Forecasting using Bayesian and information theoretic model averaging: an application to UK inflation

*George Kapetanios**

*Vincent Labhard***

and

Simon Price†

Working Paper no. 268

* Queen Mary College, University of London.
Email: g.kapetanios@qmul.ac.uk

** Bank of England.
Email: vincent.labhard@bankofengland.co.uk

† Bank of England.
Email: simon.price@bankofengland.co.uk

This paper represents the views and analysis of the authors and should not be thought to represent those of the Bank of England or Monetary Policy Committee members. This paper was finalised on 31 May 2005.

Copies of working papers may be obtained from Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH; telephone 020 7601 4030, fax 020 7601 3298, e-mail mapublications@bankofengland.co.uk

Working papers are also available at www.bankofengland.co.uk/wp/index.html

The Bank of England's working paper series is externally refereed.

ISSN 1368-5562

Contents

Abstract

In recent years there has been increasing interest in forecasting methods that utilise large data sets, driven partly by the recognition that policymaking institutions need to process large quantities of information. Factor analysis is a popular way of doing this. Forecast combination is another, and it is on this that we concentrate. Bayesian model averaging methods have been widely employed in this area, but a neglected alternative approach employed in this paper uses information theoretic based weights. We consider the use of model averaging in forecasting UK inflation with a large data set from this perspective. We find that an information theoretic model averaging scheme can be a powerful alternative both to the more widely used Bayesian model averaging scheme and to factor models.

Key words: Forecasting, inflation, Bayesian model averaging, Akaike criteria, forecast combining.

JEL classification: C110, C150, C530.

Summary

Recently, there has been increasing interest in forecasting methods that utilise large data sets. There is a huge quantity of information available in the economic arena which might be useful for forecasting, but standard econometric techniques are not well suited to extract this. In an effort to assist in this task, econometricians began assembling large macroeconomic data sets and devising ways of forecasting with them. Standard regression techniques cannot be used in this context, as the number of variables is far too large. Instead, broadly speaking there are two methodologies that can be applied: factor modelling and forecast combination. In the former, a factor structure is imposed on the data and then techniques such as principal components are used to extract the factors that are subsequently used in forecasts. This approach has been widely used in macroeconomic forecasting in recent years.

The alternative methodology is forecast combining, often of simple and probably misspecified models. This grew out of the observation by forecast practitioners in the 1960s that combining forecasts (initially by simple averaging) produced a forecast superior to any single forecast. If it were possible to identify the correctly specified model and the data generating process (DGP) is unchanging, then this approach would not be sensible. However, models may be incomplete, in different ways; they employ different information sets. Forecasts might be biased, and biases can offset each other. Even if forecasts are unbiased, there will be covariances between forecasts which should be taken into account. Thus combining misspecified models may, and often will, improve the forecast.

Despite this, combining forecasts will not in general deliver the optimal forecast, while combining information will. Nevertheless, it may not be practicable to estimate the fully encompassing model, not least because the set of variables is vast. Thus we have a justification for combining forecasts. One could call this the frequentist misspecification case. It should be clear that in this context forecast combining is viewed as mainly a stop-gap measure that works in practice but would be surpassed by an appropriate model that addressed the underlying misspecification. A further practical problem is that with standard combining methods the forecast weights can only be reliably constructed for a relatively small number of models. Nevertheless, given that the true DGP may involve a vast number of variables, it is clear that forecast combination is a route into the combining of information, and this is how it is interpreted in the literature relating to large data

sets.

Forecast combining can be also be interpreted in a Bayesian framework. Here it is assumed that there is a distribution of models. The basic problem, that a chosen model is not necessarily the correct one, can then be addressed in a variety of ways, one of which is Bayesian model averaging. A chosen model is simply the one with the best posterior odds, but posterior odds can be formed for all models under consideration and offer weights for forecast combinations.

There is an analagous frequentist information theoretic approach, on which we focus in this paper. Given we have a set of models, we can define relative model likelihood. Model weights within this framework have been suggested by Akaike in a series of papers. In practical terms such weights are easy to construct using standard information criteria. Our purpose, then, is to consider this way of model averaging as an alternative to Bayesian model averaging.

We address this in two ways. We first assess the performance of information theoretic and other model averaging techniques by means of a Monte Carlo study. We then examine how various schemes can perform in forecasting UK inflation. For this, we use a UK data set which emulates a well-known data set constructed by Stock and Watson for the United States. We find that model averaging techniques can be beneficial with the information theoretic weights performing very well. Our findings partly confirm that Bayesian model averaging can provide good inflation forecasts, but we find that the frequentist approach also works well, and dominates in a large subset of the cases we examine for UK data. It is unlikely that a single technique would be more useful than all others in all settings. Nevertheless, our work indicates that information theoretic model averaging provides a useful addition to the forecasting toolbox of macroeconomists. Indeed, we find that the information theoretic method is the most robust of those we examine.

This paper does not describe the way in which the Bank of England generates its forecasts. The findings in this paper pertain to a specific type of forecasting model which is only part of a much broader approach to forecasting applied at the Bank. The Bank does not use a single model to forecast inflation or other variables; instead it uses a 'suite' of many models ranging from purely theoretical through purely data driven to the Bank's macroeconometric model, the Bank of England Quarterly Model (BEQM). All these models are useful in a particular context: in no case will any one model provide a uniquely best forecast.

1 Introduction

In recent years there has been increasing interest in forecasting methods that utilise large data sets. There is an awareness that there is a huge quantity of information available in the economic arena which might be useful for forecasting, but standard econometric techniques are not well suited to extract this. This is not an issue of mere academic interest. Lars Svensson described what central bankers[(1)] do in practice in Svensson (2004). 'Large amounts of data about the state of the economy and the rest of the world ... are collected, processed, and analyzed before each major decision.' In an effort to assist in this task, econometricians began assembling large macroeconomic data sets and devising ways of forecasting with them: James Stock and Mark Watson (eg, Stock and Watson (1999)) were in the vanguard of this campaign.

Broadly speaking, there are two methodologies that can be applied: factor modelling, which summarises a proportion of the variation in all the data in a limited number of factors which are then used to forecast; and forecast combination, where information in many forecasting models, typically simple and incomplete, are combined in some manner.

In the first of these approaches a factor structure is imposed on the data and then techniques such as principal components (see, eg, Stock and Watson (2002)) or dynamic principal components (see, eg, Forni, Hallin, Lippi and Reichlin (2000)) are used to extract the factors and subsequently carry out forecasting. An issue with such an approach relates to the possibility of imposing a factor structure where none can be supported in the data. There is little work on the formal statistical testing of whether a factor structure is supported by the data. Nevertheless, this approach has been widely used in macroeconomic forecasting in recent years.[(2)]

A related strand of work concerns the determination of a satisfactory regression model to be used in forecasting. One of the leading approaches in this work is the widely used 'general-to-specific' approach, developed and popularised in a number of papers by David Hendry and his co-authors.[(3)] Briefly summarised, this approach involves starting from a general dynamic

(1) We do not wish to suggest that this paper describes the way in which the Bank of England generates its forecasts. The Bank does not use a single model to forecast inflation or other variables; the underlying philosophy and a description of some models in use in the recent past are set out in Bank of England (1999, 2000, 2005). The Bank's use of models is discussed in Pagan (2003).

(2) See the survey in Stock and Watson (2006), where they suggest the method has often been successful.

(3) The literature now spans three decades. One summary of the state of the art as it was in 1999 is Hendry (1999).

statistical model which captures the characteristics of the data and via sequential testing reducing the complexity of this model while retaining the congruence of the resulting model.(4) This method cannot deal, in its original form, with data sets where the number of variables exceeds the number of observations, but recent work is likely to relax this restriction.

The alternative methodology that can be applied to large data sets is forecast combining. This grew out of a different tradition from the large data set programme, after the observation by forecast practitioners that for whatever reasons, combining forecasts (initially by simple averaging) produced a forecast superior to any element in the combined set. Why would we want to have a variety of models generating different forecasts, rather than the single correct model? If it were possible to identify the correctly specified model and the data generating process (DGP) is unchanging, then the frequentist(5) answer would be that we would not. But the weight of evidence dating back to Bates and Granger (1969) and Newbold and Granger (1974) reveals that combinations of forecasts often outperform individual forecasts.(6) Models may be incomplete, in different ways; they employ different information sets. Forecasts might be biased, and biases can offset each other. Even if forecasts are unbiased, there will be covariances between forecasts which should be taken into account. Thus combining misspecified models may, and often will, improve the forecast.

Despite this, combining forecasts will not in general deliver the optimal forecast, while combining information will. Clements and Hendry (1998) therefore argue that combining is opposed to their notion of a progressive research strategy. Nevertheless, it may not be practicable to estimate the fully encompassing model, not least because the set of variables is vast. Thus we have a justification for combining forecasts. One could call this the frequentist misspecification case. It should be clear that in this context forecast combining is viewed as mainly a stop-gap measure that works in practice but would be surpassed by an appropriate model that addressed the underlying misspecification. A further practical problem is that with standard combining methods the forecast weights can only be reliably constructed for a relatively small number of models. Nevertheless,

---

(4) This methodology has been implemented in the computer package PCGETS: see also Krolzig and Hendry (2001).

(5) Statisticians are divided into two camps with different views of what a 'probability' means. The frequentist view is that a probability gives the relative frequency with which an event is observed, and in principle is knowable to any desired degree of precision, given enough data. A Bayesian believes probabilities are essentially subjective. However, even a subjective view is informed by data, and Bayes theorem shows how to update prior beliefs in the light of new evidence.

(6) Recent surveys of forecast combination from a frequentist perspective are to be found in Newbold and Harvey (2002) and Clements and Hendry (1998); see also Clements and Hendry (2002).

given that the true DGP may involve a vast number of variables, it is clear that forecast combination is a route into the combining of information: and this is how it is interpreted in the literature relating to large data sets.

But there is an alternative way of looking at this problem, most clearly seen from a Bayesian perspective. Here it is assumed that there is a distribution of models, thus delineating the concept of model uncertainty more rigorously. The basic problem, that a chosen model is not necessarily the correct one, can then be addressed in a variety of ways, one of which is Bayesian model averaging. From this point of view, a chosen model is simply the one with the best posterior odds; but posterior odds can be formed for all models under consideration, thereby suggesting a straightforward way of constructing model weights for forecast combinations. This has been used in many recent applications; for example, forecasting US inflation in Wright (2003a).

There is also a frequentist information theoretic approach in an analogous vein. In this context, information theory suggest ways of constructing model confidence sets. Given we have a set of models, we can define relative model likelihood. Model weights within this framework have been suggested by Akaike in a series of papers (see Akaike (1978, 1979, 1981, 1983)) and expounded further by Burnham and Anderson (1998). In practical terms such weights are easy to construct using standard information criteria such as Akaike's information criterion. Our purpose, then, is to consider this way of model averaging as an alternative to Bayesian model averaging.

In this paper we address this in two ways. We first assess the performance of information theoretic model averaging and other model averaging techniques by means of a Monte Carlo study. We then examine how various schemes can perform in forecasting UK inflation. For this, we use a UK data set which emulates the data constructed by Stock and Watson (2002).[7] We find that model averaging techniques can be beneficial with the information theoretic weights performing very well. Our findings partly support the findings of Wright (2003a) who concludes that Bayesian model averaging can provide superior forecasts for US inflation, but we find that the frequentist approach also works well, and dominates in a large subset of the cases we examine for UK data.

(7) In total, this data set has 131 series, comprising 20 output series, 27 labour market series, 9 retail and trade series, 6 consumption series, 6 series on housing starts, 10 series on inventories and sales, 8 series on orders, 7 stock prices, 5 exchange rate series, 7 interest rate series, and 6 monetary aggregates, 19 price indices, and an economic sentiment index. We restrict attention to the set of 58 variables, described in Appendix B, where there are at least 90 observations.

The paper is organised as follows: Section 2 discusses the various model averaging schemes we consider; Section 3 provides some Monte Carlo results; Section 4 carries out the forecasting exercise on UK CPI inflation; and finally, Section 5 concludes.

2 Forecasting using model averaging

2.1 Bayesian model averaging

The idea behind forecasting using model averaging reflects the need to account for model uncertainty in carrying out statistical analysis. From a Bayesian perspective, model uncertainty is straightforward to handle using posterior model probabilities. The use of posterior model probabilities for forecasting has been suggested, discussed and applied by, among others, Min and Zellner (1993), Koop and Potter (2003), Draper (1995) and Wright (2003a,b). Briefly, under Bayesian model averaging a researcher starts with a set of models which have been singled out as useful representations of the data. We denote this set as $\mathcal{M} = \{M_i\}_{i=1}^N$ where M_i is the i-th of the N models considered. The focus of interest is some quantity of interest for the analysis, denoted by Δ. This could be a parameter, or a forecast, such as inflation h quarters ahead. The output of a Bayesian analysis is a probability distribution for Δ given the set of models and the observed data at time t. Let us denote the relevant information set at time t by D_t. We denote the probability distribution as $pr(\Delta|D, \mathcal{M})$. This is given by

$$pr(\Delta|D_t, \mathcal{M}) = \sum_{i=1}^{N} pr(\Delta|M_i, D_t) pr(M_i|D_t) \qquad \textbf{(1)}$$

where $pr(\Delta|M_i, D_t)$ denotes the conditional probability distribution of Δ given a model M_i and the data D_t and $pr(M_i|D_t)$ denotes the conditional probability of the model M_i being the true model given the data. It is clear that implementation requires two quantities to be obtained at each point in time. First, $pr(\Delta|M_i, D_t)$ which is easily obtained from standard model specific analysis. Second, the weights, $pr(M_i|D_t)$. It is easy to see that the weights are formed as part of a stochastic process where $pr(M_i|D_t)$ is obtained from $pr(M_i|D_{t-1})$ via a number of intermediate steps. This implies the need of a prior distribution $pr(M_i|D_0) = pr(M_i)$ and for $pr(\theta_i|M_i, D_{t-1})$ to be specified.

Thus we need to obtain a number of expressions for **(1)** to be operational. First, using Bayes' theorem

$$pr(M_i|D_t) = \frac{pr(D_t|M_i, D_{t-1})pr(M_i|D_{t-1})}{pr(D_t|D_{t-1})} = \frac{pr(D_t|M_i, D_{t-1})pr(M_i|D_{t-1})}{\sum_{i=1}^{N} pr(D_t|M_i, D_{t-1})pr(M_i|D_{t-1})} \quad \textbf{(2)}$$

where $pr(D_t|M_i, D_{t-1})$ denotes the conditional probability distribution of the data given the model M_i and the previous period's data, $pr(M_i|D_{t-1})$ denotes the conditional probability of the model M_i being true, given the previous period's data.

$$pr(D_t|M_i, D_{t-1}) = \int pr(D_t|\theta_i, M_i, D_{t-1})pr(\theta_i|M_i, D_{t-1})d\theta_i \quad \textbf{(3)}$$

(3) is the likelihood of model M_i and θ_i are the parameters of model M_i. Given this, the quantity of interest is

$$E(\Delta|D_t) = \sum_{i=1}^{N} \hat{\Delta}_i pr(M_i|D_t) \quad \textbf{(4)}$$

In theory (see, eg, Madigan and Raftery (1994)) when Δ is a forecast, this sort of averaging provides better average predictive ability than single model forecasts.

2.2 Information theoretic model averaging

Model averaging is not confined to the Bayesian approach. In the context of forecasting the idea of model averaging (ie, forecast combination) has a long tradition starting with Bates and Granger (1969). The main suggestion of this line of work is to use forecasts obtained during some forecast evaluation period to determine optimal weights from which a forecast can be constructed along the lines of **(4)**. These weights are usually constructed using some regression method and the available forecasts. A problem with this class of methods arises if N is large. For example, setting N to 93 as in Wright (2003a) would require an infeasibly large forecast evaluation period.

An alternative which has received little attention in the literature, can be based on the analogue of $pr(M_i|D_t)$ for frequentist statistics. Such a weight scheme has been implied in a series of papers by Akaike and others (see, eg, Akaike (1978, 1979, 1981, 1983) and Bozdogan (1987)) and expounded further by Burnham and Anderson (1998). Akaike's suggestion derives from the Akaike information criterion (AIC). AIC is an asymptotically unbiased measure of minus twice the log likelihood of a given model. It contains a term in the number of parameters in the model, which may be viewed as a penalty for over-parameterisation. Akaike's original frequentist interpretation[8] relates to the classic mean-variance trade-off. In finite samples, when we add parameters there is a benefit (lower bias), but also a cost (increased variance); the latter is a loss of

(8) Akaike (1979) offers a Bayesian interpretation.

information. More technically, from an information theoretic point of view, AIC is an unbiased estimator of the Kullback and Leibler (1951) (KL) distance of a given model where the KL distance is given by

$$I(f,g) = \int f(x)\log\left(\frac{f(x)}{g(x|\hat{\theta})}\right)dx.$$

Here $f(x)$ is the unknown true model generating the data, $g(x|.)$ is the entertained model and $\hat{\theta}$ is the estimate of the parameter vector for $g(x|.)$. The KL distance is an influential concept in the model selection literature and forms the basis of the development of AIC. Within a given set of models, the difference of the AIC for two different models can be given a precise meaning. It is an estimate of the difference between the KL distance for the two models. Further, $exp\left(-1/2\Psi_i\right)$ is the relative likelihood of model i where $\Psi_i = AIC_i - \min_j AIC_j$ and AIC_i denotes the AIC of the i-th model in $\mathcal{M}$. Thus, $exp\left(-1/2\Psi_i\right)$ can be thought of as the odds for the i model to be the best KL distance model in $\mathcal{M}$. In other words this quantity can be viewed as the weight of evidence for model i to be the KL best model given that there is some model in $\mathcal{M}$ that is KL best as a representation of the available data. Note that there is no assumption made here about the true model belonging to $\mathcal{M}$. We are only considering the ranking of models in terms of KL distance. This may be viewed as as a crucial difference from a Bayesian analysis, in which it is assumed that a model in $\mathcal{M}$ or a weighted average of the models in $\mathcal{M}$ is the true model.

It is natural to normalise $exp\left(-1/2\Psi_i\right)$ so that

$$w_i = \frac{exp\left(-1/2\Psi_i\right)}{\sum_{i=1}^{N} exp\left(-1/2\Psi_i\right)} \qquad \textbf{(5)}$$

where $\sum_i w_i = 1$. We refer to these as AIC weights.

We note w_i are not the relative frequencies with which given models would be picked up according to AIC as the best model given $\mathcal{M}$. Since the likelihood provides a superior measure of data based weight of evidence about parameter values compared to such relative frequencies (see, eg, Royall (1997)), it is reasonable to suggest that this superiority extends to evidence about a best model given $\mathcal{M}$. The w_i can be thought of as model probabilities under non-informative priors giving a parallel to Bayesian analysis. However, this analogy should not be taken literally as these model weights are firmly based on frequentist ideas and do not make explicit reference to prior probability distributions about either parameters or models. Also, the Akaike criterion is only one criterion which can form the basis of such weights. We also consider weights based on the Schwartz information criterion, which has a similar rationale. We refer to these as SIC weights.

3 A Monte Carlo study

We now undertake a small Monte Carlo study to explore the properties of various model averaging techniques in the context of forecasting. As we discussed above, model averaging aims to address the problem of model uncertainty in small samples. There are two broad cases to consider. The first is when the model that generates the data belongs to the class under consideration. In this case it addresses the issue that the chosen model is not necessarily the true model, and by assigning probabilities to various models provides a forecast that is, to some extent, robust to model uncertainty. The second, perhaps more relevant case, is where the true model does not belong to the class of models being considered. Here there is no possibility that the chosen model will capture all the features of the true model. As a result, the motivation for model averaging becomes stronger, since forecasts from different models can inform the overall forecast in different ways. We examine this latter case.

In the experimental design, we adapt the set-up proposed in Fernandez, Ley and Steel (2001) and used more recently in Eklund and Karlsson (2004), which therefore offers a standard problem to examine. Let $\mathbf{X} = (x_1, \ldots, x_N)$ be a $T \times N$ matrix of regressors, where $x_i = (x_{i,1}, \ldots, x_{i,T})'$. The series in the first $2N/3$ columns are given by

$$x_{i,t} = \alpha_i x_{i,t-1} + \epsilon_{i,t}, i = 1, \ldots, N, \quad t = 1, \ldots, T \qquad \textbf{(6)}$$

where ϵ_t is i.i.d. $N(0, 1)$ and N is divisible by 3. The last $N/3$ series are constructed as

$$(x_{2N/3+1}, \ldots, x_N) =$$
$$(x_1, \ldots, x_{N/3})(0.3, 0.3 + (1)0.2, 0.3 + (2)0.2, \ldots, 0.3 + (N/3 - 1)0.2)'(1, \ldots, 1) + E \qquad \textbf{(7)}$$

where E is a $T \times N/3$ matrix of standard normal variates. This set-up allows for some cross-sectional correlation in the predictor variables. The true model is given by

$$y_t = 2\,x_{1,t} - x_{5,t} + 1.5\,x_{7,t} + x_{11,t} + 0.5\,x_{13,t} + 2.5\,\varepsilon_t \qquad \textbf{(8)}$$

where ε_t is i.i.d. $N(0, 1)$. The numbering of the variables is prompted partly by the size of the data set and features of the models investigated in the source references, but this is not a critical feature of the design. The important point is that none of the models considered are the true DGP.

The design in Eklund and Karlsson (2004) sets $N = 15$ and $\alpha = 0$. We generalise the set-up in two directions. First, we set $N = 60$. By increasing N we provide a closer approximation to real-world situations where large data sets are considered when forecasting models are built.

Second, we let $\alpha_i \sim U(0.5, 1)$. The α_i introduce persistence, which we allow to be random.

We evaluate eight categories of forecasts in total.

The benchmark is the forecast produced by a simple $AR(1)$ model for y_t (AR). For the remaining forecasts, we use variants of the model

$$y_{t+h} = a'x_t + by_t + \varepsilon_t \tag{9}$$

for the h-step ahead forecast, where x_t is a K-dimensional regressor set, and K takes the values of 1 or 2. The second single forecast is chosen in sample by the Akaike information criterion and can vary with h (AIC). As K is no greater than 2, this can never select the true model.

The remaining forecasts are different averages of the complete set of models of form **(9)**. The first three of these are produced by Bayesian model averaging (BMA), differing by the shrinkage parameter described below. The Bayesian weights are set following Wright (2003a). In particular, we set the model prior probabilities $P(M_i)$ to the uninformative priors $1/N$. The prior for the regression coefficients is chosen to be given by $N\left(0, \phi\sigma^2(X'X)^{-1}\right)$, conditional on σ^2, where X is the $T \times p$ regressor matrix for a given model and p is the numbers of regressors. The improper prior for σ^2 is proportional to $1/\sigma^2$. The specification for the prior of the regression coefficients implies a degree of shrinkage towards zero (which implies no predictability). The degree of shrinkage is controlled by ϕ. The rationale is that some degree of shrinkage steers away from models that may fit well in sample (by chance or because of overfitting) but have little forecasting power. There is empirical evidence that such shrinkage is beneficial for out-of-sample forecasting, but no *a priori* guidance for what values should be selected. Following Wright (2003a) we consider conventional choices of $\phi = 20, 2, 0.5$. Given the above, routine integration gives model weights which are proportional to

$$(1+\phi)^{-p/2}S^{-T/h}$$

where

$$S = Y'Y - Y'X(X'X)^{-1}X'Y\frac{\phi}{1+\phi}$$

and Y is the $T \times 1$ regressand vector.

We next consider information theoretic model averaging (ITMA), where the weights are given by **(5)**, and the individual models are again given by **(9)**. We consider two versions, one based on AIC weights (AITMA) and the other on SIC weights (SITMA).

Finally, we examine equal-weight model averaging (AV) where the weights are given by $1/N$. This last scheme is motivated by the work of Stock and Watson (2004) (see also Stock and Watson (2003)), and is a commonly used scheme often thought to work well in practice.

We set $T = 50, 100$. The forecast evaluation period for each sample is the last 30 observations. Finally, we examine the forecast horizons $h = 1, \ldots, 8$. For all model averaging techniques and AIC we consider two different classes of models over which the weighting scheme is applied. The first class is simply the class of all models with one predictor variables ($K = 1$). The second class is the class of all models with two predictor variables ($K = 2$). It is clear that neither of these two classes of models contains the true model. We do not allow for higher K for two reasons. First, most forecasting models used in practice, and found to have good performance, are parsimonious. Second, we assign weights to all members of the model class. With our set-up and $K = 2$ we have 1,711 models to consider. For (say) $K = 3$ the number of models rises to 32,509 and therefore becomes computationally intensive.[9]

The main measures of forecast performance that we use are based on RMSE. Results for RMSE itself are given in Table 1. The best forecast method in a particular row (that is, for given K, T and horizon) are indicated in bold. Variations in performance are reasonably large. It is immediately evident that for this model design the simple $AR(1)$ does not perform well, being dominated at all values of K, horizon and sample sizes by the combined forecasts. Similarly, using AIC to pick the best model is an inferior method. Using simple averaging (AV) does better than either single forecast, while the Bayesian (BMA) method works better still, dominating in many cases, especially for low horizons. It is best for a large shrinkage parameter, giving the data more weight. But the information-criteria based methods also do well, especially at longer horizons, where they dominate BMA.[10] This is a robust result across samples and choice of K. In the remainder of the paper we see if this conclusion carries over to the real data.

(9) Note that there exist methods to search the model space efficiently that bypass this problem. One is that discussed by Fernandez *et al.* (2001) and based on Markov Chain Monte Carlo algorithms. Another is by Kapetanios (2005) which uses genetic and simulated annealing algorithms to search for good models in terms of information criteria. We do not explore these methods in this paper.

(10) The two methods (AITMA and SITMA) are based on penalty factors that are numerically similar in this experiment, so the results are correspondingly close.

4 An application to forecasting UK inflation

Our main interest is practical, and in particular the practice of inflation forecasting using the model averaging schemes examined in the Monte Carlo study. The models we consider are a standard specification, as discussed in Stock and Watson (2004). We modify our Monte Carlo design by using a k lags autoregressive process augmented with a single predictor variable ($ARX(k)$). The number of lags is either set to 1 or 2 or chosen optimally for each model, each sample and each forecast horizon using the Akaike information criterion. Different models are specified for each forecasting horizon. Model i for forecasting horizon h is given by

$$\pi_{t+h} = \alpha + \sum_{j=1}^{k_1} \beta_j \pi_{t-j+1} + \sum_{j=1}^{k_2} \gamma x_{it-j+1} + \epsilon_t \qquad \textbf{(10)}$$

where π_t is UK year-on-year CPI inflation, x_{it} is the i-th predictor variable at time t and ϵ_t is the error term, with variance σ^2. We consider 58 predictor variables, where the data span 1980 Q2-2004 Q1.(11)

Where we average models, we consider Bayesian, information theoretic and equal-weight model averaging. The information theoretic weights are given by **(5)**. We include the AR forecast, making a total of 59 forecasts to combine. As in the Monte Carlo exercise, both AIC and SIC are considered. The Bayesian weights are given by the scheme discussed in the Monte Carlo section. We also consider two factor model forecasts. As discussed in the introduction, these are widely used alternatives to forecast combination in large data sets. In this case we specify models of the form given by **(10)** where the exogenous variables are replaced by either the first or the first five principal components of the data set as estimated in the full sample.

4.1 Forecast performance

We evaluate the forecasts over two post-sample periods: 1990 Q2-1997 Q1 (pre-MPC)(12) and 1997 Q2-2004 Q1 (MPC). We consider $h = 1, \ldots, 12$ and $j = 1, \ldots, 4$. The number of lags in the pair (k_1, k_2) are set either to (1, 1), (2, 2) or $(1, k)$ where k is chosen optimally for each model, each sample and each forecast horizon using the Akaike information criterion. We report three

(11) Brief descriptions including the source and the Office for National Statistics (ONS) codes (where applicable) for these variables are given in Appendix B.
(12) We use MPC as a shorthand to indicate the period where monetary policy is set by the Bank of England's Monetary Policy Comittee (MPC) in the context of an explicit inflation-targeting monetary regime. This follows the granting of monetary policy independence to the Bank of England on May 1997.

performance indicators of the forecast of the mean: (i) relative RMSE, compared to the benchmark AR model; (ii) the percentage of models of the form **(10)** which perform worse than a given model averaging scheme in terms of relative RMSE; and (iii) the proportion of periods in which the model averaging scheme has a smaller absolute forecast error than the AR model. In the relative RMSE tables we report a Diebold-Mariano test[(13)] of whether the forecast is significantly different from the benchmark AR model at the 10% level, indicated with an asterisk. In addition, we report the identity of the variables whose individual models **(10)** get the ten largest weights, averaged across the periods in the forecast evaluation period and the horizons.

These results are reported in Tables 2-31. We first consider the (chronologically later but more pertinent) MPC forecast evaluation period (Tables 2-16). The key tables are 2, 7 and 12, which give RMSE, and these are the ones we focus on in this section.[(14)] It is striking that the factor models do not work well, contrary to the commonly received wisdom.[(15)] For many horizons and specifications they underperform relative to the benchmark ($RMSE > 1$). Moreover, in many cases the forecasts are significantly worse for the $ARX(1)$, $ARX(2)$ and $ARX(k)$ models than the benchmark. In this sample, the five-factor model tends to do a little better than the single-factor model, but is still outperformed by most other schemes at most horizons. The Akaike information theory based AITMA does not work well for the $ARX(1)$ and $ARX(2)$ models at short forecasting horizons. However, it works better than all methods across all AR specifications by a significant margin for longer horizons, and at almost all horizons for $ARX(k)$ models. Moreover, for the longer horizons not only does it perform well but is also significantly better than the benchmark.

The performance of the SIC based scheme SITMA is practically identical to the AITMA, with AITMA having a slight edge in the $ARX(k)$ case. The Bayesian BMA scheme works best for high ϕ (giving the data a high weight) but is inferior to AITMA overall. It follows that high ϕ dominates the simple averaging scheme AV, which amounts to setting $\phi = 0$.[(16)] Disregarding the short-horizon $ARX(1)$ case, AITMA (or SITMA) dominates both the BMA and other methods in almost all cases.

(13) Diebold and Mariano (1995).
(14) The tables giving the proportion of models beating ARX models on RMSE and periods beating AR models on absolute error give equivalent rankings.
(15) For example, see Stock and Watson (1999).
(16) This suggests that structural breaks, thought to favour simple averaging over other methods, are not the main source of forecast error in this set.

Over the pre-MPC period, then, there is strong support for using Akaike weights for forecast combinations. The superiority of the information criteria emerges most clearly in the $ARX(k)$ cases. In only one case (horizon 5) is the best model not based on an information criterion. Despite this RMSE dominance, only half the AITMA forecasts are significantly different from the benchmark; this indicates that the information-based forecasts are more volatile than those based on Bayesian averaging, especially with a low value for the shrinkage factor.(17)

Tables 6, 11 and 16 list the top-ten ranked models for the Bayesian and information theoretic schemes. Generally, the information theoretic schemes put more weight on a smaller number of variables. The AITMA and SITMA rankings are not identical but are extremely close. For the one-lag models 85% of the information theoretic weights is on S21 (weekly hours in manufacturing), which is also given the highest weight in the BMA schemes, but at a lower level: 11% in the $\phi = 20$ case. S23 (manufacturing trade balance) comes in a distant second at 4%. This is the case where AITMA does relatively badly at short horizons. For two lags (Table 9), the AR (with no additional explanatory variable) is given the highest weight in the BMA cases. For AITMA, S21 is still ranked highest, but with a lower weight (55%), while S11 (inverse unemployment rate) comes in at 15%, S24 (consumers' total expenditure) at 8%, followed by S45 (the one-year interest rate spread) at 7% and S28 (services consumption) at 5%. One should not read too much into what series have been chosen as this is an atheoretic method, the series are transformed to stationarity and all the models enter with a positive weight, but the fact that two labour market quantities and consumption are in the top three does not ring alarm bells. S21 tops the ranking in the Akaike selected lag length case as well (Table 16): S23 returns with a 4% weight in the fourth rank.

The broad conclusions remain the same in the pre-MPC forecast evaluation period (Tables 17-31: the key tables being 17, 22 and 27). In fact the AITMA emerges stronger in this period. In almost all horizons AITMA dominates all other models, including the best BMA model (once again, $\phi = 20$). On the basis of these results one would unambiguously choose AITMA as the most robust method. Although their failure is less marked than in the post-MPC period, factor forecasts once again do not justify their reputation in this data set. They are hardly better than AR forecasts at any horizon and forecast evaluation period, and in some cases they are worse. The information

(17) However, forecast predictive tests have notoriously low power. As Ashley (1998) concludes, 'a model which cannot provide at least a 30% MSE improvement over that of a competing model is not likely to appear significantly better than its competitor over post sample periods of reasonable size.'

criteria methods emerge most strongly in the $ARX(2)$ and $ARX(k)$ comparisons. In this case, it is the $ARX(2)$ comparison which gives the strongest relative performance, although once again the results are not all significantly better than the benchmark.

Tables 21, 26 and 31 again list the top-ten ranked models for the Bayesian and information theoretic schemes. In this period, while the AITMA continues to place more weight on fewer variables, the weights are less concentrated. For the $ARX(1)$ case, S23 (manufacturing trade balance) enters first with a weight of 17%, followed by a succession of labour market quantity variables (S21 (manufacturing weekly hours), weight 15%; S11 (employment rate), weight 11%; S12 (non-agricultural employment), weight 10%; S13 (private employment), 7%), then another trade balance, labour market variables and an interest rate spread. Roughly the same set of variables is juggled in the other cases.

4.2 Forecast density evaluation

Recent work in the forecasting literature has moved away from focusing on point forecasts and towards forecasting the whole probability distribution of the variable(s) of interest. Interest in this aspect of forecasting has partly arisen out of the introduction of inflation-targeting regimes. Risks to the forecast evolution of inflation translate into an interest on the whole forecast distribution. The Bank of England has been at the forefront of this development, pioneering the publication of 'fan charts' in its quarterly *Inflation Reports* precisely because it is misleading to ignore the variance and possibly other moments of the forecast distribution. Nevertheless little work has been carried out on the possibility of constructing forecast density combinations. To partly amend for this we address the issue of density combining in this subsection.

In this context it is important to compare the forecast probability distribution with the actual data distribution. There is considerable work on how such comparisons can be carried out. We briefly outline the main ideas. Assume there exists a true conditional distribution for inflation using information available at time $t-s, s>0$. Let us denote this by $F_{t|s}(\pi) = Prob(\pi_t < \pi | \mathcal{I}_{t-s})$, where $\mathcal{I}_{t-s}$ is the information set available at time $t-s$. Its estimate is denoted by $\hat{F}_{t|s}(\pi)$. We assume that this distribution is differentiable and denote its derivative (ie, the conditional probability density function) by $f_{t|s}(\pi)$. Note that $f_{t|s}(\pi)$ is allowed to vary with t as well as s. So inflation is assumed to be generated by $f_{t|s}(\pi)$. We denote the estimate of $f_{t|s}(\pi)$ by $\hat{f}_{t|s}(\pi)$. We wish to test the hypothesis that $\hat{f}_{t|s}(\pi) = f_{t|s}(\pi)$ for all t, s, π: ie, that the estimate of the true distribution does not differ significantly from the truth.

This may at first seem a difficult task. However, we note that for any inflation realisation pi_t, $F_{t|s}(\pi_t)$ is a uniform $U(0, 1)$ random variable. $F_{t|s}(\pi_t)$ is referred to as the probability integral transform. Diebold, Gunther and Tay (1998) prove that for a series of realisations $\{\pi_1, \ldots, \pi_T\}$, $\{F_{t|s}(\pi_t)\}_1^T, s > 0$ is a set of i.i.d. $U(0, 1)$ realisations. Further, if the null hypothesis is true, then $\{\hat{F}_{t|s}(\pi_t)\}_1^T, s > 0$, is also a set of $U(0, 1)$ realisations, which we denote z_T. Standard tests may then be used to test this hypothesis. This result is valid even if the true distribution of inflation $f_{t|s}(\pi)$ varies with t.(18) A standard test in the literature for comparing two distributions is the Kolmogorov-Smirnov (KS) test. It simply looks at the maximum difference between the assumed and empirical distributions. This is used as a summary statistic of how similar the two are to each other, and determines whether the gap is statistically different from zero. Formally, the test takes the form

$$KS_s = max_{t=1,\ldots,T} \left| 1/T \sum_{i=1}^{T} \left[\hat{F}_{i|s}(\pi_i) < \hat{F}_{t|s}(\pi_t) \right] - F_U(\hat{F}_{t|s}(\pi_t)) \right|$$
$$= max_{t=1,\ldots,T} \left| 1/T \sum_{i=1}^{T} \left[\hat{F}_{i|s}(\pi_i) < \hat{F}_{t|s}(\pi_t) \right] - \hat{F}_{t|s}(\pi_t) \right|$$

since the uniform cumulative density function, $F_U(.)$, is the identity function. Large values for the test indicate a significant difference between the assumed and empirical distributions, and therefore rejection of the null hypothesis that they are the same. It can be shown that the significance level is given by

$$Q\left(\left[\sqrt{T} + 0.12 + 0.11/\sqrt{T}\right] KS_s\right)$$

where $Q(\lambda) = 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2\lambda^2}$. Following the work of Diebold *et al.* (1998) there has been a considerable amount of work in the area, efficiently summarised by Corradi and Swanson (2006).

Our innovation is to construct forecast densities via model averaging. To do this we need the component empirical distributions. We obtain these using stochastic simulations on **(10)**. These use the estimated parameters for **(10)**, resampling the residuals of **(10)** over the estimation period. $B = 1,000$ stochastic simulations are carried out for each model, each point in the evaluation period and each horizon. Then the $10/B\%$ quantile of the combined forecast distribution, denoted

(18) The null then clearly requires that $\hat{f}_{t|s}$ varies with t as well.

$q_{10/B}$, is obtained as

$$q_{10/B} = \sum_{i=1}^{N} w_i q_{10/B,i}$$

where w_i are the model weights and $q_{10/B,i}$ are the relevant quantiles for the i-th model.

Tables 3, 8 and 13 present KS tests for the MPC period that the forecast and actual distributions are equal. We note that this may be a hard test to pass, not least because it is highly probable that all the component models are misspecified. Informally, the criteria used to combine the models are all loosely designed to mimimise mean-square error, not match higher moments.

The results are striking. Almost all the models fail the test at very high levels of significance. We cannot reject equality of the five-factor model at $h = 1$ for $ARX(1)$, at $h = 1, 2$ for $ARX(2)$, or at $h = 1, 23, 4$ and 7 for $ARX(k)$. But in no cases can we not reject at 5% for the forecasts not based on AITMA or SITMA. By contrast, for AITMA and SITMA in the $ARX(k)$ case (Table 13), we cannot reject at 5% for horizons up to and including 7. Moreover, for longer horizons the rejection significance is lower than in all other cases. Similar results hold for the $ARX(1)$ case, and while there are more rejections for $ARX(2)$ the test performance greatly exceeds that of the other schemes. For the pre-MPC period (Tables 18, 23 and 28), the information criteria again strongly dominate. Although there are less rejections than for the MPC period and the factor models perform relatively well, in most cases the forecasts not based on AITMA or SITMA reject at 5%, while 8 out of 12 of the AITMA cases do not reject at 5% and a further 3 at 1% for the $ARX(k)$ results, and there are even less rejections for the $ARX(1)$ and $ARX(2)$ cases. By contrast, in all cases for the other schemes rejections outnumber non-rejections. The conclusion to be drawn from this analysis, then, is that in this data set the information-theoretic methods generate forecasts which distributed more similarly to the data than those produced by other combining methods.

5 Conclusions

In recent years there has been rapid growth of interest in forecasting methods that utilise large data sets, driven partly by the recognition that policymaking institutions process large quantities of information. While large information sets might be helpful in the construction of forecasts, standard econometric methods are not well suited to this task. A growing set of techniques have been developed to deal with large data sets, including factor models and model averaging.

This paper focuses on model averaging as a means to improve forecast accuracy. We consider two averaging schemes. The first is Bayesian model averaging. This averaging scheme has been used in a variety of forecasting applications in economics with encouraging results. The second is an information theoretic scheme that is motivated from the concept of relative model likelihood developed by Akaike. Although this model averaging scheme has not received much attention in economics, our results based on a Monte Carlo study and a forecasting application to UK CPI inflation indicate that it has the potential to provide a useful forecasting technique. As there is a clear correspondence with Bayesian averaging, inasmuch as both are based on model performance, it would be odd if the alternative scheme were not also useful. But our work shows that it may outperform Bayesian weights in some cases. The forecast distributions are also better matched to that of the actual data than with other methods.

It is important to realise that it is highly unlikely that a single technique would be more useful than all others in all settings. Nevertheless, our work indicates that information theoretic model averaging provides a useful addition to the forecasting toolbox of macroeconomists. Indeed, in this paper we find that the information theoretic method is the most robust of those we examine.

Appendix A: Tables

Table 1: Monte Carlo study: RMSE of various model averaging schemes

K	T	Horizon	AR		AIC BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV
		1	4.379	4.672	4.354	**4.326**	4.335	4.371	4.370	4.344
		2	4.849	5.020	**4.701**	4.717	4.740	4.767	4.767	4.753
		3	5.171	5.291	**4.956**	4.989	5.013	4.992	4.992	5.027
1	50	4	5.387	5.376	5.146	5.181	5.205	**5.117**	**5.117**	5.219
		5	5.477	5.337	5.195	5.230	5.252	**5.079**	**5.079**	5.266
		6	5.544	5.329	5.249	5.281	5.302	**5.130**	**5.130**	5.315
		7	5.542	5.391	5.245	5.266	5.282	**5.119**	**5.119**	5.293
		8	5.555	5.426	5.257	5.273	5.287	**5.136**	**5.136**	5.296
		1	4.346	4.338	**4.266**	4.286	4.319	4.267	4.267	4.331
		2	4.825	4.832	**4.713**	4.751	4.780	4.722	4.722	4.793
		3	5.054	5.028	4.934	4.970	4.991	**4.920**	**4.920**	5.003
1	100	4	5.197	5.229	**5.088**	5.112	5.129	5.109	5.109	5.139
		5	5.291	5.364	**5.179**	5.197	5.211	5.224	5.224	5.220
		6	5.373	5.358	5.242	5.263	5.278	**5.241**	**5.241**	5.288
		7	5.468	5.393	5.323	5.344	5.360	**5.286**	**5.286**	5.370
		8	5.535	5.516	5.381	5.400	5.415	**5.361**	**5.361**	5.425
		1	4.349	4.935	4.360	**4.300**	4.304	4.393	4.393	4.312
		2	4.808	5.287	**4.630**	4.649	4.673	4.758	4.758	4.687
		3	5.039	5.394	**4.739**	4.786	4.818	4.841	4.841	4.839
2	50	4	5.166	5.471	**4.813**	4.853	4.881	4.909	4.909	4.900
		5	5.271	5.430	**4.841**	4.882	4.911	4.874	4.874	4.931
		6	5.368	5.487	4.901	4.937	4.962	**4.855**	**4.855**	4.979
		7	5.411	5.483	4.910	4.943	4.967	**4.876**	**4.876**	4.983
		8	5.463	5.541	4.947	4.979	5.003	**4.896**	**4.896**	5.019
		1	4.363	4.466	**4.270**	4.281	4.322	4.278	4.277	4.339
		2	4.841	4.946	**4.706**	4.740	4.769	4.737	4.737	4.784
		3	5.140	5.336	**4.986**	5.016	5.039	5.055	5.055	5.053
2	100	4	5.343	5.434	**5.164**	5.196	5.220	5.210	5.210	5.235
		5	5.442	5.457	**5.235**	5.271	5.297	5.249	5.249	5.313
		6	5.568	5.521	5.329	5.364	5.390	**5.291**	**5.291**	5.407
		7	5.662	5.608	5.418	5.454	5.479	**5.358**	**5.358**	5.496
		8	5.719	5.557	5.452	5.485	5.510	**5.298**	**5.298**	5.527

bold indicates best forecast in row

Table 2: Relative RMSE of out-of-sample CPI forecasts using ARX(1) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	1.022	1.016*	1.016*	1.118	1.116	1.016*	1.048*	1.156*
2	**0.982**	0.991	0.991	1.257	1.257	0.992	1.164*	1.036
3	**0.910***	0.975*	0.981	1.121	1.121	0.983	1.161*	1.027
4	**0.860***	0.957*	0.969*	0.991	0.991	0.973*	1.181*	0.979
5	**0.846***	0.947*	0.960*	0.984	0.984	0.964*	1.234*	0.968
6	**0.844***	0.944*	0.959*	0.874	0.874	0.963*	1.268*	1.016
7	0.852*	0.940*	0.954*	**0.793**	**0.793**	0.958*	1.289*	0.988
8	0.870*	0.936*	0.947*	**0.725**	**0.725**	0.951*	1.252*	0.938
9	0.894*	0.939*	0.947*	**0.689***	**0.689***	0.950*	1.227*	0.921
10	0.911*	0.940*	0.946*	**0.675***	**0.675***	0.948	1.184*	0.869*
11	0.919*	0.939*	0.944*	**0.665***	**0.665***	0.946*	1.169*	0.870
12	0.925*	0.940*	0.944*	**0.662***	**0.662***	0.946*	1.150*	0.855

* 10% rejection of Diebold-Mariano test that the forecast differs from the benchmark

bold indicates best forecast in row

Table 3: Probability values of forecast density KS tests for out-of-sample CPI forecasts using ARX(1) models (period: 1997 Q2-2004 Q1)

Horizon	AR	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.024	0.030	0.016	0.015	0.339	0.354	0.019	0.016	0.316
2	0.000	0.001	0.000	0.000	0.997	0.996	0.000	0.000	0.017
3	0.000	0.001	0.000	0.000	0.509	0.731	0.000	0.000	0.040
4	0.000	0.007	0.000	0.000	0.484	0.493	0.000	0.000	0.053
5	0.000	0.007	0.000	0.000	0.675	0.775	0.000	0.000	0.014
6	0.000	0.003	0.000	0.000	0.475	0.549	0.000	0.000	0.013
7	0.000	0.001	0.000	0.000	0.068	0.194	0.000	0.000	0.020
8	0.000	0.000	0.000	0.000	0.023	0.036	0.000	0.000	0.004
9	0.000	0.000	0.000	0.000	0.010	0.021	0.000	0.000	0.002
10	0.000	0.000	0.000	0.000	0.020	0.017	0.000	0.000	0.002
11	0.000	0.000	0.000	0.000	0.007	0.010	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.013	0.003	0.000	0.000	0.002

Table 4: Proportion of individual models with higher relative RMSE for out-of-sample CPI forecasts using ARX(1) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.397	0.483	0.500	0.172	0.172	0.500	0.328	0.138
2	0.759	0.759	0.759	0.103	0.103	0.759	0.155	0.310
3	0.931	0.776	0.776	0.172	0.172	0.776	0.138	0.345
4	0.931	0.810	0.793	0.759	0.759	0.793	0.121	0.793
5	0.931	0.793	0.776	0.741	0.741	0.776	0.103	0.776
6	0.931	0.828	0.759	0.897	0.897	0.759	0.103	0.397
7	0.879	0.793	0.759	0.914	0.914	0.759	0.052	0.741
8	0.862	0.776	0.776	0.914	0.914	0.776	0.034	0.776
9	0.862	0.793	0.776	0.983	0.983	0.776	0.017	0.828
10	0.845	0.793	0.776	0.983	0.983	0.759	0.017	0.897
11	0.793	0.793	0.776	0.983	0.983	0.776	0.017	0.879
12	0.793	0.793	0.776	0.983	0.983	0.776	0.017	0.862

Table 5: Proportion of periods in which model has smaller absolute forecast error than AR model for out-of-sample CPI forecasts using ARX(1) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.344	0.406	0.406	0.469	0.469	0.406	0.313	0.406
2	0.500	0.563	0.563	0.438	0.438	0.531	0.313	0.594
3	0.719	0.656	0.625	0.531	0.531	0.625	0.250	0.563
4	0.719	0.656	0.656	0.688	0.688	0.656	0.219	0.500
5	0.813	0.781	0.781	0.656	0.656	0.781	0.219	0.531
6	0.875	0.781	0.813	0.688	0.688	0.813	0.250	0.469
7	0.844	0.781	0.781	0.750	0.750	0.781	0.250	0.563
8	0.813	0.781	0.781	0.844	0.844	0.781	0.250	0.594
9	0.813	0.813	0.813	0.906	0.906	0.813	0.250	0.594
10	0.906	0.844	0.844	0.906	0.906	0.844	0.219	0.656
11	0.938	0.875	0.875	0.906	0.906	0.875	0.219	0.719
12	0.938	0.906	0.906	0.938	0.938	0.906	0.250	0.813

Table 6: Individual variables with top 10 average weights in *BMA* and *AITMA* model averaging using ARX(1) models (period: 1997 Q2-2004 Q1)

$BMA(\phi=20)$		$BMA(\phi=2)$		$BMA(\phi=0.5)$		$AITMA$		$SITMA$	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
S21	0.111	S21	0.027	AR	0.020	S21	0.846	S21	0.846
S45	0.048	AR	0.027	S21	0.019	S23	0.039	S23	0.039
AR	0.047	S45	0.022	S45	0.018	S45	0.031	S45	0.030
S44	0.037	S44	0.021	S44	0.018	S11	0.028	S11	0.027
S23	0.034	S9	0.021	S9	0.018	S44	0.016	S44	0.016
S9	0.030	S10	0.020	S10	0.018	S10	0.007	S10	0.007
S10	0.029	S23	0.020	S23	0.018	S22	0.006	S22	0.006
S22	0.022	S18	0.019	S18	0.018	S15	0.004	S15	0.004
S18	0.022	S46	0.019	S46	0.017	S12	0.003	S12	0.003
S11	0.021	S19	0.018	S19	0.017	S13	0.002	S13	0.002

Table 7: Relative RMSE of out-of-sample CPI forecasts using ARX(2) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi=20$)	BMA ($\phi=2$)	BMA ($\phi=0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.989	0.992	0.992	0.990	**0.988**	0.992	1.095	1.065
2	**0.937***	0.964*	0.967*	0.971	0.971	0.967*	1.104*	0.961
3	0.875*	0.943*	0.951*	**0.832**	**0.832**	0.954*	1.147*	0.824*
4	**0.870***	0.934*	0.944*	1.020	1.020	0.948*	1.219*	0.880
5	**0.868***	0.925*	0.936*	1.081	1.081	0.940*	1.253*	0.922
6	**0.864***	0.918*	0.930*	1.029	1.029	0.934*	1.289*	0.945
7	**0.860***	0.911*	0.923*	0.953	0.953	0.928*	1.272*	0.912
8	0.867*	0.907*	0.918*	**0.824**	**0.824**	0.922*	1.254*	0.884
9	0.885	0.912	0.921	**0.737**	**0.737**	0.924	1.222*	0.860*
10	0.896	0.914	0.921	**0.706***	**0.706***	0.925	1.208*	0.869*
11	0.900	0.915	0.922	**0.708***	**0.708***	0.925	1.191*	0.866*
12	0.907	0.918	0.924	**0.673***	**0.673***	0.927	1.157	0.845

* 10% rejection of Diebold-Mariano test that the forecast differs from the benchmark

bold indicates best forecast in row

Table 8: Probability values of forecast density KS tests for out-of-sample CPI forecasts using ARX(2) models (period: 1997 Q2-2004 Q1)

Horizon	AR	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.093	0.109	0.106	0.119	0.112	0.103	0.109	0.007	0.047
2	0.001	0.005	0.002	0.002	0.602	0.638	0.002	0.001	0.127
3	0.000	0.004	0.000	0.000	0.332	0.409	0.000	0.000	0.055
4	0.000	0.001	0.000	0.000	0.397	0.430	0.000	0.000	0.020
5	0.000	0.000	0.000	0.000	0.049	0.067	0.000	0.000	0.006
6	0.000	0.000	0.000	0.000	0.016	0.013	0.000	0.000	0.001
7	0.000	0.000	0.000	0.000	0.006	0.003	0.000	0.000	0.007
8	0.000	0.000	0.000	0.000	0.010	0.013	0.000	0.000	0.001
9	0.000	0.000	0.000	0.000	0.006	0.005	0.000	0.000	0.001
10	0.000	0.000	0.000	0.000	0.002	0.004	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 9: Proportion of individual models with higher relative RMSE for out-of-sample CPI forecasts using ARX(2) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.810	0.793	0.793	0.793	0.810	0.793	0.207	0.276
2	0.828	0.793	0.793	0.759	0.759	0.793	0.155	0.810
3	0.897	0.845	0.845	0.931	0.931	0.828	0.155	0.931
4	0.879	0.828	0.810	0.345	0.345	0.810	0.103	0.862
5	0.862	0.793	0.793	0.155	0.155	0.793	0.103	0.793
6	0.845	0.776	0.776	0.328	0.328	0.776	0.052	0.776
7	0.793	0.759	0.759	0.724	0.724	0.759	0.034	0.759
8	0.776	0.759	0.759	0.828	0.828	0.759	0.034	0.759
9	0.793	0.759	0.759	0.931	0.931	0.759	0.017	0.810
10	0.793	0.759	0.759	0.966	0.966	0.741	0.017	0.810
11	0.776	0.724	0.724	0.966	0.966	0.724	0.017	0.810
12	0.741	0.741	0.741	0.983	0.983	0.741	0.017	0.845

Table 10: Proportion of periods in which model has smaller absolute forecast error than AR model for out-of-sample CPI forecasts using ARX(2) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.469	0.500	0.500	0.594	0.594	0.500	0.375	0.469
2	0.625	0.688	0.688	0.531	0.531	0.656	0.219	0.563
3	0.719	0.813	0.813	0.719	0.719	0.781	0.406	0.594
4	0.750	0.750	0.750	0.531	0.531	0.750	0.313	0.719
5	0.719	0.750	0.750	0.594	0.594	0.750	0.281	0.625
6	0.781	0.781	0.781	0.594	0.594	0.781	0.219	0.594
7	0.781	0.813	0.781	0.719	0.719	0.813	0.063	0.625
8	0.813	0.844	0.844	0.750	0.750	0.844	0.063	0.688
9	0.875	0.875	0.875	0.844	0.844	0.875	0.125	0.688
10	0.938	0.906	0.906	0.906	0.906	0.906	0.063	0.719
11	0.969	0.938	0.938	0.906	0.906	0.938	0.094	0.750
12	0.969	0.938	0.906	0.969	0.969	0.906	0.219	0.813

Table 11: Individual variables with top 10 average weights in *BMA* and *AITMA* model averaging using ARX(2) models (period: 1997 Q2-2004 Q1)

$BMA(\phi = 20)$		$BMA(\phi = 2)$		$BMA(\phi = 0.5)$		$AITMA$		$SITMA$	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
AR	0.161	AR	0.044	AR	0.024	S21	0.553	S21	0.552
S21	0.074	S21	0.025	S21	0.019	S11	0.152	S11	0.151
S11	0.050	S11	0.023	S11	0.018	S24	0.079	S24	0.078
S45	0.042	S45	0.021	S45	0.018	S45	0.068	S45	0.067
S44	0.034	S9	0.020	S9	0.018	S28	0.045	S28	0.045
S9	0.029	S44	0.020	S44	0.018	S44	0.038	S44	0.037
S10	0.025	S10	0.020	S10	0.018	S8	0.016	S8	0.014
S28	0.024	S28	0.020	S28	0.018	S23	0.013	S23	0.013
S23	0.023	S18	0.019	S18	0.018	S10	0.005	AR	0.007
S18	0.020	S19	0.018	S19	0.017	S34	0.004	S10	0.005

Table 12: Relative RMSE of out-of-sample CPI forecasts using ARX(k) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	1.128*	1.026	1.001	**0.946**	1.068	0.992	0.939	0.969
2	0.948	0.963*	0.964*	**0.773***	0.831*	0.964*	0.974	0.808*
3	0.871*	0.960*	0.966*	**0.705***	**0.705***	0.966*	1.023*	0.840*
4	0.995	0.980	0.972*	**0.849**	0.893	0.966*	1.074*	0.906
5	**0.821***	0.915*	0.931*	0.924	0.932	0.934*	1.080	0.839
6	0.879	0.928	0.937	**0.858**	0.868	0.939	1.151*	0.910
7	0.818*	0.911*	0.928*	0.797	**0.794**	0.934*	1.213*	0.889
8	0.822*	0.906*	0.923	**0.718**	**0.718**	0.928	1.175*	0.853*
9	0.867	0.925	0.935	**0.706***	**0.706***	0.938	1.176*	0.854*
10	0.919	0.943*	0.948*	**0.694***	**0.694***	0.949*	1.160*	0.840*
11	0.924*	0.940*	0.944	**0.689***	**0.689***	0.946	1.145*	0.840*
12	0.930*	0.942*	0.944*	0.684*	**0.683***	0.945*	1.126*	0.836*

* 10% rejection of Diebold-Mariano test that the forecast differs from the benchmark

bold indicates best forecast in row

Table 13: Probability values of forecast density KS tests for out-of-sample CPI forecasts using ARX(k) models (period: 1997 Q2-2004 Q1)

Horizon	AR	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.002	0.003	0.006	0.006	0.129	0.052	0.005	0.015	0.697
2	0.000	0.003	0.001	0.000	0.226	0.239	0.000	0.004	0.570
3	0.000	0.003	0.000	0.000	0.249	0.168	0.000	0.000	0.174
4	0.000	0.000	0.000	0.000	0.320	0.273	0.000	0.000	0.085
5	0.000	0.007	0.000	0.000	0.836	0.874	0.000	0.000	0.018
6	0.000	0.001	0.000	0.000	0.456	0.453	0.000	0.000	0.040
7	0.000	0.004	0.000	0.000	0.224	0.138	0.000	0.000	0.078
8	0.000	0.001	0.000	0.000	0.016	0.009	0.000	0.000	0.013
9	0.000	0.000	0.000	0.000	0.012	0.008	0.000	0.000	0.007
10	0.000	0.000	0.000	0.000	0.004	0.007	0.000	0.000	0.002
11	0.000	0.000	0.000	0.000	0.005	0.009	0.000	0.000	0.002
12	0.000	0.000	0.000	0.000	0.002	0.001	0.000	0.000	0.003

Table 14: Proportion of individual models with higher relative RMSE for out-of-sample CPI forecasts using ARX(k) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.138	0.414	0.672	0.897	0.241	0.759	0.914	0.879
2	0.828	0.776	0.776	0.948	0.931	0.776	0.690	0.931
3	0.931	0.845	0.845	0.948	0.948	0.845	0.259	0.931
4	0.690	0.707	0.759	0.931	0.879	0.776	0.172	0.879
5	0.914	0.793	0.793	0.793	0.793	0.776	0.155	0.897
6	0.862	0.793	0.793	0.862	0.862	0.793	0.121	0.828
7	0.879	0.776	0.759	0.914	0.914	0.759	0.052	0.828
8	0.862	0.793	0.759	0.931	0.931	0.759	0.052	0.845
9	0.845	0.776	0.776	0.983	0.983	0.776	0.034	0.845
10	0.810	0.793	0.793	0.983	0.983	0.776	0.017	0.897
11	0.776	0.759	0.759	0.983	0.983	0.759	0.017	0.897
12	0.776	0.776	0.776	0.983	0.983	0.776	0.017	0.897

Table 15: Proportion of periods in which model has smaller absolute forecast error than AR model for out-of-sample CPI forecasts using ARX(k) models (period: 1997 Q2-2004 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.375	0.406	0.438	0.563	0.313	0.531	0.625	0.406
2	0.594	0.719	0.750	0.656	0.656	0.719	0.656	0.719
3	0.688	0.750	0.688	0.688	0.688	0.750	0.406	0.656
4	0.531	0.531	0.531	0.750	0.688	0.594	0.344	0.469
5	0.688	0.750	0.750	0.719	0.719	0.781	0.406	0.656
6	0.656	0.688	0.688	0.719	0.719	0.719	0.281	0.563
7	0.656	0.750	0.750	0.719	0.719	0.781	0.313	0.594
8	0.750	0.750	0.750	0.781	0.781	0.781	0.313	0.656
9	0.813	0.781	0.813	0.844	0.844	0.813	0.281	0.656
10	0.906	0.844	0.844	0.875	0.875	0.844	0.250	0.719
11	0.938	0.875	0.875	0.906	0.906	0.875	0.250	0.781
12	0.938	0.906	0.906	0.938	0.938	0.906	0.281	0.844

Table 16: Individual variables with top 10 average weights in *BMA* and *AITMA* model averaging using ARX(k) models (period: 1997 Q2-2004 Q1)

$BMA(\phi = 20)$		$BMA(\phi = 2)$		$BMA(\phi = 0.5)$		$AITMA$		$SITMA$	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
S21	0.126	S21	0.031	S21	0.021	S21	0.704	S21	0.695
S23	0.064	S23	0.027	AR	0.020	S45	0.115	S45	0.108
S45	0.050	S9	0.026	S23	0.020	S44	0.070	S44	0.065
S9	0.046	S10	0.025	S9	0.019	S23	0.036	S23	0.054
S10	0.046	AR	0.025	S10	0.019	S34	0.018	S11	0.020
S44	0.042	S45	0.023	S22	0.019	S15	0.016	S15	0.014
S11	0.039	S11	0.023	S11	0.019	S46	0.006	S34	0.011
AR	0.036	S22	0.023	S45	0.019	S11	0.006	S46	0.005
S22	0.032	S44	0.022	S19	0.019	S47	0.006	S10	0.005
S19	0.024	S19	0.022	S44	0.018	S10	0.005	S47	0.005

Table 17: Relative RMSE of out-of-sample CPI forecasts using ARX(1) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.965*	0.970*	0.970*	**0.925**	0.926	0.970*	1.009	1.031
2	0.921*	0.946*	0.948*	**0.831***	**0.831***	0.949*	0.978	1.012
3	0.892*	0.930*	0.934*	**0.790**	**0.790**	0.935*	0.975	0.929
4	0.870*	0.920*	0.927*	**0.758**	**0.758**	0.929*	0.950	0.970
5	0.846*	0.905*	0.913*	**0.691***	**0.691***	0.916*	0.898*	0.919
6	0.827*	0.890*	0.899*	**0.674***	0.675*	0.902*	0.836*	0.829
7	0.822*	0.882*	0.892*	**0.658***	**0.658***	0.895*	0.763*	0.797
8	0.843	0.891	0.899	**0.656**	**0.656**	0.901	0.752	0.828
9	0.872	0.904	0.909	**0.756**	**0.756**	0.911	0.773	0.817*
10	0.902	0.922	0.925	**0.823**	**0.823**	0.927	0.835	0.892
11	0.914	0.928	0.931	0.882	0.882	0.932	**0.864**	0.893*
12	0.926	0.936	0.938	0.928	0.928	0.939	**0.868**	0.908

* 10% rejection of Diebold-Mariano test that the forecast differs from the benchmark

bold indicates best forecast in row

Table 18: Probability values of forecast density KS tests for out-of-sample CPI forecasts using ARX(1) models (period: 1990 Q2-1997 Q1)

Horizon	AR	$BMA(20)$ ($\phi = 20$)	$BMA(2)$ ($\phi = 2$)	$BMA(0.5)$ ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.885	1.000	0.997	0.998	0.907	0.913	0.999	0.999	0.570
2	0.646	0.995	0.985	0.989	0.551	0.668	0.983	0.668	0.660
3	0.066	0.188	0.124	0.114	0.673	0.716	0.105	0.301	0.137
4	0.029	0.159	0.070	0.060	0.670	0.670	0.055	0.063	0.242
5	0.012	0.012	0.008	0.008	0.226	0.161	0.010	0.032	0.071
6	0.021	0.009	0.007	0.006	0.194	0.143	0.007	0.041	0.035
7	0.009	0.006	0.006	0.006	0.075	0.082	0.006	0.015	0.020
8	0.010	0.006	0.005	0.005	0.115	0.106	0.006	0.015	0.010
9	0.003	0.005	0.004	0.004	0.010	0.011	0.004	0.001	0.011
10	0.009	0.006	0.004	0.004	0.117	0.149	0.005	0.090	0.019
11	0.010	0.006	0.005	0.004	0.286	0.311	0.004	0.005	0.067
12	0.007	0.007	0.006	0.006	0.018	0.018	0.006	0.015	0.167

Table 19: Proportion of individual models with higher relative RMSE for out-of-sample CPI forecasts using ARX(1) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.862	0.862	0.862	0.914	0.914	0.862	0.483	0.121
2	0.862	0.828	0.828	0.983	0.983	0.828	0.776	0.345
3	0.862	0.828	0.828	0.966	0.966	0.810	0.741	0.828
4	0.879	0.793	0.793	0.966	0.966	0.776	0.759	0.707
5	0.862	0.776	0.776	0.983	0.983	0.776	0.776	0.759
6	0.879	0.776	0.776	1.000	1.000	0.759	0.879	0.879
7	0.879	0.793	0.776	1.000	1.000	0.776	0.914	0.897
8	0.862	0.810	0.810	1.000	1.000	0.810	0.931	0.879
9	0.862	0.845	0.845	0.948	0.948	0.845	0.931	0.914
10	0.897	0.862	0.845	0.948	0.948	0.845	0.931	0.897
11	0.862	0.828	0.828	0.914	0.914	0.828	0.948	0.914
12	0.845	0.810	0.810	0.845	0.845	0.810	0.966	0.897

Table 20: Proportion of periods in which model has smaller absolute forecast error than AR model for out-of-sample CPI forecasts using ARX(1) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.750	0.750	0.750	0.719	0.719	0.750	0.594	0.563
2	0.750	0.750	0.750	0.719	0.719	0.750	0.594	0.531
3	0.813	0.844	0.844	0.719	0.719	0.813	0.625	0.625
4	0.875	0.875	0.844	0.750	0.750	0.844	0.656	0.594
5	0.875	0.906	0.906	0.813	0.813	0.906	0.719	0.563
6	0.906	0.875	0.875	0.813	0.813	0.875	0.750	0.656
7	0.813	0.844	0.844	0.813	0.813	0.844	0.781	0.656
8	0.781	0.781	0.781	0.813	0.813	0.781	0.719	0.531
9	0.750	0.750	0.750	0.719	0.719	0.750	0.750	0.563
10	0.813	0.813	0.813	0.750	0.750	0.813	0.719	0.625
11	0.781	0.781	0.781	0.719	0.719	0.781	0.688	0.719
12	0.781	0.781	0.781	0.781	0.781	0.781	0.656	0.719

Table 21: Individual variables with top 10 average weights in *BMA* and *AITMA* model averaging using ARX(1) models (period: 1990 Q2-1997 Q1)

$BMA(\phi = 20)$		$BMA(\phi = 2)$		$BMA(\phi = 0.5)$		$AITMA$		$SITMA$	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
AR	0.058	AR	0.028	AR	0.020	S23	0.174	S23	0.174
S23	0.035	S23	0.019	S23	0.017	S21	0.149	S21	0.149
S11	0.032	S21	0.019	S21	0.017	S11	0.107	S11	0.107
S12	0.030	S11	0.019	S22	0.017	S12	0.102	S12	0.102
S22	0.030	S22	0.019	S11	0.017	S13	0.071	S13	0.071
S13	0.030	S12	0.019	S12	0.017	S22	0.069	S22	0.069
S21	0.030	S13	0.019	S13	0.017	S10	0.065	S10	0.065
S15	0.021	S15	0.018	S15	0.017	S44	0.031	S44	0.031
S10	0.021	S10	0.018	S10	0.017	S58	0.027	S58	0.027
S9	0.019	S9	0.018	S9	0.017	S9	0.025	S9	0.025

Table 22: Relative RMSE of out-of-sample CPI forecasts using ARX(2) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.985*	0.983*	0.982*	0.996	0.996	0.982*	**0.979**	1.101
2	0.952*	0.960*	0.961*	**0.933**	**0.933**	0.961*	0.999	1.075
3	0.920*	0.936*	0.938*	**0.892**	**0.892**	0.939*	0.951	1.089
4	0.881*	0.918*	0.924*	**0.739***	**0.739***	0.926*	0.905	1.114
5	0.851*	0.896*	0.904*	**0.666***	**0.666***	0.906*	0.836*	0.997
6	0.836*	0.879	0.887	**0.635**	**0.635**	0.890	0.751*	0.910
7	0.841*	0.876*	0.883*	**0.663**	**0.663**	0.886*	0.730	0.909
8	0.860*	0.880*	0.886*	**0.722***	**0.722***	0.888	0.749	0.897
9	0.888	0.897	0.901	**0.800**	**0.800**	0.902	0.804	0.935
10	0.902	0.907	0.909	**0.867**	**0.867**	0.911	0.866	0.923
11	0.905	0.908	0.911	**0.875***	**0.875***	0.912	0.879	0.901
12	0.921	0.920	0.922	**0.867**	**0.867**	0.922	0.897	0.965

* 10% rejection of Diebold-Mariano test that the forecast differs from the benchmark

bold indicates best forecast in row

Table 23: Probability values of forecast density KS tests for out-of-sample CPI forecasts using ARX(2) models (period: 1990 Q2-1997 Q1)

Horizon	AR	$BMA(20)$ ($\phi = 20$)	$BMA(2)$ ($\phi = 2$)	$BMA(0.5)$ ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.990	0.997	1.000	1.000	0.986	0.996	1.000	0.807	0.631
2	0.484	0.687	0.745	0.773	0.422	0.397	0.764	0.549	0.029
3	0.062	0.286	0.124	0.102	0.181	0.123	0.102	0.027	0.049
4	0.035	0.065	0.053	0.052	0.034	0.038	0.056	0.148	0.016
5	0.018	0.012	0.008	0.009	0.023	0.033	0.010	0.122	0.004
6	0.006	0.008	0.009	0.008	0.122	0.129	0.007	0.126	0.001
7	0.012	0.010	0.008	0.008	0.229	0.229	0.007	0.033	0.000
8	0.003	0.004	0.004	0.004	0.216	0.200	0.004	0.004	0.005
9	0.010	0.026	0.007	0.007	0.136	0.057	0.006	0.001	0.009
10	0.009	0.022	0.018	0.015	0.118	0.103	0.015	0.033	0.007
11	0.005	0.008	0.007	0.006	0.098	0.135	0.006	0.001	0.020
12	0.002	0.002	0.002	0.002	0.053	0.044	0.002	0.009	0.002

Table 24: Proportion of individual models with higher relative RMSE for out-of-sample CPI forecasts using ARX(2) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.810	0.828	0.828	0.690	0.690	0.828	0.862	0.000
2	0.862	0.862	0.845	0.862	0.862	0.845	0.690	0.086
3	0.845	0.845	0.828	0.862	0.862	0.828	0.810	0.017
4	0.845	0.793	0.776	0.966	0.966	0.776	0.828	0.000
5	0.845	0.776	0.759	0.983	0.983	0.759	0.862	0.552
6	0.845	0.793	0.759	0.983	0.983	0.759	0.948	0.724
7	0.828	0.793	0.759	0.983	0.983	0.759	0.931	0.707
8	0.810	0.793	0.776	0.948	0.948	0.759	0.931	0.759
9	0.810	0.793	0.759	0.931	0.931	0.759	0.931	0.707
10	0.828	0.810	0.793	0.879	0.879	0.793	0.879	0.759
11	0.828	0.828	0.828	0.897	0.897	0.828	0.879	0.862
12	0.862	0.862	0.862	0.931	0.931	0.862	0.897	0.724

Table 25: Proportion of periods in which model has smaller absolute forecast error than AR model for out-of-sample CPI forecasts using ARX(2) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.656	0.688	0.688	0.469	0.469	0.688	0.563	0.438
2	0.750	0.750	0.781	0.656	0.656	0.781	0.531	0.469
3	0.750	0.813	0.813	0.688	0.688	0.813	0.500	0.531
4	0.875	0.938	0.938	0.750	0.750	0.938	0.594	0.500
5	0.906	0.906	0.875	0.781	0.781	0.875	0.625	0.594
6	0.813	0.844	0.844	0.781	0.781	0.875	0.719	0.563
7	0.844	0.844	0.844	0.813	0.813	0.844	0.688	0.563
8	0.781	0.781	0.781	0.688	0.688	0.781	0.750	0.625
9	0.781	0.813	0.813	0.688	0.688	0.813	0.750	0.563
10	0.813	0.813	0.813	0.719	0.719	0.813	0.750	0.594
11	0.781	0.781	0.781	0.750	0.750	0.781	0.688	0.781
12	0.844	0.813	0.781	0.813	0.813	0.781	0.688	0.844

Table 26: Individual variables with top 10 average weights in *BMA* and *AITMA* model averaging using ARX(2) models (period: 1990 Q2-1997 Q1)

$BMA(\phi = 20)$		$BMA(\phi = 2)$		$BMA(\phi = 0.5)$		$AITMA$		$SITMA$	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
AR	0.212	AR	0.047	AR	0.025	S11	0.239	S11	0.239
S11	0.031	S11	0.019	S11	0.017	S23	0.123	S23	0.122
S23	0.026	S23	0.018	S23	0.017	S21	0.093	S21	0.092
S12	0.025	S12	0.018	S13	0.017	S12	0.070	S12	0.070
S13	0.025	S13	0.018	S12	0.017	S22	0.068	S22	0.067
S21	0.022	S21	0.018	S21	0.017	S13	0.066	S13	0.065
S22	0.022	S22	0.018	S22	0.017	S10	0.046	S10	0.046
S15	0.020	S15	0.018	S15	0.017	S24	0.040	S24	0.040
S10	0.018	S10	0.018	S10	0.017	S58	0.040	S58	0.039
S9	0.016	S9	0.017	S9	0.017	S44	0.027	S44	0.027

Table 27: Relative RMSE of out-of-sample CPI forecasts using ARX(k) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	1.000	0.987	0.983	**0.964**	0.968	0.982	1.021	1.090
2	0.980	0.973	0.971	**0.886**	0.904	0.971	1.027	1.137
3	0.871*	0.922*	0.935*	**0.817**	0.819	0.941*	0.982	1.017
4	0.837*	0.898*	0.916*	0.746*	**0.743***	0.924*	0.956	0.982
5	0.832*	0.888*	0.906*	**0.699***	0.703*	0.914*	0.920*	0.937
6	0.811	0.877	0.900	**0.695**	**0.695**	0.912	0.887	0.866
7	0.770*	0.864	0.892	0.797*	**0.717***	0.906	0.799	0.846
8	0.811*	0.876*	0.890	0.744*	**0.707***	0.893	0.763	0.843
9	0.893*	0.914	0.912	0.781	**0.758**	0.906	0.782*	0.851
10	0.932*	0.935*	0.925	0.845*	**0.825**	0.913	0.829	0.886
11	0.955*	0.946	0.933	**0.851***	0.882	0.919	0.892	0.919
12	0.955	0.952	0.943	0.935	0.939	0.932	**0.885**	0.927

* 10% rejection of Diebold-Mariano test that the forecast differs from the benchmark

bold indicates best forecast in row

Table 28: Probability values of forecast density KS tests for out-of-sample CPI forecasts using ARX(k) models (period: 1990 Q2-1997 Q1)

Horizon	AR	$BMA(20)$ ($\phi = 20$)	$BMA(2)$ ($\phi = 2$)	$BMA(0.5)$ ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.228	0.580	0.590	0.551	0.773	0.697	0.505	0.580	0.115
2	0.083	0.438	0.345	0.303	0.929	0.253	0.270	0.449	0.007
3	0.032	0.164	0.127	0.117	0.887	0.919	0.131	0.062	0.102
4	0.022	0.068	0.019	0.020	0.194	0.269	0.019	0.094	0.192
5	0.081	0.028	0.018	0.021	0.071	0.085	0.021	0.054	0.043
6	0.014	0.021	0.010	0.009	0.152	0.129	0.009	0.038	0.029
7	0.013	0.024	0.005	0.005	0.010	0.026	0.004	0.007	0.007
8	0.006	0.005	0.005	0.005	0.041	0.084	0.005	0.012	0.004
9	0.002	0.001	0.005	0.004	0.080	0.018	0.005	0.003	0.055
10	0.004	0.009	0.004	0.005	0.017	0.072	0.005	0.005	0.020
11	0.004	0.001	0.002	0.004	0.083	0.032	0.007	0.006	0.026
12	0.005	0.003	0.002	0.002	0.015	0.032	0.004	0.038	0.074

Table 29: Proportion of individual models with higher relative RMSE for out-of-sample CPI forecasts using ARX(k) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.707	0.793	0.793	0.879	0.879	0.793	0.431	0.034
2	0.810	0.810	0.828	0.983	0.931	0.828	0.310	0.000
3	0.897	0.845	0.810	0.948	0.948	0.810	0.724	0.207
4	0.862	0.845	0.828	0.983	0.983	0.828	0.741	0.655
5	0.879	0.828	0.828	0.983	0.983	0.810	0.793	0.759
6	0.931	0.828	0.793	0.983	0.983	0.759	0.828	0.828
7	0.914	0.845	0.810	0.897	0.966	0.793	0.897	0.897
8	0.897	0.828	0.810	0.931	0.966	0.810	0.914	0.862
9	0.845	0.828	0.828	0.948	0.948	0.845	0.948	0.897
10	0.828	0.810	0.862	0.931	0.948	0.862	0.931	0.897
11	0.793	0.810	0.828	0.966	0.931	0.862	0.931	0.862
12	0.810	0.845	0.862	0.879	0.862	0.879	0.966	0.879

Table 30: Proportion of periods in which model has smaller absolute forecast error than AR model for out-of-sample CPI forecasts using ARX(k) models (period: 1990 Q2-1997 Q1)

Horizon	BMA ($\phi = 20$)	BMA ($\phi = 2$)	BMA ($\phi = 0.5$)	$AITMA$	$SITMA$	AV	1 Factor	5 Factors
1	0.656	0.719	0.688	0.688	0.563	0.688	0.563	0.469
2	0.719	0.688	0.688	0.656	0.719	0.688	0.531	0.438
3	0.781	0.750	0.781	0.781	0.750	0.781	0.594	0.531
4	0.875	0.813	0.844	0.688	0.719	0.906	0.656	0.563
5	0.688	0.719	0.781	0.688	0.688	0.844	0.594	0.531
6	0.781	0.781	0.813	0.594	0.625	0.813	0.531	0.625
7	0.844	0.875	0.906	0.656	0.688	0.906	0.563	0.719
8	0.844	0.844	0.844	0.688	0.688	0.844	0.656	0.656
9	0.688	0.688	0.719	0.625	0.656	0.750	0.750	0.656
10	0.781	0.813	0.813	0.719	0.656	0.813	0.688	0.656
11	0.781	0.781	0.781	0.781	0.781	0.781	0.594	0.688
12	0.656	0.719	0.719	0.625	0.656	0.719	0.625	0.719

Table 31: Individual variables with top 10 average weights in *BMA* and *AITMA* model averaging using ARX(k) models (period: 1990 Q2-1997 Q1)

$BMA(\phi = 20)$		$BMA(\phi = 2)$		$BMA(\phi = 0.5)$		$AITMA$		$SITMA$	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
S11	0.063	S11	0.026	AR	0.020	S23	0.222	S23	0.209
S23	0.058	AR	0.026	S22	0.019	S12	0.090	S11	0.105
S22	0.052	S22	0.025	S11	0.019	S11	0.087	S12	0.088
S12	0.051	S23	0.025	S23	0.019	S21	0.076	S21	0.086
S13	0.050	S12	0.025	S12	0.019	S15	0.071	S22	0.071
AR	0.044	S13	0.025	S13	0.019	S13	0.062	S15	0.069
S21	0.038	S21	0.023	S21	0.019	S22	0.061	S13	0.063
S10	0.029	S14	0.021	S14	0.018	S58	0.048	S58	0.042
S14	0.027	S16	0.020	S19	0.018	S10	0.034	S10	0.040
S16	0.024	S10	0.020	S16	0.018	S44	0.025	S9	0.023

Appendix B: Data definitions

In this appendix, we provide a list of the series used in Section 4 to forecast UK inflation. These series come from a data set which has been constructed to match the set used by Stock and Watson (2002). In total, this data set has 131 series, comprising 20 output series, 25 labour market series, 9 retail and trade series, 6 consumption series, 6 series on housing starts, 12 series on inventories and sales, 8 series on orders, 7 stock price series, 5 exchange rate series, 7 interest rate series and 6 monetary aggregates, 19 price indices and an economic sentiment index. We retained the 58 series with at least 90 observations. For each series used in Section 4 the list gives the FAME alias, a brief description, seasonal adjustment (SA), the transformation applied to the series to ensure stationarity and the first available observation. The transformations applied to the series are: 1 = no transformation; 2 = first difference; 3 = second difference; 4 = logarithm; 5 = first difference of logarithm; 6 = second difference of logarithm. Series 3, 4, 5, 10, 11, 12, 13, 21 and 32 are derived series, described below. The series are grouped under 10 categories.

Series 1 to 8: Real output and income

- S1: ABMI: Gross Domestic Product: chained volume measures: SA 5 Q1:1955
- S2: CKYY IOP: Manufacturing SA 5 Q1:1948
- S3: IOP: Durable Manufacturing SA 5 Q1:1948
- S4: IOP: Semi-durable Manufacturing SA 5 Q1:1948; constructed as CKZB (IOP: Industry DB: Manuf of textile & textile products) plus CKZC (IOP: Industry DC: Manuf of leather & leather products) plus CKZG (IOP: Industry DG: Manuf of chemicals & man-made fibres) plus CKZH (IOP: Industry DH: Manuf of rubber & plastic products)
- S5: IOP: Non-durable Manufacturing SA 5 Q1:1948; constructed as CKZA (IOP: Industry DA: Manuf of food, drink & tobacco) plus CKZE (IOP: Industry DE: Pulp/paper/printing/publishing industries) plus CKZF (IOP: Industry DF: Manuf coke/petroleum prod/nuclear fuels)
- S6: CKYX IOP: Mining & quarrying SA 5 Q1:1948
- S7: CKYZ IOP: Electricity, gas and water supply SA 5 Q1:1948
- S8: NRJR: Real households disposable income SA 5 Q1:1955

Series 9 to 21: Employment and hours

- S9: DYDC: UK Workforce jobs: Total SA 5 Q2:1959
- S10: Employed, Nonagric. Industries SA 5 Q2:1978; constructed as DYDC (UK Workforce jobs

(SA) : Total) minus LOLI (UK Workforce jobs (SA): Total - A,B Agriculture & fishing) minus LOMJ (UK Workforce jobs (SA): Total - G-Q Total services)

- S11: Employment Rate: All NSA 1 Q1:1971; concatenate MGRZ and MGRZ_EXP (LFS: In employment: UK: All: Aged 16), concatenate MGSL and MGSL_EXP (LFS: Population aged 16+: UK: All), then compute 1-MGRZ/MGSL
- S12: Employees on nonag. Payrolls: Total SA 5 Q2:1978; constructed as BCAJ (UK Employee jobs: Total (SA)) minus YEHU (UK Employee jobs (SA): All jobs Agriculture, hunting, forestry & fishing)
- S13: Employees nonag. Payrolls: Total: private SA 5 Q2:1978; constructed as S12 minus LOKS (UK Employee jobs (SA): Public admin. & defence)
- S14: YEJF Employee jobs: All jobs: Production Inds. SA 5 Q2:1978
- S15: YEHX Employee jobs: All jobs: Construction SA 5 Q2:1978
- S16: YEHW Employee jobs: All jobs: Manufacturing SA 5 Q2:1978
- S17: LOKL Employee jobs: Wholesale & retail trade SA 5 Q2:1978
- S18: YEIA Employee jobs: Banking, finance & ins. SA 5 Q2:1978
- S19: YEID Employee jobs: Total services SA 5 Q2:1978
- S20: LOKS Employee jobs: Public admin. & defence SA 5 Q2:1978
- S21: Avg. weekly hrs. prod. wkrs.: manuf. SA 1 Q1:1971; constructed from YBUS and YBUS_EXP (LFS: Total actual weekly hours worked (millions): UK: All), MGRZ and MGRZ_EXP (LFS: In employment: UK: All: Aged 16+ SA), as YBUS/MGRZ

Series 22 to 23: Trade

- S22: BOKI BOP: Balance: Total Trade in Goods SA 5 Q1:1955
- S23: ELBJ BOP: Balance: Manufactures SA 5 Q1:1970

Series 24 to 29: Consumption

- S24: ABJR Household final consumption expenditure SA 5 Q1:1955
- S25: UTID Durable goods: Total SA 5 Q1:1964
- S26: UTIT Semi-durable goods: Total SA 5 Q1:1964
- S27: UTIL Non-durable goods: Total SA 5 Q1:1964
- S28: UTIP Services: Total SA 5 Q1:1964
- S29: TMMI Purchase of vehicles SA 5 Q1:1964

Series 30 to 35: Real inventories and inventories sales

- S30: CDQN Change in Inventories: Manufacturing SA 5 Q4:1954
- S31: CDQZ Change in Inv: Manuf: Textiles & Leather SA 5 Q4:1954

- S32: Manuf & Trade Invent: Nondurable Goods SA 5 Q4:1954; constructed as CDQP (Change in Inventories: Manufacturing: Fuels) plus CDQX (Change in Inventories: Manufacturing: Food, Drink & Tobacco) plus CDQT (Change in Inventories: Manufacturing: Chemicals)
- S33: FAJX Change in Inventories: Wholesale SA 5 Q1:1959
- S34: FBYN Change in Inventories: Retail SA 5 Q1:1955
- S35: FAPF Ratio for Mfg & Trade: Inventory/Output SA 2 Q1:1955

Series 36 to 38: Stock prices

- S36: FTALLSH_PI FTSE All Share Price Index 5 Q1:1980
- S37: FTSE100_PI FTSE 100 5 Q1:1980
- S38: FTALLSH_DY FTSE All Share Dividend Yield 1 Q1:1980

Series 39 to 43: Exchange rates

- S39: A_GBG Sterling - Effective SA 5 Q1:1979
- S40: A_ERS EURO / £ SA 5 Q1:1979; constructed from A_DMS (MTH AVE - DEUTSCHEMARK /£) and fixed conversion rate of 1.95583
- S41: A_SFS SWISS FRANC /£ SA 5 Q1:1979
- S42: A_JYS JAPANESE YEN /£ SA 5 Q1:1979
- S43: A_USS UNITED STATES DOLLAR /£ SA 5 Q1:1979

Series 44 to 47: Interest rates

- S44: Spread 6-months 1
- S45: Spread 1-year 1
- S46: Spread 5-years 1
- S47: Spread 10-years 1

Series 48 to 50: Monetary and quantity credit aggregates

- S48: AUYN Money stock: M4 SA 6 Q2:1963
- S49: AVAE M0 wide monetary base SA 6 Q2:1969
- S50: AEFI BOE: reserves & other accounts outstanding NSA 6 Q1:1975

Series 51 to 57: Price indices

- S51: PLLU PPI: Output of manufactured products NSA 6 Q1:1974
- S52: LCPI Long Run CPI NSA 6 Q1:1975
- S53: ABJS Implicit Price Deflator: H'old final cons exp SA 6 Q1:1955
- S54: UTKT Durable goods: Total IDEF SA 6 Q1:1964
- S55: UTLB Semi-durable goods: Total IDEF SA 6 Q1:1964
- S56: UTKX Non-durable goods: Total IDEF SA 6 Q1:1964

- S57: UTKZ Services: Total IDEF SA 6 Q1:1964

Series 58: Surveys

- S58: MORI MORI General Economic Optimism index SA 1 Q3:1979

References

Akaike, H (1978), 'A Bayesian analysis of the minimum AIC procedure', *Annals of the Institute of Statistical Mathematics*, Vol. 30.

Akaike, H (1979), 'A Bayesian extension of the minimum AIC procedure of autoregressive model fitting', *Biometrika*, Vol. 66.

Akaike, H (1981), 'Modern development of statistical methods', in Eykhoff, P (ed), *Trends and progress in system identification*, Pergamon Press Paris, pages 169–84.

Akaike, H (1983), 'Information measures and model selection', *International Statistical Institute*, Vol. 44.

Ashley, R (1998), 'A new technique for postsample model selection and validation', *Journal of Economic Dynamics and Control*, Vol. 22, pages 647–65.

Bank of England, (1999), *Economic models at the Bank of England*, London: Bank of England.

Bank of England, (2000), *Economic models at the Bank of England: September 2000 update*, London: Bank of England.

Bank of England, (2005), *The Bank of England Quarterly Model*, London: Bank of England.

Bates, J M and Granger, C W J (1969), 'The combination of forecasts', *Operations Research Quarterly*, Vol. 20, pages 451–68.

Bozdogan, H (1987), 'Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions', *Psychometrika*, Vol. 52, No. 3, pages 345–70.

Burnham, K P and Anderson, D R (1998), *Model selection and inference*, Berlin: Springer Verlag.

Clements, M P and Hendry, D F (1998), *Forecasting economic time series*, Cambridge: CUP.

Clements, M P and Hendry, D F (2002), 'Pooling of forecasts', *Econometrics Journal*, Vol. 5, pages 1–26.

Corradi, V and Swanson, N (2006), 'Predictive density evaluation', *The handbook of economic forecasting* (forthcoming).

Diebold, F X, Gunther, T A and Tay, A S (1998), 'Evaluating density forecasts with application to financial risk management', *International Economic Review,* Vol. 39, No. 4, pages 863–83.

Diebold, F X and Mariano, R S (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics*, Vol. 13, pages 253–63.

Draper, D (1995), 'Assessment and propagation of model uncertainty', *Journal of the Royal Statistical Society Series B*, Vol. 57, pages 45–97.

Eklund, J and Karlsson, S (2004), 'Forecast combination and model averaging using predictive measures', *Stockholm School of Economics*.

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

Forni, M , Hallin, M , Lippi, M and Reichlin, L (2000), 'The generalised factor model: identification and estimation', *Review of Economics and Statistics*, Vol. 82, pages 540–54.

Hendry, D F (1999), *Dynamic econometrics*, Oxford: Oxford University Press.

Kapetanios, G (2005), 'Variable selection using non-standard optimisation of information criteria', Queen Mary and Westfield College, University of London, Mimeo, Vol. 10.

Koop, G and Potter, S (2003), 'Forecasting in large macroeconomic panels using Bayesian model averaging', *Federal Reserve Bank of New York Report 163*.

Krolzig, H M and Hendry, D F (2001), 'Computer automation of general-to-specific model selection procedures', *Journal of Economic Dynamics and Control*, Vol. 25, No. 6–7, pages 831–66.

Kullback, S and Leibler, R A (1951), 'On information and sufficiency', *Annals of Mathematical Statistics*, Vol. 22, pages 79–86.

Madigan, D and Raftery, A E (1994), 'Model selection and accounting for model uncertainty in graphical models using Occam's window', *Journal of the American Statistical Association*, Vol. 89, pages 1,535–46.

Min, C and Zellner, A (1993), 'Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates', *Journal of Econometrics*, Vol. 56, pages 89–118.

Newbold, P and Granger, C W J (1974), 'Experience with forecasting univariate time series and the combination of forecasts', *Journal of the Royal Statistical Society, Series A*, Vol. 137, pages 131–65.

Newbold, P and Harvey, D I (2002), 'Forecast combination and encompassing', in Clements, M and Hendry, D F (eds), *A companion to economic forecasting*, Oxford: Basil Blackwell, chap. 12.

Pagan, A (2003), *Report on modelling and forecasting at the Bank of England*, London: Bank of England.

Royall, R M (1997), *Statistical evidence: a likelihood paradigm*, New York: Chapman and Hall.

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

Stock, J H and Watson, M W (2002), 'Macroeconomic forecasting using diffusion indices', *Journal of Business and Economic Statistics*, Vol. 20, pages 147–62.

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

Stock, J H and Watson, M W (2004), 'Combination forecasts of output growth in a seven country dataset', *Journal of Forecasting*, Vol. 23, pages 405–30.

Stock, J H and Watson, M W (2006), 'Forecasting with many predictors', in G Elliot, C W J G and Timmermann, A (eds), *The handbook of economic forecasting* (forthcoming), Amsterdam: Elsevier.

Svensson, L E O (2004), 'Monetary policy with judgment: forecast targeting', unpublished.

Wright, J H (2003a), 'Bayesian model averaging and exchange rate forecasts', *Board of Governors of the Federal Reserve System, International Finance Discussion Papers No 779*.

Wright, J H (2003b), 'Forecasting US inflation by Bayesian model averaging', *Board of Governors of the Federal Reserve System, International Finance Discussion Papers No 780*.