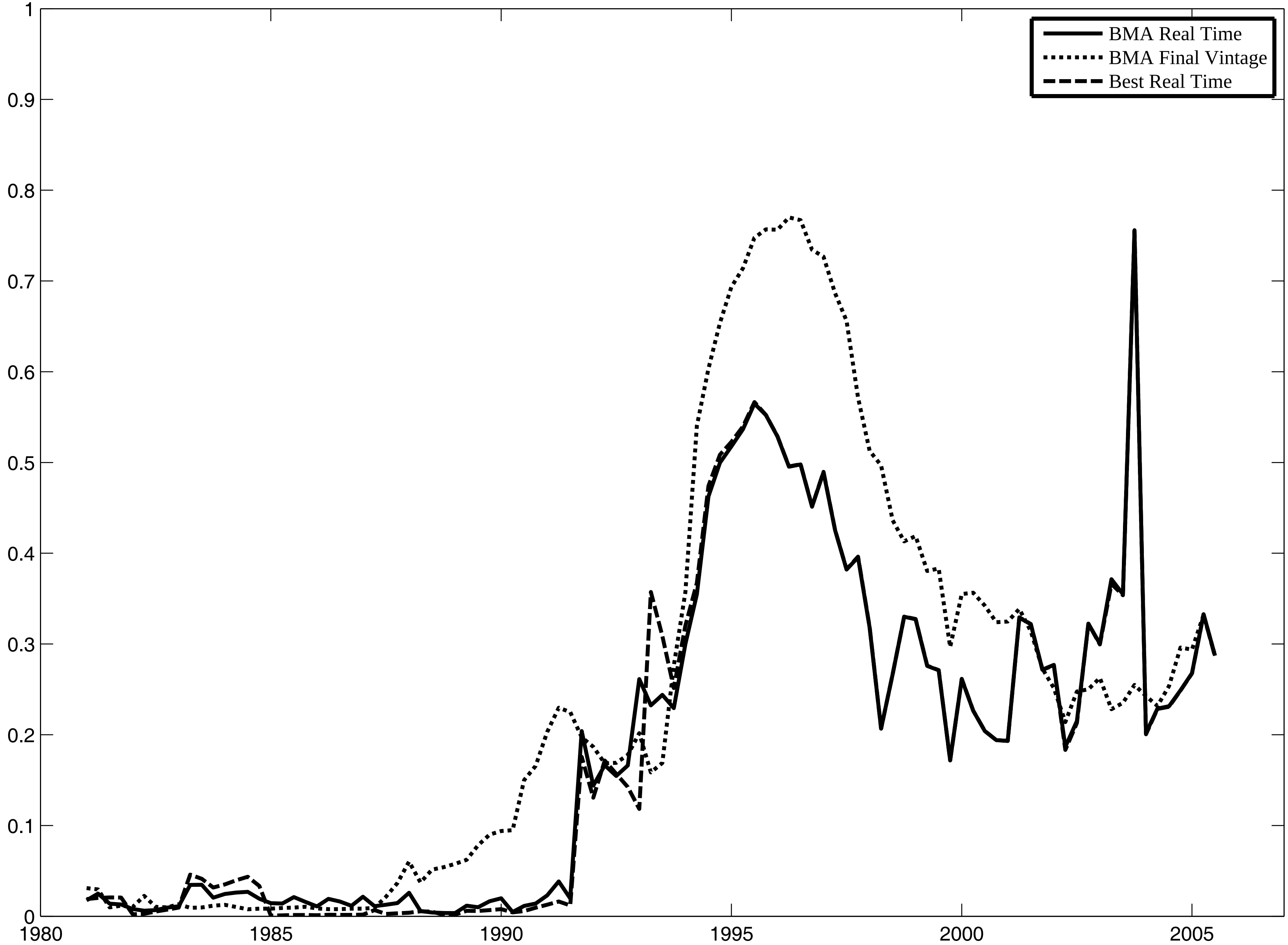

Figure 8: Probability M0 Predicts Inflation

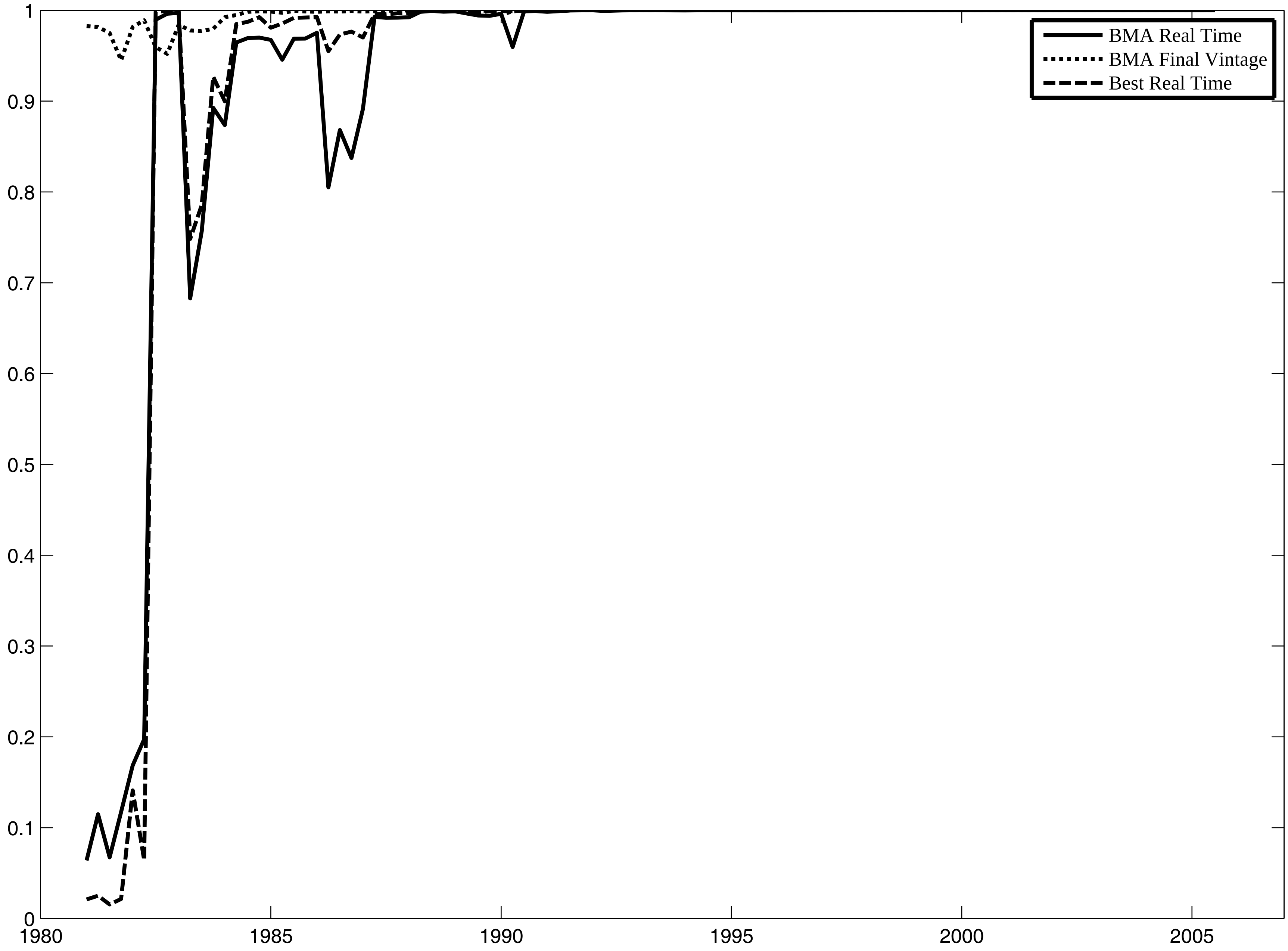

Figure 9: Probability M4 Predicts Output Growth

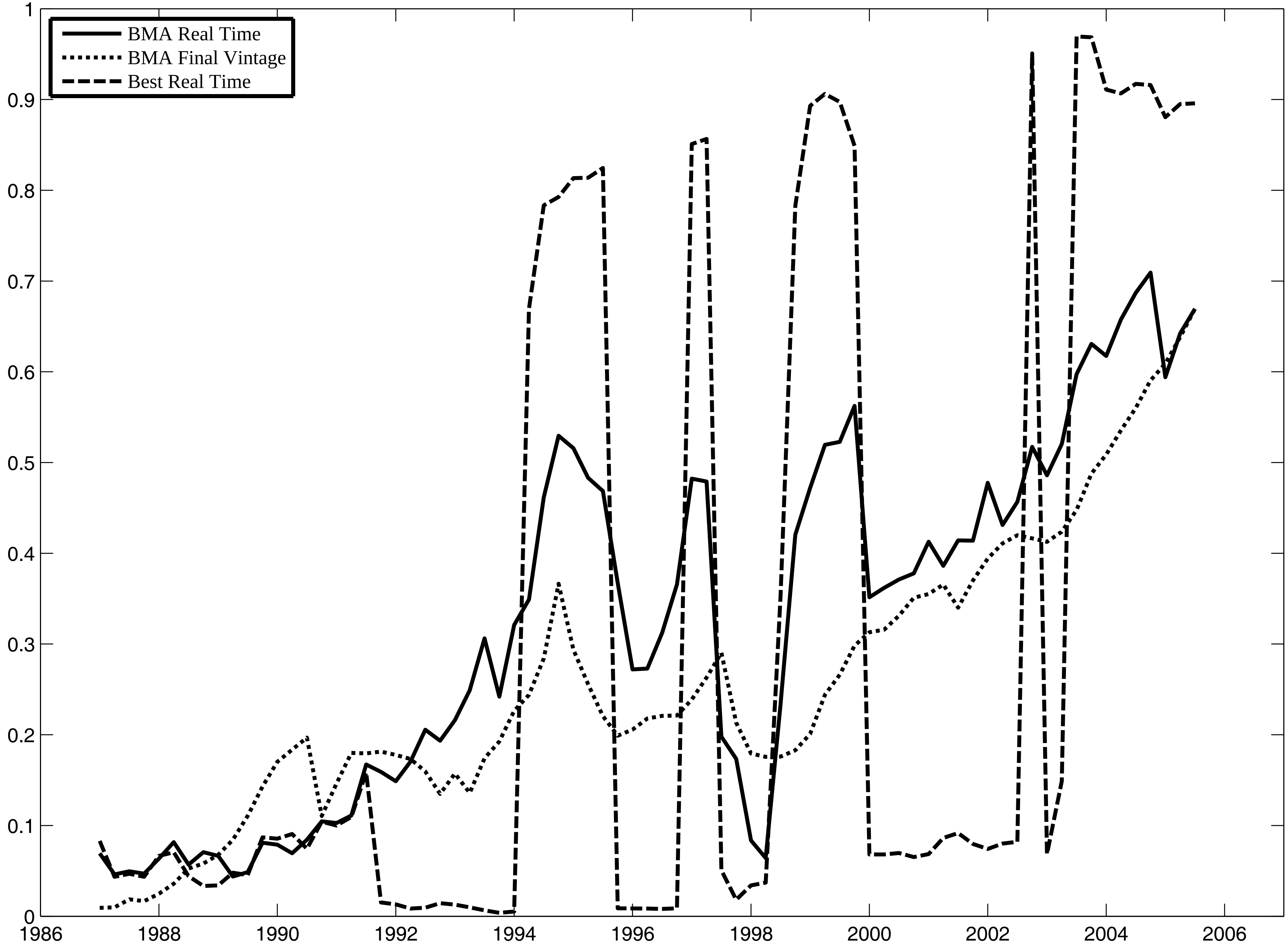

Figure 10: Probability M4 Predicts Inflation