Forecasting Inflation Using Dynamic Model Averaging*

Gary Koop†
University of Strathclyde

Dimitris Korobilis
Université Catholique de Louvain

June 2009, revised November 2010 and June 2011

Abstract

We forecast quarterly US inflation based on the generalized Phillips curve using econometric methods which incorporate dynamic model averaging. These methods not only allow for coefficients to change over time, but also allow for the entire forecasting model to change over time. We find that dynamic model averaging leads to substantial forecasting improvements over simple benchmark regressions and more sophisticated approaches such as those using time varying coefficient models. We also provide evidence on which sets of predictors are relevant for forecasting in each period.

Keywords: Bayesian, State space model, Phillips curve

JEL Classification: E31, E37, C11, C53

*We would like to thank three referees, the editor of this journal, participants at the 20th EC^2 Conference in Real-Time Econometrics, the 6th EUROSTAT Colloquim on Modern Tools for Business Cycle Analysis, the European Central Bank, the Bundesbank, the Bank of England and the Universities of Aarhus, Cambridge, Lancaster, Namur and Queen Mary for helpful comments. This research was supported by the ESRC under grant RES-062-23-2646. Both authors are Fellows at the Rimini Centre for Economic Analysis.

†Corresponding author, Department of Economics, University of Strathclyde, 130 Rottenrow, Glasgow G4 0GE, UK. Email: Gary.Koop@strath.ac.uk

1 Introduction

Forecasting inflation is one of the more important, but difficult, exercises in macroeconomics. Many different approaches have been suggested. Perhaps the most popular are those based on extensions of the Phillips curve. This literature is too voluminous to survey here, but a few representative and influential papers include Ang, Bekaert and Wei (2007), Atkeson and Ohanian (2001), Groen, Paap and Ravazzolo (2010), Stock and Watson (1999) and Stock and Watson (2008). The details of these papers differ, but the general framework involves a dependent variable such as inflation (or the change in inflation) and explanatory variables including lags of inflation, the unemployment rate and other predictors. Recursive, regression-based methods, have had some success. However, three issues arise when using such methods.

First, the coefficients on the predictors can change over time. For instance, it is commonly thought that the slope of the Phillips curve has changed over time. If so, the coefficients on the predictors that determine this slope will be changing. More broadly, there is a large literature in macroeconomics which documents structural breaks and other sorts of parameter change in many time series variables (see, among many others, Stock and Watson, 1996). Recursive methods are poorly designed to capture such parameter change. It is better to build models designed to capture it.

Second, the number of potential predictors can be large. Groen, Paap and Ravazzolo (2010) consider ten predictors. Researchers working with factor models such as Stock and Watson (1999) typically have many more than this. The existence of so many predictors can result in a huge number of models. If the set of models is defined by whether each of m potential predictors is included or excluded, then the researcher has 2^m models. This raises substantive statistical problems for model selection strategies. In light of this, many authors have turned to Bayesian methods, either to do Bayesian model averaging (BMA) or to automate the model selection process. Examples in macroeconomics and finance include Avramov (2002), Cremers (2002) and Koop and Potter (2004). Computational demands can become daunting when the researcher is facing 2^m models.

Third, the model relevant for forecasting can potentially change over time. For instance, the set of predictors for inflation may have been different in the 1970s than now. Or some variables may predict well in recessions but not in expansions. Furthermore, papers such as Stock and Watson (2008) find that Phillips curve forecasts work well in some periods, but at other periods simpler univariate forecasting strategies work better. In a application, Pesaran and Timmermann (2005) document how regressors useful for explaining stock returns change over time. Such arguments suggest that the forecasting model is changing

over time. This kind of issue further complicates an already difficult econometric exercise. That is, if the researcher has 2^m models and, at each point in time, a different forecasting model may apply, then the number of combinations of models which must be estimated in order to forecast at time τ is $2^{m\tau}$. Even in relatively simple forecasting exercises, it can be computationally infeasible to forecast by simply going through all of these $2^{m\tau}$ combinations. For this reason, to our knowledge, there is no literature on forecasting inflation with many predictors where the coefficients on those predictors may change over time and where a different forecasting model might hold at each point in time. A purpose of this paper is to fill this gap.

In this paper, we consider a strategy developed by Raftery, Karny and Ettler (2010) which they refer to as dynamic model averaging or DMA. Their approach can also be used for dynamic model selection (DMS) where a single (potentially different) model can be used as the forecasting model at each point in time. DMA or DMS seem ideally suited for the problem of forecasting inflation since they allow for the forecasting model to change over time while, at the same time, allowing for coefficients in each model to evolve over time. They involve only standard econometric methods for state space models such as the Kalman filter but (via some empirically-sensible approximations) achieve vast gains in computational efficiency so as to allow DMA and DMS to be done in real time despite the computational problem described in the preceding paragraph.

We use these methods in the context of a forecasting exercise with quarterly US data from 1960Q1 through 2008Q4. We use two measures of inflation and fourteen predictors and compare the forecasting performance of DMA and DMS to a wide variety of alternative forecasting procedures. DMA and DMS indicate that the set of good predictors for inflation changes substantially over time. Due to this, we find DMA and DMS to forecast very well (in terms of forecasting metrics such as log predictive likelihoods, MSFEs and MAFEs), in most cases leading to large improvements in forecast performance relative to alternative approaches.

2 Forecasting Inflation

2.1 Generalized Phillips curve models

Many forecasting models of inflation are based on the Phillips curve in which current inflation depends only on the unemployment rate and lags of inflation and unemployment. Authors such as Stock and Watson (1999) include additional predictors leading to the so-called generalized Phillips curve. We take as a starting point, on which all models used

in this paper build, the following generalized Phillips curve:

$$y_t = \phi + x'_{t-1}\beta + \varepsilon_t \tag{1}$$

where y_t is inflation which we define as $100 \ln\left(\frac{P_t}{P_{t-1}}\right)$, with P_t being a price index, and x_t is a vector of predictors (including lagged inflation). This equation is relevant for forecasting at time t given information through time $t-1$. When forecasting $h > 1$ periods ahead, the direct method of forecasting can be used and $y_t = (100/h) \ln\left(\frac{P_t}{P_{t-h}}\right)$ and x_{t-1} becomes x_{t-h} in (1).[1] Following, inter alia, Stock and Watson (2011), we include p lags of the log first difference of the the price index, regardless of the choice of h. So, for instance, when forecasting annual inflation ($h = 4$) with $p = 2$, the dependent variable is $(100/4) \ln\left(\frac{P_t}{P_{t-4}}\right)$, the lags are $100 \ln\left(\frac{P_{t-4}}{P_{t-5}}\right)$ and $100 \ln\left(\frac{P_{t-5}}{P_{t-6}}\right)$ and the other predictors are all dated $t-4$.

In this paper we use real time quarterly data. We provide results for inflation as measured by the GDP deflator and Personal Consumption Expenditure (PCE) deflator. As predictors, authors such as Stock and Watson (1999) consider measures of real activity including the unemployment rate. Various other predictors (e.g. cost variables, the growth of the money supply, the slope of term structure, etc.) are suggested by economic theory. Finally, authors such as Ang, Bekaert and Wei (2007) have found surveys of inflation expectations to be useful predictors. These considerations suggest the following list of potential predictors which we use in this paper. Precise definitions and sources are given in the Data Appendix.

- UNEMP: unemployment rate.
- CONS: the percentage change in real personal consumption expenditures.
- INV: the percentage change in private residential fixed investment.
- GDP: the percentage change in real GDP.
- HSTARTS: the log of housing starts (total new privately owned housing units).
- EMPLOY: the percentage change in employment (All Employees: Total Private Industries, seasonally adjusted).

[1] Our justification for using direct rather than iterated forecasts for $h > 1$ is largely practical. Iterated forecasting is difficult when the model includes exogenous regressors and would require predictive simulation which, in the context of a model space of the magnitude we consider here, is computationally infeasible. With direct forecasts the errors could have an MA(h-1) structure. In practice, we assume uncorrelated errors (and include a time-varying intercept in our state equation which allows for an MA(1) component), but include lags of the dependent variable so as to make this assumption empirically reasonable.

- PMI: the change in the Institute of Supply Management (Manufacturing): Purchasing Manager's Composite Index.
- TBILL: three month Treasury bill (secondary market) rate.
- SPREAD: the spread between the 10 year and 3 month Treasury bill rates.
- DJIA: the percentage change in the Dow Jones Industrial Average.
- MONEY: the percentage change in the money supply (M1).
- INFEXP: University of Michigan measure of inflation expectations.
- COMPRICE: the change in the commodities price index (NAPM commodities price index).
- VENDOR: the change in the NAPM vendor deliveries index.

This set of variables is a wide one reflecting the major theoretical explanations of inflation as well as variables which have found to be useful in forecasting inflation in other studies.

2.2 Time Varying Parameter Models

Research in empirical macroeconomics often uses time varying parameter (TVP) models which are estimated using state space methods such as the Kalman filter. A standard specification can be written, for $t = 1, .., T$, as

$$y_t = z_t\theta_t + \varepsilon_t \tag{2a}$$
$$\theta_t = \theta_{t-1} + \eta_t. \tag{2b}$$

In our case, y_t is inflation, $z_t = [1, x_{t-h}]$ is a $1 \times m$ vector of predictors for inflation (defined as after equation 1), θ_t is an $m \times 1$ vector of coefficients (states), $\varepsilon_t \overset{ind}{\sim} N(0, H_t)$ and $\eta_t \overset{ind}{\sim} N(0, Q_t)$. The errors, ε_t and η_t, are assumed to be mutually independent at all leads and lags. Examples of recent papers which use such models (or extensions thereof) in macroeconomics include Cogley and Sargent (2005), Cogley, Morozov and Sargent (2005), Groen, Paap and Ravazzolo (2010), Koop, Leon-Gonzalez and Strachan (2009), Korobilis (2009) and Primiceri (2005).

The model given by (2a) and (2b) is an attractive one that allows for empirical insights which are not available with traditional, constant coefficient models (even when the latter

are estimated recursively). However, when forecasting, they have the potential drawback that the same set of explanatory variables is assumed to be relevant at all points in time. Furthermore, if the number of explanatory variables in z_t is large, such models can often over-fit in-sample and, thus, forecast poorly.

Popular extensions of (2a) and (2b) such as TVP-VARs also include the same set of explanatory variables at all times and suffer from the same problems. Innovative extensions such as that of Groen, Paap and Ravazollo (2010) involve a treatment of predictor uncertainty, but not as general a treatment as is allowed for our approach. In an inflation forecasting exercise, they use a model which modifies the measurement equation to be:

$$y_t = \sum_{j=1}^{m} s_j \theta_{jt} z_{jt} + \varepsilon_t,$$

where θ_{jt} and z_{jt} denote the j^{th} elements of θ_t and z_t. The key addition to their model is $s_j \in \{0, 1\}$. Details of the exact model used for s_j are provided in Groen, Paap and Ravazollo (2010). For present purposes, the important thing to note is that it allows for each predictor for inflation to either be included (if $s_j = 1$) or excluded (if $s_j = 0$), but that s_j does not vary over time. That is, this model either includes a predictor at all points in time or excludes it at all points in time. It does not allow for the set of predictors to vary over time. It is the treatment of this latter issue which is the key addition provided by DMA. In our empirical work, we find strong evidence for model change and show how substantially improves forecast performance.

Groen, Paap and Ravazollo (2010) use posterior model probabilities to average across models. If their approach is used recursively in a forecasting exercise, then their posterior model probabilities will change over time allowing for some time-variation in the forecasting procedure (i.e. when forecasting at time τ, the posterior model probabilities will be calculated using information through time τ and τ will change as the forecaster moves through time). However, their approach weights equally all data from $t = 1, .., \tau$ and, as τ gets larger, posterior model probabilities will typically change only slightly as new data points are added. In contrast, as we shall see, DMA weights more recent forecast performance more heavily than forecast performance in the more distant past and allows for more abrupt changes in weights attached to different forecasting models.

Another recent related contribution is Hoogerheide, Kleijn, Ravazzolo, van Dijk and Verbeek (2009). This paper adopts a model of the form

$$y_t = \sum_{j=1}^{K} s_{jt} y_{jt} + \varepsilon_t,$$

where their y_{jt} are forecasts arising from different models and s_{jt} are time-varying weights associated with each forecast. The adoption of time-varying weights makes this model attractive and it shares this feature with our approach. However, it differs from our approach in several ways, the most important of which is that it allows the time-varying weights to evolve according to a random walk. This approach may not work well when there is a rapid switch between one forecasting model to another. Furthermore, when m is huge (as in our forecasting exercise) estimating all the parameters associated with the random walk evolution of the weights using MCMC methods will be computationally daunting or infeasible. As we shall see, DMA manages to surmount these computational difficulties.

2.3 Dynamic Model Averaging

To define what we do in this paper, suppose that we have a set of K models which are characterized by having different subsets of z_t as predictors. Denoting these by $z^{(k)}$ for $k = 1, .., K$, our set of models can be written as:

$$\begin{aligned} y_t &= z_t^{(k)} \theta_t^{(k)} + \varepsilon_t^{(k)} \\ \theta_{t+1}^{(k)} &= \theta_t^{(k)} + \eta_t^{(k)}, \end{aligned} \tag{3}$$

$\varepsilon_t^{(k)}$ is $N\left(0, H_t^{(k)}\right)$ and $\eta_t^{(k)}$ is $N\left(0, Q_t^{(k)}\right)$. Let $L_t \in \{1, 2, .., K\}$ denote which model applies at each time period, $\Theta_t = \left(\theta_t^{(1)\prime}, .., \theta_t^{(K)\prime}\right)'$ and $y^t = (y_1, .., y_t)'$. The fact that we are letting different models hold at each point in time and will do model averaging justifies the terminology "dynamic model averaging". To be precise, when forecasting time t variables using information through time $t-1$, DMA involves calculating $\Pr\left(L_t = k | y^{t-1}\right)$ for $k = 1, .., K$ and averaging forecasts across models using these probabilities. DMS involves selecting the single model with the highest value for $\Pr\left(L_t = k | y^{t-1}\right)$ and using this to forecast.[2] Details on the calculation of $\Pr\left(L_t = k | y^{t-1}\right)$ will be provided below.

Specifications such as (3) are potentially of great interest in empirical macroeconomics since they allow for the set of predictors for inflation to change over time as well as allowing the marginal effects of the predictors to change over time. The problems with such a framework are that many of the models can have a large number of parameters (and, hence, risk being over-parameterized) and the computational burden which arises when K

[2]Both DMA and DMS have strong statistical foundations. However, policymakers may dislike the fact that DMS can result in the forecasting model constantly switching and find the gradual re-weighting which will occur with DMA more appealing.

is large implies that estimation can take a long time (a potentially serious drawback when forecasting in real time).

To understand the source and nature of these problems, consider how the researcher might complete the model given in (3). Some specification for how predictors enter/leave the model in real time is required. A simple way of doing this would be through a transition matrix, P, with elements $p_{ij} = \Pr(L_t = i|L_{t-1} = j)$ for $i, j = 1, .., K$. Beginning with Hamilton (1989) such Markov switching processes have been commonly-used in macroeconomics. Bayesian inference in such a model is theoretically straightforward, but will be computationally infeasible since P will typically be an enormous matrix. Consider the case where we have m potential predictors and our models are defined according to whether each is included or excluded. Then we have $K = 2^m$ and P is a $K \times K$ matrix. Unless m is very small, P will have so many parameters that inference will be very imprecise and computation very slow.[3] Thus, a full Bayesian approach to DMA can be quite difficult. In this paper, we use approximations suggested by Raftery, Karny and Ettler (2010) in an industrial application. These approximations have the huge advantage that standard state space methods (e.g. involving the Kalman filter) can be used, allowing for fast real time forecasting.

The framework given in (3) is related to the switching linear Gaussian state space model described in Fruhwirth-Schnatter (2006, pages 393-394 and 406-410) who provides several citations, mostly from the engineering literature, of papers which have used such models. However, due to the need to run recursive algorithms such as the Kalman filter K^T times, exact treatments in the literature are rare unless T and K are small (see, e.g., Schervish and Tsay, 1988, for an empirical application of this sort). In most cases approximations have been used. In econometrics, the approximate method proposed by Kim (1994), which requires the Kalman filter to be run K^2 times, has perhaps been the most influential. Kim (1994) investigates both the computational speed and accuracy of his approximation relative to exact methods. Great advantages in speed and only a small loss of accuracy relative exact likelihood methods are found. However, this method proceeds conditionally on P which, as we have seen, is of too high-dimension in the present case. Accordingly, we turn to the approximations of Raftery et al (2010) which, although they do not explicitly specify a transition matrix such as P, do have sensible properties, are computationally simple and seem to work well in practice. In the context of switching linear Gaussian state space models, it is worth noting that the structure of (3) implies that the entire state vector, Θ_t, breaks into blocks (with one block for each model) which are

[3] See, for instance, Chen and Liu (2000) who discuss related models and how computation time up to t typically involves mixing over K^t terms.

independent of one another (i.e. the predictive density depends on $\theta_t^{(k)}$ only conditionally on $L_t = k$). This property is an important one in the derivations of Raftery et al (2010) which result in an accurate approximation which only involves running the Kalman filter K times.

The approximations used by Raftery et al (2010) involve two parameters, λ and α, which they refer to as *forgetting factors* and fix to numbers slightly below one. To explain the role of these forgetting factors, first consider the standard state space model in (2a) and (2b) and ignore the role of model uncertainty. For given values of H_t and Q_t, standard filtering results can be used to carry out recursive estimation or forecasting. That is, Kalman filtering begins with the result that

$$\theta_{t-1}|y^{t-1} \sim N\left(\widehat{\theta}_{t-1}, \Sigma_{t-1|t-1}\right) \tag{4}$$

where formulae for $\widehat{\theta}_{t-1}$ and $\Sigma_{t-1|t-1}$ are standard (and are provided below for the case considered in this paper). Note here only that these formulae depend on H_t and Q_t. Then Kalman filtering proceeds using:

$$\theta_t|y^{t-1} \sim N\left(\widehat{\theta}_{t-1}, \Sigma_{t|t-1}\right), \tag{5}$$

where

$$\Sigma_{t|t-1} = \Sigma_{t-1|t-1} + Q_t.$$

Raftery et al (2010) note that things simplify substantially if this latter equation is replaced by:

$$\Sigma_{t|t-1} = \frac{1}{\lambda}\Sigma_{t-1|t-1} \tag{6}$$

or, equivalently, $Q_t = \left(1 - \lambda^{-1}\right)\Sigma_{t-1|t-1}$ where $0 < \lambda \leq 1$. Such approaches have long been used in the state space literature going back to Fagin (1964) and Jazwinsky (1970). In econometrics the forgetting factor approach allowed the implementation of time-varying parameter VARs using the limited computing power available in the 1980s; see Doan, Litterman, and Sims (1984). The name "forgetting factor" is suggested by the fact that this specification implies that observations j periods in the past have weight λ^j. An alternative way of interpreting λ is to note that it implies an effective window size of $\frac{1}{1-\lambda}$. It is common to choose a value of λ near one, suggesting a gradual evolution of coefficients. Raftery et al (2010) set $\lambda = 0.99$. For quarterly macroeconomic data, this implies observations five years ago receive approximately 80% as much weight as last period's observation. This is

the sort of value consistent with fairly stable models where coefficient change is gradual. With $\lambda = 0.95$, observations five years ago receive only about 35% as much weight as last period's observations. This suggests substantial parameter instability where coefficient change is quite rapid. This seems to exhaust the range of reasonable values for λ and, accordingly, in our empirical work we consider $\lambda \in (0.95, 0.99)$. $\lambda = 0.99$ will be our benchmark choice and most of our empirical results will be reported for this (although we also include an analysis of the sensitivity to this choice).

An important point to note is that, with this simplification, we no longer have to estimate or simulate Q_t. Instead, all that is required (in addition to the Kalman filter) is a method for estimating or simulating H_t (something which we will discuss below).

Estimation in the one model case is then completed by the updating equation:

$$\theta_t | y^t \sim N\left(\widehat{\theta}_t, \Sigma_{t|t}\right), \tag{7}$$

where

$$\widehat{\theta}_t = \widehat{\theta}_{t-1} + \Sigma_{t|t-1} z_t \left(H_t + z_t \Sigma_{t|t-1} z_t'\right)^{-1} \left(y_t - z_t \widehat{\theta}_{t-1}\right) \tag{8}$$

and

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} z_t \left(H_t + z_t \Sigma_{t|t-1} z_t'\right)^{-1} z_t \Sigma_{t|t-1}. \tag{9}$$

Recursive forecasting is done using the predictive distribution

$$y_t | y^{t-1} \sim N\left(z_t \widehat{\theta}_{t-1}, H_t + z_t \Sigma_{t|t-1} z_t'\right). \tag{10}$$

We stress that, conditional on H_t, these results are all analytical and, thus, no Markov chain Monte Carlo (MCMC) algorithm is required. This greatly reduces the computational burden.

The case with many models, (3), uses the previous approximation and an additional one. To explain this, we now switch to the notation for the multiple model case in (3) and let Θ_t denote the vector of all the coefficients. In the standard single model case, Kalman filtering is based on (4), (5) and (7). In the multi-model case, for model k, these three equations become:

$$\Theta_{t-1} | L_{t-1} = k, y^{t-1} \sim N\left(\widehat{\theta}_{t-1}^{(k)}, \Sigma_{t-1|t-1}^{(k)}\right), \tag{11}$$

$$\Theta_t | L_t = k, y^{t-1} \sim N\left(\widehat{\theta}_{t-1}^{(k)}, \Sigma_{t|t-1}^{(k)}\right) \tag{12}$$

and

$$\Theta_t|L_t = k, y^t \sim N\left(\widehat{\theta}_t^{(k)}, \Sigma_{t|t}^{(k)}\right), \tag{13}$$

where $\widehat{\theta}_t^{(k)}, \Sigma_{t|t}^{(k)}$ and $\Sigma_{t|t-1}^{(k)}$ are obtained via Kalman filtering in the usual way using (8), (9) and (6), except with (k) superscripts added to denote model k. To make clear the notation in these equations, note that, conditional on $L_t = k$, the prediction and updating equations will only provide information on $\theta_t^{(k)}$ and not the full vector Θ_t. Hence, we have only written (11), (12) and (13) in terms of the distributions which hold for $\theta_t^{(k)}$.

The previous results were all conditional on $L_t = k$, and we need a method for unconditional prediction (i.e. not conditional on a particular model). In theory, a nice way of doing this would be through specifying a transition matrix, P, such as that given above and using MCMC methods to obtain such unconditional results. However, for the reasons discussed previously, this will typically be computationally infeasible and empirically undesirable due to the resulting proliferation of parameters. In this paper, we follow the suggestion of Raftery et al (2010) involving a forgetting factor for the state equation for the models, α, comparable to the forgetting factor λ used with the state equation for the parameters. The derivation of Kalman filtering ideas begins with (4). The analogous result, when doing DMA, is

$$p\left(\Theta_{t-1}|y^{t-1}\right) = \sum_{k=1}^{K} p\left(\theta_{t-1}^{(k)}|L_{t-1} = k, y^{t-1}\right) \Pr\left(L_{t-1} = k|y^{t-1}\right), \tag{14}$$

where $p\left(\theta_{t-1}^{(k)}|L_{t-1} = k, y^{t-1}\right)$ is given by (11). To simplify notation, let $\pi_{t|s,l} = \Pr\left(L_t = l|y^s\right)$ and thus, the final term on the right hand side of (14) is $\pi_{t-1|t-1,k}$.

If we were to use the unrestricted matrix of transition probabilities in P with elements p_{kl} then the model prediction equation would be:

$$\pi_{t|t-1,k} = \sum_{l=1}^{K} \pi_{t-1|t-1,l} p_{kl},$$

but Raftery et al (2010) replace this by:

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{l=1}^{K} \pi_{t-1|t-1,l}^{\alpha}}, \tag{15}$$

where $0 < \alpha \leq 1$ is set to a fixed value slightly less than one and is interpreted in a similar manner to λ. Raftery et al (2010) argue that this is an empirically sensible simplification and, in particular, is a type of multiparameter power steady model used elsewhere in the

literature. See also Smith and Miller (1986) who work with a similar model and argue approximations such as (15) are sensible and not too restrictive.

The huge advantage of using the forgetting factor α in the model prediction equation is that we do not require an MCMC algorithm to draw transitions between models nor a simulation algorithm over model space.[4] Instead, simple evaluations comparable to those of the updating equation in the Kalman filter can be done. In particular, we have a model updating equation of:

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} p_k\left(y_t|y^{t-1}\right)}{\sum_{l=1}^{K} \pi_{t|t-1,l} p_l\left(y_t|y^{t-1}\right)}, \tag{16}$$

where $p_l\left(y_t|y^{t-1}\right)$ is the predictive density for model l (i.e. the Normal density in (10) with (l) superscripts added) evaluated at y_t.

Recursive forecasting can be done by averaging over predictive results for every model using $\pi_{t|t-1,k}$. Therefore, DMA point predictions are given by:

$$E\left(y_t|y^{t-1}\right) = \sum_{k=1}^{K} \pi_{t|t-1,k} z_t^{(k)} \widehat{\theta}_{t-1}^{(k)}.$$

DMS proceeds by selecting the single model with the highest value for $\pi_{t|t-1,k}$ at each point in time and simply using it for forecasting.

To understand further how the forgetting factor α can be interpreted, note that this specification implies that the weight used in DMA which is attached to model k at time t is:

$$\begin{aligned} \pi_{t|t-1,k} &\propto \left[\pi_{t-1|t-2,k} p_k\left(y_{t-1}|y^{t-2}\right)\right]^{\alpha} \\ &= \prod_{i=1}^{t-1} \left[p_k\left(y_{t-i}|y^{t-i-1}\right)\right]^{\alpha^i}. \end{aligned}$$

Thus, model k will receive more weight at time t if it has forecast well in the recent past (where forecast performance is measured by the predictive density, $p_k\left(y_{t-i}|y^{t-i-1}\right)$). The interpretation of "recent past" is controlled by the forgetting factor, α and we have the same exponential decay at the rate α^i for observations i periods ago as we had associated with λ. Thus, if $\alpha = 0.99$ (our benchmark value and also the value used by Raftery et al, 2010), forecast performance five years ago receives 80% as much weight as forecast per-

[4] Examples of simulation algorithms over model space include the Markov chain Monte Carlo model composition (MC^3) algorithm of Madigan and York (1995) or the reversible jump MCMC algorithm of Green (1995).

formance last period (when using quarterly data). If $\alpha = 0.95$, then forecast performance five years ago receives only about 35% as much weight. These considerations suggest that, as with λ, we focus on the interval $\alpha \in (0.95, 0.99)$.

Note also that, if $\alpha = 1$, then $\pi_{t|t-1,k}$ is simply proportional to the marginal likelihood using data through time $t-1$. This is what standard approaches to BMA would use. If we further set $\lambda = 1$, then we obtain BMA using conventional linear forecasting models with no time variation in coefficients. In our empirical work, we include BMA in our set of alternative forecasting procedures and implement this by setting $\alpha = \lambda = 1$.

We stress that, conditional on H_t, the estimation and forecasting strategy outlined above only involves evaluating formulae such as those in the Kalman filter. All the recursions above are started by choosing a prior for $\pi_{0|0,k}$ and $\theta_0^{(k)}$ for $k = 1, .., K$.

The preceding discussion is all conditional on H_t. Raftery et al (2010) recommend a simple plug in method where $H_t^{(k)} = H^{(k)}$ and is replaced with a consistent estimate. When forecasting inflation, however, it is likely that the error variance is changing over time. In theory, we could use a stochastic volatility or ARCH specification for $H_t^{(k)}$. However, to do this would greatly add to the computational burden. Thus, we use an Exponentially Weighted Moving Average (EWMA) estimate of $H_t^{(k)}$:

$$\widehat{H}_t^{(k)} = \sqrt{(1-\kappa)\sum_{j=1}^{t}\kappa^{j-1}\left(y_j - z_j^{(k)}\widehat{\theta}_j^{(k)}\right)^2}. \tag{17}$$

EWMA estimators are commonly used to model time-varying volatilities in finance; see Riskmetrics (1996) for the properties of EWMA estimators. κ is called a decay factor, and Riskmetrics proposes setting 0.97 for monthly data and 0.94 for daily data. We have quarterly data, so we expect to have a slower decay of volatility so we set $\kappa = 0.98$. An attractive feature of the EWMA specification is that it can be approximated by a recursive form, which can be used to obtain volatility forecasts. The period $t+1$ forecast given data up to time t takes the form.

$$\widehat{H}_{t+1|t}^{(k)} = \kappa\widehat{H}_{t|t-1}^{(k)} + (1-\kappa)\left(y_t - z_t^{(k)}\widehat{\theta}_t^{(k)}\right)^2.$$

In our preliminary empirical experimentation, we compared the resulting estimated volatility with that estimated using the popular UC-SV model of Stock and Watson (2007). We found the patterns of the UC-SV and EWMA volatilities to be very similar (although the UC-SV approach requires the use of MCMC methods).

3 Empirical Work

In order to evaluate DMA and DMS we use two measures of inflation: one based on the GDP deflator and the other on the PCE deflator. Our dataset is completed with the 14 potential predictors described in Section 2 and the Data Appendix. Our forecasting is done in real time in the sense that, for all our variables, we use the value which would have been available to the forecaster at the time the forecast was being made. The full sample runs from 1960:Q1 to 2008:Q4.

To be clear on our timing convention, note that for any variable (say, X_t) we will have the observed value of it at different vintages. Let X_t^v be the value of X_t as known at time v. We will refer to X_t^T as final vintage data and X_t^{t+1} as the initial release.[5] For variables which are subject to revision, forecasts made of inflation at time $\tau + h$ given information through time τ will use data of vintage τ (i.e. the dependent variable will have dating convention X_t^τ for $t = 1, .., \tau$ whereas the explanatory variables will be dated X_{t-h}^τ). Since, for some of the variables, real time versions are only available beginning in 1969Q4, we start the forecast evaluation period after this.[6] Since the initial release of time t data is made in time $t+1$, we start our h-step ahead forecasts in 1969Q4+h+1. When evaluating forecast performance in a real time forecasting exercise, the researcher must choose which realization of the data to use. We use final vintage data for this purpose as being most likely to reflect the true value of inflation. We consider forecast horizons of $h = 1, 4$ and 8.

Our empirical work is divided into two sub-sections. The first of these sub-sections present results using DMA and DMS, implemented in our preferred way. This involves setting $\alpha = 0.99$, $\lambda = 0.99$, a noninformative prior over the models (i.e. $\pi_{0|0,k} = \frac{1}{K}$ for $k = 1, .., K$ so that, initially, all models are equally likely) and a very diffuse prior on the initial conditions of the states: $\theta_0^{(k)} \sim N(0, 100 I_{n_k})$, where n_k is the number of variables in model k, for $k = 1, .., K$. The first sub-section presents evidence on which variables are good for predicting inflation over time. The second sub-section investigates forecast performance by comparing DMA forecasts to those produced by several alternative forecasting strategies. We also present evidence on the sensitivity of our results to the choice of the forgetting factors. All of our models include an intercept and two lags of the dependent variable.[7]

[5] Several of our variables are not subject to revisions and, hence, we just use final vintage data for them (see the appendix for a list of which variables are subject to revisions and which are not).

[6] Prior to 1969Q4, if real time data is not available we use the final vintage data.

[7] Preliminary experimentation with lag lengths up to four indicated two lags leads to the best forecast performance for both our measures of inflation.

3.1 Which Variables are Good Predictors for Inflation?

In theory, DMA has a large potential benefit over other forecasting approaches in that it allows the forecasting model to change over time. It has a second benefit in that many of the models under consideration are parsimonious and, if DMA attached a great deal of weight to such models, it can avoid over-fitting problems. Of course, in a particular empirical application, these benefits may or may not be achieved. Accordingly, we begin by presenting evidence that, when forecasting inflation, DMA is favoring parsimonious models and that the forecasting model is changing over time.

One striking feature of all of our empirical results is that, although we have 14 potential predictors (and, thus, tens of thousands of models), most probability is attached to very parsimonious models with only a few predictors. If we let $Size_{k,t}$ be the number of predictors in model k at time t (not including the intercept and AR lags which are common to all models) then

$$E\left(Size_t\right) = \sum_{k=1}^{K} \pi_{t|t-1,k} Size_{k,t}$$

can be interpreted as the expected or average number of predictors used in DMA at time t. Figure 1 plots this for our six empirical exercises (i.e. two definitions of inflation and three forecast horizons).

For the short forecast horizon ($h = 1$), the shrinkage of DMA is particularly striking. Virtually everywhere it includes (in an expected value sense) between zero and three of the 14 predictors listed in Section 2 for both our definitions of inflation. At the longer horizons of $h = 4$ and $h = 8$, slightly more predictors are included, but almost never are more than 4 predictors included, showing that DMA is strongly favoring parsimonious models.

Figure 1 shows clear evidence that DMA will shrink forecasts and provides some evidence that the way this shrinkage is done changes over time. But it does not tell us which predictors are important and how the predictors are changing over time. It is to these issues we now turn.

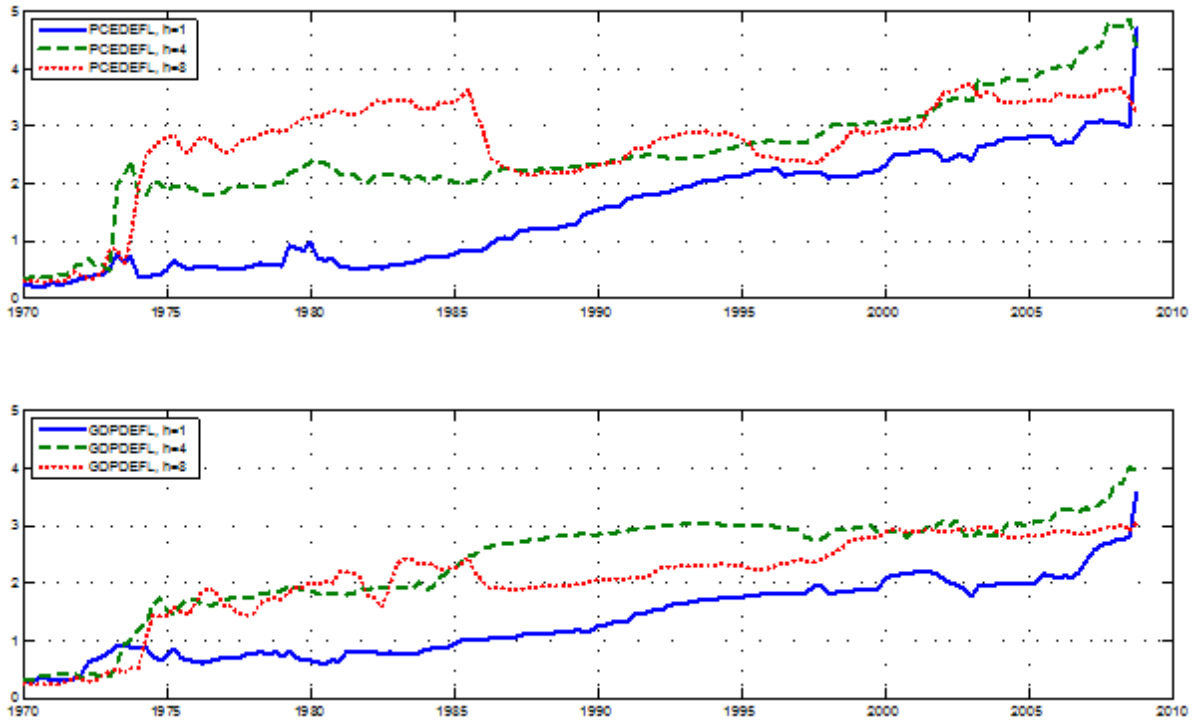

Figure 1: Expected Number of Predictors in Each Forecasting Exercise.

Figures 2 through 4 shed light on which predictors are important at each point in time for each of our six empirical exercises. These graphs contain posterior inclusion probabilities. That is, they are the probability that a predictor is useful for forecasting at time t. Equivalently, they are the weight used by DMA attached to models which include a particular predictor. To keep the figures readable, we only present posterior inclusion probabilities for predictors which are important at least one point in time. To be precise, any predictor where the inclusion probability is never above 0.5 is excluded from the appropriate figure.

These figures confirm that DMS is almost always choosing parsimonious models and the weights in DMA heavily reflect parsimonious models. That is, it is rare for DMS to choose a model with more than two or three predictors.

Another important result is that for both measures of inflation and for all forecast horizons, we are finding strong evidence of model change. That is, the set of predictors in the forecasting model is changing over time. Furthermore, DMA can be seen to allow for both gradual or abrupt changes in the role of a predictor. That is, there are many cases where the posterior inclusion probability associated with a predictor increases or decreases gradually over time (see, e.g., the gradual change in the inclusion probability of UNEMP

in panel (a) of Figure 3). But there are also several abrupt changes where a posterior inclusion probability changes abruptly from near zero to near one (or vice versa) within a quarter or two (see, e.g., INFEXP in panel (b) of Figure 3 or TBILL as a predictor for GDP deflator inflation for $h = 1$ and $h = 4$, in Figures 2 and 3). A TVP regression model of the form given in (2a) and (2b) using all the predictors would not allow the role of individual predictors to switch so rapidly.

The interested reader can examine Figures 2 through 4 for any particular variable of interest. Here we note only a few main findings. Note first that half of our potential explanatory variables come through as being important at some time, for some forecast horizon for some measure of inflation. These variables are: INFEXP, TBILL, HSTARTS, UNEMP, MONEY, SPREAD and EMPLOY with the first three of these variables being usually of particular importance. But it is clearly the case that there is a large variation over time, over forecast horizons and over measures of inflation in relation to what is a good predictor for inflation.

Results for both inflation measures for $h = 1$ are particularly striking. In both cases, DMA puts most weight on the same predictors and these show similar patterns. For instance, the INFEXP variable is a poor predictor early in the sample, but becomes the dominant predictor after the mid-1980s. A story consistent with this pattern is that surveys of inflation expectations became more reliable after the Great Moderation of the business cycle when inflation became less volatile. Another finding is that TBILL is a good predictor in the 1970s, but DMA abruptly drops the predictor around 1980. These patterns are also found when $h = 4$, but do not come through so clearly when $h = 8$.

Unemployment, a variable conventionally used in the Phillips' curve, is not found to be an important predictor when $h = 1$, but is of some importance at longer horizons (particularly early in the sample). The housing starts variable is found to be a much more important predictor at medium and long horizons than when $h = 1$. We hesitate to provide too many economic stories from a reduced-form forecasting exercise such as ours. But Figures 2 through 4 illustrate an important benefit of DMA and DMS: that they will pick up good predictors automatically as the forecasting model evolves over time.

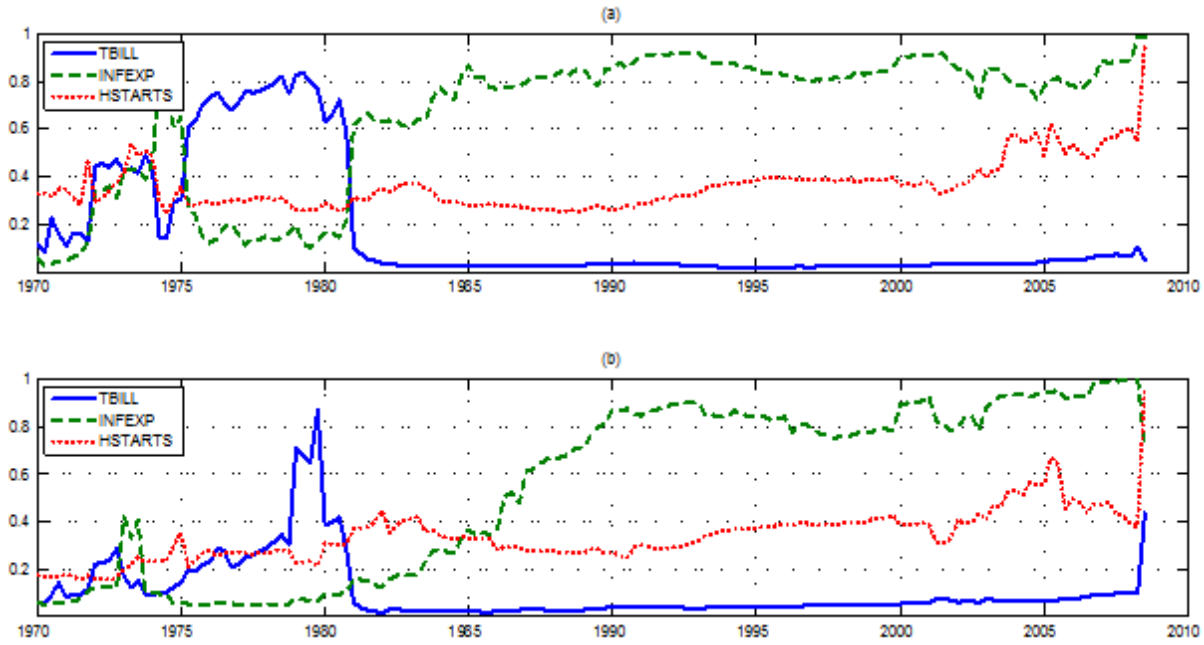

Figure 2: Posterior Probability of Inclusion of Predictors, $h = 1$. Panel (a) refers to GDP deflator inflation, and panel (b) to PCE deflator inflation.

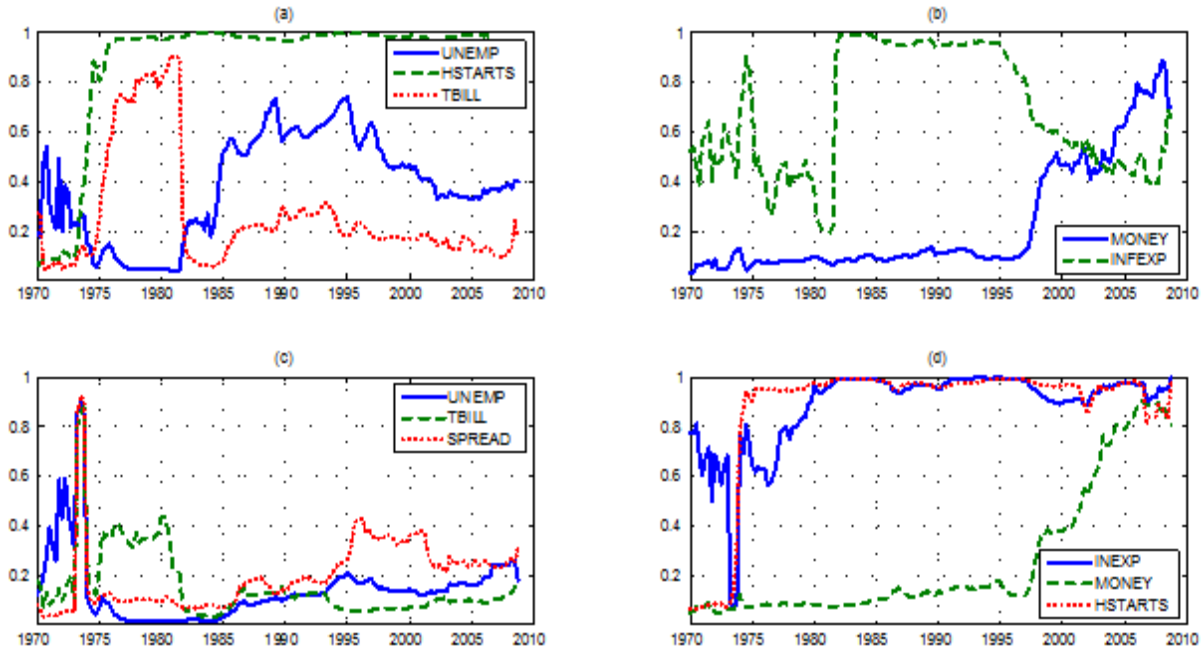

Figure 3: Posterior Probability of Inclusion of Predictors, $h = 4$. Panels (a) - (b) refer to GDP deflator inflation, and panels (c) - (d) to PCE deflator inflation.

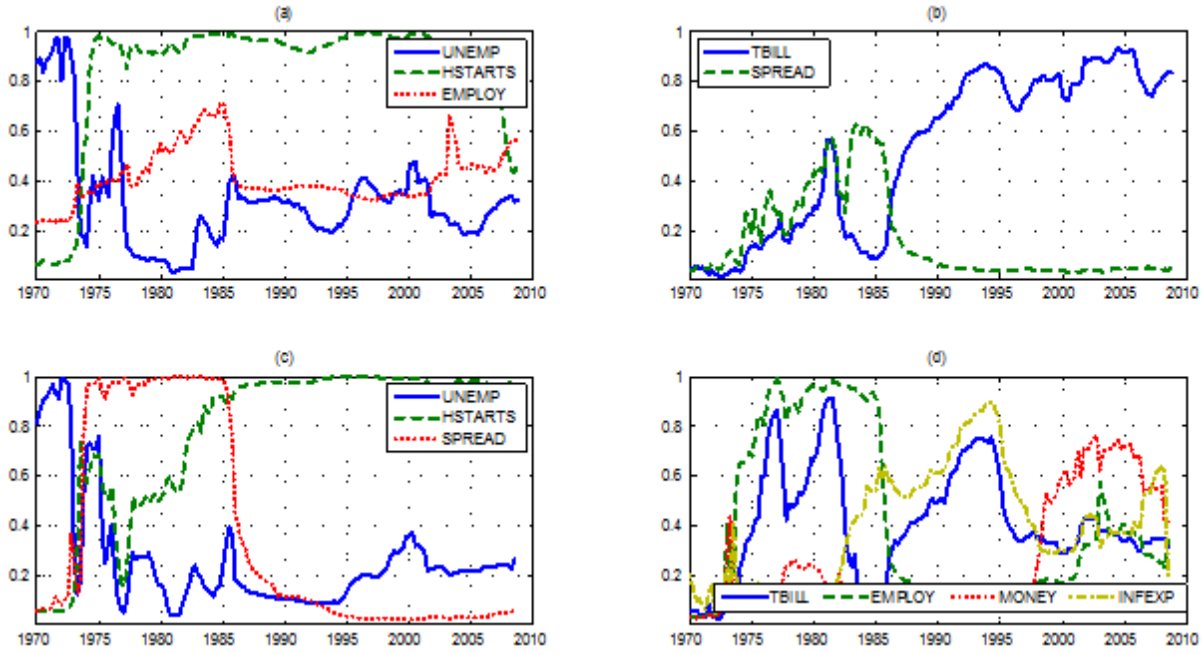

Figure 4: Posterior Probability of Inclusion of Predictors, $h = 8$. Panels (a) - (b) refer to GDP deflator inflation, and panels (c) - (d) to PCE deflator inflation.

3.2 Forecast Performance

3.2.1 DMA versus Other Forecast Procedures

There are many metrics for evaluating forecast performance and many alternative forecasting methodologies that we could compare our DMA and DMS forecasts to. In this paper, we present two forecast comparison metrics involving point forecasts. These are mean squared forecast error (MSFE) and mean absolute forecast error (MAFE). We also present a forecast metric which involves the entire predictive distribution: the sum of log predictive likelihoods. Predictive likelihoods are motivated and described in many places such as Geweke and Amisano (2010). The predictive likelihood is the predictive density for y_t (given data through time $t-1$) evaluated at the actual outcome (we remind the reader that we are using the final vintage data for the actual outcome). The formula for the one-step ahead predictive density in model l was denoted by $p_l\left(y_t|y^{t-1}\right)$ above and can be calculated as described in Section 2.3. We use the direct method of forecasting and, hence, the log predictive density for the h-step ahead forecast is the h-period extension of this. We use the sum of log predictive likelihoods for forecast evaluation, where the sum begins in 1969Q4+h+1. MSFEs and MAFEs are also calculated beginning in 1969Q4+h+1.

In terms of forecasting methods, we present results for:

- Forecasts using DMA with $\alpha = \lambda = 0.99$.
- Forecasts using DMS with $\alpha = \lambda = 0.99$.
- Forecasts using DMA with $\alpha = \lambda = 0.95$.
- Forecasts using DMS with $\alpha = \lambda = 0.95$.
- Forecasts using DMA, but where the coefficients do not vary over time in each model (i.e. this is a special case of DMA where $\lambda = 1$). We set $\alpha = 0.99$.
- Forecasts using BMA as a special case of DMA (i.e. we set $\lambda = \alpha = 1$).
- Forecasts using a single model containing an intercept, two lags of inflation, and all the predictors, but with time varying parameters (i.e. this is a special case of DMA or DMS where 100% of the prior weight is attached to the model with all the predictors, but all other modelling choices are identical including $\lambda = 0.99$). This is labelled TVP-AR(2)-X in the tables.
- Forecasts as for TVP-AR(2)-X model, but only including the intercept and two lags of inflation. This is labelled TVP-AR(2) in the tables.
- Forecasts done recursively using traditional g-prior BMA estimated with MCMC with $g = \frac{1}{T}$.[8]
- Forecasts using the unobserved components with stochastic volatility (UC-SV) model of Stock and Watson (2007).[9]
- Recursive OLS forecasts using an AR(p) model with an intercept where p is selected at each point in time using BIC. Maximum lag length is eight.
- Recursive OLS forecasts using an AR(2) model with an intercept and all of the predictors.
- Rolling OLS forecasts using an AR(p) model with an intercept. We use a window of 40 quarters.

[8] We implement this exactly as in Fernandez, Ley and Steel (2001). This approach to BMA is the most popular one in cross-sectional regressions, but has been used in time series applications in, e.g., Koop and Potter (2004).

[9] The exact specification is on page 16 of Stock and Watson (2007). We estimate the model exactly as they do. Following Stock and Watson (2007) we set the coefficient they call γ to 0.2.

- Rolling OLS forecasts using an AR(2) model with an intercept and all of the predictors. We use a window of 40 quarters.

- Random walk forecasts.

The final five methods are not Bayesian, so no predictive likelihoods are presented for these cases.

Tables 1 and 2 present results for our forecasting exercise for our two different measures of inflation.

The overall story is a clear and strong one: DMA and DMS forecast well. In most cases much better than other forecasting methods and in no case much worse than the best alternative method. We elaborate on these points below.

Consider first the log predictive likelihoods (the preferred method of Bayesian forecast comparison). These always indicate that DMA or DMS forecasts best, often much better than the other forecasting strategies used in our comparison. Note, in particular, the excellent forecast performance of DMS with $\alpha = \lambda = 0.95$ for both measures of inflation at all horizons. This value for the forgetting factors allows for rapid change in both coefficients and in models. Versions of DMA or DMS which impose more gradual model change do slightly worse. Note, however, that conventional BMA forecasts poorly. The TVP-AR(2) and UC-SV models also have substantially lower predictive likelihoods than the DMA or DMS approaches. Of the non-DMA and non-DMS approaches, the UC-SV approach of Stock and Watson (2007) consistently is the best performer. Another message coming out of Tables 1 and 2 is that simply using a TVP model with all predictors tends to forecast poorly. Of course, we are presenting results for only a single empirical exercise. But TVP models such as TVP-VARs are gaining increasing popularity in macroeconomics and the poor forecast performance of TVP models found in Tables 1 and 2 should serve as a caution to users of such models (at least in forecasting exercises). Clearly, we are finding that the shrinkage provided by DMA or DMS is of great value in forecasting.

In most cases, predictive likelihoods also indicate that DMS forecasts a bit better than DMA (although this result does not carry over to MAFEs and MSFEs where in some cases DMA does better). DMS and DMA can be interpreted as doing shrinkage in different ways. DMS puts zero weight on all models other than the one best model, thus "shrinking" the contribution of all models except a single one towards zero. It could be that this additional shrinkage provides some additional forecast benefits over DMA. Furthermore, in times of rapid change, DMS will tend to switch more quickly that DMA since it can select an entirely new model as opposed to adjusting the weights on all the models.

If we turn our attention to results using MSFE and MAFE, we can see that the previous

picture still broadly holds: DMS and DMA, particularly for the case where $\alpha = \lambda = 0.95$, always forecast best (although DMA does somewhat better relative to DMS than we found using predictive likelihoods). In addition, we can say that naive forecasting methods such as using an AR(2) or random walk model are usually inferior to DMA and DMS for both measures of inflation at all forecast horizons. However, results are not as strong as with predictive likelihoods and there are some differences with the patterns noted above. The reason results are not as strong is that the MSFEs and MAFEs are only using the point forecasts, whereas predictive likelihoods are using the entire predictive distribution (i.e. when working with the predictive, the tails of the density matter and having a few realizations in the tails of the predictive density can have a big negative impact on forecast performance).

Table 1. Comparing Different Forecasting Methods: GDP deflator inflation

	h=1			h=4			h=8		
	MAFE	MSFE	log(PL)	MAFE	MSFE	log(PL)	MAFE	MSFE	log(PL)
DMA ($\alpha = \lambda = 0.99$)	0.248	0.306	-0.292	0.269	0.349	-0.421	0.333	0.413	-0.583
DMS ($\alpha = \lambda = 0.99$)	0.256	0.318	-0.277	0.277	0.361	-0.406	0.338	0.423	-0.578
DMA ($\alpha = \lambda = 0.95$)	0.248	0.310	-0.378	0.255	0.334	-0.455	0.293	0.379	-0.570
DMS ($\alpha = \lambda = 0.95$)	0.235	0.297	-0.237	0.249	0.316	-0.307	0.295	0.385	-0.424
DMA ($\lambda = 1$, $\alpha = 0.99$)	0.249	0.306	-0.300	0.277	0.355	-0.445	0.346	0.423	-0.626
BMA (DMA with $\alpha = \lambda = 1$)	0.256	0.316	-0.320	0.282	0.363	-0.463	0.364	0.449	-0.690
TVP-AR(2) ($\lambda = 0.99$)	0.260	0.327	-0.344	0.320	0.401	-0.480	0.398	0.502	-0.662
TVP-AR(2)-X ($\lambda = 0.99$)	0.309	0.424	-0.423	0.336	0.453	-0.508	0.410	0.532	-0.701
BMA-MCMC ($g = \frac{1}{T}$)	0.234	0.303	-0.369	0.285	0.364	-0.503	0.319	0.401	-0.667
UC-SV ($\gamma = 0.2$)	0.256	0.332	-0.320	0.311	0.396	-0.473	0.350	0.465	-0.613
Recursive OLS - AR(BIC)	0.251	0.326	-	0.344	0.433	-	0.436	0.516	-
Recursive OLS - All Preds	0.265	0.334	-	0.302	0.376	-	0.369	0.441	-
Rolling OLS - AR(2)	0.251	0.325	-	0.328	0.425	-	0.380	0.464	-
Rolling OLS - All Preds	0.252	0.327	-	0.273	0.349	-	0.325	0.398	-
Random Walk	0.262	0.349	-	0.333	0.435	-	0.428	0.598	-

Note: The dependent variable is defined as $(100/h)\ln\left(\frac{P_t}{P_{t-h}}\right)$ and the forecast evaluation period is 1969:Q4+h+1 through 2008:Q4

Table 2. Comparing Different Forecasting Methods: PCE deflator inflation

	h=1			h=4			h=8		
	MAFE	MSFE	log(PL)	MAFE	MSFE	log(PL)	MAFE	MSFE	log(PL)
DMA ($\alpha = \lambda = 0.99$)	0.253	0.322	-0.451	0.311	0.406	-0.622	0.357	0.448	-0.699
DMS ($\alpha = \lambda = 0.99$)	0.259	0.326	-0.430	0.330	0.431	-0.631	0.369	0.469	-0.699
DMA ($\alpha = \lambda = 0.95$)	0.267	0.334	-0.519	0.290	0.382	-0.652	0.317	0.403	-0.673
DMS ($\alpha = \lambda = 0.95$)	0.236	0.295	-0.348	0.288	0.353	-0.499	0.293	0.371	-0.518
DMA ($\lambda = 1$, $\alpha = 0.99$)	0.250	0.317	-0.444	0.315	0.412	-0.636	0.366	0.458	-0.733
BMA (DMA with $\alpha = \lambda = 1$)	0.259	0.331	-0.464	0.325	0.429	-0.668	0.397	0.490	-0.779
TVP-AR(2) ($\lambda = 0.99$)	0.280	0.361	-0.488	0.355	0.459	-0.668	0.450	0.573	-0.837
TVP-AR(2)-X ($\lambda = 0.99$)	0.347	0.492	-0.645	0.378	0.556	-0.764	0.432	0.574	-0.841
BMA-MCMC ($g = \frac{1}{T}$)	0.269	0.352	-0.489	0.307	0.414	-0.633	0.357	0.454	-0.788
UC-SV ($\gamma = 0.2$)	0.269	0.341	-0.474	0.340	0.443	-0.651	0.406	0.528	-0.774
Recursive OLS - AR(BIC)	0.310	0.439	-	0.390	0.513	-	0.463	0.574	-
Recursive OLS - All Preds	0.303	0.421	-	0.325	0.442	-	0.378	0.481	-
Rolling OLS - AR(2)	0.316	0.430	-	0.378	0.510	-	0.428	0.540	-
Rolling OLS - All Preds	0.289	0.414	-	0.313	0.422	-	0.338	0.436	-
Random Walk	0.294	0.414	-	0.407	0.551	-	0.531	0.698	-

Note: The dependent variable is defined as $(100/h)\ln\left(\frac{P_t}{P_{t-h}}\right)$ and the forecast evaluation period is 1969:Q4+h+1 through 2008:Q4

When we look at MSFEs and MAFEs, it is rolling OLS forecasts using all the predictors which forecast best among all of the OLS-based methods. DMS and DMA with $\alpha = \lambda = 0.95$ always lead to lower MSFEs and MAFEs than rolling OLS with all the predictors. However, there are some cases (particularly at longer forecast horizons) where rolling OLS with all the predictors leads to lower MSFEs and MAFEs than some other implementations of DMA or DMS. This reflects the fact that, in this data set, there are times where rapid change is necessary to give good forecast performance. DMA and DMS with $\alpha = \lambda = 0.95$ achieve this rapid change and forecast well. To a lesser extent, rolling OLS can also achieve this by having a short window width. However, we do stress that there are many cases where DMA or DMS does much better than rolling OLS and in all cases the other OLS benchmarks do poorly.

3.2.2 Comparison to Greenbook Forecasts

It is also of interest to compare the forecast performance of DMA to those provided in the Greenbooks produced by the Federal Reserve Board of Governors. These are for GDP deflator inflation and are published with a lag and we have only the forecasts made up to 2003Q4 (which provide us with $h = 1$ forecasts up to 2004Q1 and $h = 4$ up to 2004Q4). One quarter ahead forecasts are available since 1970Q1, while one year ahead forecasts were only made starting in 1974Q1. Two year ahead forecasts are available only for a considerably shorter period of time with many missing observations, hence we will not make comparisons for $h = 8$. Since the Greenbook forecasts are point forecasts, we cannot present predictive likelihoods. Accordingly, in Table 3 we present MAFEs (MSFEs give qualitatively similar results) for GDP deflator inflation for $h = 1$ and $h = 4$. For the case of this comparison we use the simple random walk as a base model, which means that the MAFEs of all models are relative to the MAFE of this base model. We present DMA results for values of the forgetting factors (α, λ): $(0.99, 0.99)$ and $(0.95, 0.95)$. Both DMA models are doing excellent for the short term forecasts compared to the Greenbook. For $\alpha = \lambda = 0.99$ DMA improves over Greenbook forecasts by 12% and for $\alpha = \lambda = 0.95$ the improvement is as high as 15%. However, for $h = 4$ the Greenbook forecasts are 10% better than those provided by DMA with $\alpha = \lambda = 0.99$. It is only the DMA with $\alpha = \lambda = 0.95$ (i.e. allow faster switching of models and parameters) which gives a very good performance for this forecast horizon, comparable with the Greenbook forecasts and much better than the random walk.

Table 3. Comparison of DMA with Greenbook forecasts: MAFE

	$h = 1$	$h = 4$
Greenbook forecasts	0.91	0.84
DMA $\alpha = \lambda = 0.99$	0.80	0.94
DMA $\alpha = \lambda = 0.95$	0.77	0.83

Notes: MAFEs are relative to random walk benchmark. Forecast evaluation period differs from Tables 1 and 2 and over h as described in text.

3.2.3 Sensitivity Analysis

Our previous DMA and DMS results were for $\alpha = \lambda = 0.99$ and $\alpha = \lambda = 0.95$. As discussed previously, researchers in this field choose pre-selected values for α and λ and the interval $(0.95, 0.99)$ is the sensible one for most empirical applications. It would be possible to choose α and λ in a data-based fashion, but this is typically not done for computational reasons. For instance, the researcher could select a grid of values for these two forgetting factors and then do DMA at every possible combination of values for α and λ. Some metric (e.g. an information criteria or the sum of log predictive likelihoods through time $t-1$) could be used to select the preferred combination of α and λ at each point in time. However, this would turn an already computationally demanding exercise to one which was g^2 times as demanding (where g is the number of values in the grid). Accordingly, researchers such as Raftery et al (2010) simply set $\alpha = \lambda = 0.99$ and argue that results will be robust to reasonable changes in these factors. In order to investigate such robustness claims, Tables 4 and 5 present results for our forecasting exercise using different combinations of the forgetting factors.

Overall, Tables 4 and 5 reveal a high degree of robustness to choice of α and λ within the range we suggested previously, i.e. $\alpha, \lambda \in (0.95, 0.99)$. If anything, these tables emphasize the benefits of DMA in that measures of forecast performance are sometimes a bit better than those in Tables 1 and 2 and rarely much worse. In Tables 1 and 2 we found that the combination $\alpha = \lambda = 0.95$ led to the best forecast performance. Particularly at short and medium forecast horizons, we can sometimes forecast a bit better by setting $\alpha = 0.95$ and $\lambda = 0.99$. Remember that the value $\alpha = 0.95$ allows for quite rapid change in forecasting model over time, but by setting $\lambda = 0.99$ less change in the coefficients is allowed for. This is consistent with a story that that allowing for models to change over time is more important in improving forecast performance than allowing for parameters to change (at least in our data set).

Allowing both forgetting factors to become as low as $\alpha = \lambda = 0.80$, and hence favoring faster change in both parameters and models, has no obvious benefits for forecasting as

is shown in Tables 4 and 5. Allowing the forgetting factors to be reduced even more ($\alpha = \lambda = 0.50$), results in complete forecast breakdown of the time-varying parameters models (results not presented in the tables). Reducing the forgetting factors to very low levels results in: i) models with very flexible parameters so that each single model overfits the data in sample, and ii) extremely fast switches among models, where there is a new "best" model almost every single quarter. Both these facts certainly do not assist in favorable forecast performance, something that explains why we (and the previous literature) are focusing on a modest range for the forgetting factors, like $\alpha, \lambda \in (0.95, 0.99)$.

Table 4. Sensitivity Analysis: GDP deflator inflation

Forecast Method	MAFE	RMSFE	log PL
		$h = 1$	
DMA, $\alpha = 0.99, \lambda = 0.95$	0.245	0.311	-0.307
DMS, $\alpha = 0.99, \lambda = 0.95$	0.251	0.321	-0.300
DMA, $\alpha = 0.95, \lambda = 0.99$	0.247	0.304	-0.324
DMS, $\alpha = 0.95, \lambda = 0.99$	0.233	0.287	-0.202
DMA, $\alpha = 0.80, \lambda = 0.80$	0.268	0.347	-0.463
DMS, $\alpha = 0.80, \lambda = 0.80$	0.255	0.325	-0.364
		$h = 4$	
DMA, $\alpha = 0.99, \lambda = 0.95$	0.256	0.336	-0.400
DMS, $\alpha = 0.99, \lambda = 0.95$	0.263	0.344	-0.390
DMA, $\alpha = 0.95, \lambda = 0.99$	0.263	0.342	-0.429
DMS, $\alpha = 0.95, \lambda = 0.99$	0.254	0.327	-0.298
DMA, $\alpha = 0.80, \lambda = 0.80$	0.277	0.362	-0.449
DMS, $\alpha = 0.80, \lambda = 0.80$	0.270	0.359	-0.406
		$h = 8$	
DMA, $\alpha = 0.99, \lambda = 0.95$	0.299	0.390	-0.497
DMS, $\alpha = 0.99, \lambda = 0.95$	0.310	0.407	-0.491
DMA, $\alpha = 0.95, \lambda = 0.99$	0.314	0.397	-0.579
DMS, $\alpha = 0.95, \lambda = 0.99$	0.297	0.381	-0.436
DMA, $\alpha = 0.80, \lambda = 0.80$	0.431	0.538	-0.788
DMS, $\alpha = 0.80, \lambda = 0.80$	0.394	0.497	-0.613

Table 5. Sensitivity Analysis: PCE deflator inflation

Forecast Method	MAFE	RMSFE	log PL
		$h = 1$	
DMA, $\alpha = 0.99, \lambda = 0.95$	0.265	0.339	-0.475
DMS, $\alpha = 0.99, \lambda = 0.95$	0.271	0.351	-0.484
DMA, $\alpha = 0.95, \lambda = 0.99$	0.253	0.318	-0.462
DMS, $\alpha = 0.95, \lambda = 0.99$	0.228	0.287	-0.338
DMA, $\alpha = 0.80, \lambda = 0.80$	0.282	0.373	-0.568
DMS, $\alpha = 0.80, \lambda = 0.80$	0.275	0.358	-0.511
		$h = 4$	
DMA, $\alpha = 0.99, \lambda = 0.95$	0.303	0.387	-0.603
DMS, $\alpha = 0.99, \lambda = 0.95$	0.322	0.410	-0.601
DMA, $\alpha = 0.95, \lambda = 0.99$	0.299	0.389	-0.615
DMS, $\alpha = 0.95, \lambda = 0.99$	0.292	0.373	-0.494
DMA, $\alpha = 0.80, \lambda = 0.80$	0.345	0.444	-0.612
DMS, $\alpha = 0.80, \lambda = 0.80$	0.336	0.432	-0.587
		$h = 8$	
DMA, $\alpha = 0.99, \lambda = 0.95$	0.329	0.421	-0.623
DMS, $\alpha = 0.99, \lambda = 0.95$	0.347	0.443	-0.613
DMA, $\alpha = 0.95, \lambda = 0.99$	0.336	0.420	-0.676
DMS, $\alpha = 0.95, \lambda = 0.99$	0.323	0.411	-0.555
DMA, $\alpha = 0.80, \lambda = 0.80$	0.401	0.513	-0.764
DMS, $\alpha = 0.80, \lambda = 0.80$	0.368	0.474	-0.7059

4 Conclusions

This paper has investigated the use of DMA and DMS methods for forecasting US inflation. These extend conventional approaches by allowing for the set of predictors for inflation to change over time. When you have K models and a different one can potentially hold at each of T points in time, then the resulting K^T combinations can lead to serious computational and statistical problems (regardless of whether model averaging or model selection is done). As shown in this paper, DMA and DMS handle these problems in a simple, elegant and sensible manner.

In our empirical work, we present evidence indicating the benefits of DMA and DMS. In particular, it does seem that the best predictors for forecasting inflation are changing considerably over time. By allowing for this change, DMA and DMS lead to substantial improvements in forecast performance.

References

Ang, A. Bekaert, G. and Wei, M., "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?," *Journal of Monetary Economics* 54 (2007), 1163-1212.

Atkeson, A. and Ohanian, L., "Are Phillips Curves Useful for Forecasting Inflation?," *Federal Reserve Bank of Minneapolis Quarterly Review* 25 (2001), 2-11.

Avramov, D., "Stock Return Predictability and Model Uncertainty," *Journal of Financial Economics* 64 (2002), 423-458.

Chen, R. and Liu. J., "Mixture Kalman Filters," *Journal of the Royal Statistical Society, Series B* 62 (2000), 493-508.

Cogley, T. and Sargent, T., "Drifts and Volatilities: Monetary Policies and Outcomes in the post WWII U.S.," *Review of Economic Dynamics* 8 (2005), 262-302.

Cogley, T., Morozov, S. and Sargent, T., "Bayesian Fan Charts for U.K. inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System," *Journal of Economic Dynamics and Control* 29 (2005), 1893-1925.

Cremers, K., "Stock Return Predictability: A Bayesian Model Selection Perspective," *Review of Financial Studies* 15 (2002), 1223-1249.

Doan, T., R. Litterman, and C. Sims, "Forecasting and conditional projections using a realistic prior distribution", *Econometric Reviews* 3 (1984), 1-100.

Fagin, S. "Recursive Linear Regression Theory, Optimal Filter Theory, and Error Analyses of Optimal Systems," *IEEE International Convention Record Part i*, (1964) 216-240.

Fernandez, C., Ley, E. and Steel, M. "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics* 16 (2001), 563-576.

Fruhwirth-Schnatter, S. *Finite Mixture and Markov Switching Models* (New York: Springer, 2006).

Geweke, J. and Amisano, G., "Hierarchical Markov Normal Mixture Models with Applications to Financial Asset Returns," *Journal of Applied Econometrics* forthcoming (2010).

Green, P., "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika* 82 (1995), 711-732.

Groen, J., Paap, R. and Ravazzolo, F., "Real-time Inflation Forecasting in a Changing World," *Federal Reserve Bank of New York Staff Report Number 388* (2010).

Jazwinsky, A., *Stochastic Processes and Filtering Theory* (New York: Academic Press, 1970).

Hamilton, J., "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica* 57 (1989), 357-384.

Hoogerheide, L., Kleijn, R., Ravazzolo, F., van Dijk, H. and Verbeek, M., "Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time-Varying Weights," *Tinbergen Institute Discussion Paper 2009-061/4*, 2009.

Kim, C.J., "Dynamic Linear Models with Markov Switching," *Journal of Econometrics* 60 (1994), 1-22.

Koop, G., Leon-Gonzalez, R., Strachan, R., "On the Evolution of the Monetary Policy Transmission Mechanism," *Journal of Economic Dynamics and Control* 33 (2009), 997-1017.

Koop, G. and Potter, S., "Forecasting in Dynamic Factor Models using Bayesian Model Averaging," *The Econometrics Journal* 7 (2004), 550-565.

Korobilis, D., "Assessing the Transmission of Monetary Policy Shocks using Dynamic Factor Models," Discussion Paper 9-14, University of Strathclyde, 2009.

Madigan, D., York, J., "Bayesian Graphical Models for Discrete Data," *International Statistical Review* 63 (1995), 215-232.

Pesaran, M.H., Timmermann, A., "A Recursive Modelling Approach to Predicting UK Stock Returns," *The Economic Journal* 110 (2000), 159-191.

Primiceri. G., "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies* 72 (2005), 821-852.

Raftery, A., Karny, M. and Ettler, P., "Online Prediction Under Model Uncertainty Via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics*, 52 (2010), 52-66.

RiskMetrics, "Technical Document - Fourth Edition", (1996) available at http://www.riskmetrics.com/system/files/private/td4e.pdf.

Schervish, M. J. and Tsay, R., "Bayesian Modeling and Forecasting in Autoregressive Models," in J. C. Spall (Ed.), *Bayesian Analysis of Time Series and Dynamic Linear Models*, pp. 23–53. (New York: Marcel Dekker, 1988).

Smith, J. and Miller, J., "A Non-Gaussian State Space Model and Application to Prediction Records," *Journal of the Royal Statistical Society, Series B* 48 (1986), 79-88.

Stock, J. and Watson, M., "Evidence on Structural Instability in Macroeconomic Time Series Relations," J*ournal of Business and Economic Statistics* 14 (1996), 11-30.

Stock, J. and Watson, M., "Forecasting Inflation," *Journal of Monetary Economics* 44 (1999), 293-335.

Stock, J. and Watson, M., "Why Has U.S. Inflation Become Harder to Forecast?," *Journal of Monetary Credit and Banking* 39 (2007), 3-33.

Stock, J. and Watson, M., "Phillips Curve Inflation Forecasts," NBER Working Paper No. 14322, 2008.

Stock, J. and Watson, M., "Generalized Shrinkage Methods for Forecasting Using Many Predictors," manuscript available at http://www.princeton.edu/~mwatson/papers/stock_watson_gene (2011).

Data Appendix

The variables used in this study were taken from the sources in the table below. PHIL refers to the "Real-Time Data Set for Macroeconomists" database of the Philadelphia Federal Reserve Bank. FRED refers to the FRED database of the St. Louis Federal Reserve Bank. The inflation expectations index is maintained by the University of Michigan and can be found in http://www.sca.isr.umich.edu/main.php. All series are seasonally adjusted, where applicable, and run from 1960:Q1 to 2008:Q4. All variables are transformed to be approximately stationary. In particular, for the inflation variables we apply the transformation (column Tcode in the table): 0 - $y_{t+h} = \frac{400}{h}\left(\log\left(P_{t+h}\right) - \log\left(P_t\right)\right)$ where P_t is the original price index series (GDP or Personal consumption expenditure), and h is the forecast horizon. For the exogenous predictors, if we denote by $z_{i,t}$ the original untransformed series, we apply the transformations (column Tcode in the table): 1 - no transformation (levels), $x_{i,t} = z_{i,t}$; 2 - first difference, $x_{i,t} = z_{i,t} - z_{i,t-1}$; 4 - logarithm, $x_{i,t} = \log z_{i,t}$; 5 - first difference of logarithm, $x_{i,t} = 100\left(\log z_{i,t} - \log z_{i,t-1}\right)$.

#	Mnemonic	Tcode	Description	Source
1	PGDP	0	Inflation based on GDP deflator	PHIL
2	PCONX	0	Inflation based on PCE deflator	PHIL
3	UNEMP	1	Unemployment Rate	PHIL
4	CONS	5	Real Personal Consumption Expenditures: Total	PHIL
5	INV	5	Real Gross Private Domestic Investment: Residential	PHIL
6	GDP	5	Real Gross Domestic Product	PHIL
7	HSTARTS	4	Housing Starts	PHIL
8	EMPLOY	5	Nonfarm Payroll Employment	PHIL
9	PMI	2	ISM Manufacturing: PMI Composite Index	FRED
10	TBILL	1	3-Month Treasury Bill: Secondary Market Rate	FRED
11	SPREAD	1	Spread 10-year T-Bond yield / 3-month T-Bill (GS10 -TB3MS)	FRED
12	DJIA	5	Dow Jones Industrial Average	Bloomberg
13	MONEY	5	M1 Money Stock	PHIL
14	INFEXP	1	University of Michigan Inflation Expectations	Uni. of Mich.
15	COMPRICE	2	NAPM Commodity Prices Index (Percent)	Bloomberg
16	VENDOR	2	NAPM Vendor Deliveries Index (Percent)	Bloomberg