Classification of Monetary and Fiscal Dominance Regimes Using Machine Learning Techniques

Natascha Hinterlang*, Josef Hollmayr†

Deutsche Bundesbank‡

November 12, 2020

Abstract

This paper identifies U.S. monetary and fiscal dominance regimes using machine learning techniques. The algorithms are trained and verified by employing simulated data from Markov-switching DSGE models, before they classify regimes from 1968-2017 using actual U.S. data. All machine learning methods outperform a standard logistic regression concerning the simulated data. Among those the Boosted Ensemble Trees classifier yields the best results. We find clear evidence of fiscal dominance before Volcker. Monetary dominance is detected between 1984-1988, before a fiscally led regime turns up around the stock market crash lasting until 1994. Until the beginning of the new century, monetary dominance is established, while the more recent evidence following the financial crisis is mixed with a tendency towards fiscal dominance.

Keywords: Monetary-fiscal interaction, Machine Learning, Classification, Markov-switching DSGE

JEL Codes: C38, E31, E63

---

*Deutsche Bundesbank, DG Economics, Public Finance Division, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany, Natascha.Hinterlang@bundesbank.de. This paper is part of my Ph.D. thesis supervised by Prof. Dr. Uwe Hassler and Prof. Volker Wieland, Ph.D. at Goethe University Frankfurt. We are grateful for helpful comments from them, Pablo Burriel, Daniel Ollech, Nora Traum, from the participants of the Hot Topics in Econometrics Research Seminar 2019 at Goethe University Frankfurt, the 27th Meeting on Public Economics, the 76th Annual Congress of the International Institute of Public Finance, the 82th Symposium of the Society for Nonlinear Dynamics and Econometrics and the German Economic Association Annual Conference 2020.

†Deutsche Bundesbank, DG Economics, Public Finance Division

‡The views expressed in this paper are those of the authors; they do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

1 Introduction

Since the last financial crisis, U.S. total debt-to-GDP ratio has increased by 40 percentage points from about 64% in 2008 to 104% in 2019. Moreover, the deficit-to-GDP ratio in 2009 was the highest since the Korean war in the 1950s. At the same time, the conduct of monetary policy was characterized by the nominal interest rate being stuck at the zero lower bound and unconventional asset purchase programs. This situation brings back concerns of what Leeper (1991) calls the fiscal theory of the price level (FTPL). It states that if the central bank is *passive* (does not fight *actively* inflationary pressures), the price level might increase to stabilize real outstanding government debt that is not backed by future primary surpluses. Hence, price level changes would not be under control of the monetary authority anymore - a situation that is usually called a fiscal dominance (FD) regime. In contrast, a situation where the monetary authority responds *actively* to inflation while the fiscal authority commits to adjust *passively* its primary balance to stabilize public debt is known as monetary dominance (MD).

The concepts of MD and FD are of a purely theoretical nature. In reality, the true dominant regime is unobservable and unknown. Distinguishing between both regimes therefore serves as a useful classification application for machine learning techniques since these are shown to be good classifiers in other areas like engineering.

We contribute to the literature first established by Sargent and Wallace (1981) by providing and applying a new approach for classifying MD/FD regimes in reality. Specifically, the approach consists of four steps. First, we simulate both regimes using a simple Markov-switching dynamic stochastic general equilibrium (MS-DSGE) model as the data generating process (DGP). Second, we use the simulated data to train different machine learning classifiers. Third, we evaluate the predictive performance of the trained classifiers and their robustness with respect to changes in the underlying DGP. Fourth, the trained classifiers are used to identify regimes from 1968 to 2017 with actual U.S. data. We find that all machine learning methods outperform a standard logistic regression with respect to in- and (pseudo-)out-of-sample prediction accuracy using simulated MS-DSGE data. Among the machine learning methods, it is the Boosted Ensemble Trees (AdaBoost) classifier that produces the most reliable predictions with an accuracy rate of about 90%. By applying this trained classifier, we identify historical U.S. regimes from 1968-2017 that support the existing literature. In the literature there exist several papers that try to classify historical periods into FD or MD regimes using different approaches.

Favero and Monacelli (2005) estimate fiscal policy rules by Markov-switching (MS) regressions for the period 1960-2002. Davig and Leeper (2011) also estimate Markov-switching fiscal and monetary policy rules over the period from 1948-2008 and incorporate the results in a calibrated DSGE model in order to investigate government spending multipliers. Martin (2015) approaches the interplay between fiscal and monetary dominance from a different perspective. He uses the number of meetings and official phone conversations between the U.S. President and the Fed Chairman as a proxy for central bank independence. Kliem et al. (2016a,b) analyze the monetary-fiscal policy interaction by the low-frequency relationship between inflation and the fiscal stance. The majority of related papers, however, tackles the question by estimating MS-DSGE models with Bayesian methods (see e.g. Bianchi (2012); Bianchi and Ilut (2017); Chen et al. (2019)). Lately, some studies also account for the zero lower bound constraint. Gonzalez-Astudillo (2018), for example, estimates censored MS policy rules and Bianchi and Melosi (2017) estimate a MS-DSGE adding a fiscally-led zero lower bound regime.

Within this literature, there is a broad consensus of a FD regime during the 1970s. However, there is mixed evidence on the switching point. For example, Davig and Leeper (2011) and Bianchi and Ilut (2017) find an explosive regime with both policies being *active* after the appointment of Volcker as the Fed chairman in 1979. MD is established only in early to mid 1980s. Chen et al. (2019), however, find that fiscal policy was *active* until 1995 with monetary policy behaving *passive* between 1988 and 1992. The era under U.S. president Clinton is usually associated with an *active* monetary regime accomodated by the fiscal authority (see e.g. Davig and Leeper (2011), Bianchi (2012), Chen et al. (2019)). In the early to mid 2000s, fiscal policy is found to be *active* again in Davig and Leeper (2011) and Chen et al. (2019). Concerning the more recent periods, Gonzalez-Astudillo (2018) and Bianchi and Melosi (2017) provide evidence for FD after the financial crisis.

We also find clear evidence of FD pre-Volcker with our method, while MD is finally established only in 1984 until 1988. The FD regime is further found to be in place around the stock market crash and the early 1990s recession and after the Dot-com-Bubble crisis in the early 2000s. The evidence for the periods thereafter is mixed with a tendency to FD after the financial crisis.

Machine learning gains more and more attention in economics. Recent applications include forecasting of macroeconomic variables (e.g. Teräsvirta et al. (2005)), early warning predictions for financial crises (e.g. Beutel et al. (2018) and Alessi and Detken (2018)), for recessions (e.g. Ng (2014)) or for default risks (e.g. Badia et al. (2020) and Khandani et al. (2010). However, to the best of our knowledge, this paper is the first one applying machine learning techniques to classify an unobserved economic state using simulated DSGE data. Concerning our new approach, we would like to highlight four advantages with respect to the existing literature. First, it allows an easier and faster real-time classification of the current regime since the trained algorithm is ready to predict within seconds, given new data of the explanatory variables is available. Second, our classifier is trained and the performance is verified in the first place by using large simulated MS-DSGE datasets, where we know when each regime is in place. Third, and relatedly, due to the simulation, there does not exist a curse of dimensionality because we are not restricted by the time span of our series. Fourth, our procedure focuses on the (nonlinear) interactions of all endogenous variables, while other approaches usually employ only a subset in order to estimate policy-rule parameters and transition probabilities. Thus, given the same DSGE model structure and taken this model as "the truth", it is possible that our preferred machine learning classifier predicts different regimes than would be implied by directly estimating the model by e.g. Bayesian methods. By experimenting with the number of variables included as predictors, we show that using more information yields overall a better performance in terms of classification accuracy. Our preferred classifier, AdaBoost, is the one that best exploits all given information. Moreover, we can show which variables are especially important to distinguish between both regimes. The AdaBoost classifier attributes a relatively equal importance to all variables slightly favoring interest rate and debt.

The remainder of the paper is organized as follows. Section 2 describes the machine learning methods we employ to classify MD/FD regimes. In Section 3 we present the DSGE models used to simulate data for both regimes as well as the actual U.S. data. In Section 4 we show and discuss the results including robustness checks and variable importance. Section 5 concludes.

2 Methodology

In this section, we describe our approach to classify MD and FD regimes. Specifically, we explain the idea of supervised learning and the different classification methods as well as our hyperparameter choices. It is a brief overview based on James et al. (2013) and Friedman et al. (2001).

2.1 Supervised Learning - Classification

Our approach is based on machine (statistical) learning. Statistical learning can be distinguished into unsupervised and supervised learning. Unsupervised learning is used for finding relationships between variables or observations when no dependent, also called response variable, is given (e.g. principal component analysis, clustering). Supervised learning can be applied when the response variable is given. Generally, it comprises different methods to estimate the function f of $y = f(x) + \varepsilon$, where y is the response variable and x are independent/ predictor variables, also called "features". Since we want to identify (predict) FD/ MD regimes, our response variable takes on categorical values. Hence our task is a binary classification problem instead of a regression problem, where the response variable is numerical.[1] While the exact form of f needs to be known for inference about the relationship between x and y, it is not of interest when the focus lies solely on prediction (i.e. $\hat{y} = \hat{f}(x)$). We therefore apply parametric as well as nonparametric methods.

In total, we consider 6 different methods (Logistic, K-Nearest Neighbors, Decision Trees, Random Forest, Boosting and SVM).[2] Among them, the logistic model is the most widespread approach in the economics literature. Still, there exist several papers with economic applications of the other classification methods as well: For example, Khandani et al. (2010) employ decision trees to forecast consumer credit default risks, while Badia et al. (2020) rely on the random forest technique to predict fiscal crises. Ng (2014) explores boosting as a tool for indicating recessions and to identify relevant predictors. Alessi and Detken (2018) base their financial crisis early warning system on random forest, whereas Beutel et al. (2018) compare the performance of all mentioned methods (except boosting) for the same task and find that the logistic model outperforms the others. Since the best performing method is not known in advance and surely depends on the specific application, we include all of these most common models in our comparison.

In the following, we want to give a brief overview of these methods, focusing on the role of hyperparameters and pointing out advantages and drawbacks.

[1] We restrict our analysis to the binary response. The related literature distinguishes four regimes (i.e. both policies active/passive and only one active, one passive) in total. However, the two missing regimes refer to regions where the MS-DSGE model is explosive or indeterminate. Generally, our precedure can be extended to the multinomial case.

[2] We also considered artificial neural networks with a single hidden layer structure. However, since their accuracy was not superior to the other machine learning methods, we do not include them in the paper.

Logistic Model (Logit)

The logistic model is the standard model for a binary response variable y.[3] It relies on the assumption that y is driven by a latent process y^* that depends linearly on the explanatory variables, i.e. $y^* = X\beta + \varepsilon$. Moreover, the estimation errors ε are assumed to follow a logistic distribution. The conditional probability $P(y = 1|X) = \frac{\varepsilon^{X\beta}}{1+\varepsilon^{X\beta}}$ then gives class probabilities, where $\hat{\beta}$ can be estimated using nonlinear least squares or maximum likelihood for prediction. Class assignment then follows the largest probability. While being based on a clear statistical model and easy to interpret, the logistic model has the main drawback of being restricted to that pre-specified functional form.

K-Nearest Neighbors (KNN)

The KNN method is a two-step approach by Cover and Hart (1967). First, given a positive integer K and test observation x_0, it identifies the K closest (nearest) observations of the training sample. The closeness between two observations is usually measured by the Euclidean distance. We define the resulting neighborhood as N_0. Second, it estimates the conditional probability $P(y = 1|X = x_0)$ as a fraction of points in N_0 whose response values equal 1, i.e. $\widehat{P}(y = 1|X = x_0) = \frac{1}{K} \sum_{i \in N_0} I(y_i = 1)$, where I is the indicator function. The test observation is then assigned to the class with the largest probability. The choice of the hyperparameter K is related to a bias-variance trade-off. Selecting $K = 1$ yields an overly flexible model with low bias, but high variance when using different data (vice versa for $K \to \infty$). In contrast to the logistic model, the KNN method has the advantage that no pre-specified functional form nor an assumption on the underlying distribution is necessary. However, it suffers strongly from the curse of dimensionality.

Decision Trees (Tree)

The idea of decision trees by Breiman et al. (1984) is to split the predictor space into smaller regions by binary choices. Graphically, these trees consist of a root, interior nodes (branches) and final nodes (leafs). At the root and at each interior node observations are assigned to the following left or right subtree according to a decision rule. The rule simply compares the value of a single explanatory variable x_i to a threshold τ_i. By repeating the same procedure, observations are passed down the tree until they reach a final node. At every leaf, the class probability is then given by the respective fraction of assigned

[3]Strongly speaking, the logistic regression also belongs to the statistical learning classifiers (cf. James et al. (2013)). However, since it does not involve a hyperparameter choice and does not require much computational effort, we do not call it a statistical learning classifier. As a widely spread method, it rather serves as our benchmark model.

observations from the training sample. Jointly choosing the variables x and thresholds τ is a computationally infeasible task, such that approximation algorithms are used. The so-called "recursive binary splitting" approach is a top-down approach. It chooses the splits such that the purity (measured for example by the Gini index or deviance) of the subtrees (the gain from each considered split) is maximized. The algorithm is further subject to potential stopping criteria, that generally limit the complexity of the tree. These are e.g. a minimal number of observations per leaf or a maximum number of leafs. This hyperparameter again involves a bias-variance trade-off. Higher tree complexity with more splits yields a lower bias but higher variance and vice versa. One advantage of decision trees is as for KNN its non-parametric nature. Additionally, the method includes an automatic variable selection by choosing the predictors to split on. At the same time, decision trees are so-called "weak learners", i.e. they are generally instable across time and across different samples.

Ensemble Trees

Ensemble techniques were developed to improve weak learners. They are based on the idea of reducing prediction errors by averaging over a large number of different trees. Thereby, the heterogeneity between trees should be large to decrease variance, while maintaining homogeneity within trees. The two most popular approaches of combining trees are so-called Random forests (RF) (cf. Breiman (1996, 2001)) and Boosting (AdaBoost) (cf. Freund and Schapire (1999)).

Random Forests (RF)

RF consist of two parts. The first one is called Bagging (Bootstrap aggregation). It means that B individual trees are grown, where B refers to the number of drawn bootstrap samples from the data. Each individual tree is large (that means low bias, but high variance) and the variance is reduced by averaging over all individual trees. The second part involves a further decorrelation of the individual trees by randomly considering only a subset of m variables for the splitting decision at each node. Thus, the hyperparameters of this method are the following. First, the number of bootstraps (trees) B has to be chosen, which should be sufficiently large in order to guarantee convergence. The second is the number of predictors m to perform the split. Third, usual hyperparameters controlling the complexity of individual trees (e.g. minimal number of observations per leaf) have to be selected.

Boosting (AdaBoost)

Boosting refers to growing trees sequentially. Instead of bootstrapping, each tree is grown based on information from previously grown trees. Thereby, a decision tree is fitted to the residuals from the current given model. The new decision tree is then added into the previous fitted model and the residuals are updated again. Each of the trees is usually rather small determined by the hyperparameter d, the maximum number of splits in each tree. Together with the learning rate parameter λ , it makes boosting a rather slow learning approach.[4] The third tuning parameter is the number of trees T. In contrast to RF, boosting can overfit using too many trees. However, a very small learning rate needs a large number of trees.

Support Vector Machines (SVM)

The idea of support vector machines is to separate data linearly into 2 classes such that the distance between classes gets large. Depending on the dimensionality of the data, the decision boundary is a point (1-d data), a line (2-d data) or a hyperplane (3-d or higher data). Given the data is linearly separable, there is an infinite space of possible hyperplanes. In the end, the approach boils down to a constrained optimization problem: It chooses the hyperplane that maximizes the distance between the closest observation of each class to the hyperplane subject to the constraint that each observation lies on the correct side of the hyperplane. The described distance is called "margin" (maximal margin classifier) and the closest observations are called "support vectors" since these determine the solution to the optimization problem solely.

Most of the time, however, linear separability does not hold. Then, the approach of support vector classifiers is to enlarge the feature space by nonlinear transformations, e.g. by higher-order polynomial terms or interactions. However, enlarging the feature space in that way increases the number of parameters immensely. To overcome this issue, SVM use kernel functions instead, which is also known as the "kernel-trick" in the literature (cf. Shalev-Shwartz and Ben-David (2014)). Basically, kernels are functions that describe the similarity between two observations. The default kernel function is the radial (Gauß) kernel, $K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2)$, where x_i and x_i' are two (distinct) observations and p denotes the number of features. The hyperparamater γ determines the complexity with larger values leading to a lower bias and higher variance, vice versa. Another hyperparameter comes in when allowing for misclassifications by adding a penalty term for

[4] We use the adaptive boosting method (AdaBoostM1 in Matlab) for training the ensemble. The learning rate $\lambda \in (0, 1]$ controls the shrinkage in each iteration step.

misclassified observations to the loss function, also called “soft margin constraint”. The hyperparameter C then determines the cost of violations to this constraint. The trade-off involved here again consists of lower bias (i.e. smaller margins, less misclassifications) and higher variance vs. higher bias (i.e. larger margins, more misclassifications) and lower variance.

SVM are quite robust with respect to outliers and noise in the data. At the same time, their performance crucially hinges on the choice of the hyperparameters. Further drawbacks are the lack of interpretability and the large memory usage in terms of computational effort.

As outlined in this section, the models’ complexities depend crucially on the choice of hyperparameters. In all cases, this choice involves a bias-variance trade-off, i.e. more flexible models (less bias) vs. simpler models (less variance). Appendix A describes how and which hyperparameters are selected in our case.

3 Data

Supervised learning involves splitting available data (response and predictors) into training and validation/ test sets. The training set is then used to estimate the respective parameters of each approach and the validation set is employed to choose hyperparameters and/ or to estimate the error rate. When the trained algorithms produce satisfying results, they can be used for prediction with new data (with unknown responses).

When dealing with FD and MD, we face the problem that there is no actual data for the response variable. The concept of FD and MD regimes is a rather theoretical one. It stems from a shared government budget balance of both authorities. However, in reality we cannot observe the true regime. Still, in state-of-the art DSGE models, we can explicitly differentiate between both regimes by determining corresponding parameters. Hence, we can use these Markov-Switching-DSGE (MS-DSGE) models as our data generating process and thereby gather training data for the supervised learning classifiers.

In this section, we shortly present the MS-DSGE model that is used to simulate training data. Then we describe the actual U.S. data employed for the identification of historical regimes.

3.1 Data Generating Process

This section lays out a conventional new Keynesian model that serves as our data generating process.[5] Further we present the benchmark calibration and how the FD and MD regimes are simulated using the model.

3.1.1 Model

Households:

The representative household maximizes its expected life-time utility where the period utility function is given by

$$U_t = \left(\ln(C_t) - \frac{N_t^{1+\phi}}{1+\phi}) \right). \tag{1}$$

The household derives utility from consumption C_t and disutility from labor N_t. The households' budget constraint is

$$C_t + B_t = W_t N_t (1 - \tau_t^l) + \frac{R_{t-1} B_{t-1}}{\pi_t} + T_t, \tag{2}$$

with B_t denoting government bonds, W_t wages, τ_t^l labor income taxes, R_t gross nominal interest rate, π_t inflation and T_t profits from the firm.

Firms:

The production side consists of a continuum of competitive firms, where each firm produces its good j according to the production function

$$Y_t(j) = \exp(A_t) N_t(j), \tag{3}$$

where Y_t denotes the output produced with a given level of technology A_t and hours worked $N_t(j)$ as the only input factor. Technology evolves according to an exogenous AR(1) process:

$$\ln(A_t) = \rho_a \ln(A_{t-1}) + \varepsilon_t^A, \quad \varepsilon_t^A \sim \mathcal{N}(0, \sigma_a^2). \tag{4}$$

The final goods are sold by monopolistically competitive retailers, where price setting is subject to nominal rigidities. Following Calvo (1983), each period only a fraction (1 - θ)

[5] We consider a very basic model without including e.g. labor market or financial frictions. How such restrictions would influence the classification problem is left for future research at this stage.

of all retailers is allowed to reset optimally their prices ($P_t(j)$). There is no indexation of those retailers who cannot reoptimize their prices. Profits of firm j are then given by (in nominal terms)

$$T_t(j) = (P_t(j) - MC_t(j)) \left(\frac{P_t(j)}{P_t}\right)^{-\epsilon^d} Y_t(j). \tag{5}$$

Real marginal costs are given by the following expression

$$MC_t(j) = \frac{W_t(j)}{A_t}, \tag{6}$$

while the demand for good j is expressed by

$$Y_t(j) = Y_t \left(\frac{p_t(j)}{P_t}\right)^{-\frac{1+\epsilon^d}{\epsilon^d}}, \tag{7}$$

where P_t denotes the aggregate price level and ϵ^d is the demand elasticity.

Government:

The governments' budget constraint takes the following form:

$$B_t = \left(\frac{B_{t-1}R_{t-1}}{\pi_t}\right) + G_t - \tau_t^l W_t N_t \tag{8}$$

The government can accumulate debt B_t in form of government bonds. Every year, debt is increased by interest payments, i.e. the previous years' debt multiplied by the previous years' nominal interest rate which is given by R_{t-1}. The inflation rate is given by π_t and government spending in the current period by G_t. Labor income taxes (with τ_t^l as the tax rate applied to wages W_t multiplied by hours worked N_t) reduce the deficit and hence government debt. In this model, the tax rate follows a simple AR(1) process for simplicity, while government spending responds in a rule-based manner. We focus on the government spending rule for now. It is persistent and has an anticyclical component which is linked to last period's debt level. If debt is higher than its long run trend b^*, government spending is accordingly cut back, in order to return to the long-run equilibrium path. If government debt (denoted in deviations from its own steady state and in absolute terms, not relative to GDP[6]) is below its long-term equilibrium value, government spending can be increased.[7]

[6]The analysis does not change if the debt-to-GDP ratio is included as a target variable or if another variable such as GDP is included in the government spending rule.

[7]It is also assumed that the long-run structural growth of the economy is zero, so the interest growth differential is assumed to be positive.

The rule can be expressed in log-linearized form arount the steady state (denoted by small case letters) as follows:

$$g_t = \rho_g g_{t-1} - \delta_b(s_t)(b_{t-1}) + \varepsilon_t^g, \quad \varepsilon_t^g \sim \mathcal{N}(0, \sigma_g^2), \tag{9}$$

where ε_t^g is a fiscal policy shock. The parameters ρ_g and $\delta_b \geq 0$ denote the intensity of the response of government spending to its own lag and the deviation of debt from its long-run trend, respectively, The parameter δ_b depends on the regime s_t in period t, which will be specified in the next subsection. Monetary policy is conducted by the central bank which follows a Taylor-type rule and reacts to its own nominal interest rate lag as well as to deviations of inflation and output from its respective target. The coefficient on inflation ϕ_π is regime dependent similar to the coefficient on debt in the government spending rule. Hence, the log-linearized monetary policy rule around the steady state is given by:

$$r_t = \rho_r r_{t-1} + \phi_\pi(s_t)\pi_t + \phi_y y_t + \varepsilon_t^r, \quad \varepsilon_t^r \sim \mathcal{N}(0, \sigma_r^2), \tag{10}$$

with ε_t^r being a monetary policy shock.

Market clearing and Monetary Policy:
Total demand by the government and by households in the form of consumption must fully absorb the output of the firms:

$$Y_t = C_t + G_t. \tag{11}$$

Market clearing in the bond market implies that all bonds issued by the government are bought by the households in the economy.

3.1.2 Calibration and Solution

In order to quantify the point beyond which debt is no longer stationary we have to calibrate all model parameters which will be explained in this section. The model is calibrated to a quarterly frequency where most parameters are taken from the literature. Table 1 presents the respective parameters. The Calvo-parameter is chosen to be 0.75, which means that 25% of all firms can choose to reset their prices each quarter. The discounting parameter β is calibrated to 0.99 to arrive at an annual real interest rate of 4%. The autoregressive parameters are all set to uniform values of 0.9 and the standard deviations of all shocks to 0.01.

The coefficient on output in the Taylor rule is fixed at 0.5. However, the two key parameters for our analysis are the state-dependent policy rule parameters. Specifically, MD and FD are defined in the model by the pair-values of $(\delta_b(s_t), \phi_\pi(s_t))$. Under MD ($s_t = 1$), we assume $\delta_b = 0.1$ and $\phi_\pi = 1.5$. That means, fiscal policy reacts to last periods' government debt with a coefficient larger than the real interest rate, while monetary policy adjusts the nominal interest rate more than one-for-one with inflation. Under FD ($s_t = 2$), $\delta_b = 0$ and $\phi_\pi = 0.5$, i.e. the fiscal authority does not stabilize its debt level and the central bank responds only sluggishly to inflation.

Table 1: Calibrated Parameters of the model

Description	Parameter	Value
Impatience	β	0.99
Disutility of labor	ϕ	1
Calvo prices	θ	0.75
Steady state tax rate	τ_{ss}	0.3
Coeff. on inflation in TR	ϕ_π	$s_t = 1 : 1.5$
		$s_t = 2 : 0.5$
Coeff. on output in TR	ϕ_y	0.5
Coeff. on debt in gov.spending	$\delta_b(s)$	$s_t = 1 : 0.1$
		$s_t = 2 : 0$
AR parameter tax	ρ_t	0.7
AR parameter gov. spending	ρ_g	0.7
AR parameter technology	ρ_a	0.7
AR parameter interest rate	ρ_r	0.7
Steady state ratios	C_{ss}/Y_{ss}	0.7
	G_{ss}/Y_{ss}	0.3
	G_{ss}/B_{ss}	0.6
Std.deviation technology	σ_a	0.01
Std.deviation gov. spending	σ_g	0.01
Std.deviation interest rate	σ_r	0.01
Std.deviation labor tax rate	σ_{τ^l}	0.01

Note: This table summarizes all calibrated parameters. Most of them are taken from the literature. δ_b and ϕ_π can take take two values depending on the prevalent regime.

All log-linearized model equations are stated in the Appendix B. The solution of the Markov-switching problem is achieved using the code developed by Farmer et al. (2009). The initial scale is set to 0 and the convergence criterion to $1 * 10^{-9}$.

3.1.3 MS-DSGE Model Simulation

Our model allows for endogenous regime switches between MD and FD. The corresponding transition probabilities between the two regimes are given by the matrix

$$P = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}, \tag{12}$$

where 1 and 2 refer to MD and FD, respectively. Expressed in general analytical terms, the two regimes follow a Markov chain which is described as

$$p_{ij} = P[s_t = j | s_{t-1} = i] \tag{13}$$

with $i, j = 1, 2$. It means that given the previous period's regime, the probability of remaining in the same regime this period is 90%. The conditional probability of switching to the other regime is 10%. The probabilities p_{11} and p_{22} are directly related to the average length of time for which each policy is being pursued. This relation is given by $1/(1-p_{ii}) = 1/0.1 = 10, i = 1, 2$. Hence, the FD and MD regimes are assumed to be quite persistent with an average duration of 10 periods.

In order to generate data, we draw randomly from the four shocks given their distribution. Each period, households and firms expect fiscal and monetary policy to either remain the same or change given the transition probabilities. The fact that a random switch might occur is entering their decisions-making. In contrast to reality, households and firms observe the current regime. In total, we have 11 endogenous variables in our model. For these variables, we simulate time series over 10000 periods, where 5000 build the training dataset and the other half is used for testing the predictive performance.[8]

3.2 Actual U.S. Data

All statistical learning algorithms are trained using 5000 periods of the simulated MS-DSGE data. In order to use the trained models for classifying new data, we need data on the same predictor variables as in the training set. We use the following endogenous variables of the MS-DSGE model in our estimation: Output, Consumption, Labor, Wages, Technology, Nominal Interest Rate, Inflation, Government Spending, Taxes, Government Debt. Marginal Cost is also an endogenous variable of the DSGE model. However, we do not include it in the training set since there is no actual U.S. data available for that variable. Table 7 in the Appendix C presents the corresponding actual U.S. data that we employ for classifying historical regimes. We use quarterly U.S. data from 1966:Q1-2017:Q4, which is determined by data availability.

The training data is simulated by exposing the MS-DSGE model economy to different shocks. Hence, the variables represent deviations from its steady state values. Since training data and new data that is used for prediction must be comparable, we have to

[8] By using simulated time series, we have the advantage that we are not confronted with a curse of dimensionality that might exist using actual macroeconomic time series of much shorter length.

transform the actual U.S. data into deviations from steady state. Therefore, we employ the Hamilton (2018) filter (with lag length $p = 4$ and forecast horizon $h = 8$) in order to extract the cyclical component of each time series. As pointed out by Hamilton (2018), this procedure allows a better match between DSGE simulated data and actual data compared to the Hodrick-Prescott filter.[9] Tables 8 and 9 in Appendix C show standard deviations and correlations of the variables inflation, interest rate, debt and output from the actual data and compare it to the implied moments of the MS-DSGE model in total and conditional on each regime. Inflation and interest rate volatility are matched best by the MS-DSGE conditional on the MD regime. The MS-DSGE implies a larger volatility in both cases though. Conditional on FD, the volatility of debt is close but below the one of the actual data. The standard deviation of output implied by the model is clearly undersized compared to the actual data. Concerning the implied correlation between the four variables, the model performs quite well in general, except for the pair of inflation and output.

4 Results

4.1 Performance on Simulated MS-DSGE Data

4.1.1 In-sample

In order to train the different machine learning algorithms, we use half of the simulated periods from our baseline MS-DSGE model. The sample consists of 47% MD and 53% FD regimes. Table 2 reports the accuracy for this in-sample data across all trained classifiers. The best performance is indicated in bold.

The standard logit model performs poorly. It correctly classifies only about 53% of the training observations. The KNN method improves on that with a validation accuracy about 87%, which can be further increased to 92% and 98% by using the decision tree method and SVM, respectively. Both ensemble tree methods, RF and AdaBoost, perfectly classify the training data with an accuracy rate of 100%.

4.1.2 (Pseudo-)Out-of-sample

Since the trained classifiers shall be used for prediction, the focus lies on the out-of-sample rather than in-sample performance. Hence, we use four simulated test datasets in order

[9] As a cross-check, however, we also try the Hodrick-Prescott filter (with $\lambda = 1600$). See Section 4.4.

Table 2: In-sample Accuracy

Method	**Accuracy**
Logit	0.53
KNN	0.87
Tree	0.92
RF	**1.00**
AdaBoost	**1.00**
SVM	0.98

Note: Accuracy on training data set from baseline MS-DGSE model.

to check the classifiers' generalization ability.

The first one comprises further simulated periods of the baseline MS-DSGE model, that were not used for training. However, DSGE models can never fully represent reality. We need to account for parameter uncertainty in our data generating process and to check the classifiers robustness with respect to changes in the data. Therefore, the three additional test sets stem from modified versions of the baseline MS-DSGE model. The first modification is a lower Calvo parameter of $\theta = 0.7$. The second one assumes a sharper response of the central bank with respect to inflation under the MD regime, i.e. $\phi_\pi(s_1) = 2.5$. For the third test dataset we change the transition probabilities to $p_{11} = 0.92$ and $p_{22} = 0.88$. Each test dataset consists of 5000 simulated periods.[10] All resulting confusion matrices across methods and for each test sample are shown in Tables 8 and 9 in the Appendix D. Table 3 summarizes the results by presenting different performance measures averaged over the four test samples. The AdaBoost method outperforms the others over all categories[11], while RF is overall the second best performing model. The standard logistic model yields a remarkably low specificity of 0.39, i.e. of all true fiscal dominance regimes, it correctly classifies (on average) only 39%.

[10] The share of MD regimes in the four test sets is 50%, 47%, 47% and 56%, respectively.

[11] AUC refers to the area under the receiver operating characteristic curve (ROC). It plots the sensitivity against against (1-specificity) for different conditional probability classification thresholds. AUC takes on values between 0 and 1, where 1 means perfect performance and 0.5 corresponds to an uninformative classifier.

Table 3: Averaged Out-of-Sample Performance Measures

Method	**Accuracy**	**TPR**	**TNR**	**PPV**	**NPV**	**AUC**
Logit	0.50	0.39	0.61	0.50	0.50	0.50
KNN	0.86	0.97	0.75	0.79	0.96	0.95
Tree	0.83	0.88	0.78	0.80	0.87	0.91
RF	0.90	0.95	0.86	0.87	0.94	0.98
AdaBoost	**0.95**	**0.97**	**0.92**	**0.92**	**0.97**	**0.99**
SVM	0.86	0.84	0.87	0.87	0.85	0.95

Note: Different performance measures averaged over the four test samples. TPR, TNR, PPV, NPV, AUC denote true positive rate, true negative rate, positive predictive value, negative predictive value and area under the curve, respectively, with positive corresponding to MD negative to FD. Bold numbers emphasize the best performing classifier according to each performance measure.

4.2 Classification of Historical U.S. Regimes

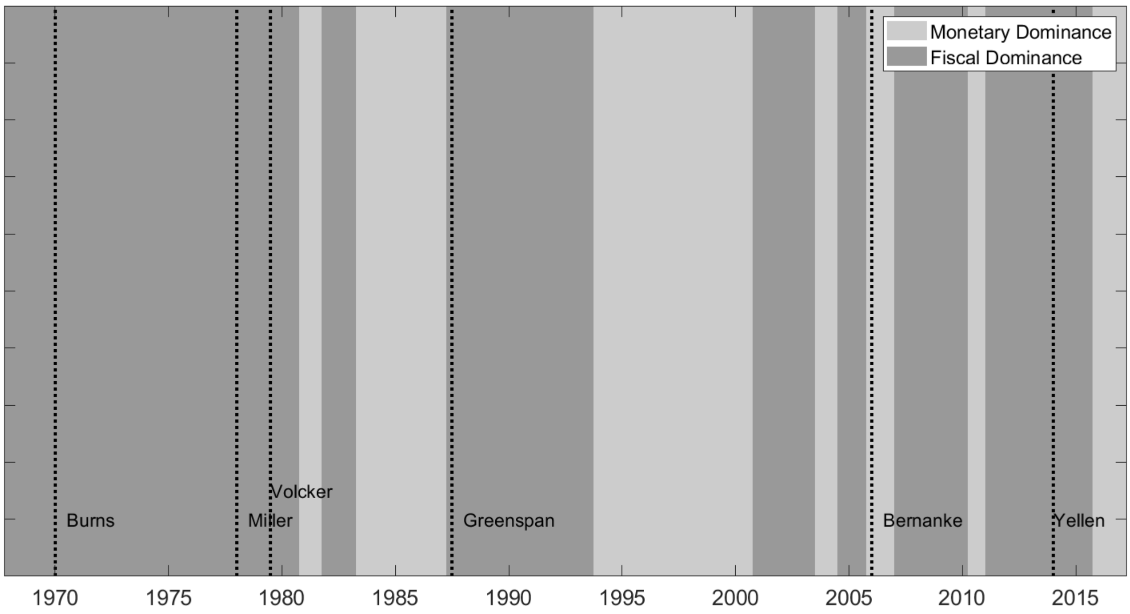
This section presents our results on the classification of historical U.S. regimes and compares it to the existing literature as well as general narratives. Since AdaBoost yields the best predictive performance using simulated MS-DSGE test data, we trust this trained classifier the most when applying it to the actual U.S. data.[12] Given the data of Table 7, we predict the historical regimes from 1968:Q4 to 2017:Q4. Figure 1 shows the smoothed results by the AdaBoost classifier.[13]

We find clear evidence of FD at the beginning of our sample until mid 1980s. This period covers completely the eras of the Fed Chairmen Burns (1970-1977) and Miller (1978-1979) under the U.S. presidents Ford (1975-1976) and Carter (1977-1980). The appointment of Volcker as the Fed Chairman in January 1980 and his success in bringing down inflation in the following years is usually found to be a turning point in the literature. Meltzer (2011) for example states that Volcker played a major role in rebuilding central banks' independence while it accepted its role as a junior partner in the two previous decades. Still, our preferred classifier finds that MD was finally established only at the beginning of 1984 until 1988. The first half of the presidency of Reagan (1981-1988) with his expansionary tax reforms is still partly found to be a FD regime. Davig and Leeper (2011) and Bianchi and Ilut (2017) find both policies being active from 1980-1983. The results of Chen et al. (2019) imply that fiscal policy was active from the end of 1981

[12]The predicted regimes of the second and third best performing classifiers (RF and SVM) are shown in Appendix E.

[13]We smooth the results by restricting a regime change to occur only if the new regime lasts for at least 4 quarters in a row.

Figure 1: Predicted U.S. Regimes by the AdaBoost Classifier



Note: Predicted U.S. Regimes according to the trained Boosted Trees classifier. Dark-shaded areas correspond to the fiscal dominance regime, while light-shaded areas belong to the monetary dominance regime. The black-dotted vertical lines represent the appointment date of the respective Fed Chairman.

until even 1995, while monetary policy acted less conservatively until 1983 and around 1988-1991. Overall, our results largely coincide with the ones of Chen et al. (2019) for the period of 1980-1994, which falls into the Fed chairmanship of Greenspan (1988-2005) and U.S. presidency of Bush sen. (1989-1992). The end of this period was characterized by a recession following the stock market crash in 1989. Starting only one year after the appointment of Clinton in 1993, our preferred classifier finds clear evidence for MD until the third quarter of 2001, which might be a result of the Deficit Reduction Act. The Dot-Com Bubble led to expansionary tax reforms during the first half of the first half of the 2000s under U.S. president Bush jun. The AdaBoost classifier predicts these periods as a FD regime. We find mixed evidence for the periods prior to the crisis with one year (2005:Q3-2006:Q2) of FD and MD until the beginning of the crisis. Since then, FD is found to be the prevalent regime until mid of 2016. This period overlaps with the Feds' large scale asset purchase programs and the zero lower bound (ZLB) phase. These quantitative easing (QE) programs are often critized as monetization of government debt. Since our data generating MS-DSGE model neither incorporates a ZLB nor any QE measures the results have to be interpreted with caution. Still, the results are similar to the ones of Gonzalez-Astudillo (2018) and Bianchi and Melosi (2017). Gonzalez-Astudillo (2018) estimate MS policy rules where the Taylor rule is allowed to be censored at the ZLB and finds FD after the financial crisis until the end of its sample in 2015. Bianchi and Melosi (2017) estimate a three-regime MS-VAR, where the third regime refers to a fiscally-led ZLB regime. They find this regime to be in place starting with the financial crisis until their sample end in 2014. Our sample ends with a classified MD regime until the fourth quarter of 2017. Table 10 in the Appendix compares our results to the findings of related literature.

4.3 Inspecting the Classification Mechanism

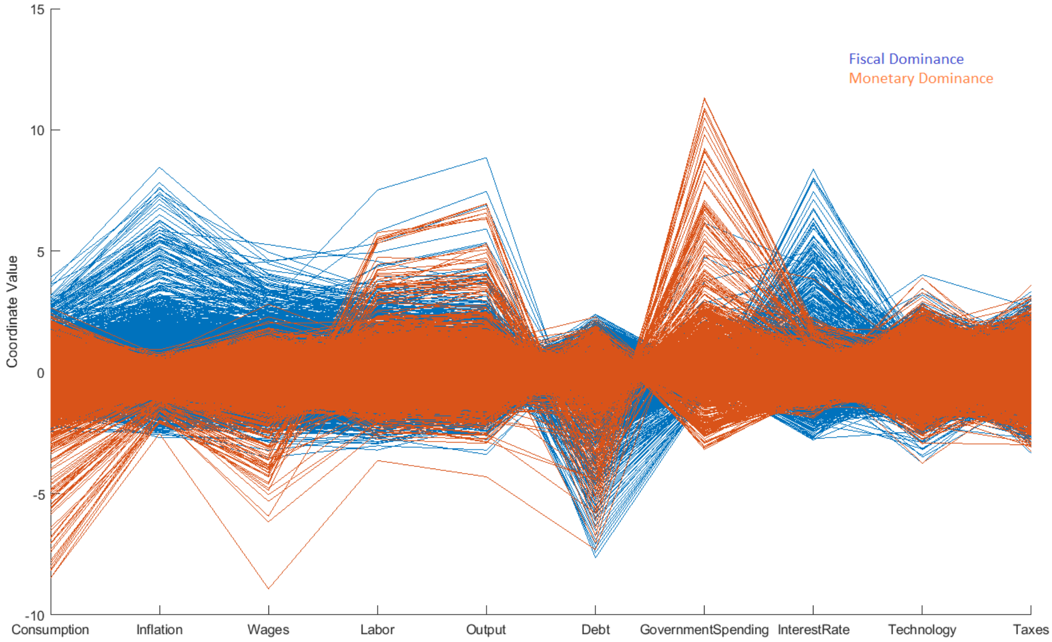
In this section, we take a deeper look at the classification mechanism of the trained AdaBoost model. Specifically, we analyze which variables are important for the algorithm to distinguish MD from FD.

We start by visualizing the data that was used for training the classifiers. Figure 2 presents the Parallel Coordinates Plot (PCP) of the training dataset. The PCP allows to inspect multidimensional data by plotting each observation as a sequence of its (standardized) coordinate values against their coordinate indices (variable names). In Figure 2, blue (orange) sequences belong to observations from the FD (MD) regime. This visualization shows that variables like interest rate, debt or inflation have high informational content for distinguishing between FD and MD. High interest and inflation rates, for example, are only found under the FD regime. In contrast, technology and taxes do not help a lot for differentiating the two regimes. So far, this finding is not really surprising since the variables appearing in the policy reaction functions (9) and (10) should of course have predictive content since the two regimes are defined in these equations. Still, because of the endogeneity of all variables in the DSGE model, there is more information in the other variables as well, that can be used by the classifiers.

Figures 3 and 4 present the relative variable importance for AdaBoost and RF, respectively. The predictor importance is estimated by dividing the sum of changes in impurity due to splits on the respective variable by the number of interior nodes (branches). Node impurity is measured here by the Gini Diversity Index (GDI), i.e. $1 - \sum_i p^2(i)$, where $p(i)$ is the fraction of observations at that node belonging to class i. The GDI is equal to 0 for a node with only one class and positive otherwise. The results are shown in relative terms, where the predictor with the largest variable importance estimate is set to 100%.

The most important predictor for the AdaBoost classifier is the nominal interest rate followed by debt and consumption. Output, taxes and inflation, government spending and

Figure 2: Parallel Coordinates Plot of Training Data



Note: Parallel Coordinates Plot of baseline MS-DSGE training dataset. Blue (orange) observations belong to the Fiscal (Monetary) Dominance Regime.

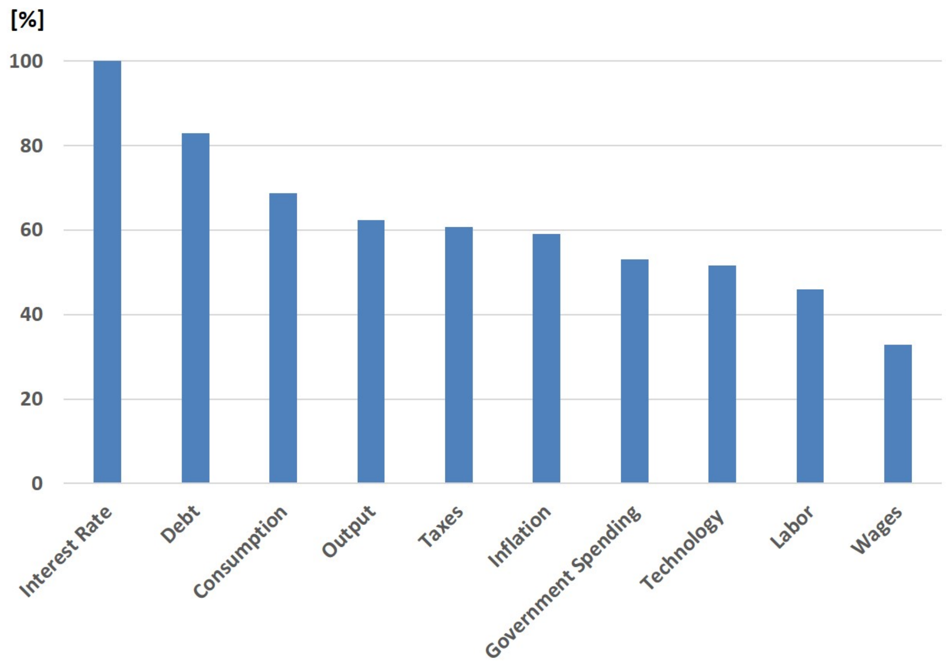
technology are almost equally important ranging from ca. 52-62% of the benchmark importance measure. Labor and wages are the two least important variables for AdaBoost. Comparing this result to the PCP, it is especially interesting that taxes are middle ranked since one would actually not expect high informational content by eyeballing. For the RF classifier, the most important variable is inflation, followed by debt, government spending and the interest rate. Missing only output, these are exactly the variables from the policy reaction functions. Interestingly, all other variables are rather unimportant for the RF classifier with taxes and wages ranked last. In contrast, for AdaBoost the variable importance is distributed more uniformly.

In order to check how the predictive performance depends on the included variables, we also experiment with the number of predictors. Specifically, we once train the classifiers with only 2 predictors (i.e. inflation and debt) and once with 4 predictors (i.e. inflation, debt, interest rate and government spending).[14] We compare the averaged accuracy over the four test datasets (see Section 4.1.2) in Figure 5.

Note, that the logistic classifier is not included in the Figure since it yields a poor averaged accuracy of around 50% irrespective of the predictor subset. Using only inflation and debt

[14]See Tables 5 and 6 in Appendix A for the chosen hyperparameters in these cases.

Figure 3: AdaBoost: Relative Variable Importance



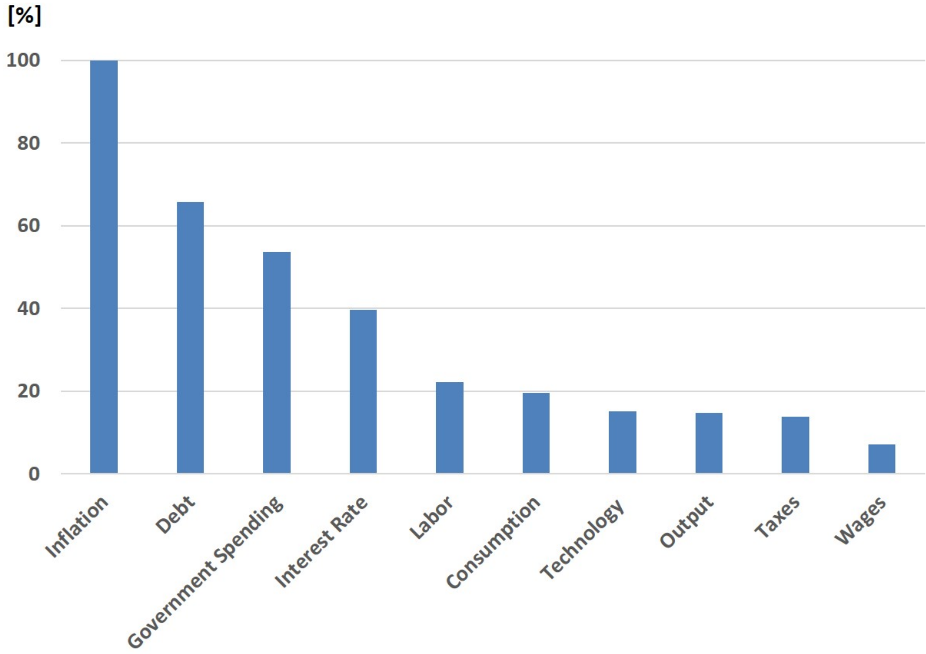
Note: Relative variable importance of the AdaBoost classifier. Variable importance is estimated by the average change of node impurity due to splits on the corresponding predictor. Node impurity is measured by the Gini diversity index. The variable with the largest importance measure is set as a benchmark (100%) in order to get relative measures.

as predictors, all machine learning classifiers perform remarkably well with an average accuracy of ca. 75%. Increasing the number of predictors to four by further including the nominal interest rate and government spending rises the average accuracy by about 7 percentage points (PP). Interestingly, the smaller the predictor subset, the more equal is the performance across classifiers. The AdaBoost classifier is the one that benefits most from including all 10 endogenous variables with an increase in average accuracy by 13 PP. This result supports the view that the AdaBoost classifier is the one that best exploits all given information in the predictors to distinguish MD from FD regimes.

4.4 Robustness

In the last section inflation was described as one key variable to differentiate between both regimes. To check the robustness of our results with respect to the inflation measure, we repeat the classification exercise of the actual regimes using the GDP deflator and Personal Consumption Expenditure (PCE) instead of CPI inflation. All other variables stay unchanged. Figures 6 and 12 (in Appendix E) present the predicted U.S. regimes

Figure 4: RF: Relative Variable Importance



Note: Relative variable importance of the RF classifier. Variable importance is estimated by the average change of node impurity due to splits on the corresponding predictor. Node impurity is measured by the Gini diversity index. The variable with the largest importance measure is set as a benchmark (100%) in order to get relative measures.

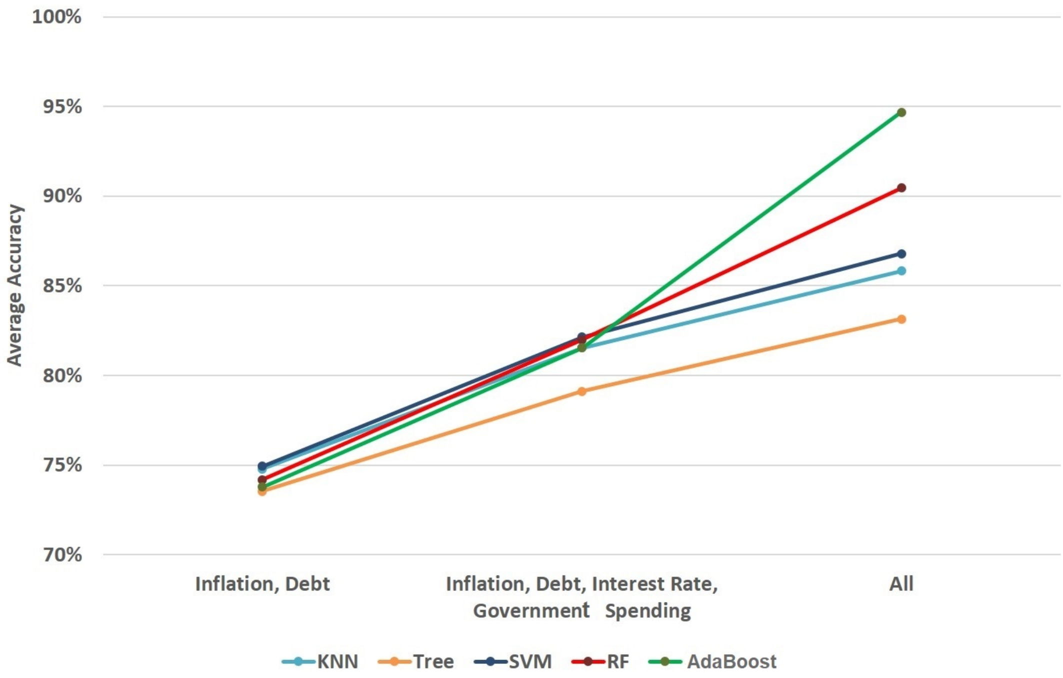
according to AdaBoost with GDP deflator and PCE inflation, respectively.

Overall, the predicted U.S. regimes are robust with respect to the inflation measure. We only see small deviations for the periods between 1985-1988 and around the financial crisis. Using the GDP deflator, we get more FD regimes for the first and more MD regimes for the latter period.

Additionally, we check robustness with respect to the filter method used to extract the cyclical component. Figure 7 shows the predicted regimes by AdaBoost employing the Hodrick-Prescott (HP) filter (with $\lambda = 1600$) to the actual data instead of the Hamilton filter. Using the HP filter allows us to classify periods going back to 1966 since there are no excluded lags as with the Hamilton filter.

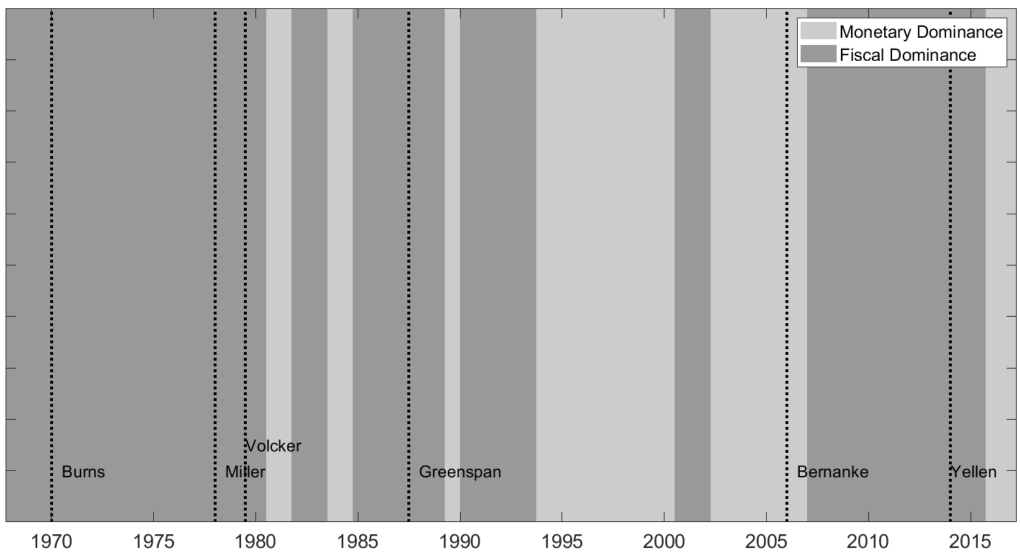
Before 1970, we find mainly MD to be the prevalent regime. Although, using the HP filter the AdaBoost classifier predicts shorter MD regimes for the rest of the sample, the timing roughly coincides with our baseline result. The predicted regimes using the two different filter methods only clearly diverge at the end of the sample. This could be due to the HP filters' usual problem of biasedness at the interval borders.

Figure 5: Averaged Out-of-Sample Accuracy for Different Predictor Subsets



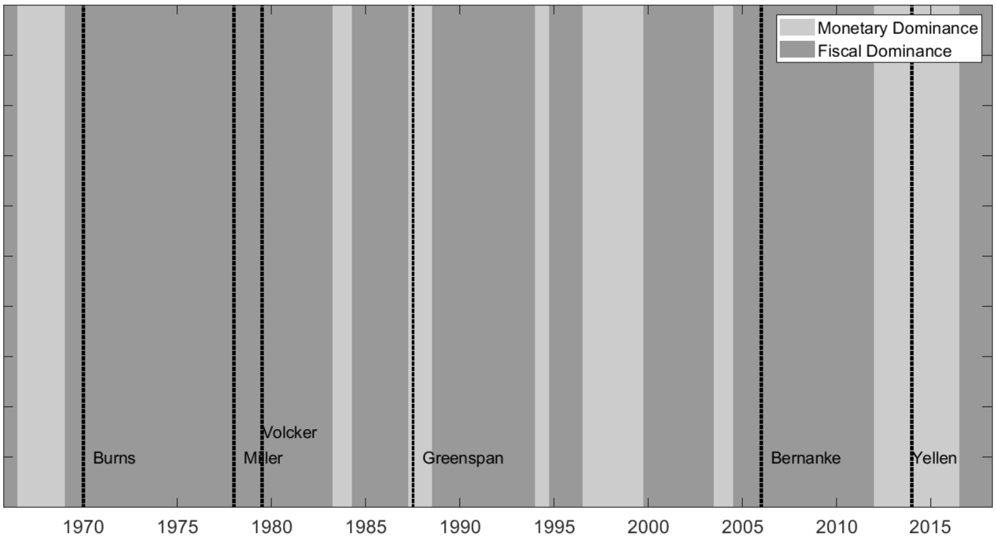
Note: Averaged accuracy over the four test samples for each classifier given different predictor subsets.

Figure 6: Predicted U.S. Regimes by the AdaBoost Classifier with GDP Deflator



Note: Predicted U.S. Regimes according to the trained Boosted Trees classifier using GDP deflator as the inflation measure. Dark-shaded areas correspond to the fiscal dominance regime, while light-shaded areas belong to the monetary dominance regime. The black-dotted vertical lines represent the appointment date of the respective Fed Chairman.

Figure 7: Predicted U.S. Regimes by the AdaBoost Classifier with HP-Filter



Note: Predicted U.S. Regimes according to the trained Boosted Trees classifier using the HP filter to extract the cyclical component. Dark-shaded areas correspond to the fiscal dominance regime, while light-shaded areas belong to the monetary dominance regime. The black-dotted vertical lines represent the appointment date of the respective Fed Chairman.

5 Conclusion

Due to its non-observability, the exact interaction of monetary and fiscal policy is both an interesting research area as well as an area of mere speculation. Specifically in periods with high levels of government debt, central banks' independence might not be guaranteed. Our paper contributes to this literature and provides a new technique that allows to classify regimes in real-time using machine learning techniques. It is trained to *understand* state-of the art Markov-switching DSGE models with respect to the underlying regime and its predictions are based on the (nonlinear) interactions of (all) endogenous variables. Applying it to U.S. data, we corroborate the finding of fiscal dominance in the 1970s until early 1980s. The period around the stock market crash determines the next turning point from monetary to fiscal dominance, before the monetary dominance regime is prevalent from 1994 until mid 2001. We find mixed evidence for the more recent years. There is ample room for further research using machine learning techniques. Particular interesting avenues concerning this paper's application would be to incorporate the time-series dynamics of the data and to expand the analysis on other countries or currency unions.

References

L. Alessi and C. Detken. Identifying Excessive Credit Growth and Leverage. *Journal of Financial Stability*, 35:215–225, 2018.

M. M. Badia, F. Ohnsorge, P. Gupta, and Y. Xiang. Debt is not Free. *IMF Working Paper*, 2020.

J. Beutel, S. List, and G. Von Schweinitz. An Evaluation of Early Warning Models for Systemic Banking Crisis: Does Machine Learning Improve Predictions? *Deutsche Bundesbank Discussion Paper*, 2018.

S. Bhattarai, J. W. Lee, and W. Y. Park. Policy Regimes, Policy Shifts, and U.S. Business Cycles. *Review of Economics and Statistics*, 98(5):968–983, 2016.

F. Bianchi. Evolving Monetary/Fiscal Policy Mix in the United States. *American Economic Review*, 102(3):167–172, 2012.

F. Bianchi and C. Ilut. Monetary/Fiscal Policy Mix and Agents' Beliefs. *Review of Economic Dynamics*, 26:113–139, 2017.

F. Bianchi and L. Melosi. Escaping the Great Recession. *American Economic Review*, 107 (4):1030–1058, 2017.

L. Breiman. Bagging Predictors. *Machine Learning*, 24(2):123–140, 1996.

L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Chapman & Hall, 1984.

G. Calvo. Staggered Prices in a Utility Maximizing Model. *Journal of Monetary Economics*, 12(3):383–398, 1983.

X. Chen, E. M. Leeper, and C. Leith. US Monetary and Fiscal Policies - Conflict or Cooperation? Sire discussion papers, Scottish Institute for Research in Economics (SIRE), 2019.

T. Cover and P. Hart. Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

T. Davig and E. M. Leeper. Monetary - Fiscal Policy Interactions and Fiscal Stimulus. *European Economic Review*, 55(2):211–227, 2011.

R. E. Farmer, D. F. Waggoner, and T. Zha. Understanding Markov-switching Rational Expectations Models. *Journal of Economic Theory*, 144(5):1849–1867, 2009.

C. A. Favero and T. Monacelli. Fiscal Policy Rules and Regime (In)Stability: Evidence from the U.S. *IGIER Working Paper*, 2005.

Y. Freund and R. E. Schapire. A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5):771–780, 1999.

J. Friedman, T. Hastie, and R. Tibshirani. *The Elements of Statistical Learning*, volume 1. Springer series in statistics New York, 2001.

M. Gonzalez-Astudillo. Identifying the Stance of Monetary Policy at the Zero Lower Bound: A Markov-Switching Estimation Exploiting Monetary-Fiscal Policy Interdependence. *Journal of Money, Credit and Banking*, 50(1):115–154, 2018.

J. D. Hamilton. Why You Should Never Use the Hodrick-Prescott Filter. *Review of Economics and Statistics*, 100(5):831–843, 2018.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning*. Springer, 2013.

A. E. Khandani, A. J. Kim, and A. W. Lo. Consumer Credit-risk Models via Machine-learning Algorithms. *Journal of Banking & Finance*, 34(11):2767 – 2787, 2010.

M. Kliem, A. Kriwoluzky, and S. Sarferaz. On the Low-Frequency Relationship between Public Deficits and Inflation. *Journal of Applied Econometrics*, 31(3):566–583, 2016a.

M. Kliem, A. Kriwoluzky, and S. Sarferaz. Monetary âĂŞ fiscal policy interaction and fiscal inflation: A tale of three countries. *European Economic Review*, 88:158–184, 2016b.

E. M. Leeper. Equilibria Under Active and Passive Monetary and Fiscal Policies. *Journal of Monetary Economics*, 27(1):129–147, 1991.

F. M. Martin. Debt, Inflation and Central Bank Independence. *European Economic Review*, 79:129–150, 2015.

A. H. Meltzer. Politics and the Fed. *Journal of Monetary Economics*, 58(1):39–48, 2011.

S. Ng. Viewpoint: Boosting Recessions. *Canadian Journal of Economics*, 47(1):1–34, 2014.

T. J. Sargent and N. Wallace. Some Unpleasant Monetarist Arithmetic. *Quarterly Review*, 5:1–17, 1981.

S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

T. Teräsvirta, D. Van Dijk, and M. C. Medeiros. Linear Models, Smooth Transition Autoregressions, and Neural Networks for Forecasting Macroeconomic Time Series: A Re-examination. *International Journal of Forecasting*, 21(4):755–774, 2005.

N. Traum and S.-C. S. Yang. Monetary and Fiscal Policy Interactions in the Post-war U.S. *European Economic Review*, 55:140–164, 2011.

Appendix

A Choice of Hyperparameters

We select the hyperparameters by 10-fold cross-validation of the training data and Matlabs' Bayesian optimization routine. Tables 4- summarizes the chosen hyperparameters of each method.

Table 4: Optimized Hyperparameters

Method	**Hyperparameter**	**Value**
KNN	Number of neighbors K	10
Tree	Max. number of splits	232
RF	Number of bootstraps B	500
	Number of predictors to split on	9
	Min. leaf size	1
AdaBoost	Number of trees T	499
	Learning rate λ	0.56496
	Max. number of splits	144
SVM	Cost of violations to soft margin constraint C	988.21
	Kernel scale γ	3.2373

Note: Chosen hyperparameters by 10-fold cross-validation and Bayesian optimization. The number of boostraps B was chosen exogenously since it does not involve an overfitting issue.

Table 5: Optimized Hyperparameters with Inflation and Debt as the Only Predictors

Method	**Hyperparameter**	**Value**
KNN	Number of neighbors K	159
Tree	Max. number of splits	78
RF	Number of bootstraps B	500
	Number of predictors to split on	2
	Min. leaf size	53
AdaBoost	Number of trees T	11
	Learning rate λ	0.0013069
	Max. number of splits	77
SVM	Cost of violations to soft margin constraint C	1.5678
	Kernel scale γ	0.35848

Note: Chosen hyperparameters by 10-fold cross-validation and Bayesian optimization, when only inflation and debt are included as predictors. The number of boostraps B was chosen exogenously since it does not involve an overfitting issue.

Table 6: Optimized Hyperparameters with Inflation, Debt, Interest Rate and Government Spending as the Only Predictors

Method	**Hyperparameter**	**Value**
KNN	Number of neighbors K	40
Tree	Max. number of splits	132
RF	Number of bootstraps B	500
	Number of predictors to split on	4
	Min. leaf size	25
AdaBoost	Number of trees T	84
	Learning rate λ	0.98072
	Max. number of splits	14
SVM	Cost of violations to soft margin constraint C	1.1414
	Kernel scale γ	1.168

Note: Chosen hyperparameters by 10-fold cross-validation and Bayesian optimization, when only inflation, debt, interest rate and government spending are included as predictors. The number of boostraps B was chosen exogenously since it does not involve an overfitting issue.

B Log-Linearized Model Equations

$$
\begin{aligned}
c_t &= c_{t+1} - (r_t - E_t \pi_{t+1}) \\
n_t &= w_t - c_t \\
w_t &= y_t + mc_t - n_t \\
n_t &= y_t - a_t + \frac{\tau^l_{ss}}{1-\tau^l_{ss}} \tau_t \\
a_t &= \rho_a a_{t-1} + \epsilon^a_t \\
\pi_t &= \beta \pi_{t+1} + \frac{(1-\theta)(1-\beta\theta)}{\theta} mc_t \\
y_t &= \frac{C_{ss}}{Y_{ss}} c_t + \frac{G_{ss}}{Y_{ss}} g_t \\
b_t &= \frac{1}{\beta}(b_{t-1} - \pi_t + r_{t-1}) + \frac{G_{ss}}{B_{ss}} g_t - \frac{\tau^l_{ss} W_{ss} N_{ss}}{B_{ss}} (\tau^l_t + w_t + n_t) \\
\tau^l_t &= \rho_t \tau^l_{t-1} + \epsilon^t_t \\
g_t &= \rho_g g_{t-1} - \delta_y y_t - \delta_b b_{t-1} + \epsilon^g_t \\
r_t &= \rho_r r_{t-1} + \phi_\pi \pi_t + \phi_y y_t + \epsilon^r_t
\end{aligned}
$$

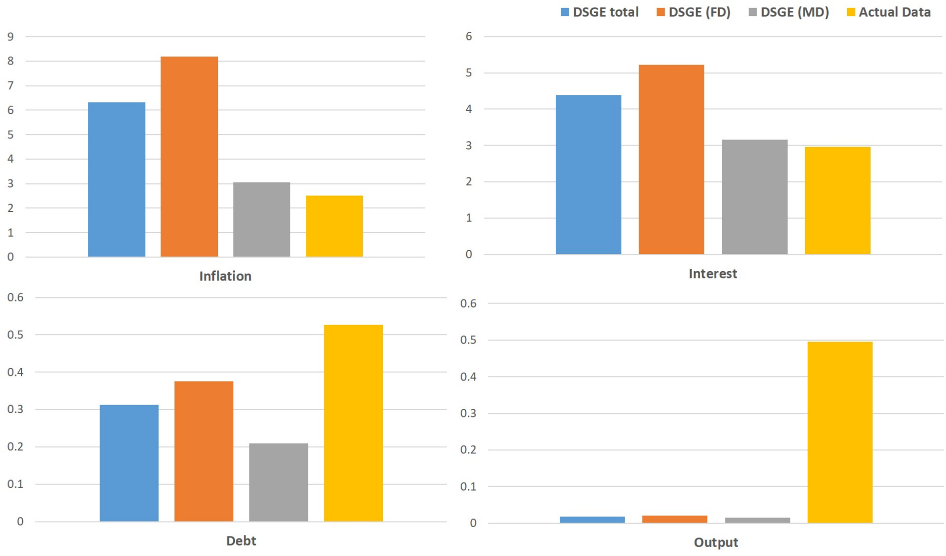
C Data

Table 7: Actual U.S. Data

Variable	Data	Description/ Transformations	Source
Output	Real GDP	Bill. of Chained 2012 Dollars, Seasonally Adjusted	FRED/ BEA
Consumption	Personal Consumption Expenditures	Bill. of Dollars, Seasonally Adjusted	FRED/ BEA
Labor	Weekly Hours worked: Manufacturing for the U.S.	Quarterly Averages	FRED/ OECD
Wages	Compensation of Employees, Received: Wage and Salary Disbursements	Billions of Dollars, Seasonally Adjusted	FRED/ BEA
Techn.	Total Factor Productivity (TFP)	Annual Data linearly interpolated	AMECO
Interest Rate	Effective Federal Funds Rate	Percent	FRED
Inflation	Consumer Price Index: Total All Items for the U.S.	Growth Rate Same Period Previous Year	FRED / OECD
	GDP: Implicit Price Deflator	Index 2012=100, Seasonally Adjusted	
	Personal Consumption Expenditures: Chain-type Price Index	Change, Index 2012=100, Seasonally Adjusted	
Gov. Spend.	Gov. Total Expenditures	Bill. of Dollars, Seasonally Adjusted	FRED/ BEA
Taxes	Federal Gov. Current Tax Receipts	Bill. of Dollars, Seasonally Adjusted	FRED/ BEA
Gov. Debt	Federal Debt: Total Public Debt	Percent of GDP Seasonally Adjusted	FRED/ OMB

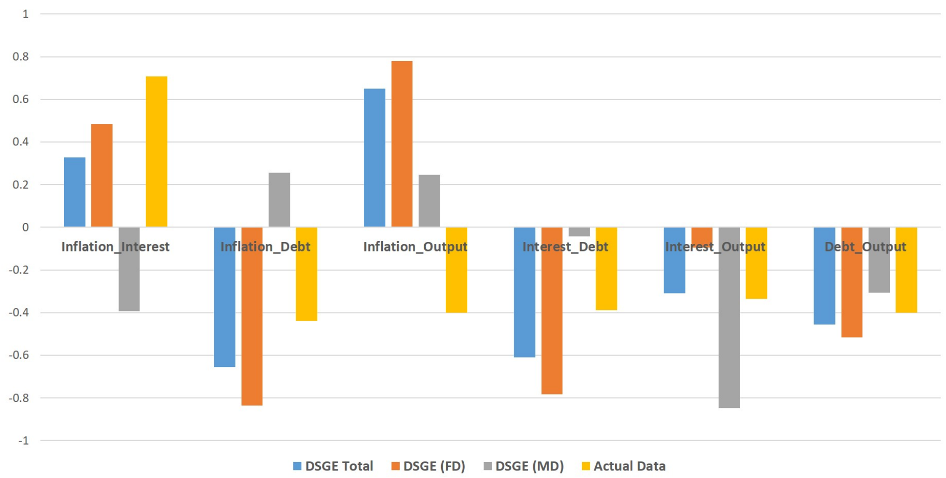
Note: The table represents actual U.S. data that is used for classifying historical MD/FD regimes. The sample period is 1966:Q1-2017:Q4.

Figure 8: Standard Deviations of Selected Variables from Actual and Simulated Data



Note: The figures show standard deviations of inflation, interest rate, debt and output from actual U.S. data (yellow) and from simulated MS-DSGE data. Orange (grey) bars belong to simulated MS-DSGE data conditional on the fiscal (monetary) dominance regime, while the blue ones correspond to the total simulated data set. All variables represent deviations from steady state.

Figure 9: Correlations between Selected Variables from Actual and Simulated Data



Note: The figures show correlations between the variables inflation, interest rate, debt and output from actual U.S. data (yellow) and from simulated MS-DSGE data. Orange (grey) bars belong to simulated MS-DSGE data conditional on the fiscal (monetary) dominance regime, while the blue ones correspond to the total simulated data set. All variables represent deviations from steady state.

D Confusion Matrices for Test Datasets

Baseline

Logit

True		Predicted FD	Predicted MD
True	FD	0.50	0.03
True	MD	0.44	0.03

KNN

True		Predicted FD	Predicted MD
True	FD	0.41	0.12
True	MD	0.01	0.46

Tree

True		Predicted FD	Predicted MD
True	FD	0.46	0.07
True	MD	0.01	0.46

RF

True		Predicted FD	Predicted MD
True	FD	0.53	0.00
True	MD	0.00	0.47

AdaBoost

True		Predicted FD	Predicted MD
True	FD	0.53	0.00
True	MD	0.00	0.47

SVM

True		Predicted FD	Predicted MD
True	FD	0.51	0.02
True	MD	0.00	0.47

Table 8: Confusion matrices across methods (rows) corresponding to test dataset from baseline MS-DSGE model (column).

Logit

	Calvo (Predicted)		Sharper Response (Predicted)		Transition (Predicted)	
True	FD	MD	FD	MD	FD	MD
FD	0.26	0.27	0.26	0.27	0.22	0.22
MD	0.23	0.24	0.23	0.24	0.30	0.26

KNN

	Calvo (Predicted)		Sharper Response (Predicted)		Transition (Predicted)	
True	FD	MD	FD	MD	FD	MD
FD	0.40	0.13	0.40	0.13	0.34	0.10
MD	0.01	0.46	0.01	0.46	0.02	0.54

Tree

	Calvo (Predicted)		Sharper Response (Predicted)		Transition (Predicted)	
True	FD	MD	FD	MD	FD	MD
FD	0.43	0.10	0.43	0.10	0.34	0.10
MD	0.06	0.41	0.05	0.42	0.06	0.50

RF

	Calvo (Predicted)		Sharper Response (Predicted)		Transition (Predicted)	
True	FD	MD	FD	MD	FD	MD
FD	0.47	0.06	0.47	0.06	0.39	0.06
MD	0.04	0.43	0.01	0.46	0.02	0.53

B-Trees

	Calvo (Predicted)		Sharper Response (Predicted)		Transition (Predicted)	
True	FD	MD	FD	MD	FD	MD
FD	0.49	0.03	0.50	0.03	0.41	0.03
MD	0.00	0.47	0.00	0.47	0.03	0.53

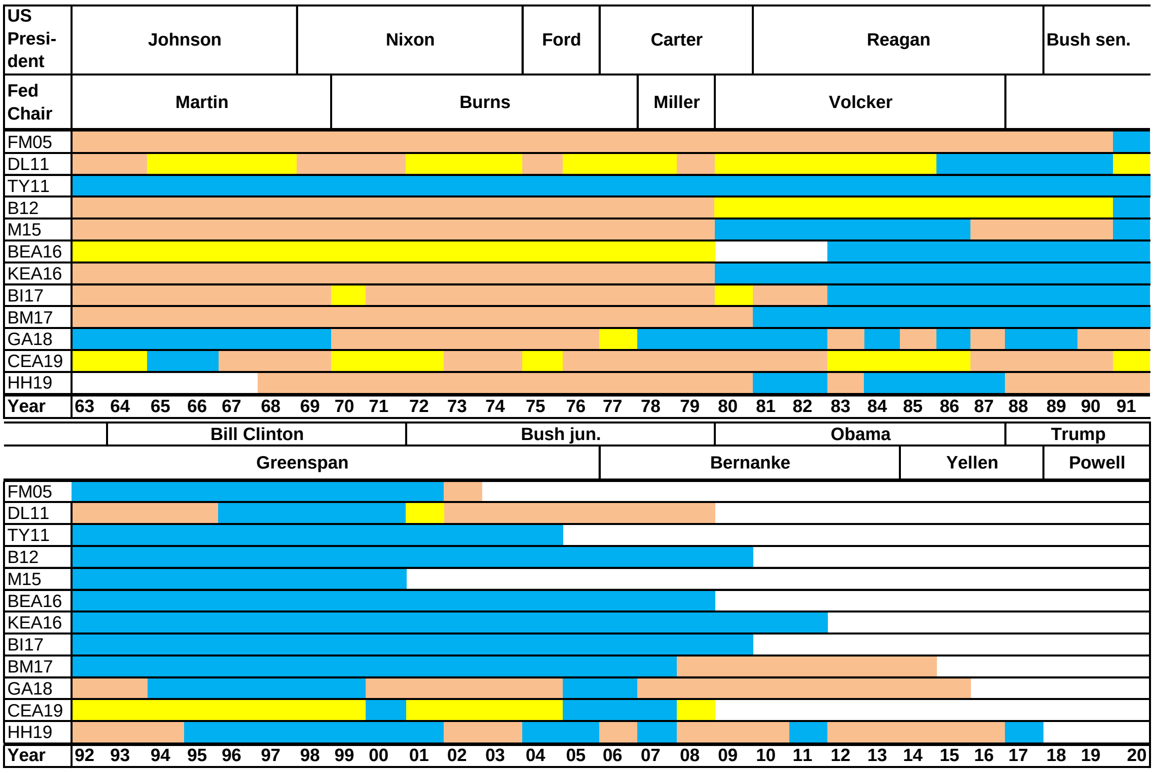
SVM

	Calvo (Predicted)		Sharper Response (Predicted)		Transition (Predicted)	
True	FD	MD	FD	MD	FD	MD
FD	0.42	0.11	0.43	0.10	0.43	0.02
MD	0.00	0.47	0.01	0.46	0.29	0.26

Table 9: Confusion matrices across methods (rows) corresponding to different test datasets from modified MS-DSGE models (columns).

E Predicted U.S. Regimes

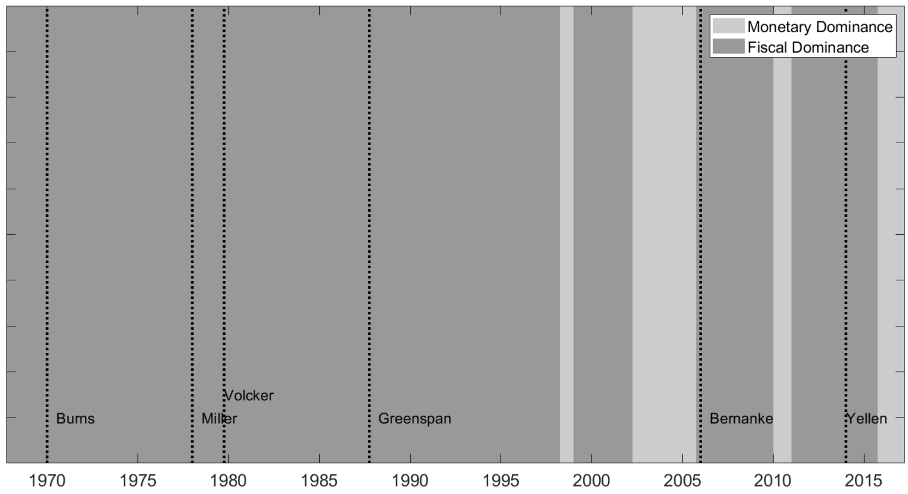
Table 10: Comparison of U.S. Regimes Found in the Literature



Note: This table compares the historical U.S. regimes found in the literature over the period from 1963-2020. It neither claims completeness nor an exact representation of the results since some of them are drawn by eye-balling of the respective graphs or aggregated over quarterly results. Blue-marked (orange-marked) periods denote monetary (fiscal) dominance. Yellow-marked periods comprise either indetermined, explosive or optimal fiscal policy regimes, while white means that there are no results available. Besides the U.S. presidents and Fed chairs, each row corresponds to the findings of a different paper. The abbreviations denote:

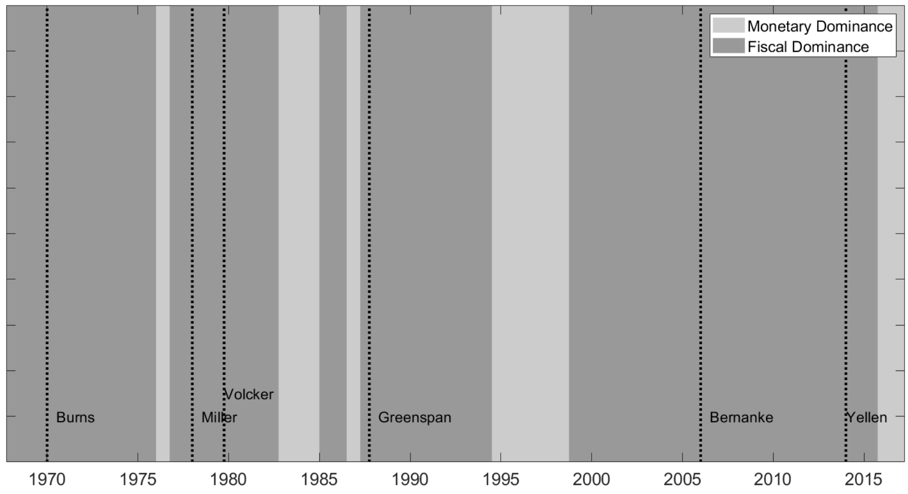
FM05 = Favero and Monacelli (2005),
DL11 = Davig and Leeper (2011),
TY11 = Traum and Yang (2011),
B12= Bianchi (2012),
M15 = Martin (2015),
BEA16 = Bhattarai et al. (2016),
KEA16 = Kliem et al. (2016a),
BI17 = Bianchi and Ilut (2017),
BM17 = Bianchi and Melosi (2017),
GA18 = Gonzalez-Astudillo (2018),
CEA19 = Chen et al. (2019) and
HH19 = Hinterlang and Hollmayr (2019)(the findings of this paper).

Figure 10: Predicted U.S. Regimes by the RF Classifier



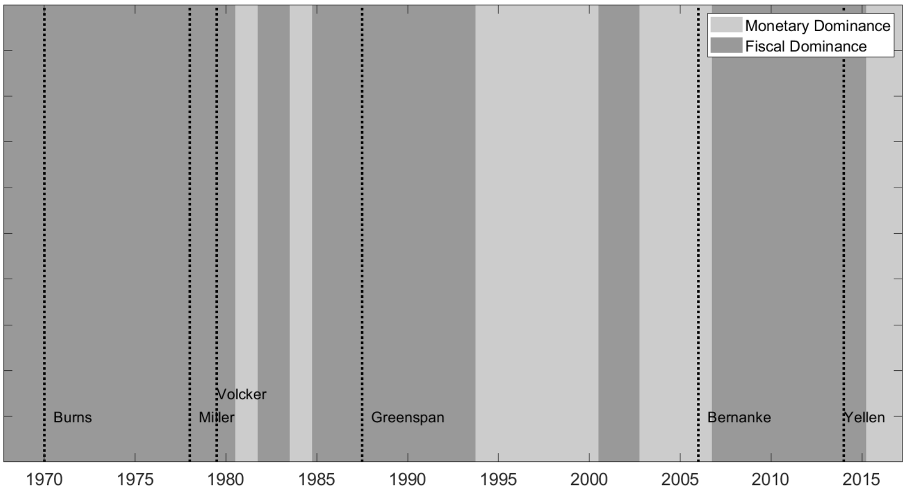
Note: Predicted U.S. Regimes according to the trained Random Forest classifier. Dark-shaded areas correspond to the fiscal dominance regime, while light-shaded areas belong to the monetary dominance regime. The black-dotted vertical lines represent the appointment date of the respective Fed Chairman.

Figure 11: Predicted U.S. Regimes by the SVM Classifier



Note: Predicted U.S. Regimes according to the trained SVM classifier. Dark-shaded areas correspond to the fiscal dominance regime, while light-shaded areas belong to the monetary dominance regime. The black-dotted vertical lines represent the appointment date of the respective Fed Chairman.

Figure 12: Predicted U.S. Regimes by the AdaBoost Classifier with PCE Inflation



Note: Predicted U.S. Regimes according to the trained Support Vector Machine classifier using PCE as the inflation measure. Dark-shaded areas correspond to the fiscal dominance regime, while light-shaded areas belong to the monetary dominance regime. The black-dotted vertical lines represent the appointment date of the respective Fed Chairman.