DP2008/13

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

Anthony Garratt, Gary Koop, Emi Mise and Shaun Vahey

September 2008

JEL classification: C11, C32, C53, E51, E52

www.rbnz.govt.nz/research/discusspapers/

Discussion Paper Series

ISSN 1177-7567

DP2008/13

Real-time Prediction with UK Monetary Aggregates in the Presence of Model Uncertainty*

Anthony Garratt, Gary Koop, Emi Mise and Shaun Vahey†

Abstract

A popular account for the demise of the UK's monetary targeting regime in the 1980s blames the fluctuating predictive relationships between broad money and inflation and real output growth. Yet ex post policy analysis based on heavily-revised data suggests no fluctuations in the predictive content of money. In this paper, we investigate the predictive relationships for inflation and output growth using both real-time and heavily-revised data. We 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. We use Bayesian model averaging (BMA) to demonstrate that real-time monetary policymakers faced considerable model uncertainty. The in-sample predictive content of money fluctuated during the 1980s as a result of data revisions in the presence of model uncertainty. This feature is only apparent with real-time data as heavily-revised data obscure these fluctuations. Out of sample predictive evaluations rarely suggest that money matters for either inflation or real output. We conclude that both data revisions and model uncertainty contributed to the demise of the UK's monetary targeting regime.

---

* The views expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Reserve Bank of New Zealand. We are grateful to Todd Clark, Dean Croushore, Domenico Giannone, Marek Jarocinski, James Mitchell, Serena Ng, Simon van Norden, Adrian Pagan, Barbara Rossi, Glenn Rudebusch, Yongcheol Shin, Tom Stark, Christie Smith, Norman Swanson, Allan Timmermann and two referees for helpful comments. We acknowledge financial support from the ESRC (Research Grant No RES-000-22-1342). Alistair Cunningham (Bank of England) kindly provided the GDP data.

† Anthony Garratt, Birkbeck, University of London. Gary Koop, University of Strathclyde. Emi Mise, University of Leicester. Shaun Vahey, Melbourne Business School, Reserve Bank of New Zealand, and Norges Bank. Correspondence: a.garratt@bbk.ac.uk.

ISSN 1177-7567 ©Reserve Bank of New Zealand

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 fluctuating predictive 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. One account for this demise is that the relationships between money and output growth, or inflation, broke down or changed markedly over time.

In the empirical analysis that follows, we assess whether the predictive content of broad money fluctuated through the period of monetary targeting, using both UK real-time and heavily-revised final vintage data. To be precise, we investigate the predictability of UK inflation and output growth using the monetary aggregate $M3$ (the monetary target preferred by UK policymakers in the 1980s). We carry out a recursive analysis of whether the predictive content of money varies over time.

In terms of our set of models, we adopt a similar Vector Error Correction Model framework to 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).

Using BMA to allow for model uncertainty, we demonstrate that UK monetary policymakers faced considerable model uncertainty in real-time. In particular, there was ambiguity over the number of long-run relationships within the VECM systems. The in-sample predictive power of broad money is sensitive to the number of cointegrating terms. We demonstrate that data revisions repeatedly shifted the posterior model probabilities so that the overall in-sample predictive content of money fluctuated during the 1980s. This feature is apparent with real-time data, but not with heavily-revised data. That is, subsequent data revisions have removed the fluctuations in predictability that hindered contemporary policymakers. Out of sample predictive evaluations using heavily-revised data rarely suggest that money matters for either inflation or output growth. We conclude that both data revisions and model uncertainty contributed to the demise of UK monetary targeting. Subsequent to the discussion of the breakdown in UK monetary targeting (see, Leigh-Pemberton, 1986), the Governor of the Bank of England drew attention to the difficulties of demand management when confronted by data revisions; see, Leigh-Pemberton (1990).

Although, of course, the Bank of England in the 1980s was not using BMA, we feel that our BMA results might be a reasonable approximation of how UK monetary policymakers updated their views in the presence of model uncertainty. We contrast our BMA findings with those based on: (1) a frequentist recursive selection of the best model in each period; and, (2) a structural restriction of the model space motivated by the long-run money demand relationship. Both of these approaches are difficult to reconcile with the behavior of UK policymakers. Because the identity of the best model varies through time, this particular frequentist strategy generates very unstable beliefs. The structural approach we consider rules out the model space in which money matters.

Our paper relates to the large literature which investigates whether money has predictive power for inflation and output growth. Numerous studies have assessed money causation, conditional on other macroeconomic variables. Nevertheless, the evidence on the extent of the marginal predictive content of money remains mixed. For example, (among many others) Feldstein and Stock (1994), Stock and Watson (1989), Swanson (1998) and Armah and Swanson (2007) argued that US money matters for output growth; and Friedman and Kuttner (1992), and Roberds and Whiteman (1992) argued that it does not. Stock and Watson (1999, 2003) and Leeper and Roush (2003) 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. 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 frequentist econometric methodology of Amato and Swanson (2001) but also find a weak predictive relationship between money and output growth.

Whether model uncertainty compounds the real-time difficulties of assessing the predictive properties of money has not previously been studied. Egginton, Pick and Vahey (2002), Faust, Rogers and Wright (2005), Garratt and Vahey (2006) and Garratt, Koop and Vahey (2008) have shown that initial measurements to UK macroeconomic variables have at times been subject to large and systematic revisions. The phenomenon was particularly severe during the 1980s. But these papers do not discuss the predictability of money for inflation or output growth.

The main contribution of this paper is the use of Bayesian methods to gauge the model uncertainty apparent to real-time monetary policymakers. We demonstrate that UK policymakers in the 1980s faced a substantial degree of model uncertainty and that their assessments of whether money matters were clouded by data revisions. Setting aside our treatment of model uncertainty and its policy implications, our methodology remains close to Amato and Swanson (2001). We augment their core set of variables (comprising money, real output, prices, and the short-term interest rate) with the exchange rate to match the open economy setting of our UK application. We consider a similar set of VARs and VECMs; and, we use the real-time data in exactly the same fashion, to draw a contrast with the results from heavily-revised data.

The remainder of the paper is organized as follows. Section 2 discusses the econometric methods. Section 3 discusses real-time data issues. Section 4 presents our empirical results. In the final section, we draw some conclusions.

2 Econometric Methods

2.1 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 growth), then the Bayesian is interested in $\Pr(z|Data)$. The rules of conditional probability imply:

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

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 $\Pr\left(M_i|Data\right)$ (that is, the probability that model M_i generated the data) and $\Pr\left(z|Data, M_i\right)$ (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:

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

where $\Pr\left(Data|M_i\right)$ is referred to as the marginal likelihood and $\Pr\left(M_i\right)$ 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, $\Pr\left(M_i\right)$ 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 $\Pr\left(Data|M_i\right)$ which do not require the researcher to subjectively elicit a prior (see, eg, 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:

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

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 it 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 $\Pr\left(z|Data, M_i\right)$, 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 growth, and (ii) inflation. An explanation for how the predictive densities are calculated is given in Appendix A.

Remaining econometric details can be found in the working paper version, Garratt, Koop, Mise and Vahey (2007) [GKMV].

2.2 The Models

The models we examine are VECMs (and without error correction terms these become Vector Autoregressions (VARs)). 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 $\boldsymbol{\alpha} = \boldsymbol{I}_n$ and all the series do not have unit roots (i.e. this usually means they are stationary to begin with). Here 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 the monetary aggregate (m_t). Hence, $\boldsymbol{x}_t = (y_t, p_t, i_t, e_t, m_t)'$ (where we have taken the natural logarithm of all variables). When cointegration does not occur, we have a VAR (which we refer to generically as M_{var}). Consider the equation for real output:

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

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

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 growth" 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 inflation equation, and the "probability that money has no predictive content for inflation" can be obtained as described in the preceding paragraph.

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

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

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

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

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

To summarize, for every single (unrestricted) model, $M_1, \ldots, M_q$, we calculate the probability that money has no predictive power for output growth (or inflation) using (6) or (8). The probability that money has predictive power for output growth is one minus this. Our goal is to use BMA to assess whether overall "money has predictive power for output" by averaging over all the models. We achieve this by using the BICs for the unrestricted models as described above. 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 frequentist 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.

We emphasize that we use BMA over all the (unrestricted) models (denoted $M_1, \ldots, M_q$) and that, following Amato and Swanson (2001), our model space includes forecasting specifications with more than one cointegrating relationship, $r > 1$. In contrast, some studies of the predictive properties of money have restricted attention to models with a single long-run vector, $r = 1$, motivated by the (long-run) money demand relationship with constant velocity. Of course, in modeling the relationships between money and inflation and output growth, there are many candidate restrictions and variable transformations; Orphanides and Porter (2000) and Rudebusch and Svensson (2002) provide some examples which they associate with the views of contemporary US policymakers. For illustrative purposes, in the sections that follow, we consider one such "structural" specification. We report model averaged results (over each lag length for $p = 1, 2, ..., 8$) for the one cointegrating vector systems, with the long-run parameters imposed at the values implied by the money demand relationship. That is, we use the following long-run parameters (in the same order as in $\boldsymbol{x}_t$) of $(-1, -1, \beta_3, 0, 1)$, with the interest rate coefficient $\beta_3 > 0$ in each recursion. Since these models assume a constant money velocity of money, we refer to them simply as "VEL" models.

Both our BMA and VEL results utilize model averaging. We also use a popular method of recursive model selection in which we select a single best model in each time period using the BIC. We refer to this methodology as choosing the "best" models (i.e. one, possibly different, model is seleted in each time period).

As a digression, this "best" model selection strategy can be given either a Bayesian or a frequentist econometric interpretation as a model selection strategy (i.e. since BIC is commonly-used by frequentist econometricians). Note also that the Bayesian uses posterior model probabilities (i.e. $p\left(M_i|Data\right)$) to select models or use Bayesian model averaging. This holds true regardless of whether we are averaging over a set of possibly non-nested models (as we are doing when we do BMA) or calculating the probability that a restriction holds (as when we calculate the probability that money has in-sample predictive power for output growth or inflation). In the frequentist econometric literature, there has recently been concern about the properties of hypothesis testing procedures with repeated tests (as in a recursive testing exercise). See, among many others, Inoue and Rossi (2005) and Corradi and Swanson (2006). In particular, there has been concern about getting the correct critical values for such tests. In our Bayesian approach, such considerations are irrelevant, as the Bayesian approach does not involve critical values. For instance, in our in-sample results, we simply calculate the probability that money does not cause output growth at each point in time. We are not carrying out a frequentist econometric hypothesis test and problems caused with sequential use of hypothesis tests are not an issue. Similarly, with our recursive forecasting exercise, we are deriving the predictive density at each point in time and then calculating various functions of the resulting densities. Issues relating to differential power of in-sample versus out of sample power of frequentist hypothesis testing procedures (see, eg Inoue and Kilian, 2005) are not applicable. We are simply calculating the posterior or predictive probability of some feature of interest.

3 Data Issues

The case for using real-time data as the basis for policy analysis has been made forcefully by (among others) Orphanides (2001), Bernanke and Boivin (2003) and Croushore and Stark (2003). Since revising and re-basing of macro data are common phenomena, the heavily-revised measurements available currently from a statistical agency typically differ from those used by a policymaker in real time.

There are good reasons to suspect that the empirical relationships between money and inflation, and output growth might be sensitive to data measurement issues in the UK. Subsequent to the discussion of the break down in UK monetary targeting (see Leigh-Pemberton, 1986), the Governor of the Bank of England drew attention to the magnitude of revisions to UK demand-side macroeconomic indicators, including the national accounts; see Leigh-Pemberton (1990). An official scrutiny published in April 1989, known as the 'Pickford Report', recommended wide-ranging reforms to the data reporting processes. Wroe (1993), Garratt and Vahey (2006) and Garratt, Koop and Vahey (2008) discuss these issues in detail.

In our recursive analysis of the predictive content of money, we consider two distinct data sets. The first uses heavily-revised 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. We work with a "publication lag" of two quarters—a vintage dated time t includes time series observations up to date $t-2$. We use the sequence of real time vintages in exactly the same fashion as Amato and Swanson (2001), and like them, draw comparisons between our results using real-time and heavily-revised (final vintage) data.

Our data set of five variables in logs: y_t (seasonally adjusted real GDP), p_t (the seasonally adjusted implicit price deflator), i_t (the 90 day Treasury bill average discount rate), e_t (the sterling effective exchange rate), m_t (seasonally adjusted M3), runs from to 1963Q1 through 1989Q4. After this last date, M3 was phased out. Our data sources are the Bank of England's on-line real-time database and the Office of National Statistics. GKMV, the working paper version of this paper, provides much more motivation for, and explanation of, our data set and its sources.

Our empirical analysis assesses the predictive content of money for inflation and output growth, both in and out of sample, taking into account model uncertainty, as outlined in Section 2. Each model is defined by the cointegrating rank, r, and the lag length, p. We consider $r = 0, \ldots, 4$ and $p = 1, \ldots, 8$. Thus, we consider $5 \times 8 = 40$ models, $q = 40$ (where this model count excludes the VEL models). 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 heavily-revised final vintage data. We present results using the model averaging strategy described in Section 2. For comparative purposes, we also present results for the model with the highest BIC in each time period (which can be interpreted as a frequentist model selection strategy) and for the VEL models with theoretical restrictions imposed (a structural approach).

4 Empirical Results

Before beginning our Bayesian analysis, it is worth mentioning that a frequentist econometric analysis using the entire sample of final vintage data reveals strong evidence of model uncertainty. Using the likelihood ratio (LR) test, 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 cointegrating rank of 0, but if we choose $p = 1$, the trace test indicates a rank of 1 and the maximum eigenvalue test a rank of 2.

We present our empirical results in three sections. The first section examines the in-sample behavior of the various models. The second section focuses on whether money matters, with the systems recursively estimated from 1965Q4 to t, for t =1978Q4, . . . ,1989Q2 (43 recursions). We evaluate the probability that money has predictive power for output growth and inflation for each t. The third section examines out of sample prediction, for recursive estimation based on the sample 1965Q4 to t, where t =1981Q1, . . . ,1989Q2 (34 recursions). Remember that $M3$ was phased out in the 1989Q4 vintage–the last recursion uses data up to 1989Q2 (given the publication lag of two quarters). GKMV contains results using $M0$ and $M4$ instead of $M3$.

4.1 Model Comparison

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 the models in our recursive BMA exercise based on the real-time data. For each time period, the same three models almost always receive the vast majority of the posterior probability. These three preferred models differ in the number of cointegrating relationships but all have the same lag length, $p = 1$, and hence, we focus on the uncertainty about the number of cointegrating relationships.

Figure 1 plots the probability of $r = 1, 2$ and 3 cointegrating relationships (the probability of $r = 0$ is approximately zero). 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 1 is that the number of cointegrating vectors varies over time: the degree of model uncertainty that confronted policymakers in real time is substantial. It is rare for a single model to dominate (for example, one value of r hardly ever receives more than 90% of the posterior model probability), and the model with highest probability varies 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, during the mid 1980s, the $r = 2$ and

$r = 1$ models are often equally likely, with several switches in the identity of the preferred model.

Faced with this considerable real-time model uncertainty, it seems reasonable that a monetary policymakers may wish to (implicitly or explicitly) average over different models, as is done by BMA. However, if a policymaker ignores the evidence from less preferred systems, then a strategy of simply selecting the sequence of best models is possible. In our recursive exercise, the best model varies over time. So the commonly-used frequentist approach of selecting a single model using the BIC, and then basing forecasts on MLE's involves many shifts between the $r = 1$ and $r = 2$ specifications. Typically, the evidence in Figure 1 suggests the most preferred model received between 90% and 50% weight. For a couple of periods, the best model receives less than 50% weight.

Turning to the more structural approach where we restrict the long-run parameters to be those implied by the money demand relationship (the VEL models), it is clear that restricting attention to specifications with $r = 1$ has a substantial impact on the empirical relevance of the models. Figure 1 demonstrates that for much of our sample, the evidence rarely provides very strong support the $r = 1$ specification. Even though the $r = 1$ models were typically preferred before 1985, they received less that 90% of the posterior probability in nearly all periods. During 1986—when monetary targeting was abandoned—the probability that the restriction holds was roughly 50%. The $r = 1$ model receives less than 40% weight from 1987 to the end of the sample.

4.2 In-sample Prediction

In this section, we examine the in-sample ability of money to predict inflation and output growth in our recursive exercise. Figure 2 plots the probability that money can predict output growth; and Figure 3 shows the corresponding plot for predicting inflation. All probabilities are calculated using the approach described in Section 2. 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. Hence, two of the lines use real-time data, and the third line has the advantage of hindsight. Results for the best model using final vintage data exhibit a similar pattern to BMA results, but are slightly more erratic and are omitted to clarify the graphs. 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 2 shows that $M3$ has no predictive power for output growth, 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 2 ends.

Figure 3 presents the probability that money has predictive power for inflation. The lines in this plot exhibit fluctuations. These are particularly pronounced for the best models with real-time data. The probability that money can predict inflation swings rapidly from near zero to near one several times in the 1980s. A policymaker who simply selects the best model at each point in time could conclude that money mattered in some periods, but then next quarter it did not matter all. These switches between times when money matters, and when it does not, occur with embarrassing frequency.

Model averaging yields a less volatile pattern. In this case, there are repeated fluctuations, but the shifts in probability are not as sharp as with the best model selection strategy. The probability that money matters for inflation increased for much of the 1980s, but there were quarters with rapid declines. In particular, 1980 and 1984 saw distinct drops in predictability before the official demise of monetary targeting in 1986. Furthermore, in 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 data, fell to around 0.4 (plotted as 1986Q2) from approximately 0.8 in the previous quarter. We conclude that the real-time evidence in support of a relationship between money and inflation was prone to fluctuations.

With final vintage data, however, the BMA approach reveals no marked deterioration in the predictive power of money during the mid-1980s. The fluctuations in predictability discussed by the Governor of the Bank of England (Leigh-Pemberton, 1986) are absent. Instead, following a fall to around 0.7 in the early 1980s, the probability that money matters for inflation rises back to approximately one by mid-1984. Comparing the real-time and heavily-revised (final vintage) evidence, we conclude that subsequent data revisions have removed the fluctuations in predictability that so concerned contemporary UK policymakers in real time. The results presented in GKMV confirm that in-sample fluctuations in probabilities were not confined to the $M3$ monetary aggregate, the relationship with inflation, or the period of monetary targeting. Data revisions in the presence of model uncertainty caused large and repeated fluctuations in the predictability of money for other UK macro indicators.

The fluctuations with real-time data displayed in Figure 3 result from the changes in the model weights, plotted in Figure 1. The peaks and troughs in prediction probabilities almost exactly match the turning points for $r = 2$ weights. The two cointegrating relationships system gives very high probabilities that money matters for inflation in almost all periods. The single long run relationship case suggests no relationship. (Hence, we have not plotted the VEL case in Figures 2 and 3: the probabilities are zero for almost every period.)

It has been noted that using heavily-revised data could distort the predictive ability of broad money since data revisions might be aimed at strengthening the link between macroeconomic variables (see Diebold and Rudebusch, 1991, and Amato and Swanson, 2001). There were minor changes to UK $M3$ during the 1980s (for example, due to the status of institutions recording money, and various privatizations). However, these small changes (typically smaller than 0.2% of the level) do not coincide with upward movement in the probabilities reported in Figures 2 and 3. It was not revisions to money, but rather revisions to UK National Accounts that caused the divergence between the real-time and final vintage probabilities displayed in Figure 3.

4.3 Out of Sample Prediction

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, 2006, 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 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. 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 for real-time out of sample prediction, although this makes evaluations less timely.

With these issues in mind, and given the short sample of UK data available with the $M3$ definition of money, 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 heavily-revised (final-vintage) data by comparing them to the "actual" outcome. We use final vintage data for this "actual" outcome for the results presented (although we have experimented with other measures of "actual" outcome, discussed briefly below). We emphasize that these measures of predictive performance are not timely indicators—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 model averaging (BMA), the strategy of selecting the single best model (Best) and the structural approach (VEL). Separate sets of comparisons are made for real-time and final vintage data.

Technical details of the forecasting methodology are provided in the appendix. 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 $\Pr\left(\Delta p_{t+h}|Data_t\right)$ which averages over all the models. This is what the BMA results below are based on. Our best model selection strategy provides us with a predictive density $\Pr\left(\Delta p_{t+h}|Data_t, M_{Best}\right)$ 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.). The structural results using the VEL models presented here use model averaging over the $r = 1$ space (with some long-run parameters fixed) using the methods outlined in Section 2 (selecting the best structural model gives similar results).

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) present results for output growth using real-time use final vintage data respectively.

We begin by discussing the results relating to inflation 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 relative to the RMSE from BMA without money. A number less than one indicates an improved forecast performance relative to this case (ie 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, select the best model, or take the structural approach.

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. 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 realization is also 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 for inflation 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. However, when money is included the hit rate rises to 65%. In contrast, with final vintage data, the inclusion of money causes little change in the hit rates, except at $h = 1$.

In summary, we find 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, the best single model and the VEL approach.

The row labelled $\Pr(\overline{DM} > 0)$ contains a Bayesian variant of the popular DM statistic of Diebold and Mariano (1995). Details are given in Appendix B. The basic idea is that we calculate a statistic which is a measure of

Table 1
Evaluation of Out of Sample Forecast Performance

Horizon, h		No Money 1	4	8	With Money 1	4	8
(a) Inflation Real Time							
BMA	RMSE	1.00	1.00	1.00	1.19	1.40	1.50
	Hit Rate, $Pr(\Delta p_{t+h} < 5.0\%)$	0.68	0.53	0.65	0.74	0.65	0.65
	$\Pr(\overline{DM} > 0)$				0.49	0.48	0.43
Best	RMSE	1.00	1.00	1.00	1.21	1.41	1.50
	Hit Rate, $Pr(\Delta p_{t+h} < 5.0\%)$	0.68	0.53	0.65	0.71	0.62	0.65
	$\Pr(\overline{DM} > 0)$				0.49	0.49	0.43
VEL	RMSE	1.00	1.00	1.00	1.02	1.14	1.30
	Hit Rate, $Pr(\Delta p_{t+h} < 5.0\%)$	0.68	0.65	0.65	0.65	0.65	0.65
	$\Pr(\overline{DM} > 0)$				0.49	0.49	0.47
(b) Inflation Final Vintage							
BMA	RMSE	1.00	1.00	1.00	1.14	1.43	1.78
	Hit Rate, $Pr(\Delta p_{t+h} < 5.0\%)$	0.68	0.62	0.65	0.56	0.65	0.65
	$\Pr(\overline{DM} > 0)$				0.49	0.48	0.38
Best	RMSE	1.00	1.00	1.00	1.21	1.46	1.58
	Hit Rate, $Pr(\Delta p_{t+h} < 5.0\%)$	0.68	0.62	0.65	0.56	0.65	0.65
	$\Pr(\overline{DM} > 0)$				0.49	0.49	0.39
VEL	RMSE	1.00	1.00	1.00	1.29	1.07	1.31
	Hit Rate, $Pr(\Delta p_{t+h} < 5.0\%)$	0.65	0.65	0.65	0.62	0.65	0.65
	$\Pr(\overline{DM} > 0)$				0.49	0.49	0.46

Table 1 (cont'd)

Horizon, h		No Money 1	4	8	With Money 1	4	8
(c) Output Growth Real Time							
BMA	RMSE	1.00	1.00	1.00	1.06	1.09	1.12
	Hit Rate, $Pr(\Delta y_{t+h} < 2.3\%)$	0.41	0.68	0.38	0.35	0.68	0.47
	$\Pr(\overline{DM} > 0)$				0.49	0.49	0.48
Best	RMSE	1.00	1.00	1.00	1.07	1.11	1.13
	Hit Rate, $Pr(\Delta y_{t+h} < 2.3\%)$	0.41	0.68	0.38	0.38	0.68	0.47
	$\Pr(\overline{DM} > 0)$				0.49	0.49	0.49
VEL	RMSE	1.00	1.00	1.00	1.02	1.06	1.18
	Hit Rate, $Pr(\Delta y_{t+h} < 2.3\%)$	0.41	0.68	0.38	0.38	0.68	0.35
	$\Pr(\overline{DM} > 0)$				0.49	0.49	0.48
(d) Output Growth Final Vintage							
BMA	RMSE	1.00	1.00	1.00	1.08	1.10	1.25
	Hit Rate, $Pr(\Delta y_{t+h} < 2.3\%)$	0.47	0.71	0.38	0.59	0.65	0.47
	$\Pr(\overline{DM} > 0)$				0.49	0.49	0.48
Best	RMSE	1.00	1.00	1.00	1.11	1.13	1.25
	Hit Rate, $Pr(\Delta y_{t+h} < 2.3\%)$	0.47	0.71	0.38	0.59	0.65	0.47
	$\Pr(\overline{DM} > 0)$				0.48	0.49	0.48
VEL	RMSE	1.00	1.00	1.00	0.96	1.21	1.08
	Hit Rate, $Pr(\Delta y_{t+h} < 2.3\%)$	0.47	0.71	0.38	0.47	0.68	0.35
	$\Pr(\overline{DM} > 0)$				0.50	0.49	0.48

Notes: RMSE denotes Root Mean Square Forecast Error, relative to the benchmark without money. The Hit Rate is the proportion of correctly forecast events, with a correct forecast defined by the probability of the outturn greater than 0.5. The probability $Pr(\overline{DM} > 0)$ is described in Appendix B.

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 it is positive then models with money are forecasting better than models without money. Hence, if $\Pr(\overline{DM} > 0) > 0.5$ (that is a positive value of $\overline{DM}$ is more likely than not) we find evidence in favor of of money having predictive power for inflation (output growth). As shown in Table 1, we find no evidence that the inclusion of money improves forecast performance for inflation. This statement holds true regardless of whether we are using BMA, selecting a single model or using the more structural VEL models, and regardless of whether we are using final vintage or real-time data.

We turn now to output growth. In panels (c) and (d) of Table 1, we report the same set of forecast evaluation statistics for output growth. Our hit rates are based on the probability $\Pr\left(\Delta y_{t+h} < a | Data_t\right)$ 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. As with inflation, there is a strong similarity of the RMSE results for the statistical strategies of BMA, selecting the single best models and using the structural approach VEL.

The same conclusion, that the inclusion of money makes relatively small differences, can be drawn from the hit rates. Depending on the horizon, we see a slight worsening through to a slight improvement. For example using BMA and real-time data, we see that including money 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.

Our Bayesian DM statistic provides more evidence for the story that including money does not improve output growth forecasts. For all of our models and approaches, the probability that this statistic is positive (that is, models with money have better forecast performance) is always very near to 0.5, indicating roughly equal forecast performance between models with and without money.

We note that the performance of the various modelling approaches and data types are typically robust to the alternative definitions of the outturn which we use to evaluate forecast performance. In Table 1, we use final vintage measurements as the "actual" outturn. For the sake of brevity, we do not include results for other outturns (eg the first release outturn, or the outturn three years later). But results (available on request) are very similar to those in Table 1. An exception to this is that the RMSE performance with money improves with first-release outturns. However, this exception is limited to the RMSE measure—beyond point forecasts, the results suggest that money does not matter.

The overall impression from our out of sample prediction analysis with the $M3$ monetary aggregate is that the 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 makes substantial differences to out of sample prediction, either with real-time or final vintage data.

5 Conclusions

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

Our empirical results are divided into in and out of sample components. With regards to in-sample results, using the real-time data, we find that the predictive content of $M3$ fluctuates throughout the 1980s. 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 fluctuate substantially during the 1980s. The $M3$ monetary aggregate provides little help in predicting output growth at any point. But we stress that results using final vintage data require the benefit of hindsight about data revisions. With the data that would have been available at the time, the salient feature is the fluctuations in the probability that money can predict inflation.

Our out of sample forecasting analysis suggests no strong evidence that $M3$ matters for inflation or output growth, either with real-time or final vintage data.

UK policymakers have argued that fluctuations in the relationships between broad monetary aggregates and inflation and output growth undermined their faith in the UK's monetary targeting regime. The results in this paper indicate that they were right to worry about this issue. Issues relating to model uncertainty and data revisions have an important influence on beliefs about whether money matters.

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. (2007). "Predictive inference under model misspecification with an application to assessing the marginal predictive content of money for output", in Forecasting in the Presence of Structural Breaks and Model Uncertainty, eds. Mark Wohar, Elsevier, Amsterdam..

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

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.

Clark, T.E. and McCracken, M.W. (2006). "The predictive content of the output gap for inflation: Resolving in-sample and out of sample evidence", *Journal of Money, Credit, and Banking*, 38, No. 5, pp. 1127-1148.

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. and Stark, T. (2003). "A real-time data set for macroeconomists: does the data vintage matter?", *Review of Economics and Statistics*, 85, pp. 605-617.

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.

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., Mise, E. and S.P. Vahey (2007), "Real-time prediction with monetary aggregates in the presence of model uncertainty", *Birkbeck working paper No. 0714*, available at http://www.ems.bbk.ac.uk/research/wp.

Garratt, A., Koop, G. and Vahey, S.P. (2008). "Forecasting substantial data revisions in the presence of model uncertainty", *Economic Journal* forthcoming and RBNZ Discussion Paper 2006/02.

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.

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

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. (1990). “Monetary policy in the second half of the 1980s”, *Bank of England Quarterly Bulletin*, November.

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

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

Orphanides A. and Porter, R. (2000). “P* revisited: money-based inflation forecasts with a changing equilibrium velocity”, *Journal of Economics and Business* 52, pp. 87–100.

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.

Rudebusch, G.D. and L.E.O. Svensson, (2002). “Eurosystem monetary targeting: lessons from US data”, *European Economic Review*, 46, 3, pp. 417-442.

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

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.

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.

Appendices

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 and deterministic terms. 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 (see Zellner, 1971, chapter 8 for details). 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 $\Pr(\Delta p_{t+h} < 5.0|Data_t)$ 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 assumes Z is known, which is straight forward in the case of one period ahead prediction, $h = 1$. 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 for the case of h period ahead prediction, where $h > 1$, then Z is not known. However, following common practice, we can simply estimate a different VAR for each forecast horizon h. So for $h = 1$ we can work with a standard VAR as described above, but for $h > 1$, we can 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 $t-h$ (ie let X contain lags of explanatory variables lagged at least

h periods). In the notation used in the main text, for each forecast horizon $h = 1, 2, ...8$ and recursion we estimate the following equation for output growth (or an equation for inflation) as part of our set of VECM's (VAR when $r = 0$) defined over p and r :

$$\Delta y_t = b_0 + \sum_{i=0}^{p-1} b_{1i} \Delta y_{t-h-i} + \sum_{i=0}^{p-1} b_{2i} \Delta p_{t-h-i} + \sum_{i=0}^{p-1} b_{3i} \Delta i_{t-h-i} + \sum_{i=0}^{p-1} b_{4i} \Delta e_{t-h-i} + \sum_{i=0}^{p-1} b_{5i} \Delta m_{t-h-i} + \sum_{i=1}^{r} \alpha_i \xi_{i,t-h-1} + \varepsilon_t. \tag{A.3}$$

Leading the above equation by h time periods gives Δy_{t+h} as a function of known variables where their 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 (ie 40 VARs and VECMs, each of which has a variant with money and a variant without money), and two different data combinations (ie we have real-time and final vintage versions of our variables). For each of the different data combinations we have to do a recursive prediction exercise. Furthermore, we have to do all this for $h = 1, 4$ and 8. In total, our empirical results involve posterior and predictive results for tens of thousands of 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.

B A Bayesian Variant of the Diebold-Mariano Statistic

To develop a Bayesian interpretation of various frequentist 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 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 frequentist 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. This is what we do so in the relevant BMA rows of Table 1.

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 could simply use the DM test in exactly the same way as they do. Our statistical methods would then be a combination of Bayesian methods (to produce the predictions) and frequentist methods (to evaluate the quality of the predictions). GKMV contains results using such an approach.

The fully Bayesian procedure used in this paper 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 realization, 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 (ie 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:

$$p(d_t).$$

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

$$\Pr(d_t > 0 | Data_t)$$

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:

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

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 $\Pr(\overline{DM} > 0)$.

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