

# Forecasting US Inflation using Markov Dimension Switching

Jan Prüser<sup>a,b</sup>

<sup>a</sup>University of Duisburg-Essen\*    <sup>b</sup>Ruhr Graduate School in Economics<sup>†</sup>

May 16, 2018

This study considers Bayesian variable selection in the Phillips curve context by using the Bernoulli approach of Korobilis (2013a). 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. The MDS and Bernoulli model reveal that inflation expectations, the growth rate of the oil price and the Treasury bill rate are the most important variables for one quarter inflation. For one year inflation the unemployment rate, inflation expectations, the Treasury bill rate and the number of newly built houses turn out to be the most important variables. Furthermore, 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 a forecasting exercise the MDS model compares favorably to the Bernoulli model for one quarter and one year ahead inflation.

**Keywords:** Phillips Curve, Fat Data, Variable Selection, Model Change

**JEL classification:** C11, C32, C53, E37

---

\*Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany.

<sup>†</sup>RWI - Leibniz Institute for Economic Research, Hohenzollernstrasse 1-3, D-45128 Essen, Germany, e-mail address: [jan.prueser@rgs-econ.de](mailto:jan.prueser@rgs-econ.de).

# 1. Introduction

The Phillips curve has served as an important tool in macroeconomics for explaining and forecasting inflation in the US 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), Atkeson and Ohanian (2001), Ang et al. (2007), Stock and Watson (2007) and Groen et al. (2013). Forecasting inflation is crucial, e.g., 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 (2013a) provides 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, *inter alia*, Stock and Watson (2010), who find that most predictors for inflation improve 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 to allow for this, as they only test whether a restriction holds for all time periods or never.<sup>1</sup> 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

---

<sup>1</sup>Furthermore, the Bayesian methods used in this paper have the advantage that they allow for a formal treatment of model uncertainty. Using hypothesis tests to select a parsimonious model ignores model uncertainty, as the selected model is assumed to be the one which generated the data. Treating one model as if it were the “true” model and ignoring the huge number of other potential models may be seen as problematic.

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 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 and the number of newly built houses turn out to be the most important predictors for one year inflation. The relevant variables show a sizable degree of time variation, which the Bernoulli approach can not 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 and house prices for the one year horizon changes abrupt over time, which would be difficult to capture for existing methods which assume a gradually change of the relevance of predictors. From an economic perspective it is particular interesting that the relevance of unemployment rate changes that rapidly as it has long been assumed that economic policymakers face a trade-off between unemployment and inflation. This result however suggests that this inverse relation might not be stable over time and that a break down of the Phillips curve may only be temporary. Furthermore, this paper investigates the forecasting performance of both approaches. It turns out that the MDS approach exhibits a better forecasting performance than the Bernoulli approach for one quarter and one year inflation. An additional finding is that the forecasting performance of the MDS approach is competitive in comparison with a range of other plausible approaches.

The remainder of this paper is organized as follows. Section 2 lays out and discusses the econometric framework. Section 3 presents the empirical findings and the last section concludes.

## 2. 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

$$\pi_{t+h} = \alpha + \sum_{j=0}^{p-1} \phi_j \pi_{t-j} + \mathbf{x}_t \boldsymbol{\beta} + \epsilon_{t+h}, \quad (1)$$

where  $\mathbf{x}_t$  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). Both, imprecise estimation and overfitting translate into inaccurate out-of-sample predictions. 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

$$\pi_{t+h} = (\mathbf{z}_t \odot \boldsymbol{\gamma})\boldsymbol{\theta} + \epsilon_{t+h}, \quad (2)$$

where  $\mathbf{z}_t = (1, \pi_t, \dots, \pi_{t-p+1}, \mathbf{x}_t)$ ,  $\boldsymbol{\theta} = (\alpha, \phi_0, \dots, \phi_{p-1}, \boldsymbol{\beta}')$  and  $\odot$  denotes elementwise multiplication. Hence, if  $\gamma_i = 1$ , the  $i$ th 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  $\gamma_i$  follows a first-order Markov-switching process  $S_{i,t}$  and therefore  $\boldsymbol{\gamma}$  now has a time index  $t$ :

$$\pi_{t+h} = (\mathbf{z}_t \odot \boldsymbol{\gamma}_t)\boldsymbol{\theta} + \epsilon_{t+h}, \quad (3)$$

where  $\boldsymbol{\gamma}_t = (S_{1,t}, \dots, S_{m,t})$ . Each Markov switching process  $S_{i,t}$  can take on the value one or zero and is characterized by a  $2 \times 2$  transition matrix  $\boldsymbol{\mu}_i$ , where  $\mu_{kj,i} = \Pr(S_{i,t+1} = j | S_{i,t} = k)$ ,  $k = 0, 1$  and  $j = 0, 1$ .<sup>2</sup> If  $S_{i,t} = 1$ , the  $i$ th variable is included in the model at period  $t$  and if  $S_{i,t} = 0$ , it is not. Therefore, the means of the posterior draws of  $S_{i,t}$  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 bears a high risk of overfitting and can empirically only poorly approximate changing predictors by allowing coefficients to be estimated as

---

<sup>2</sup>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 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.

being approximately zero. Furthermore, models with time-varying parameters typically assume a gradually change in parameters and therefore are not well suited to capture abrupt changes in the relevance of predictors.

## 2.1. Gibbs Sampler

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

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

$$\boldsymbol{\theta} | \boldsymbol{\gamma}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}_{1+h:T+h}, \sigma_{1+h:T+h}^2 \sim N(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\Omega}}), \quad (4)$$

with

$$\begin{aligned} \bar{\boldsymbol{\theta}} &= \bar{\boldsymbol{\Omega}} \left( \mathbf{V}(\hat{\boldsymbol{\theta}}_{OLS}) \hat{\boldsymbol{\theta}}_{OLS} + \sum_{t=1}^T (\mathbf{z}_t \odot \boldsymbol{\gamma}_t)' \sigma_{t+h}^{-2} \boldsymbol{\pi}_{t+h} \right), \\ \bar{\boldsymbol{\Omega}} &= \left( \mathbf{V}(\hat{\boldsymbol{\theta}}_{OLS}) + \sum_{t=1}^T (\mathbf{z}_t \odot \boldsymbol{\gamma}_t)' \sigma_{t+h}^{-2} (\mathbf{z}_t \odot \boldsymbol{\gamma}_t) \right)^{-1}. \end{aligned}$$

For the prior, the OLS estimate of the full model is used. When one variable is omitted from the model for the full sample period, the parameter of this predictor is drawn from the prior. 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 trade-off in a convenient way by placing weights on both estimates in a data-driven way.

2. Sample  $\boldsymbol{\gamma}_t$ :

- If  $\gamma_i$  is constant, sample it from

$$\gamma_i | \boldsymbol{\gamma}_{-i}, \boldsymbol{\pi}_{1+h:T+h}, \mathbf{z}_{1:T}, \boldsymbol{\theta}, \sigma_{1+h:T+h}^2 \sim \text{Bernoulli} \left( \frac{l_{1i}}{l_{1i} + l_{0i}} \right), \quad (5)$$

with

$$l_{1i} = \exp \left( -\frac{1}{2} \sum_{t=1}^T \left( \frac{\pi_{t+h} - (\mathbf{z}_t \odot \boldsymbol{\gamma}_{[\gamma_i=1]}) \boldsymbol{\theta}}{\sigma_{t+h}^2} \right)^2 \right) p(\gamma_i = 1),$$

$$l_{0i} = \exp \left( -\frac{1}{2} \sum_{t=1}^T \left( \frac{\pi_{t+h} - (\mathbf{z}_t \odot \boldsymbol{\gamma}_{[\gamma_i=0]}) \boldsymbol{\theta}}{\sigma_{t+h}^2} \right)^2 \right) p(\gamma_i = 0),$$

where  $p(\gamma_i = 1) = 0.5$ .

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

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

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

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 is 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, i.e. a variable is excluded or included 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 setting only  $u_{00} = T$ . All three prior parametrizations, i.e. the uniform, the smooth and the sparse prior, are considered in the empirical part.

### 3. Sample $\sigma_t^{-2}$ :

- In the case of homoscedastic errors where  $\sigma_t^2 = \sigma^2$ , sample from the density

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

where  $a = T + a_0$  and  $b = b_0 + \sum_{t=1}^T (\pi_{t+h} - (\mathbf{z}_t \odot \boldsymbol{\gamma}_t) \boldsymbol{\theta})^2$ .

The hyperparameters  $a_0$  and  $b_0$  are set to zero.

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

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

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

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

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  $\kappa_1$  and  $\kappa_2$  are set to 3 and 0.0001.

## 2.2. 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 et al. (2009) find that shrinking parameters leads to improved forecasts in large 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 which are estimated from the data, resulting in data-driven shrinkage. For example, Korobilis (2013b) shows that hierarchical shrinkage is useful for macroeconomic forecasting using many predictors. In a Phillips curve context, Belmonte et al. (2014) use the lasso prior in a time-varying parameter (TVP) model. The lasso prior in their model automatically decides which parameter is 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 et al. (2012) consider dimension switching in a TVP framework using the algorithm of Gerlach et al. (2000). However, in their forecasting study, they only consider models with no predictors, a single predictor or all  $m$  predictors. In other words,  $\gamma$  can only take on  $m + 2$  values and not  $2^m$  as this would be computationally infeasible for the algorithm they used. To consider all variable combinations, dynamic model averaging (DMA) can be applied, using approximations in form of so called forgetting factors (sometimes also called discount factors) as proposed by Raftery et al. (2010). Koop and Korobilis (2012) find that DMA leads to substantial improvements in forecasting inflation over simple benchmark models and more sophisticated approaches. DMA assigns time-varying weights over the set of  $2^m$  possible TVP models.<sup>3</sup>

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

---

<sup>3</sup>In the empirical application, only the intercept and the first lag are always included.

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 model 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. 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.

Moreover, in contrast to DMA, the MDS model does not need approximations. It can easily be estimated using Gibbs sampling and thereby take full parameter uncertainty into account. Another potential drawback is that in the DMA approach all model combinations have to be estimated in a deterministic fashion, while MDS uses a stochastic search algorithm. The stochastic search is still feasible when the model space is too large to be assessed in a deterministic manner by visiting only the most probable models in a stochastic manner. 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. Forecasting Inflation

### 3.1. Data

This study forecasts core inflation as measured by the Personal Consumption Expenditure (PCE) deflator for 1978Q2 through 2016Q4. The period from 1992Q1 to 2016Q4 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 which 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 ten year and three month Treasury bill rates.
- TBILL: three 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.

### 3.2. 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. In order to assess whether the MDS or the Bernoulli approach is useful to avoid overfitting, their forecast performance is compared with an AR(1) model with intercept and a multiple regression model containing all variables. Furthermore, the uniform, the smooth and the sparse prior for the transition probabilities are compared. All these models are applied with a constant and a stochastic variance specification as described in the description of the Gibbs Sampler.

In a second step the forecasting performance of the MDS model is compared with two modeling approaches which have been found useful in inflation forecasting. These approaches are DMA proposed by Koop and Korobilis (2012) and the hierarchical shrinkage

in TVP-models proposed by Belmonte et al. (2014). 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 and Korobilis (2012) for details). Setting these forgetting factors to one leads to the special case of constant coefficients, constant variance and a 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 they find 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 non-informative 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, the last two models are estimated using the same two specifications for the variance as for the MDS models.

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. However, these 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  $\pi_{t+h}$  (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 (2010).

Table 1 contains the results for the one quarter and one year ahead forecasting performance. Overall, it turns out that the MDS models forecast quite well. For one quarter ahead inflation the forecasting performance of the Bernoulli model is similar to the forecasting performance of the AR(1) model and the full model containing all predictors. For one year ahead inflation the full model seems to overfit the data, as it forecasts poorly. 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. Fore-

Table 1: Forecasting performance for one quarter and one year inflation

Model	Variance	$h = 1$			$h = 4$		
		RMSFE	MAFE	PL	RMSFE	MAFE	PL
MDS uniform	constant	0.34	0.22	4.05	1.25	0.99	2.91
MDS uniform	stochastic	0.34	0.22	4.09	1.17	0.92	2.98
MDS sparse	constant	0.33	0.21	4.08	1.19	0.90	2.84
MDS sparse	stochastic	0.33	0.21	4.07	1.10	0.83	3.04
MDS smooth	constant	0.34	0.22	4.09	1.15	0.90	2.87
MDS smooth	stochastic	0.33	0.22	4.19	1.14	0.90	2.93
Bernoulli	constant	0.35	0.24	2.25	1.35	1.08	2.44
Bernoulli	stochastic	0.35	0.24	2.87	1.36	1.08	2.52
AR(1)	constant	0.37	0.24	2.84	1.35	0.95	2.63
AR(1)	stochastic	0.37	0.24	2.91	1.35	0.95	2.66
Full model	constant	0.37	0.24	0.70	1.40	1.14	2.27
Full model	stochastic	0.37	0.24	1.58	1.39	1.13	2.19
LASSO	constant	0.36	0.24	3.13	1.39	1.12	2.55
LASSO	stochastic	0.36	0.24	3.60	1.34	1.10	2.88
TVP-shrink	constant	1.56	1.01	2.14	2.68	1.95	1.68
TVP-shrink	stochastic	0.42	0.29	3.50	1.42	1.09	2.73
DMA (0.95)	stochastic	0.35	0.23	3.91	1.06	0.81	2.88
DMA (0.99)	stochastic	0.35	0.22	4.03	1.18	0.86	2.82
DMS (0.95)	stochastic	0.37	0.25	4.08	1.18	0.92	2.96
DMS (0.99)	stochastic	0.35	0.22	4.05	1.24	0.91	2.80

The table shows the RMSFE and MAFE in percentage points and the mean log predictive likelihood (PL).

casting 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 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 very similar forecasting performance for one quarter and one year ahead inflation. The specification of the variance turns out to be not crucial. An exception is the TVP regression model, which forecasts 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. Furthermore, the hierarchical shrinkage in TVP and constant coefficient regression produces less precise forecasts than the MDS models. Only the DMA and DMS (which also allows for changing predictors) approach shows a similar forecasting performance compared to the MDS models. This finding stresses the importance of allowing for changing predictors over time using the Phillips curve to forecast inflation.

### 3.3. 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 B.1 and B.2 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 inclusion probabilities are shown for 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 reveal 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, sometimes 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 inflation *INEXP*, *OIL* and *TBILL* turn out to be important in all approaches and for one year inflation *INEXP*, *HSTARTS*, *UNEMP* and *TBILL* turn out to be important. In particular for one year inflation these variables show a sizeable degree of time variation. In particular the inclusion probabilities of *INEXP*, *HSTARTS* and *UNEMP* switch very rapid over time. This shows that the relevance of predictors does not always change gradually, like it is assumed for example in TVP models. From an economic perspective it is particular interesting that the relevance of *UNEMP* changes that rapidly as it has long been assumed that economic policymakers face a trade-off between unemployment and inflation. These results however suggest that this relation might not be stable over time.

## 4. 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 which 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 Markov Dimension Switching (MDS) approach. In the MDS approach the set of predictors is allowed to change over time. The empirical application shows 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 the unemployment rate, the Treasury bill rate 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 sizeable 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 for one quarter and one year ahead inflation. In addition, the paper demonstrates that the forecasting performance of the MDS model is competitive 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.

## References

- Ang, A., Bekaert, G., and Wei, M. (2007). Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better? *Journal of Monetary Economics*, 54(4):1163–1212.
- Atkeson, A. and Ohanian, L. (2001). Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minneapolis*, 25(1).
- Bañbura, M., Goammpe, D., and Reichlin, L. (2009). Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics*, 25(1):71–92.
- Belmonte, M., Koop, G., and Korobilis, D. (2014). Hierarchical Shrinkage in Time-Varying Parameter Models. *Journal of Forecasting*, 33(1):80–94.
- Chan, J. C., Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2012). Time Varying Dimension Models. *Journal of Business & 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.
- Gerlach, R., Carter, C., and Kohn, R. (2000). Efficient Bayesian Inference in Dynamic Mixture Models. *Journal of American Statistical Association*, 95(451):819–828.
- Geweke, J. and Amisano, G. (2010). Hierarchical Markov Normal Mixture Models with Application to Financial Asset Returns. *Journal of Applied Econometrics*, 26(1):1–29.
- Groen, J., Paap, R., and Ravazzolo, F. (2013). Real-Time Inflation Forecasting in a Changing World. *Journal of Business & 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., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65(3):361–393.
- Koop, G. and 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 Prior for Dynamic Regressions with Many Predictors. *International Journal of Forecasting*, 29(1):43–59.
- Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of American Statistical Association*, 103(482):681–686.

- Raftery, A., Kárný, M., and Ettlér, P. (2010). Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics*, 52(1):52–66.
- Stock, J. and Watson, M. (1999). Forecasting Inflation. *Journal of Monetary Economics*, 44(2):293–335.
- Stock, J. and Watson, M. (2007). Why has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, 39(7):3–33.
- Stock, J. and Watson, M. (2010). Model Inflation After the Crisis. *National Bureau of Economic Research Working Paper*, (16488).

## Appendix A. Gibbs sampling in Markov switching models

This paper considers Markov switching for each variable. Each Markov switching process  $S_t$  can take on the value one or zero and is characterized by a  $2 \times 2$  transition matrix  $\boldsymbol{\mu}$  where  $\mu_{kj} = \Pr(S_{t+1} = j | S_t = k)$ ,  $k = 0, 1$  and  $j = 0, 1$ .<sup>4</sup> 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:

$$\Pr(S_0 = 0) = \frac{1 - \mu_{11}}{2 - \mu_{11} - \mu_{00}},$$

$$\Pr(S_0 = 1) = \frac{1 - \mu_{00}}{2 - \mu_{11} - \mu_{00}}.$$

2. Given  $\Pr(S_{t-1} = k | \psi_{t-1})$ , where  $\psi_{t-1}$  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  $\psi_t$  update the probabilities as

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

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

4. Sample  $S_T$  using  $\Pr(S_t = T | \psi_T)$ .
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.

---

<sup>4</sup>For a simplified notation the index  $i$  is omitted and the general case of a two state Markov process is considered.

# Appendix B. Figures

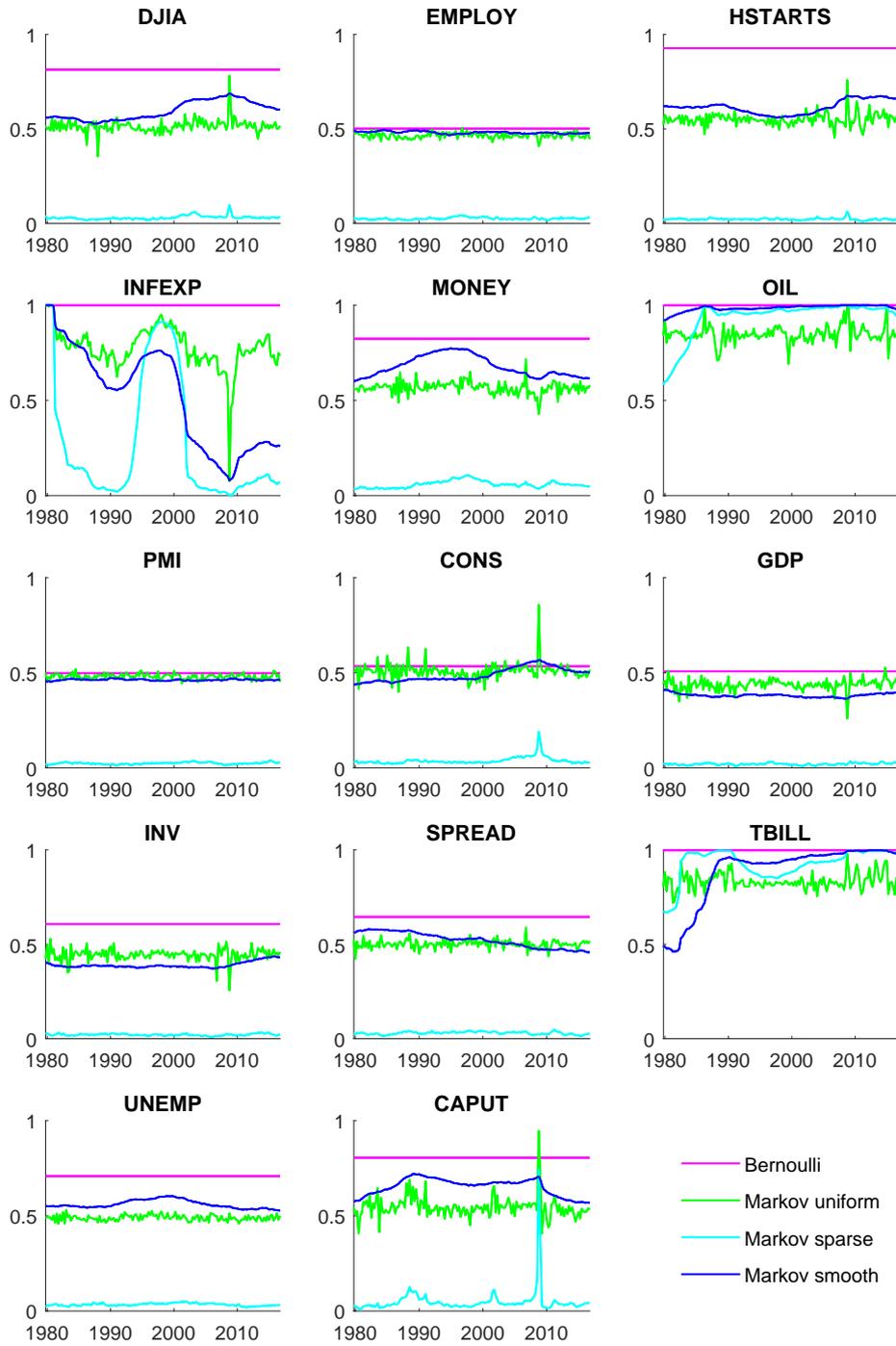


Figure B.1: Variable inclusion probabilities for one quarter inflation.

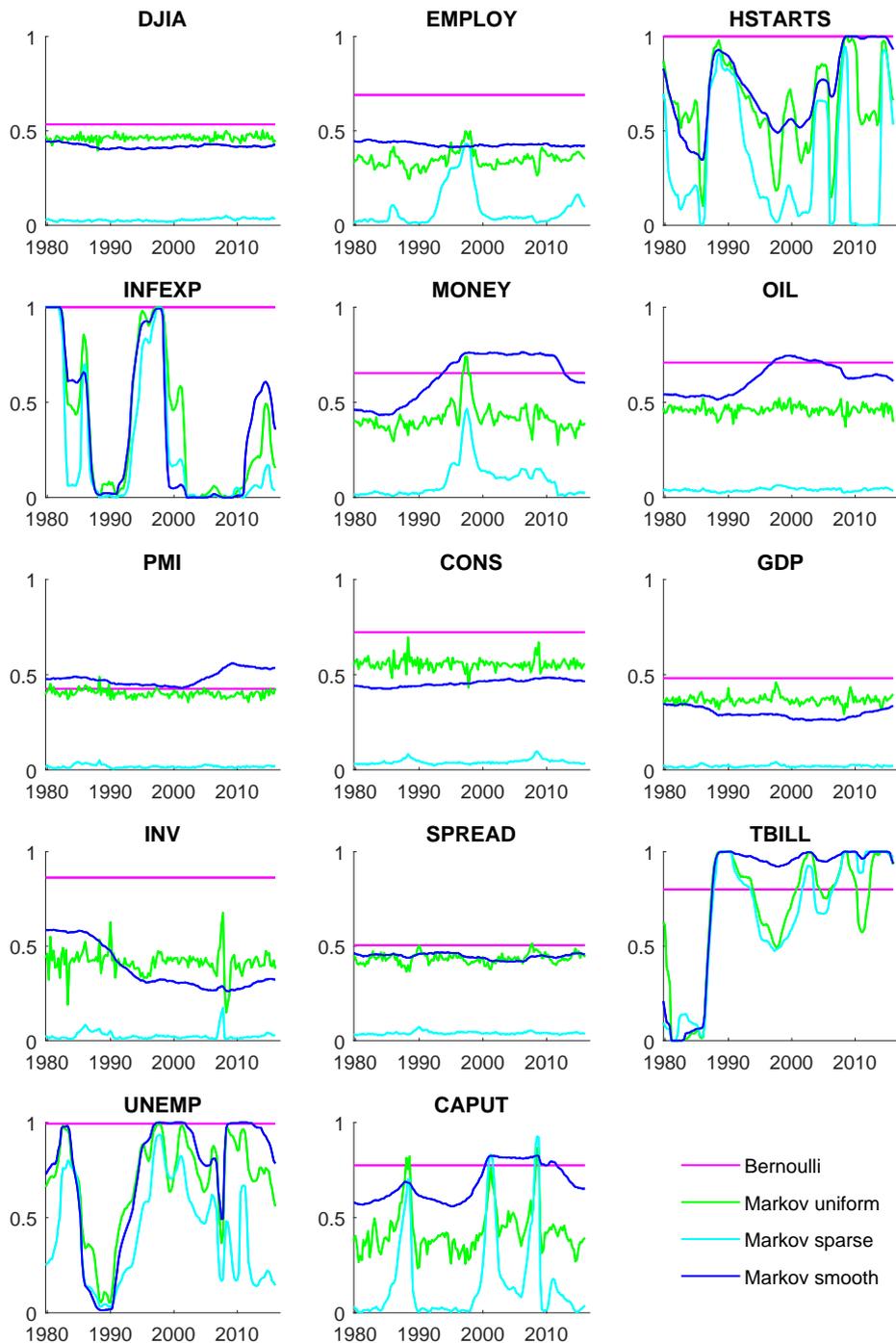


Figure B.2: Variable inclusion probabilities for one year inflation.