RUHR ECONOMIC PAPERS

Jan Prüser

Forecasting US inflation using Markov Dimension Switching

Ruhr Graduate School in Economics
UNIVERSITÄT DUISBURG ESSEN

#710

Imprint

Ruhr Economic Papers

Published by

RWI Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #710

Responsible Editor: Volker Clausen

ISSN 1864-4872 (online) – ISBN 978-3-86788-828-8

Ruhr Economic Papers #710

Jan Prüser

Forecasting US inflation using Markov Dimension Switching

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/86788828
ISSN 1864-4872 (online)
ISBN 978-3-86788-828-8

Jan Prüser[1]

Forecasting US inflation using Markov Dimension Switching

Abstract

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 over time. 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 the unemployment rate, the Treasury bill rate and the number of newly built houses are the most important variables in the generalized Phillips curve. Furthermore, these three predictors exhibit a sizeable 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. In addition, it turns out that the performance of MDS model forecasting is competitive in comparison with other models found to be useful in the inflation forecasting literature.

JEL Classification: C11, C32, C53, E37

Keywords: Phillips Curve; fat data; variable selection; model change

September 2017

1 Jan Prüser, RGS Econ and UDE. – All correspondence to: Jan Prüser, RWI, Hohenzollernstr. 1/3, 45128 Essen, Germany, e-mail: 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 U.S. 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 how to find the relevant predictors.

This paper addresses this question 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.[1] 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 which variables are

[1] 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.

important in determining inflation at different times.

Both the Bernoulli and the MDS approach are used to assess the importance of the predictors for one quarter and one year inflation. Most important predictors turn out to be the unemployment rate, the Treasury bill rate and the number of newly built houses. All three variables show a sizeable 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. 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

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

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). 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 $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$. The model can now be written as

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

where $\boldsymbol{z_t} = (1, \pi_t, \ldots, \pi_{t-p+1}, \boldsymbol{x}_t)$, $\boldsymbol{\theta} = (\alpha, \phi_0, \ldots, \phi_{p-1}, \boldsymbol{\beta}')'$ 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 γ_i follows a first-order Markov-switching process $S_{i,t}$ and therefore $\boldsymbol{\gamma}$ now has a time index t:

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

where $\boldsymbol{\gamma}_t = (S_{1,t}, \ldots, S_{m,t})$. Each Markov switching process $S_{i,t}$ can take on the value one or zero and is characterized by a 2×2 transition matrix $\boldsymbol{\mu}_i$, where $\mu_{kj,i} = \Pr(S_{i,t+1} = j|S_{i,t} = k)$, $k = 0, 1$ and $j = 0, 1$.[2] If $S_{i,t} = 1$, the ith 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.

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

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

   with

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

   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

[2] The Markov mixture modeling approach used here 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. The i.i.d. case is however nested as a special case of the Markov mixture approach.

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.

- or sample $\boldsymbol{\theta}$ using the algorithm of Carter and Kohn (1994), conditioning on $\boldsymbol{\gamma}_{1:T}$, $\pi_{1+h:T+h}$, $\boldsymbol{z}_{1:T}$ and $\sigma^2_{1+h:T+h}$, by writing the model in state space form as

$$\pi_{t+h} = (\boldsymbol{z}_t \odot \boldsymbol{\gamma}_t)\boldsymbol{\theta}_t + \epsilon_{t+h}, \tag{5}$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t, \tag{6}$$

where $\boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{W})$ and $\boldsymbol{W} = \mathbf{0}$.

Setting the covariance matrix $\boldsymbol{W}$ to zero leads to constant coefficients (i.e., $\boldsymbol{\theta}_t = \boldsymbol{\theta}$).[3] However, writing the model in state space form, the Kalman filter can be applied to estimate $\boldsymbol{\theta}$ (see Appendix A for details). In the MDS model, the Kalman filter has the crucial advantage that it is initialized with some prior, which receives less weight with each iteration. Thus, the OLS estimate of the full data matrix $\boldsymbol{z}_{1:T}$ can be used as a prior for the Kalman filter to give reasonable estimates when one variable is omitted from the model. But, as this initial condition receives less weight with each iteration, the filter converges to a potentially more precise estimate based on the sparse data matrix and this may lead to more accurate out-of-sample predictions.

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

- If γ_i is constant, sample it from

$$\gamma_i|\gamma_{-i}, \pi_{1+h:T+h}, \boldsymbol{z}_{1:T}, \boldsymbol{\theta}, \sigma^2_{1+h:T+h} \sim \text{Bernoulli}\left(\frac{l_{1i}}{l_{1i} + l_{0i}}\right), \tag{7}$$

with

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

[3] It would be conceptually straightforward to allow for time variation in the coefficients by estimating the covariance matrix $\boldsymbol{W}$. However, in the empirical application of this paper, I discovered convergence issues when estimating the MDS or Bernoulli model with time varying coefficients. Considering MDS in a framework with time varying parameters, however, would be interesting for future research.

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

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

- In the MDS model $S_{i,t}$ is sampled for $t = 1, \ldots, T$ 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 B for details). The transition probabilities of the ith Markov process are drawn from a Beta distribution

$$\mu_{11,i}|S_{i,1:T} \sim \text{Beta}(u_{11} + n_{11}, u_{10} + n_{10}), \quad (8)$$

$$\mu_{00,i}|S_{i,1:T} \sim \text{Beta}(u_{00} + n_{00}, u_{01} + n_{01}), \quad (9)$$

where n_{jk} counts the number of transitions from state j to k and $u_{11} = u_{00} = u_{10} = u_{01} = 1$.

3. Sample σ_t^{-2}:

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

$$\sigma^{-2}|\boldsymbol{\theta}, \pi_{1+h:T+h}, \boldsymbol{z}_{1:T}, \boldsymbol{\gamma}_{1:T} \sim \text{Gamma}(a, b^{-1}), \quad (10)$$

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

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

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \xi_t, \quad (11)$$

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

$$\zeta^{-1}|\sigma^2_{1+h:T+h} \sim \text{Gamma}(a, b^{-1}), \quad (12)$$

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.

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 predictor, a single predictor or all m predictors. In other words, $\boldsymbol{\gamma}$ 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.[4]

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 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.

[4] In the empirical application, only the intercept and the first lag are always included.

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.

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 the forecasting study different models are compared. 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 forecasting performance is compared with an AR(1) model with intercept and a multiple regression model containing all variables. All these models are applied with a constant and a stochastic variance specification and the MDS model is estimated using the conditional posterior of a regression model and using the Kalman filter for the model in state space form, as described in the Gibbs Sampler.

In a second step the forecasting performance of the MDS model is compared with different modeling approaches which have been found useful to forecast inflation in previous studies, namely DMA proposed by Koop and Korobilis (2012), hierarchical shrinkage in TVP-models proposed by Belmonte et al. (2014) and the unobserved components model with stochastic volatility (UC-SV) 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 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. In this paper, two cases are considered with all three forgetting factors set either to 0.95 or to 0.99. 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. Finally, for the UC-SV 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 and Watson (2007) for details).[5]

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 rest 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 forecasting performance of the entire predictive density. Furthermore, 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 presents results for the one quarter and one year ahead forecasting performance. Overall, it turns out that the MDS models forecast quite well. For the one quarter ahead forecasts the full model including all predictors forecasts quite poorly, overfitting the data. The variable selection in the Bernoulli model seems to be useful in avoiding overfitting as its forecasting performance in terms of point forecasts is very similar to the parsimonious AR(1) model and even delivers a slightly higher predictive likelihood. Further forecasting improvements 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. Only the DMA and DMS approach show a similar or better forecasting performance depending on the forecasting metric, all other models do worse. Especially the UC-SV yields poor forecasts. Moreover, hierarchical shrinkage in TVP and constant coefficient regression produce less precise forecasts than the Bernoulli or MDS models. The specification of the variance turns out to be less important. 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.

[5] Note that Stock and Watson (2007) do not estimate the system variances but rather set them to 0.2.

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

		$h = 1$			$h = 4$		
Model	Variance	RMSFE	MAFE	PL	RMSFE	MAFE	PL
MDS	constant	0.62	0.41	3.67	2.48	1.61	2.38
MDS	stochastic	0.62	0.41	3.65	2.39	1.63	2.57
MDS (Kalman)	constant	0.62	0.41	3.78	1.33	0.84	3.08
MDS (Kalman)	stochastic	0.61	0.41	3.74	1.34	0.88	3.00
Bernoulli	constant	0.65	0.45	3.61	2.22	1.63	2.08
Bernoulli	stochastic	0.66	0.45	3.60	2.23	1.63	2.08
AR(1)	constant	0.64	0.45	3.44	2.16	1.55	2.33
AR(1)	stochastic	0.63	0.45	3.10	2.15	1.55	2.28
Full model	constant	0.66	0.47	3.24	2.22	1.65	2.05
Full model	stochastic	0.67	0.47	3.19	2.22	1.65	2.06
LASSO	constant	0.66	0.46	3.48	2.22	1.64	2.19
LASSO	stochastic	0.65	0.46	3.49	2.22	1.62	2.28
TVP-shrink	constant	1.41	1.00	2.29	3.58	2.32	2.08
TVP-shrink	stochastic	0.66	0.48	3.32	2.43	1.67	2.45
UC-SV	stochastic	2.41	1.92	-1.96	3.51	2.59	-1.94
DMA (0.95)	stochastic	0.58	0.40	3.61	2.22	1.49	2.29
DMA (0.99)	stochastic	0.58	0.39	3.60	2.22	1.52	2.24
DMS (0.95)	stochastic	0.58	0.42	3.73	2.23	1.58	1.80
DMS (0.99)	stochastic	0.58	0.39	3.60	2.22	1.54	2.25

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

While the estimation method for the MDS models was not important for the one quarter ahead forecasts, it turns out to be very important for the one year ahead forecasts. Estimating the model using the Kalman filter leads to huge forecasting improvements. This finding indicates the benefit of putting more weight on the potentially more efficient estimate based on the sparse data matrix. All other models have difficulties beating the simple AR(1) model for one year ahead forecasts. Only the DMA and DMS approach are competitive. However, the MDS model using the Kalman filter is the clear winner in terms of RMSFE, MAFE and mean predictive likelihood for one year ahead forecasts.

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 C.1 and C.2 display the inclusion probabilities of the MDS model estimated using the conditional posterior of a regression model, the MDS model estimated using the Kalman filter and the Bernoulli model for the full sample. The results are shown with the stochastic variance specification, but the constant variance specification gives very similar results, as it was of minor relevance for the forecasting performance. 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 the inclusion probabilities show a sizeable degree of time variation, for which the Bernoulli approach cannot account. This demonstrates the usefulness of the MDS model over the Bernoulli model.

In many cases the Bernoulli model and the MDS model 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 *HSTARTS*, *UNEMP* and *TBILL* turn out to be important in all approaches. However, for *INEXP*, *OIL* and *CONS* the different approaches provide conflicting results. For one year inflation *HSTARTS*, *UNEMP* and *TBILL* turn out to be important as well but now show a greater degree of time variation. Again, the different approaches come to conflicting results for *INEXP* and *CONS*, while they now agree for *OIL*. The other variables are associated with a high degree of uncertainty as their inclusion probability is often close to 50%.

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 the unemployment rate, the Treasury bill rate and the number of newly built houses. Furthermore, these three predictors 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? *Federals Reserve Bank of Minneapolis*, 25(1).

Bańbura, M., Goammpme, 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). Hierachical Shrinkage in Time-Varying Parameter Models. *Journal of Forecasting*, 33(1):80–94.

Carter, C. K. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81(3):541–553.

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 Statisitcal Association*, 95(451):819–828.

Geweke, J. and Amisano, G. (2010). Hierachical Markov Normal Mixture Models with Application to Financal 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). Hierachical 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 Statisitcal Association*, 103(482):681–686.

Raftery, A., Kárný, M., and 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. 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 state space models

Given the model in state space form in (5) and (6) a standard Kalman filter delivers

$$\begin{aligned}
\boldsymbol{\theta}_{t|t-1} &= \boldsymbol{\theta}_{t-1|t-1} \\
\boldsymbol{P}_{t|t-1} &= \boldsymbol{P}_{t-1|t-1} + \boldsymbol{W} \\
\boldsymbol{K}_t &= \boldsymbol{P}_{t|t-1}(\boldsymbol{z_t} \odot \boldsymbol{\gamma}_t)'[(\boldsymbol{z_t} \odot \boldsymbol{\gamma}_t)\boldsymbol{P}_{t|t-1}(\boldsymbol{z_t} \odot \boldsymbol{\gamma}_t)' + \sigma^2_{t+h}]^{-1} \\
\boldsymbol{\theta}_{t|t} &= \boldsymbol{\theta}_{t|t-1} + \boldsymbol{K}_t[\pi_{t+h} - (\boldsymbol{z_t} \odot \boldsymbol{\gamma}_t)\boldsymbol{\theta}_{t|t-1}] \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t(\boldsymbol{z_t} \odot \boldsymbol{\gamma}_t)\boldsymbol{P}_{t|t-1}.
\end{aligned}$$

The Kalman Filter is initialized with $\boldsymbol{\theta}_{0|0} = \hat{\boldsymbol{\theta}}_{OLS}$ and $\boldsymbol{P}_{0|0} = \boldsymbol{V}(\hat{\boldsymbol{\theta}}_{OLS})$. Given $\boldsymbol{W} = 0$ the algorithm of Carter and Kohn (1994) corresponds to draw $\boldsymbol{\theta}$ from a normal distribution with mean $\boldsymbol{\theta}_{T|T}$ and variance $\boldsymbol{P}_{T|T}$.

Appendix B. 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 2×2 transition matrix $\boldsymbol{\mu}$ where $\mu_{kj} = \Pr(S_{t+1} = j|S_t = k)$, $k = 0, 1$ and $j = 0, 1$.[6] In order to draw S_t for $t = 1, \ldots, T$ 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

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

[6]For a simplified notation the index i is omitted and the general case of a two state Markov process is considered.

3. Given ψ_t update the probabilities as

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

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

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

5. Sample $S_{T-1}, \ldots, S_1$ 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.

Appendix C. Figures

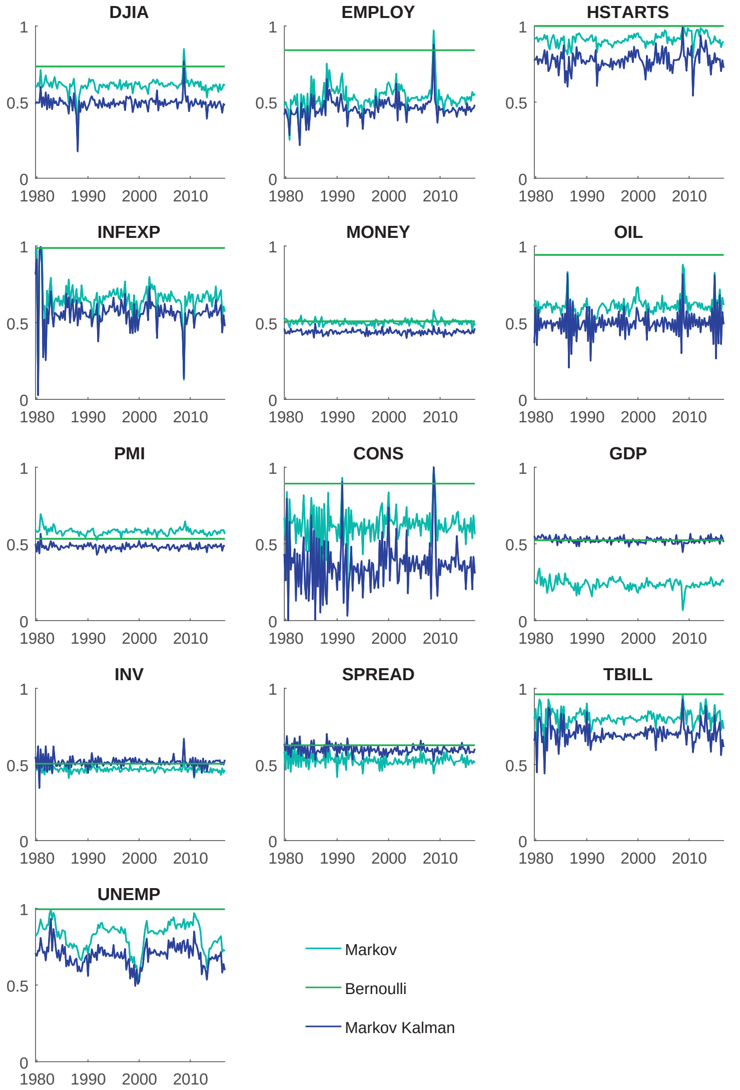

Figure C.1: Variable inclusion probabilities for one quarter inflation.

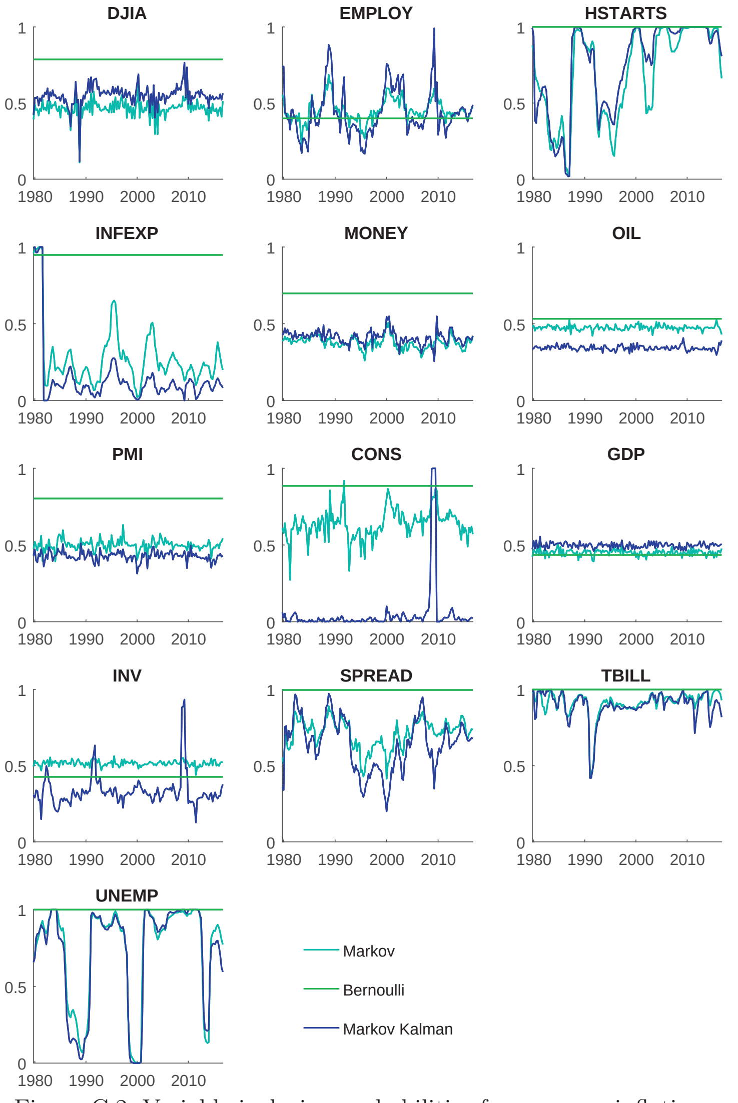

Figure C.2: Variable inclusion probabilities for one year inflation.