Received: 2 March 2020 | Revised: 27 July 2020 | Accepted: 2 August 2020

DOI: 10.1002/for.2723

RESEARCH ARTICLE

WILEY

Forecasting US inflation using Markov dimension switching

Jan Prüser

Fakultät Statistik, TU Dortmund, Dortmund, Germany

Correspondence
Jan Prüser, Fakultät Statistik, TU Dortmund, 44221 Dortmund, Germany.
Email: prueser@statistik.tu-dortmund.de

Abstract

This study considers Bayesian variable selection in the Phillips curve context by using the Bernoulli approach of Korobilis (*Journal of Applied Econometrics*, 2013, *28*(2), 204–230). The Bernoulli model, however, is unable to account for model change over time, which is important if the set of relevant predictors changes. To tackle this problem, this paper extends the Bernoulli model by introducing a novel modeling approach called Markov dimension switching (MDS). MDS allows the set of predictors to change over time. It turns out that only a small set of predictors is relevant and that the relevant predictors exhibit a sizable degree of time variation for which the Bernoulli approach is not able to account, stressing the importance and benefit of the MDS approach. In addition, this paper provides empirical evidence that allowing for changing predictors over time is crucial for forecasting inflation.

KEYWORDS

fat data, model change, Phillips curve, variable selection

JEL CLASSIFICATION

C11; C32; C53; E37

1 | INTRODUCTION

The Phillips curve has served as an important tool in macroeconomics for explaining and forecasting inflation in the USA over the past five decades. In the original Phillips curve, inflation depends on lags of inflation and the unemployment rate. In order to obtain a better understanding and potentially more precise forecasts, a large literature extends the Phillips curve with additional explanatory variables. Influential papers include Stock and Watson (1999, 2007), Atkeson and Ohanian (2001), Ang, Bekaert, and Wei (2007), and Groen, Paap, and Ravazzolo (2013). Forecasting inflation is crucial, for example for central banks, but at the same time challenging. One difficulty arises from the problem of which additional variables to include in the Phillips curve. While the original Phillips curve is likely to miss some important predictors, an augmented Phillips curve with too many predictors bears the risk of overfitting the data, leading to imprecise out-of-sample predictions. This raises the question of which predictors are relevant. However, the relevance of the predictors may change over time. In this case, only asking if a variable is important or not is not addressing the right question. A researcher may not be interested in assessing whether a variable is important, but rather when it is.

This paper addresses the question of which predictor is relevant by following Korobilis (2013a) and considers

Bayesian variable selection in the Phillips curve context. Korobilis provided an algorithm for stochastic variable selection. The key idea is to introduce an indicator for each predictor, which determines if a variable is included in the model. Each indicator is drawn from a Bernoulli distribution in a Gibbs sampler scheme. By doing so, it is possible to calculate variable inclusion probabilities to assess the importance of single predictors in determining inflation. However, a potential drawback is that the set of indicators is assumed to be constant over time. Thus the Bernoulli approach is unable to account for model change over time, which is desirable if the set of relevant predictors changes over time. The importance of changing predictors over time is documented by, among others, Stock and Watson (2010), who found that most predictors for inflation improved forecast performance only in some specific time periods. Therefore, it may be empirically important for predictors to change over time. Conventional hypothesis testing approaches designed for constant parameter models are also not capable of allowing for this, as they only test whether a restriction holds for all time periods or never. The main contribution of this paper is to tackle this problem by introducing a novel modeling approach called Markov dimension switching (MDS). The MDS model can be seen as an extension of the Bernoulli model. In the MDS model each indicator follows a Markov switching process and thus allows for changing predictors over time. Hence this approach allows for the calculation of time-varying variable inclusion probabilities to shed light on the question of which variables are important in determining inflation at different times.

The relevance of this extension is illustrated by using the Bernoulli and the MDS approach to assess the importance of the predictors for one-quarter and one-year inflation. Most important predictors for one quarter turn out to be inflation expectations, the percentage change of the oil price, and the Treasury bill rate. The unemployment rate, inflation expectations, the Treasury bill rate, capital utilization, and the number of newly built houses turn out to be the most important predictors for one-year inflation.[1]The relevant variables show a sizable degree of time variation, which the Bernoulli approach cannot account for, highlighting the benefit and importance of the proposed MDS approach of this paper. In particular, MDS reveals that the relevance of inflation expectations, unemployment, capital utilization, and house prices for the one-year horizon changes abruptly over time, which would be difficult to capture for existing methods that assume a gradual change of the relevance of predictors. A simulation study illustrates this. In addition, we provide empirical evidence that allowing for changing predictors over time is crucial for forecasting inflation. We find that the MDS approach exhibits a better forecasting performance than the Bernoulli approach. An additional finding is that the MDS approach forecasts well in comparison with a range of other plausible approaches. Approaches that allow the predictors to change over time turn out to forecast best. These findings hold for several forecasting horizons, different measures of inflation, and for point as well as for density forecasts.

The remainder of this paper is organized as follows. Section 2 compares MDS with existing literature. Section 3 lays out and discusses the econometric framework. Section 4 provides a simulation study. Section 5 presents the empirical findings and Section 6 concludes.

2 | COMPARISON WITH EXISTING LITERATURE

A growing literature works with Bayesian priors in models with many parameters, which shrink some of the parameters towards zero to ensure parsimony. For example, Bańbura, Giannone, and Reichlin (2010) found that shrinking parameters led to improved forecasts in large vector autoregression (VAR) models. There is also an increasing number of papers applying shrinkage by using hierarchical priors, such as the lasso prior introduced by Park and Casella (2008). Hierarchical priors have the advantage that the priors introducing the shrinkage depend on unknown parameters that are estimated from the data, resulting in data-driven shrinkage. For example, Korobilis (2013b) showed that hierarchical shrinkage was useful for macroeconomic forecasting using many predictors. In a Phillips curve context, Belmonte, Koop, and Korobilis (2014) used the lasso prior in a time-varying parameter (TVP) model. The lasso prior in their model automatically decided which parameter was time-varying, constant or shrunk towards zero. This approach may be well suited to model structural changes in the Phillips curve while avoiding overfitting.

Fewer papers deal with model change over time as opposed to parameter change (which empirically can only poorly approximate model change by allowing coefficients to be estimated as being approximately zero). Chan, Koop, Leon-Gonzalez, and Strachan (2012) considered dimension switching in a TVP framework using the algorithm of Gerlach, Carter, and Kohn (2000). However, in their forecasting study, they only considered models with no predictors, a single predictor, or all m predictors. In other words, Chan et al. considered $m + 2$ combinations and not 2^m, as this would be computationally infeasible for the algorithm they used. MDS addresses this

[1]Note that the results differ slightly depending on which measure of inflation is used; see Section 4.3 for details.

problem by employing one Markov process, which can take on two states, for each predictor. By doing so it is possible to explore all 2^m possible combinations. Modeling changing predictors by using a single Markov process that can take on 2^m values would result in a huge transition matrix, and hence estimation of this matrix would be infeasible. In order to avoid the estimation of such a huge matrix, Raftery, Kárný, and Ettler (2010) replaced it with time-varying model probabilities using approximations in the form of so-called forgetting factors (sometimes also called discount factors). These probabilities allow placing time-varying weight on the individual forecasts of a set of typically 2^m different models. Because the weights are allowed to change over time, this approach is called dynamic model averaging (DMA). Koop and Korobilis (2012) found that DMA led to substantial improvements in forecasting inflation over simple benchmark models and more sophisticated approaches.

In contrast to DMA or hierarchical shrinkage, the MDS model has the advantage that through the indicator variables the likelihood contains information about the relevance of every predictor at each point in time and thereby may lead to more efficient estimates. In the DMA approach each model is estimated independently and does not use the information of the time-varying weights. For example, at the beginning of the sample the most weight may be placed on models with only a few predictors, and at the end of the sample more weight may be assigned to models with a large set of predictors. However, each individual model is estimated using the same set of predictors for the whole sample ignoring this information. However, it would be useful to take this information into account when estimating the parameters, and this is exactly what the MDS model does. Moreover, DMA relies on approximations in the form of forgetting factors and only provides filtered estimates. In contrast, MDS can easily be estimated using Gibbs sampling and thereby it can use the information of the full sample (i.e., it uses filtering and smoothing) and at the same time it takes full parameter uncertainty into account. In the hierarchical shrinkage approach some parameters are shrunk towards zero (i.e., the corresponding variables are irrelevant), but this information is only contained in the prior and not in the likelihood function. Furthermore, this approach cannot account for model change over time, as it shrinks the parameters towards zero for all time periods or never. Despite the potential advantages of MDS, the assumption of constant parameters may appear restrictive. However, this assumption is less restrictive than it seems, as the time-varying inclusion probabilities introduce a time-varying data-based shrinkage on the coefficients. Therefore, MDS addresses overfitting concerns and allows for model change over time.

3 | MARKOV DIMENSION SWITCHING

The Phillips curve serves as a starting point and motivation for many models that forecast inflation. In the original Phillips curve, inflation depends only on the unemployment rate and lags of inflation. Including additional predictors, as Stock and Watson (1999) among many others do, leads to the so-called generalized Phillips curve:

$$\pi_{t+h} = \alpha + \sum_{j=0}^{p-1} \phi_j \pi_{t-j} + \boldsymbol{x_t}\beta + \epsilon_{t+h}, \tag{1}$$

where $\boldsymbol{x_t}$ is a $1 \times q$ vector of exogenous predictors, $\pi_{t+h} = \log(P_{t+h}) - \log(P_t)$, P_t denotes the price level, and $\epsilon_t \sim N(0, \sigma_t^2)$. The number of parameters may be large relative to the number of observations, as in many macroeconomic applications. Estimation of the Phillips curve in this case may cause imprecise estimation and overfitting (i.e., the model fits the noise in the data, rather than finding the pattern useful for forecasting). Hence it is important to identify the truly relevant predictors out of a set of many potentially relevant predictors. To do so, this paper follows Korobilis (2013a) and considers Bayesian variable selection in the Phillips curve context by introducing $m = q + p + 1$ indicators $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)$. The model can now be written as

$$\pi_{t+h} = (\boldsymbol{z_t} \odot \gamma)\theta + \epsilon_{t+h}, \tag{2}$$

where $\boldsymbol{z_t} = (1, \pi_t, \dots, \pi_{t-p+1}, \boldsymbol{x_t})$, $\theta = (\alpha, \phi_0, \dots, \phi_{p-1}, \beta')'$ and $\odot$ denotes elementwise multiplication. Hence, if $\gamma_i = 1$, the ith variable is included in the model and, if $\gamma_i = 0$, it is not. By sampling the indicators from their posterior, all 2^m possible variable combinations can be considered and estimated in a stochastic manner. A potential drawback, however, is that the indicators are constant over time. Thus a predictor is either included or excluded from the model for all periods, which is undesirable if the set of predictors changes over time. To address this problem, this paper introduces MDS to allow the indicator variables to change over time. In the MDS each indicator variable follows a first-order Markov switching process and therefore $\boldsymbol{\gamma}$ is replaced by $\boldsymbol{\gamma_t}$:

$$\pi_{t+h} = (\boldsymbol{z_t} \odot \gamma_t)\theta + \epsilon_{t+h}, \tag{3}$$

where $\boldsymbol{\gamma_t} = (\gamma_{1,t}, \dots, \gamma_{m,t})$. Each Markov switching process $\gamma_{i,t}$ can take on the value one or zero and is characterized by a 2×2 transition matrix $\boldsymbol{\mu_i}$, where $\mu_{kj,i} = \Pr(\gamma_{i,t+1} =$

$j|\gamma_{i,t}=k)$, $k=0,1$ and $j=0,1$.[2] If $\gamma_{i,t}=1$, the ith variable is included in the model at period t and, if $\gamma_{i,t}=0$, it is not. Therefore, the means of the posterior draws of $\gamma_{i,t}$ can be interpreted as a time-varying variable inclusion probability in this modeling context. Furthermore, note that keeping $\boldsymbol{\theta}$ constant does not imply that a certain variable has either an impact of zero or an impact given by $\boldsymbol{\theta}$. This is because the time-varying inclusion probabilities introduce a time-varying data-based shrinkage on the coefficients. Therefore, MDS may avoid overfitting and hence can be a useful tool for forecasting. In contrast, estimating $\boldsymbol{\theta}$ in a time-varying manner by models with time-varying parameters or models with Markov switching parameters bears a high risk of overfitting and may empirically only poorly approximate changing predictors by allowing coefficients to be estimated as being approximately zero. Section 4 provides a simulation study that investigates this point. Furthermore, models with time-varying parameters typically assume a gradual change in parameters and therefore are not well suited to capture abrupt changes in the relevance of predictors.

3.1 | Gibbs sampler

This section describes the Gibbs sampler, which allows us to draw from the posterior distribution of the Bernoulli and the MDS model.

1. Sample $\boldsymbol{\theta}$ from the following density:

$$\theta|\gamma_{1:T},\boldsymbol{z}_{1:T},\pi_{1+h:T+h},\sigma^2_{1+h:T+h}\sim N(\overline{\theta},\overline{\Omega}), \tag{4}$$

with

$$\begin{aligned}\overline{\theta} &= \overline{\Omega}\left(\boldsymbol{V}(\hat{\theta}_{\text{OLS}})\hat{\theta}_{\text{OLS}}+\sum_{t=1}^{T}(\boldsymbol{z}_t\odot\gamma_t)'\sigma^{-2}_{t+h}\pi_{t+h}\right),\\ \overline{\Omega} &= \left(\boldsymbol{V}(\hat{\theta}_{\text{OLS}})+\sum_{t=1}^{T}(\boldsymbol{z}_t\odot\gamma_t)'\sigma^{-2}_{t+h}(\boldsymbol{z}_t\odot\gamma_t)\right)^{-1}.\end{aligned}$$

The ordinary least squares (OLS) estimate of the full model is used as the prior. This is an empirical Bayes approach. When one variable is omitted from the model for the full sample period, the parameter of this predictor is drawn from the prior density. In order to obtain reasonable draws in this case, the OLS estimate of the model seems to be a useful choice. Then the mean of the posterior of $\boldsymbol{\theta}$ is the weighted average of the OLS estimate of the full model and the OLS estimate using only a subset of the predictors. While the OLS estimate of the full model likely has a higher variance as it is likely to include irrelevant predictors, the OLS estimate based on the sparse data matrix is more likely to suffer from omitted variables bias. Hence the posterior addresses the classic bias variance tradeoff in a convenient way by placing weights on both estimates in a data-driven way.

2. Sample $\gamma_{i,t}$ for $i=1,\dots,m$:
   - If γ_i is constant, sample it from

$$\begin{aligned}&\gamma_i|\gamma_{-i},\pi_{1+h:T+h},\boldsymbol{z}_{1:T},\theta,\sigma^2_{1+h:T+h}\\ &\sim\text{Bernoulli}\left(\frac{l_{1i}}{l_{1i}+l_{0i}}\right),\end{aligned} \tag{5}$$

with

$$\begin{aligned}l_{1i} &= \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\pi_{t+h}-(\boldsymbol{z}_t\odot\gamma_{[\gamma_i=1]})\theta}{\sigma^2_{t+h}}\right)^2\right)\\ &\quad p(\gamma_i=1),\\ l_{0i} &= \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\pi_{t+h}-(\boldsymbol{z}_t\odot\gamma_{[\gamma_i=0]})\theta}{\sigma^2_{t+h}}\right)^2\right)\\ &\quad p(\gamma_i=0),\end{aligned}$$

where $p(\gamma_i=1)=0.5$.
   - In the MDS model $\gamma_{i,t}$ is sampled for $t=1,\dots,T$ conditioning on $\boldsymbol{\gamma}_{-i,1:T}$, $\pi_{1+h:T+h}$, $\boldsymbol{z}_{1:T}$, $\boldsymbol{\theta}$, $\sigma^2_{1+h:T+h}$ and the transition probabilities of the ith Markov process $\boldsymbol{\mu}_i$, using the algorithm of Chib (1996) (see Appendix C1 for details). The transition probabilities of the ith Markov process are drawn from a Beta distribution:

$$\mu_{11,i}|\gamma_{i,1:T}\sim\text{Beta}(u_{11}+n_{11},u_{10}+n_{10}), \tag{6}$$

$$\mu_{00,i}|\gamma_{i,1:T}\sim\text{Beta}(u_{00}+n_{00},u_{01}+n_{01}), \tag{7}$$

where n_{jk} counts the number of transitions from state j to k and u_{jk} is the prior hyperparameter. Setting $u_{11}=u_{00}=u_{10}=u_{01}=1$ corresponds to the uniform prior. The posterior is not sensitive to this prior choice if none of the four possible transitions are rare. However, it is also possible to use a more informative prior. For example, a researcher may want to avoid a high frequency of regime changes and smooth the variable inclusion probability over time. Thus, once we are in a regime—that is, a variable is excluded or included

[2] The Markov mixture modeling approach allows that the probability of switching depends on the current state of the stochastic process, which is not the case for i.i.d. mixture models, but may be useful to model dependence over time and allows to formulate different prior beliefs about the frequency of dimension switching and the level of sparsity in the model (see Section 2.1). The i.i.d. case is, however, nested as a special case of the Markov mixture approach.

in the model—the regime should only be switched if there is a strong signal in the data. This prior belief can be implemented by setting $u_{11} = u_{00} = T$. Sparse models are typically known to forecast better than models with too many variables. A stronger favor for sparse models would be achieved by only setting $u_{00} = T$. All three prior specifications—that is, the uniform, the smooth, and the sparse prior—are considered in the empirical part.

3. Sample σ_t^{-2}:
   - In the case of homoskedastic errors where $\sigma_t^2 = \sigma^2$, sample from the density

$$\sigma^{-2}|\theta,\pi_{1+h:T+h},\mathbf{z}_{1:T},\gamma_{1:T} \sim \text{Gamma}(a,b^{-1}), \tag{8}$$

where $a = T + a_0$ and $b = b_0 + \sum_{t=1}^{T}(\pi_{t+h} - (\mathbf{z}_t \odot \gamma_t)\theta)^2$. The hyperparameters a_0 and b_0 are set to zero.

   - In the case of heteroskedastic errors, sample conditioning on $\boldsymbol{\theta}$, $\pi_{1+h:T+h}$, $\mathbf{z}_{1:T}$, $\boldsymbol{\gamma}_{1:T}$, using the algorithm of Kim, Shephard, and Chib (1998) by assuming that

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \xi_t, \tag{9}$$

where $\xi_t \sim N(0,\zeta)$ and ζ is sampled from

$$\zeta^{-1}|\sigma^2_{1+h:T+h} \sim \text{Gamma}(a,b^{-1}), \tag{10}$$

where $a = T + \kappa_1$ and $b = \kappa_2 + \sum_{t=1+h}^{T+h}(\log(\sigma_t) - \log(\sigma_{t-1}))^2$.
The hyperparameters κ_1 and κ_2 are set to 3 and 0.0001.

4 | SIMULATION STUDY

The aim of the simulation study is to illustrate the usefulness of the MDS approach in an environment of changing predictors over time. In particular, it is assumed that the relevance of each predictor is determined by a first-order Markov switching process (the exact form of the data generating processes (DGPs) is defined below). In principle, models with time-varying parameters or Markov switching parameters can be used to approximate changing predictors by allowing coefficients to be estimated as being approximately zero. This simulation study compares such approaches with the MDS approach. A time-varying parameter model, where the parameters are allowed to evolve according to a random walk process, and a model where each individual coefficient is allowed to switch according to an individual Markov switching process, are considered.

The following DGPs are considered:

$$y_t = (x_{1t} \times S_{1t})\beta_1 + \ldots + (x_{pt} \times S_{pt})\beta_p + \epsilon_t, \tag{11}$$

where $\epsilon_t \sim N(0,1)$, $\boldsymbol{X}_t$ is a vector of p predictor variables and S_{it} for $i = 1, \ldots, p$ follows a Markov switching process. Each Markov switching process S_{it} can take on the value one or zero and is characterized by a 2×2 transition matrix $\boldsymbol{\mu}_i$, where $\mu_{kj,i} = \Pr(S_{it+1} = j|S_{it} = k)$, $k = 0,1$ and $j = 0,1$. The diagonal elements (i.e., probabilities that the Markov switching process remains in the current regime) are set to 0.9, implying a persistent process. The coefficients, β_i $i = 1, \ldots, p$ are generated from $\beta_i \sim U(-4,4)$ and the predictor variables are generated from $\boldsymbol{X}_t \sim N(\mathbf{0},\boldsymbol{V})$, where $\boldsymbol{V}$ is a $p \times p$ matrix of correlations with elements $V_{ij} = \rho^{|i-j|}$. DGPs with different numbers of predictors p, numbers of observations T and different correlations ρ are considered. In particular, models with $p = 8{,}12$ predictors, $T = 100$, 200 observations and $\rho = 0{,}0.9$ correlations for the predictor variables are generated.

The results for all DGPs are summarized in Table 1. Entries in this table are mean squared deviations (*MSD*) averaged over 500 artificially generated data sets, T time periods, and over p. To be precise, the impact of each predictor at each time point t on y_t is given by $S_{it}\beta_i$; thus the *MSD* is calculated as

$$MSD = \frac{1}{500*T*p}\sum_{r=1}^{500}\sum_{t=1}^{T}\sum_{i=1}^{p}\left(S_{it}^{(r)}\beta_i^{(r)} - \hat{S}_{it}^{(r)}\hat{\beta}_i^{(r)}\right)^2, \tag{12}$$

where $r = 1, \ldots, 500$ denotes the number of simulation replications. For the MDS approach $\hat{S}_{it}$ and $\hat{\beta}_i$ are the posterior means obtained from the Gibbs sampler. For the time-varying parameter model $\hat{S}_{it}\hat{\beta}_i$ is replaced by $\hat{\beta}_{it}$, the posterior mean estimate of β_{it} the time-varying parameter, and in the model, where each individual coefficient is allowed to switch according to a Markov switching process, $\hat{S}_{it}\hat{\beta}_i$ is replaced by $\hat{\beta}_{S_{it}}$, the posterior mean estimate of $\beta_{S_{it}}$ the switching parameter. The simulation study reveals that MDS produces a lower *MSD* than the Markov switching parameter model and the TVP model. This is plausible as the Markov switching parameter model and TVP model can only approximate changing predictors by estimating the coefficients as approximately zero. In addition, the TVP model is not well suited to capture abrupt changes in the relevance of predictors. Hence the results illustrate that the approximation of models with time-varying or Markov switching parameters of changing predictors is less precise than directly modeling changing

TABLE 1 Simulation results

	$p = 8$ predictors				$p = 12$ predictors			
	$T = 100$		$T = 200$		$T = 100$		$T = 200$	
	$\rho = 0$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.9$	$\rho = 0$	$\rho = 0.9$
MDS	0.86	1.38	0.67	1.09	1.19	1.71	1.02	1.38
MS	1.31	1.59	1.27	1.30	1.36	1.83	1.30	1.49
TVP	1.34	2.28	1.29	1.74	1.40	2.81	1.39	2.01

Note. The table shows the *MSD* as defined in Equation 12.

predictors. How relevant changing predictors are for modeling inflation will be investigated empirically in the next section.

5 | FORECASTING INFLATION

5.1 | Data

This study forecasts US inflation as measured by the personal consumption expenditure (PCE) deflator, the gross domestic product (GDP) deflator, and the consumer price index (CPI) for 1978:Q2 to 2016:Q4.[3]The period from 1992:Q1 to 2016:Q4 is used to evaluate the out-of-sample forecast performance. A wide range of variables is considered as potential predictors, reflecting the major theoretical explanations of inflation as well as variables that have been found to be useful in forecasting inflation in other studies. The following predictors are used:

- *DJIA*: the percentage change in the Dow Jones Industrial Average.
- *EMPLOY*: the percentage change in employment.
- *HSTARTS*: the log of housing starts.
- *INFEXP*: University of Michigan survey of inflation expectations.
- *MONEY*: the percentage change in the money supply (M1).
- *OIL*: the percentage change of Spot Crude Oil Price: WTI.
- *PMI*: the change in the Institute of Supply Management (Manufacturing): Purchasing Managers Composite Index.
- *CONS*: the percentage change in real personal consumption expenditures.
- *GDP*: the percentage change in real GDP.
- *INV*: the percentage change in real gross private domestic investment (residential).
- *SPREAD*: the spread between the 10-year and 3-month Treasury bill rates.
- *TBILL*: 3-month Treasury bill (secondary market) rate.
- *UNEMP*: unemployment rate.
- *CAPUT*: the change in capital utilization (manufacturing).

The variables are obtained from the "Real-Time Data Set for Macroeconomists" database of the Philadelphia Federal Reserve Bank and from the FRED database of the Federal Reserve Bank of St. Louis. All predictors are real-time quarterly data, so that all forecasts are made using versions of the variables available at the respective time. Furthermore, all data are seasonally adjusted if necessary. If not stated otherwise, all models considered in the next section include four lags of quarterly inflation as additional predictors. This is consistent with quarterly data.

5.2 | Out-of-sample results

In this section, the forecasting performance of the MDS model is investigated. In a first step, MDS and Bernoulli models are considered in which the first lag of inflation and the intercept are always included and all other variables are allowed to be omitted from the model. The MDS model is estimated with the uniform, the smooth and the sparse prior for the transition probabilities. Furthermore, the Bernoulli model is applied with constant parameters, time-varying parameters[4] and with Markov switching parameters.[5] In order to assess whether the

[3]Owing to data availability the sample starts in 1978:Q2.

[4]In all time-varying parameter models it is assumed that the parameters follow independent random walks and for the variance of the error terms, of these random walks, inverse game priors with shape parameter of 3 and scale parameter of 0.0001 are used in order to regularize the degree of time variation.

[5]In this model, all parameters switch jointly according to one Markov switching process. This leads to better forecasting results than letting each individual parameter switch according to individual Markov switching processes.

MDS or the Bernoulli approach is useful to avoid overfitting, their forecast performance is compared with an AR(1) model, a TVP-AR(1), and a multiple regression model containing all variables. All these models are applied with a constant and a stochastic variance specification, as explained in the description of the Gibbs sampler.

In a second step the forecasting performance of the MDS model is compared with three further modeling approaches that have been found useful in inflation forecasting. These approaches are DMA proposed by Koop and Korobilis (2012), the hierarchical shrinkage (LASSO) in TVP models proposed by Belmonte et al. (2014), and the unobserved components model with stochastic volatility (UCSV) proposed by Stock and Watson (2007). For DMA, three forgetting factors have to be set by the researcher. The first controls the amount of time variation in the coefficients, the second the amount of time variation of the volatility, and the third controls the amount of time variation of the model probabilities (see Koop & Korobilis, 2012, for details). Setting these forgetting factors to one leads to the special case of constant coefficients, constant variance, and constant model probabilities. Values close to one are typically used in the literature because of overfitting concerns. Koop and Korobilis (2012) set the hyperparameter for the variance to 0.98 and set the forgetting factors for the coefficients and model probabilities to either 0.95 or 0.99, which was found to deliver a favorable forecasting performance over simple benchmark regressions and more sophisticated approaches. Thus this set of values is used to forecast inflation. Moreover, dynamic model selection (DMS) is considered next to DMA in the forecasting comparison. In the TVP model with hierarchical shrinkage the specification of the hierarchical gamma prior is crucial; see Belmonte et al. (2014) for details. In the application the shape and scale parameter of the inverse gamma prior is set to 0.1, leading to a relatively flat prior. As a special case of this model, the lasso prior by Park and Casella (2008) in a regression model with constant coefficients is also considered using the same hierarchical inverse gamma prior. Furthermore, these two models are estimated using the same two specifications for the variance as for the MDS models. Finally, for the UCSV model the same stochastic variance specification for the two system variances of the state-space model is used as for all other models (see Stock & Watson, 2007, for details).

In order to evaluate the forecast performance, the root mean squared forecast error (RMSFE) and the mean absolute forecast error (MAFE) as standard forecast metrics are used. In order to investigate whether the forecasting errors differ (statistically) significantly from those obtained from a benchmark AR(1) model, the Diebold–Mariano test (DM test) proposed by Diebold and Mariano (1995) is employed. However, the RMSFE and MAFE only evaluate the point forecasts and ignore the remaining part of the predictive distribution. This is the reason why the predictive likelihood may be preferable to evaluate the forecast performance. The predictive likelihood is the predictive density for π_{t+h} (given data through time t) evaluated at the actual outcome and as a forecast metric has the advantage of evaluating the forecast performance of the entire predictive density. Additionally, the predictive likelihood can also be used for model selection. Therefore, the mean of the log predictive likelihood is used as an additional forecast metric. For a motivation and detailed description of the predictive likelihood see Geweke and Amisano (2011).

Tables B1–B6 show the results for the forecasting performance of the different models for all three measures of inflation and four forecasting horizons ($h = 1,2,3,4$). Overall, it turns out that the MDS models forecast quite well. In particular, in terms of density forecasts they are always among the best models or the best models. The full model, including all predictors, seems to overfit as it tends to forecast poorly compared to an AR(1) benchmark model. Variable selection in the Bernoulli model delivers forecasting improvements over the full model including all predictors, but does not improve over the simple AR(1) model. An exception is the Bernoulli model with time-varying parameters, which forecasts well in some cases where the TVP-AR(1) performs well. This makes sense as the TVP-AR(1) is a special case of the TVP-Bernoulli model. But there are also cases where the TVP-AR(1) and TVP-Bernoulli model forecast poorly. Similarly, there are cases where the UCSV forecasts well, but also cases where it forecasts poorly. Forecasting improvements over the full model and a simple AR(1) model can be achieved by considering dynamic variable selection in the form of MDS. The MDS models tend to forecast better than the Bernoulli models, both in terms of point forecasts and in terms of the predictive likelihood as a forecasting metric. The different priors for the transition probabilities deliver a similar forecasting performance, but the sparse prior tends to work best. Overall, we find that they all provide a good forecasting performance. The models with stochastic variance turn out to forecast better than models with constant variance. However, the MDS model is less sensitive to this choice than TVP regression models with many predictors, which tend to forecast poorly with a constant variance specification, as the time-varying coefficients falsely fit the time-varying volatility rather than finding a pattern useful for forecasting in this case. The DMA and DMS approach (which also allows for changing predictors) is found to forecast quite well. Both the MDS and the DMA approach look to be attractive options. They often do best, but where not they do not go too far wrong. However, none of the two approaches significantly improves forecasts in comparison to the other. Nevertheless, the MDS models tend to provide better density forecasts and the DMA approach tends to provide better point forecasts. This finding stresses the importance of allowing for changing predictors over time using the Phillips curve to forecast inflation.

5.3 | Full sample results

The calculation of variable inclusion probabilities is interesting from an economic perspective, but may also provide an explanation why MDS models provide better inflation forecasts than the Bernoulli models. Figures C1–C6 display the inclusion probabilities of the MDS model with the uniform, the smooth and the sparse prior for the transition probabilities, and the Bernoulli model for the full sample. The results are shown for all three measures of inflation, namely the PCE deflator, the GDP deflator, and CPI, and the results are based on models with the stochastic variance specification. Overall, the Bernoulli approach assigns higher inclusion probabilities to the variables than the MDS models. This may be one reason why the MDS models deliver better forecasts. Another reason may be that some inclusion probabilities show a sizable degree of time variation, for which the Bernoulli approach cannot account. This demonstrates the usefulness of the MDS model over the Bernoulli model. Comparing the three different priors for the MDS models reveals that under the smooth prior the variable inclusion probabilities are less noisy, as a stronger signal is needed to obtain a regime change compared to the uniform prior. In addition, the sparse prior yields more parsimonious models, as a stronger signal in the data is needed for a variable to be included in the model. However, in some cases the signal in the data is strong enough to yield similar inclusion probabilities for the different prior specifications.

In many cases the Bernoulli model and the MDS model under the uniform prior deliver similar results. In some cases the MDS model even assigns a roughly constant inclusion probability to a variable. In other cases the MDS model also assigns a high probability to one variable, but the probability changes over time. For one-quarter PCE inflation *INEXP*, *OIL*, and *TBILL* turn out to be important, and for one-year PCE inflation *INEXP*, *CAPUT*, *HSTARTS*, and *TBILL* turn out to be important. In comparison, for one-quarter GDP inflation only *INEXP* turns out to be important, and for one-year GDP inflation *INEXP*, *HSTARTS*, and *UNEMP* turn out to be important. Finally, for one-quarter CPI inflation *INEXP*, *OIL*, and *TBILL* turn out to be important, and for one-year CPI inflation *INEXP*, *UNEMP*, *CAPUT*, and *TBILL* turn out to be important.

For one-year inflation the important variables show a sizable degree of time variation. In particular, the inclusion probabilities of *INEXP*, *HSTARTS*, *CAPUT*, and *UNEMP* switch very rapidly over time. This shows that the relevance of predictors does not always change gradually, as is assumed to be the case, for example, in TVP models. From an economic perspective it is particularly interesting that the relevance of *UNEMP* changes that rapidly as it has long been assumed that economic policymakers face a tradeoff between unemployment and inflation. These results, however, suggest that this relation might not be stable over time. Overall, the result that the relevant predictors change abruptly over time and are rather short lived are in line with results found by Koop and Korobilis (2012). Thus this paper strengthens the empirical evidence that changing predictors are important for modeling inflation.

6 | CONCLUSION

This study uses the generalized Phillips curve to forecast inflation. While the original Phillips curve is likely to miss some important predictors, a generalized Phillips curve that uses too many predictors may lead to overfitting the data and to imprecise out-of-sample predictions. Thus this paper aims to assess which variables are important in determining inflation by using the Bernoulli model. The Bernoulli model, however, is unable to account for model change over time. In order to be able to account for the possibility that the set of predictors changes over time, this paper introduces the MDS approach. In the MDS approach the set of predictors is allowed to change over time. The empirical application reveals that the most important variables in the generalized Phillips curve are inflation expectations, the percentage change of the oil price, and the Treasury bill rate for one-quarter inflation and unemployment rate, inflation expectations, the Treasury bill rate, capital utilization, and the number of newly built houses for one-year inflation. Furthermore, for one-year inflation the unemployment rate, the Treasury bill rate, and the number of newly built houses show a sizable degree of time variation for which the Bernoulli approach is not able to account, highlighting the importance and benefit of the MDS approach. This is also confirmed in a forecasting exercise, where the MDS model delivers more precise forecasts than the Bernoulli model. In addition, the paper demonstrates that the MDS model forecasts well in comparison with a range of other plausible alternatives. Taken together, the paper presents a battery of theoretical and empirical arguments for the potential benefits of the MDS approach.

ACKNOWLEDGMENTS

I thank Christoph Hanck as well as the participants of the Annual Bayesian Econometrics Workshop in Rimini, RGS Doctoral Conference in Economics in Essen, and International Conference on Computational and Methodological Statistics in London for valuable comments. Open access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the Real-Time Data Set for Macroeconomists database of the Philadelphia Federal Reserve Bank and from the FRED database of the Federal Reserve Bank of St. Louis at https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/ and https://fred.stlouisfed.org/.

ORCID

Jan Prüser https://orcid.org/0000-0001-7435-7334

REFERENCES

Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, *54*(4), 1163–1212.

Atkeson, A., & Ohanian, L. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, *25*(1), 2–11.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.

Belmonte, M., Koop, G., & Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, *33*(1), 80–94.

Chan, J. C. C., Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2012). Time varying dimension models. *Journal of Business and Economic Statistics*, *30*(3), 358–367.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, *75*(1), 79–97.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–265.

Gerlach, R., Carter, C., & Kohn, R. (2000). Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association*, *95*(451), 819–828.

Geweke, J., & Amisano, G. (2011). Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, *26*(1), 1–29.

Groen, J., Paap, R., & Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business and Economic Statistics*, *31*(1), 29–44.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*(2), 357–384.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, *65*(3), 361–393.

Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, *53*(3), 867–886.

Korobilis, D. (2013a). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, *28*(2), 204–230.

Korobilis, D. (2013b). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting*, *29*(1), 43–59.

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.

Raftery, A., Kárný, M., & Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, *52*(1), 52–66.

Stock, J., & Watson, M. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44*(2), 293–335.

Stock, J., & Watson, M. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*(1), 3–33.

Stock, J., & Watson, M. (2010). Modeling inflation after the crisis (NBER Working Paper No. 16488). Cambridge, MA: National Bureau of Economic Research.

AUTHOR BIOGRAPHY

Jan Prüser is PostDoc at the TU Dortmund University. He received his PhD at the Universityof Duisburg-Essen. His research interest include Bayesian econometrics, applied time series analysis, forecasting and modelling of high dimensional data.

How to cite this article: Prüser J. Forecasting US inflation using Markov dimension switching. *Journal of Forecasting*. 2021;40:481–499. https://doi.org/10.1002/for.2723

APPENDIX A: GIBBS SAMPLING IN MARKOV SWITCHING MODELS

This paper considers a Markov switching indicator variable (denoted by $\gamma_{i,t}$) for each variable. Each Markov switching process S_t[6] can take on the value one or zero and is characterized by a 2×2 transition matrix $\boldsymbol{\mu}$, where $\mu_{kj}=\Pr(S_{t+1}=j|S_t=k)$, $k=0{,}1$, and $j=0{,}1$. In order to draw S_t for $t=1,\dots,T$ first the Hamilton filter, proposed by Hamilton (1989), is used, followed by the simulation smoother of Chib (1996):

1. Initialize the Hamilton filter using steady-state probabilities:

$$\begin{aligned}\Pr(S_0=0) &= \frac{1-\mu_{11}}{2-\mu_{11}-\mu_{00}},\\ \Pr(S_0=1) &= \frac{1-\mu_{00}}{2-\mu_{11}-\mu_{00}}.\end{aligned}$$

2. Given $\Pr(S_{t-1}=k|\psi_{t-1})$, where ψ_{t-1} denotes the information set at time point $t-1$, calculate $\Pr(S_t=j|\psi_{t-1})$ as

$$\Pr(S_t=j|\psi_{t-1})=\sum_{k=0}^{1}\mu_{kj}\Pr(S_{t-1}=k|\psi_{t-1}).$$

3. Given ψ_t update the probabilities as

$$\Pr(S_t=j|\psi_t)=\frac{\mathrm{f}(y_t|S_t=j,\psi_{t-1})\Pr(S_t=j|\psi_{t-1})}{\sum_{j=0}^{1}\mathrm{f}(y_t|S_t=j,\psi_{t-1})\Pr(S_t=j|\psi_{t-1})},$$

where $\mathrm{f}(y_t|S_t=j,\psi_{t-1})$ denotes the likelihood function of the dependent variable.

4. Sample S_T using $\Pr(S_T=1|\psi_T)$.
5. Sample $S_{T-1},\dots,S_1$ sequentially using

$$\Pr(S_t=1|S_{t+1},\psi_t)=\frac{\Pr(S_{t+1}|S_t=1)\Pr(S_t=1|\psi_t)}{\sum_{j=0}^{1}\Pr(S_{t+1}|S_t=j)\Pr(S_t=j|\psi_t)},$$

where $\Pr(S_{t+1}|S_t=j)$ denotes the transition probability and $\Pr(S_t=j|\psi_t)$ is saved from step 3.

[6]The general case of a two-state Markov switching process is considered and the index i is dropped for a simplified notation.

APPENDIX B: TABLES

TABLE B1 Forecasting performance for one- and two-quarter PCE-deflator inflation

		$h=1$			$h=2$		
Model	**Variance**	**RMSFE**	**MAFE**	**PL**	**RMSFE**	**MAFE**	**PL**
MDS uniform	Constant	0.34*	0.22	4.05	0.70	0.47	3.51
MDS uniform	Stochastic	0.34*	0.22	4.09	0.69	0.46	3.57
MDS sparse	Constant	0.33**	0.21**	4.08	0.69	0.45	3.54
MDS sparse	Stochastic	0.33**	0.21**	4.07	0.66	0.43*	3.66
MDS smooth	Constant	0.34*	0.22	4.09	0.75	0.48	3.54
MDS smooth	Stochastic	0.33*	0.22	4.19	0.74	0.49	3.56
Bernoulli	Constant	0.35	0.24	3.25	0.75	0.52	1.97
Bernoulli	Stochastic	0.35	0.24	3.87	0.77	0.53	2.64
MS-Bernoulli	Constant	0.35	0.24	3.61	0.84	0.55	3.23
MS-Bernoulli	Stochastic	0.35	0.24	3.97	0.82	0.54	3.32
TVP-Bernoulli	Constant	0.49	0.35	2.41	0.89	0.64	2.20
TVP-Bernoulli	Stochastic	0.50	0.28	3.89	0.94	0.54	3.44
AR(1)	Constant	0.37	0.24	3.44	0.73	0.47	2.95
AR(1)	Stochastic	0.37	0.24	3.54	0.73	0.47	3.04
TVP-AR(1)	Constant	0.47	0.28	3.90	0.99	0.56	3.04
TVP-AR(1)	Stochastic	0.47	0.27	3.95	0.95	0.54	3.26
UCSV	Stochastic	0.71	0.51	1.52	1.07	0.70	1.36
Full model	Constant	0.37	0.24	2.70	0.78	0.53	1.53
Full model	Stochastic	0.37	0.24	2.98	0.78	0.53	1.93
LASSO	Constant	0.36	0.24	3.13	3.77	0.53	2.25
LASSO	Stochastic	0.36	0.24	3.60	3.77	0.52	2.87
TVP-LASSO	Constant	1.56	1.01	2.14	2.40	1.57	1.83
TVP-LASSO	Stochastic	0.42	0.29	3.50	0.80	0.58	3.16
DMA (0.95)	Stochastic	0.35	0.23	3.91	0.68	0.46	3.47
DMA (0.99)	Stochastic	0.35*	0.22**	4.03	0.70	0.44	3.44
DMS (0.95)	Stochastic	0.37	0.25	4.08	0.75	0.50	3.58
DMS (0.99)	Stochastic	0.35*	0.22*	4.05	0.67*	0.44*	3.48

Note. The table shows the RMSFE and MAFE in percentage points and the mean log predictive likelihood (PL). In addition, the DM test calculates the statistic for the null hypotheses of equal squared and absolute forecast errors against an AR(1) benchmark model. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

TABLE B2 Forecasting performance for three- and four-quarter PCE-deflator inflation

		$h = 3$			$h = 4$		
Model	**Variance**	**RMSFE**	**MAFE**	**PL**	**RMSFE**	**MAFE**	**PL**
MDS uniform	Constant	0.99**	0.71***	3.11	1.25	0.99	2.91
MDS uniform	Stochastic	0.97**	0.71***	3.13	1.17	0.92	2.98
MDS sparse	Constant	0.98**	0.69***	2.96	1.19	0.90	2.84
MDS sparse	Stochastic	0.93**	0.64***	3.18	1.10	0.83	3.04
MDS smooth	Constant	1.01*	0.73***	3.01	1.15	0.90	2.87
MDS smooth	Stochastic	1.00**	0.73***	3.10	1.14	0.90	2.93
Bernoulli	Constant	1.05	0.78	1.69	1.35	1.08	2.44
Bernoulli	Stochastic	1.04	0.77	1.92	1.36	1.08	2.52
MS-Bernoulli	Constant	1.11	0.79	1.93	1.32	0.94	2.59
MS-Bernoulli	Stochastic	1.14	0.80	2.04	1.41	1.07	2.63
TVP-Bernoulli	Constant	1.12	0.76	2.62	1.37	1.04	2.47
TVP-Bernoulli	Stochastic	0.91**	0.68***	3.29	1.29	0.88	3.01
AR(1)	Constant	1.03	0.69	2.81	1.35	0.95	2.63
AR(1)	Stochastic	1.03	0.69	2.78	1.35	0.95	2.66
TVP-AR(1)	Constant	0.94**	0.66***	2.74	1.31	0.90	2.67
TVP-AR(1)	Stochastic	0.92**	0.66***	2.94	1.28	0.89	2.69
UCSV	Stochastic	1.07	0.85	1.33	2.09	1.40	1.16
Full model	Constant	1.08	0.82	2.03	1.40	1.74	2.27
Full model	Stochastic	1.08	0.80	2.03	1.39	1.73	2.19
LASSO	Constant	1.08	0.81	3.00	1.39	1.12	2.55
LASSO	Stochastic	1.07	0.79	3.05	1.34	1.10	2.88
TVP-LASSO	Constant	2.49	1.71	2.05	2.68	1.95	1.68
TVP-LASSO	Stochastic	0.99	0.72	2.61	1.42	1.09	2.73
DMA (0.95)	Stochastic	0.93**	0.67***	3.20	1.06*	0.81*	2.88
DMA (0.99)	Stochastic	0.96**	0.64***	3.12	1.18	0.86	2.82
DMS (0.95)	Stochastic	0.97**	0.71***	3.28	1.18	0.92	2.96
DMS (0.99)	Stochastic	1.00**	0.66***	3.10	1.24	0.91	2.80

Note. The table shows the RMSFE and MAFE in percentage points and the mean log predictive likelihood (PL). In addition, the DM test calculates the statistic for the null hypotheses of equal squared and absolute forecast errors against an AR(1) benchmark model. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

TABLE B3 Forecasting performance for one- and two-quarter GDP-deflator inflation

		$h = 1$			$h = 2$		
Model	**Variance**	**RMSFE**	**MAFE**	**PL**	**RMSFE**	**MAFE**	**PL**
MDS uniform	Constant	0.19*	0.15*	4.82	0.35	0.27	4.23
MDS uniform	Stochastic	0.19*	0.15*	4.83	0.34*	0.27	4.24
MDS sparse	Constant	0.20	0.15	4.81	0.36	0.27	4.21
MDS sparse	Stochastic	0.20	0.15	4.82	0.35	0.28	4.21
MDS smooth	Constant	0.19*	0.15*	4.86	0.35*	0.27	4.25
MDS smooth	Stochastic	0.19*	0.15*	4.84	0.35	0.27	4.25
Bernoulli	Constant	0.19	0.16	4.82	0.38	0.30	4.15
Bernoulli	Stochastic	0.19	0.15	4.83	0.38	0.30	4.15
MS-Bernoulli	Constant	0.19	0.15	4.81	0.36	0.29	4.17
MS-Bernoulli	Stochastic	0.19	0.15	4.82	0.37	0.29	4.17
TVP-Bernoulli	Constant	0.29	0.21	2.43	0.40	0.32	2.52
TVP-Bernoulli	stochastic	0.21	0.15	4.24	0.40	0.29	4.01
AR(1)	Constant	0.21	0.16	4.74	0.39	0.29	4.10
AR(1)	Stochastic	0.21	0.16	4.74	0.38	0.29	4.10
TVP-AR(1)	Constant	0.22	0.16	4.47	0.40	0.29	4.05
TVP-AR(1)	Stochastic	0.21	0.15	4.36	0.38	0.28	4.07
UCSV	Stochastic	0.30	0.23	2.03	0.40	0.29	1.93
Full model	Constant	0.20	0.16	4.78	0.38	0.30	4.11
Full model	Stochastic	0.20	0.16	4.81	0.38	0.30	4.13
LASSO	Constant	0.20	0.16	4.79	0.38	0.30	4.14
LASSO	Stochastic	0.20	0.16	4.65	0.35	0.28	4.21
TVP-LASSO	Constant	1.65	1.11	2.04	1.85	1.27	1.97
TVP-LASSO	Stochastic	0.28	0.20	3.54	0.58	0.34	3.36
DMA (0.95)	Stochastic	0.18**	0.14***	4.34	0.32***	0.24***	3.85
DMA (0.99)	Stochastic	0.19**	0.15**	4.37	0.35***	0.26***	4.00
DMS (0.95)	Stochastic	0.20	0.15	4.34	0.34**	0.26**	4.01
DMS (0.99)	Stochastic	0.20**	0.15**	4.37	0.35***	0.27***	4.01

Note. The table shows the RMSFE and MAFE in percentage points and the mean log predictive likelihood (PL). In addition, the DM test calculates the statistic for the null hypotheses of equal squared and absolute forecast errors against an AR(1) benchmark model. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

TABLE B4 Forecasting performance for three- and four-quarter GDP-deflator inflation

		$h = 3$			$h = 4$		
Model	**Variance**	**RMSFE**	**MAFE**	**PL**	**RMSFE**	**MAFE**	**PL**
MDS uniform	Constant	0.53	0.42	3.77	0.69	0.54	3.56
MDS uniform	Stochastic	0.51	0.39	3.83	0.69	0.54	3.54
MDS sparse	Constant	0.52	0.39	3.73	0.63	0.52	3.63
MDS sparse	Stochastic	0.53	0.40	3.77	0.65	0.52	3.60
MDS smooth	Constant	0.48*	0.38*	3.90	0.65	0.49	3.51
MDS smooth	Stochastic	0.48*	0.38*	3.91	0.59	0.48	3.73
Bernoulli	Constant	0.60	0.47	3.47	0.88	0.69	3.13
Bernoulli	Stochastic	0.60	0.47	3.59	0.88	0.70	3.14
MS-Bernoulli	Constant	0.59	0.47	3.38	0.83	0.67	3.06
MS-Bernoulli	Stochastic	0.58	0.46	3.61	0.83	0.67	3.19
TVP-Bernoulli	Constant	0.54	0.43	2.16	2.39	1.60	1.94
TVP-Bernoulli	Stochastic	0.54	0.40	3.82	0.72	0.55	3.53
AR(1)	Constant	0.54	0.40	3.73	0.66	0.55	3.50
AR(1)	Stochastic	0.54	0.40	3.73	0.66	0.50	3.48
TVP-AR(1)	Constant	0.50	0.38	3.81	0.68	0.56	2.19
TVP-AR(1)	Stochastic	0.50	0.38	3.80	0.66	0.52	3.49
UCSV	Stochastic	0.52	0.39	2.11	0.67	0.52	1.80
Full model	Constant	0.61	0.48	3.50	0.89	0.71	3.06
Full model	Stochastic	0.62	0.49	3.51	0.89	0.71	2.18
LASSO	Constant	0.61	0.48	3.65	0.88	0.70	3.24
LASSO	Stochastic	0.55	0.44	3.75	0.78	0.62	3.45
TVP-LASSO	Constant	1.01	0.80	2.16	1.99	1.45	1.82
TVP-LASSO	Stochastic	0.57	0.38	2.17	0.72	0.55	3.04
DMA (0.95)	Stochastic	0.42**	0.32**	3.59	0.53**	0.43**	3.36
DMA (0.99)	Stochastic	0.51	0.38	3.68	0.75	0.56	3.39
DMS (0.95)	Stochastic	0.53	0.39	3.68	0.71	0.52	3.47
DMS (0.99)	Stochastic	0.54	0.39	3.67	0.81	0.60	3.35

Note. The table shows the RMSFE and MAFE in percentage points and the mean log predictive likelihood (PL). In addition, the DM test calculates the statistic for the null hypotheses of equal squared and absolute forecast errors against an AR(1) benchmark model. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

TABLE B5 Forecasting performance for one- and two-quarter CPI inflation

Model	Variance	$h = 1$			$h = 2$		
		RMSFE	MAFE	PL	RMSFE	MAFE	PL
MDS uniform	Constant	0.68	0.43	3.47	1.14*	0.75	3.00
MDS uniform	Stochastic	0.69	0.44	3.52	1.15	0.76	3.00
MDS sparse	Constant	0.65**	0.41*	3.53	1.15	0.75*	2.96
MDS sparse	Stochastic	0.65***	0.41**	3.57	1.14***	0.73*	3.05
MDS smooth	Constant	0.71	0.45	3.43	1.15*	0.75	2.96
MDS smooth	Stochastic	0.71	0.46	3.49	1.14***	0.73*	3.05
Bernoulli	Constant	0.73	0.47	3.34	1.28	0.83	2.69
Bernoulli	Stochastic	0.73	0.47	3.41	1.27	0.83	2.70
MS-Bernoulli	Constant	0.72	0.47	3.37	1.30	0.83	2.75
MS-Bernoulli	Stochastic	0.71	0.46	3.42	1.27	0.82	2.78
TVP-Bernoulli	Constant	1.30	0.75	2.55	1.76	1.15	2.26
TVP-Bernoulli	Stochastic	1.22	0.55	3.47	1.54	0.86	3.01
AR(1)	Constant	0.74	0.24	2.84	1.23	0.79	2.84
AR(1)	Stochastic	0.75	0.48	3.42	1.21	0.76	2.87
TVP-AR(1)	Constant	1.23	0.55	3.43	2.00	0.86	2.85
TVP-AR(1)	Stochastic	1.16	0.51	3.45	1.95	0.84	2.92
UCSV	Stochastic	0.78	0.53	1.59	1.09***	0.69***	1.43
Full model	Constant	0.74	0.48	3.32	1.31	0.85	2.60
Full model	Stochastic	0.74	0.48	1.30	1.30	0.84	2.67
LASSO	Constant	0.74	0.48	3.44	1.31	0.85	2.84
LASSO	Stochastic	0.74	0.47	3.48	1.15	0.80	2.93
TVP-LASSO	Constant	1.91	1.31	2.04	2.40	1.60	1.85
TVP-LASSO	Stochastic	1.21	0.60	3.26	1.15	0.80	2.20
DMA (0.95)	Stochastic	0.71	0.44	3.46	1.10**	0.71*	3.07
DMA (0.99)	Stochastic	0.71	0.42	3.43	1.09***	0.64**	2.96
DMS (0.95)	Stochastic	0.76	0.47	3.53	1.14*	0.72*	3.15
DMS (0.99)	Stochastic	0.74	0.44	3.43	1.11**	0.67*	2.96

Note. The table shows the RMSFE and MAFE in percentage points and the mean log predictive likelihood (PL). In addition, the DM test calculates the statistic for the null hypotheses of equal squared and absolute forecast errors against an AR(1) benchmark model. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

TABLE B6 Forecasting performance for three- and four-quarter CPI inflation

		$h = 3$			$h = 4$		
Model	**Variance**	**RMSFE**	**MAFE**	**PL**	**RMSFE**	**MAFE**	**PL**
MDS uniform	Constant	1.46	1.10	2.68	1.67*	1.35	2.46
MDS uniform	Stochastic	1.42*	1.07	2.72	1.66*	0.92	2.98
MDS sparse	Constant	1.39**	0.98**	2.61	1.50*	1.19*	2.42
MDS sparse	Stochastic	1.39**	0.97**	2.73	1.48**	1.19**	2.52
MDS smooth	Constant	1.42*	1.06	2.63	1.59*	1.30	2.38
MDS smooth	Stochastic	1.44	1.09	2.63	1.57*	1.28	2.45
Bernoulli	Constant	1.59	1.19	2.38	1.95	1.62	2.24
Bernoulli	Stochastic	1.58	1.18	2.39	1.96	1.63	2.25
MS-Bernoulli	Constant	1.62	1.18	2.36	1.88	1.53	2.18
MS-Bernoulli	Stochastic	1.65	1.20	2.40	1.88	1.53	2.19
TVP-Bernoulli	Constant	2.36	1.38	1.84	2.39	1.60	1.94
TVP-Bernoulli	Stochastic	1.26	0.89	2.87	1.90	1.40	1.93
AR(1)	Constant	1.67	1.08	2.62	2.09	1.45	2.41
AR(1)	Stochastic	1.57	1.04	2.68	1.94	1.37	2.48
TVP-AR(1)	Constant	1.34**	0.91**	2.72	1.97	1.30	2.62
TVP-AR(1)	Stochastic	1.34**	0.91**	2.75	1.98	1.37	2.83
UCSV	Stochastic	1.41	0.96	1.42	1.52	1.11	1.32
Full model	Constant	1.64	1.22	2.35	2.01	1.67	2.17
Full model	Stochastic	1.64	1.22	2.34	2.01	1.67	2.18
LASSO	Constant	1.63	1.22	2.55	1.98	1.65	2.39
LASSO	Stochastic	1.60	1.18	2.64	1.93	1.58	2.44
TVP-LASSO	Constant	2.48	1.67	1.47	3.56	2.43	1.58
TVP-LASSO	Stochastic	1.62	1.08	2.76	1.71	1.30	2.57
DMA (0.95)	Stochastic	1.39**	0.99**	2.89	1.43**	1.11**	2.88
DMA (0.99)	Stochastic	1.34**	0.89**	2.79	1.68*	1.24*	2.62
DMS (0.95)	Stochastic	1.61	1.06	2.94	1.59**	1.18**	2.79
DMS (0.99)	Stochastic	1.41*	0.94*	2.79	1.72*	1.28	2.62

Note. The table shows the RMSFE and MAFE in percentage points and the mean log predictive likelihood (PL). In addition, the DM test calculates the statistic for the null hypotheses of equal squared and absolute forecast errors against an AR(1) benchmark model. Asterisks (*10%, **5%, ***1%) denote the level of significance at which the null hypotheses are rejected.

APPENDIX C: FIGURES

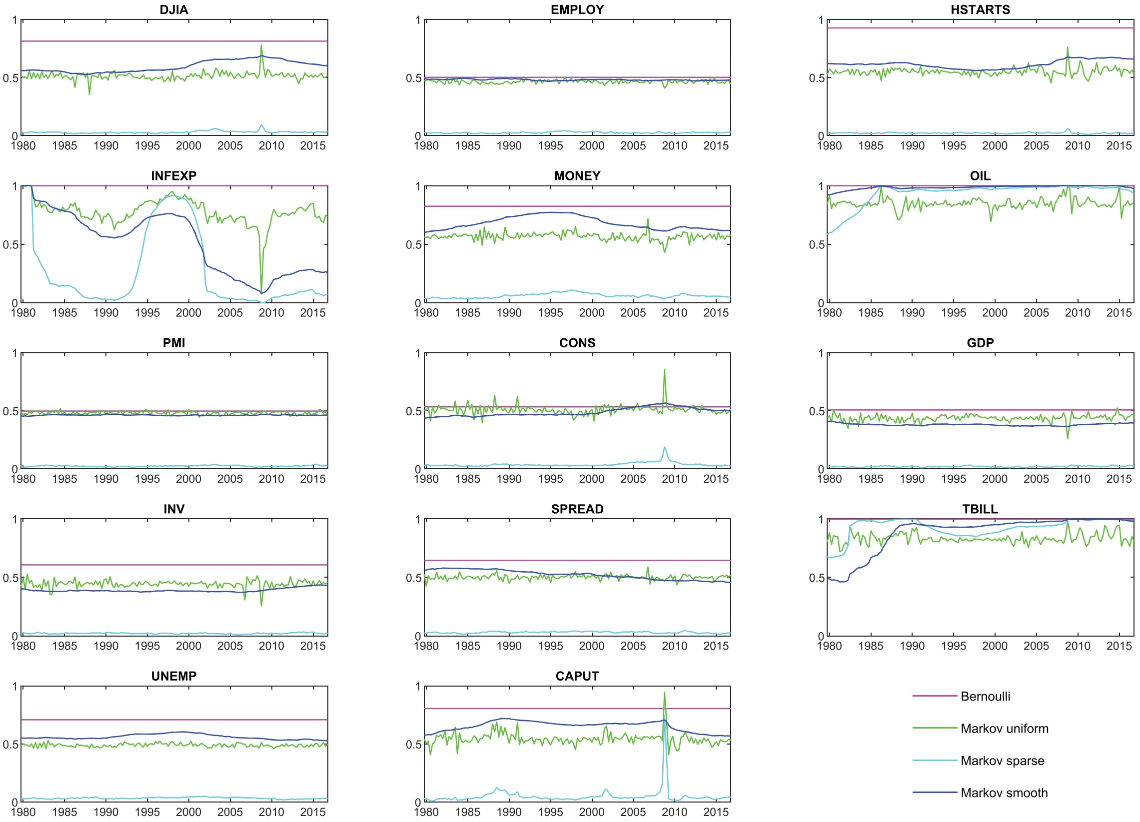

FIGURE C1 Variable inclusion probabilities for one-quarter PCE deflator inflation [Colour figure can be viewed at wileyonlinelibrary.com]

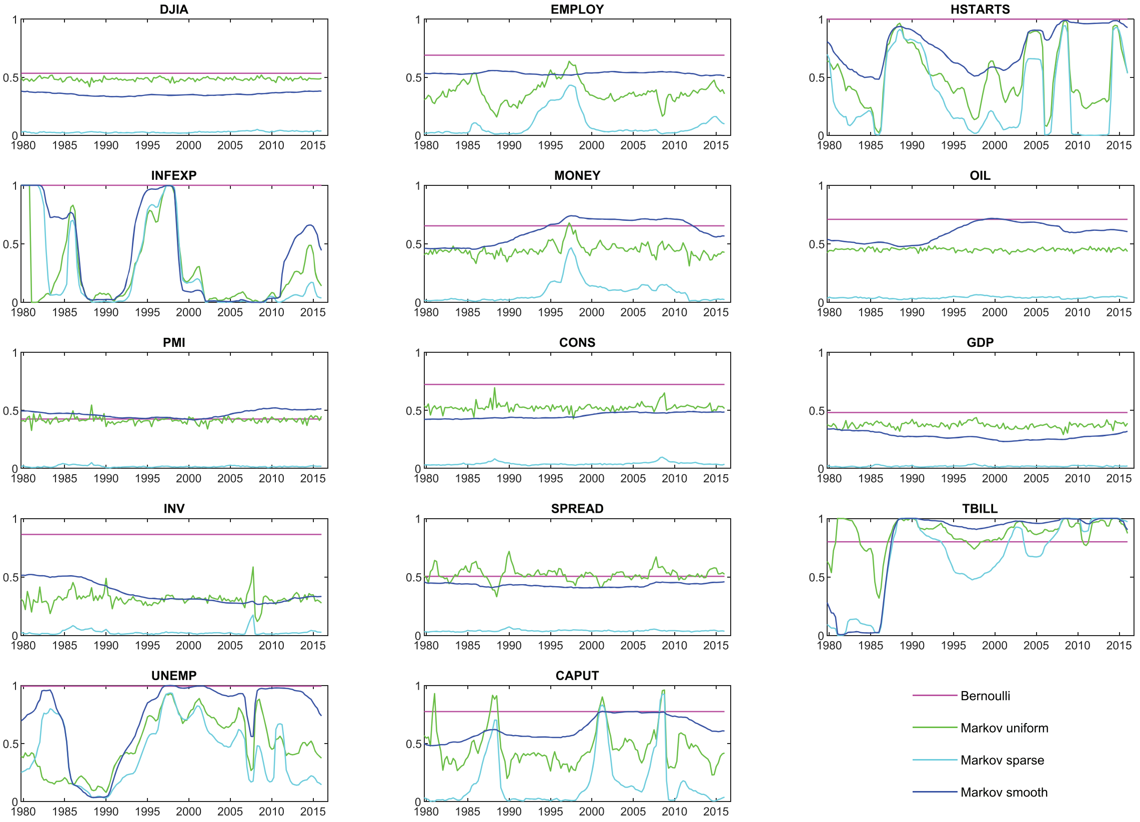

FIGURE C2 Variable inclusion probabilities for one-year PCE-deflator inflation [Colour figure can be viewed at wileyonlinelibrary.com]

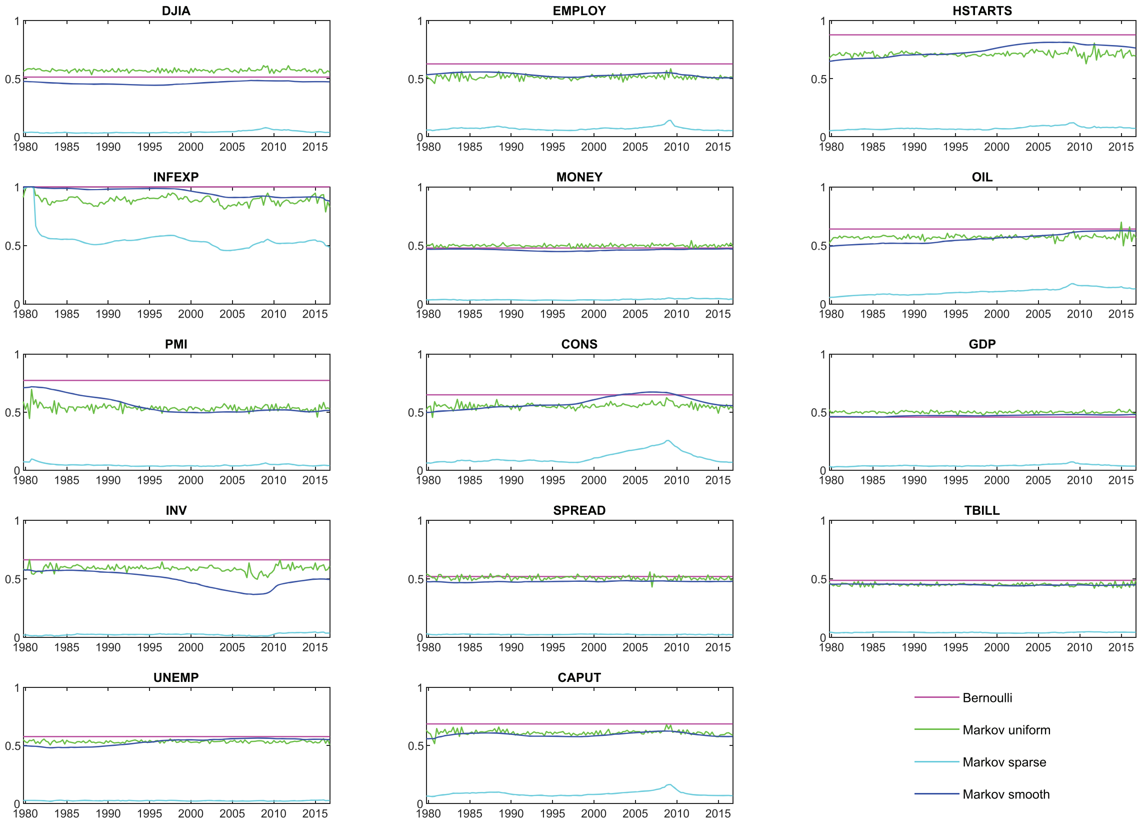

FIGURE C3 Variable inclusion probabilities for one-quarter GDP-deflator inflation [Colour figure can be viewed at wileyonlinelibrary.com]

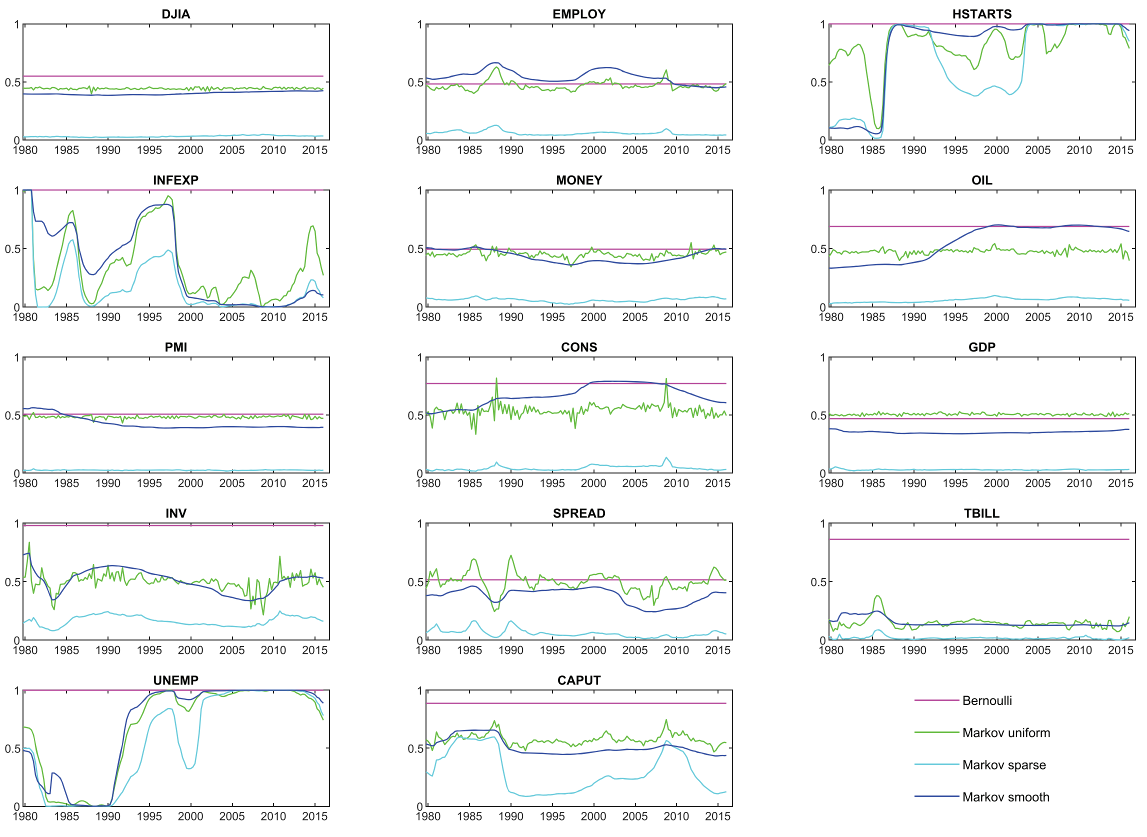

FIGURE C4 Variable inclusion probabilities for one-year GDP-deflator inflation [Colour figure can be viewed at wileyonlinelibrary.com]

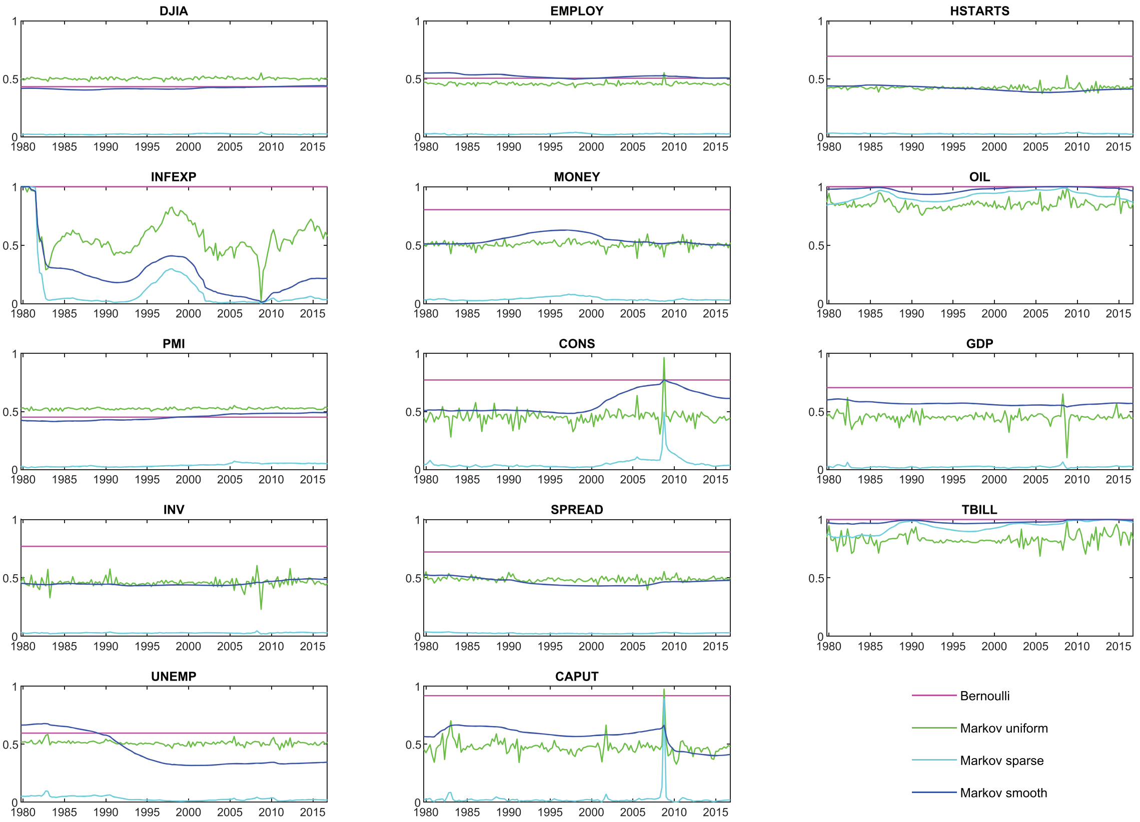

FIGURE C5 Variable inclusion probabilities for one-quarter CPI inflation [Colour figure can be viewed at wileyonlinelibrary.com]

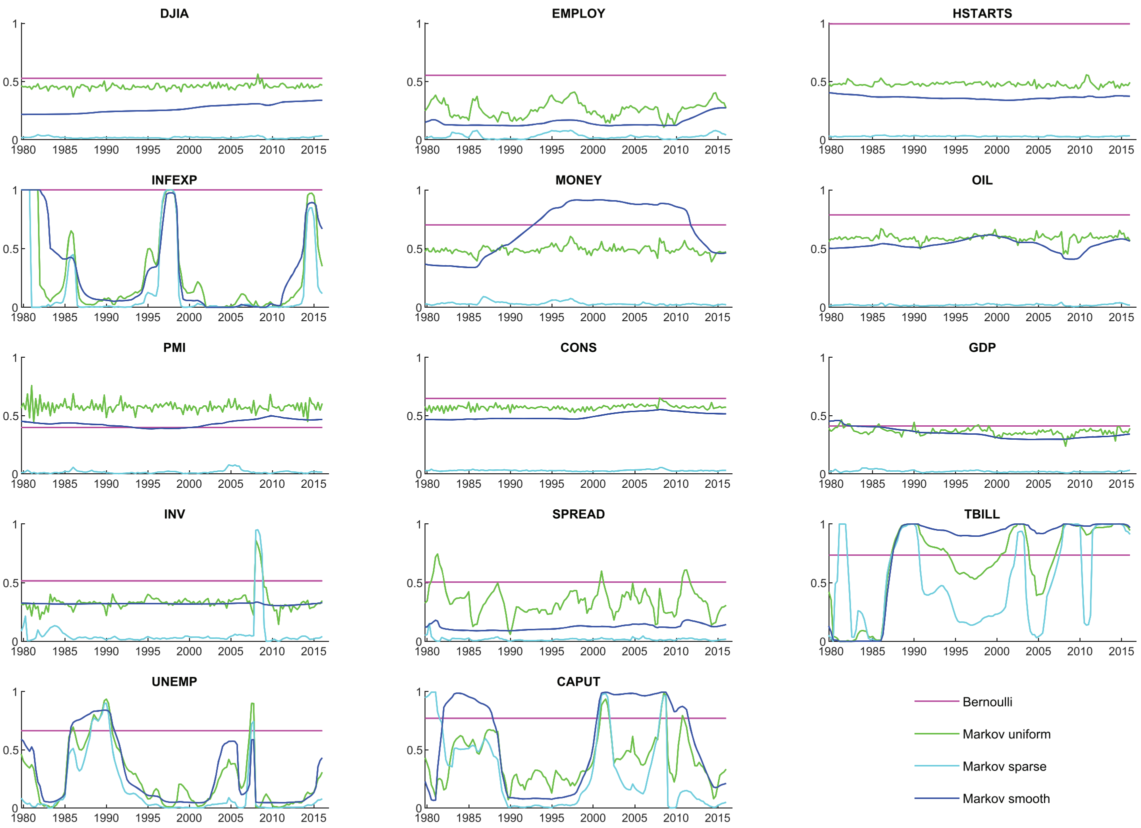

FIGURE C6 Variable inclusion probabilities for one-year CPI inflation [Colour figure can be viewed at wileyonlinelibrary.com]