DICE

DISCUSSION PAPER

NO 416

Forecasting Recessions in Germany with Machine Learning

Philip Rademacher

September 2024

hhu
Heinrich Heine
Universität Düsseldorf

IMPRINT

DICE DISCUSSION PAPER

Published by:
Heinrich-Heine-University Düsseldorf,
Düsseldorf Institute for Competition Economics (DICE),
Universitätsstraße 1, 40225 Düsseldorf, Germany
www.dice.hhu.de

Editor:
Prof. Dr. Hans-Theo Normann
Düsseldorf Institute for Competition Economics (DICE)
Tel +49 (0) 211-81-15125, E-Mail normann@dice.hhu.de

All rights reserved. Düsseldorf, Germany 2024.

ISSN 2190-9938 (online) / ISBN 978-3-86304-415-2
The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editor.

Forecasting Recessions in Germany with Machine Learning

Philip Rademacher[*]

September 2024

Abstract:

This paper applies machine learning to forecast business cycles in the German economy using a high-dimensional dataset with 73 indicators, primarily from the OECD Main Economic Indicator Database, covering a time period from 1973 to 2023. Sequential Floating Forward Selection (SFFS) is used to select the most relevant indicators and build compact, explainable, and performant models. Therefore, regularized regression models (LASSO, Ridge) and tree-based classification models (Random Forest, and Logit Boost) are used as challenger models to outperform a probit model containing the term spread as a predictor. All models are trained on data from 1973-2006 and evaluated on a hold-out-sample starting in 2006. The study reveals that fewer indicators are necessary to model recessions. Models built with SFFS have a maximum of eleven indicators. Furthermore, the study setting shows that many indicators are stable across time and business cycles. Machine learning models prove particularly effective in predicting recessions during periods of quantitative easing, when the predictive power of the term spread diminishes. The findings contribute to the ongoing discussion on the use of machine learning in economic forecasting, especially in the context of limited and imbalanced data.

Keywords: Business Cycles, Recession, Forecasting, Machine Learning

[*]Contact Information: Philip Rademacher, Heinrich-Heine University Düsseldorf, philip.rademacher@hhu.de

1. Introduction

After the German economy had just recovered from its pandemic-related downturn, the risk of a new recession has risen dramatically since the war in Ukraine. Recessions are associated with potential welfare losses, making them an act of strength for the population and policymakers. For this reason, many papers have analyzed indicators and methods that can be used to forecast business cycles at an early stage.

In recent years, machine learning (ML) techniques have been applied to model business cycles in the United States. These studies were often able to find new indicators or improve the forecasting power when compared to conventional probit models. However, the research remains diverse, with no clear consensus favoring a specific ML technique. The contributions differ in their results depending on the classification method, indicators, and validation strategy used. In addition, the question of how many indicators are necessary to produce accurate predictions is ambiguous.

In contrast to the majority of literature that focuses on the United States, this contribution relies on a high-dimensional dataset for Germany spanning from 1973 to 2023. The major objectives are to analyze ML's potential to outperform the predictive power of simple probit models, the number of indicators required to achieve reliable predictions, and the stability of these models across different business cycles.

Regarding the classification techniques, this paper examines the predictive power of LASSO, Ridge, Random Forest, and Logit Boost in comparison to a probit model. For the purpose of determining the number of indicators, a Sequential Floating Forward Selection (SFFS) is applied. This algorithm builds models starting from the strongest indicator and adds more indicators as long as the forecasts can be improved. A key advantage of this approach is that the selection of indicators is validated on the training data, which avoids a look-ahead bias. In contrast to previous contributions, the performance is evaluated on a large hold-out sample period starting in 2006, while the entire process of model optimization and indicator selection is confined to historical data spanning from 1973 to 2005. SHAP values are used to assess indicators' stability over time.

2. Literature Review

This article treats predicting recessions as a classification task. Consequently, the literature review does not cover studies that forecast recessions by modeling continuous variables, such as GDP growth.

The yield curve and the term spread, or the difference between short-term and long-term interest rates, play a crucial role in forecasting recessions. Since the seminal paper by Estrella and Hardouvelis (1991) a negative term spread is regarded as the most reliable early warning signal of an imminent recession. Their univariate probit models were built on the observation that prior to many US-recessions short-term rates were higher than long-term rates. Many research contributions confirmed the predictive power of the term spread in other countries (see e.g. Chinn & Kucko, 2015) or improved the models by adding more indicators, such as interest rates, stock prices, bond spreads, price trends or survey-based indicators on consumer or business confidence (e.g., Estrella & Mishkin, 1998, Gilchrist & Zakrajsek 2012, Christiansen et al., 2014, Pönka, 2017, Hasse & Lajaunie, 2020, Nissilä, 2020). Simultaneously, researchers discussed methodological extensions of static forecast models, such as the dynamic model (Nyberg, 2010) that accounts for the autocorrelation structure, and a business-cycle-specific model (Chauvet & Potter, 2005) that allows the variance to vary across cycles.

In economic research, probit models have traditionally been the tool of choice. However, they have the disadvantage, that they can handle only a few indicators simultaneously. Therefore, either only a few indicators are analyzed in combination, or high dimensional datasets are reduced to common factors and then included in models. Examples of this approach include the works of Christiansen et al. (2014), Chen, Iqbal & Lai (2011), and Pönka (2017).

In recent years, there has been a shift towards the application of machine-learning (ML) techniques. Numerous algorithms have been applied to business cycle forecasting, including regularized regressions, neural networks, and tree-based ensemble learners. However, the literature has so far mainly focused on the United States. We can broadly categorize the ML literature into two groups based on their goals: those that apply machine learning to discover previously unknown early warning indications and those that try to increase the prediction ability compared to probit models.

The first category's contributions build on the model's ability to focus on important indicators for their prediction. Ng (2014) employed boosting to sift through more than a hundred indicators for the USA and identify strong ones. She concluded that less than ten indicators had a high predictive power, including numerous bond spreads. However, the predictive power of single indicators varied greatly between business cycles and with the forecast horizon. More recently, Nevasalmi (2022) implemented a similar strategy, taking into account the data imbalance through cost-sensitive learning. His dataset was larger due to the inclusion of numerous lags. Among more than 1000 indicators, he could confirm the high predictive power of interest rate spreads. The contributions by Choi et al. (2023) and Cadahia Delgado et al. (2022) address a more specific research question. They investigate whether the term spread, defined as the difference between a 10-year and a 3-month term, represents the most optimal maturity combination. While Choi et al. (2023) confirmed the high predictive power of the traditionally used combination with the help of a penalized logistic regression, Cadahia Delgado et al. (2022) identified the difference between a 6- and 3-month term as most useful.

Contributions in the second category use ML-techniques to challenge the forecasts of some benchmark models. The ability of ML-techniques to improve forecasts relies on their capability to select relevant features and their higher resilience to overfitting. Non-parametric methods can also achieve improvements through better recognition of non-linearities and interaction effects between indicators. The methods used in the literature are very diverse and range from boosting (Berge 2015, Döpke et al. 2015, 2017) and support vector machines (Plakandaras et al. 2017) to a large-scale comparison of different methods (Vrontos et al. 2021, Yazdani 2020, Psimopolous 2020, Zyatkov & Krivorotko 2021). Although many contributions refer to the USA, a direct comparison of the results is not possible due to the different datasets and research settings. However, in all studies, one ML-technique outperformed the benchmark. For example, Vrontos et. al. (2021) achieved the best predictions with penalized logit regression, k-nearest neighbours, and Bayesian generalized linear models, while Yazdani (2020) with Random Forest and Zyatkov & Krivorotko (2021) with neural networks. Applications for Germany are the contributions by Döpke et al. (2015, 2017), that applied Logitboost, and Psimopoulus (2020), which implemented several tree-based ensemble methods as well as k-nearest-neighbors and support vector machines.

The contributions by Puglia & Tucker (2020, 2021) critically examine the performance of ML methods and highlight that their performance depends on the chosen validation strategy. Specifically, they explore how ML methods perform based on the number of indicators used. In an exhaustive search, they evaluated 106 combinations of indicators across six different classification methods, employing two different validation strategies. This comprehensive analysis resulted in a total of 1272 trained models. When using k-fold cross-validation, ML methods achieved better performance than the classical probit model. However, when considering a validation strategy that accounts for the time series structure, the probit model outperformed ML methods.

The general problem with k-fold cross-validation is that the time series data is shuffled, so the models receive information on future recessions during training. While this is less problematic if the trained models are subsequently evaluated completely out-of-sample, a mixed approach—where data partitions overlap for tuning and evaluation—often leads to overoptimistic results. Unfortunately, many contributions in the ML literature on recession forecasting evaluate performance entirely in-sample, relying on metrics from the validation folds. This paper considers Puglia & Tucker's critique and implements a more restrictive research setting (see chapter 5). Throughout the entire process, the models never receive future information for training.

3. Data

3.1 Recessions

The concept of recessions lacks a universally accepted definition. Heileman (2019) emphasizes an easily overlooked fact: business cycles and recessions are statistical constructs, whose temporal classifications vary greatly depending on the underlying definition. This is why numerous definitions exist side by side.

The National Bureau of Economic Research (NBER), a research institute that dates business cycles in the United States, defines a recession as "a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales." (NBER, 2008). Their definition is widely used in the empirical literature for the USA.

The present article uses the business cycles released by the German Council of Economic Experts (GCEE), whose approach is closely related to the NBER's methodology (GCEE, 2023). In Germany, they identified seven recessions for the period from 1950 to 2023. The last recession started in the last quarter of 2019 and overlapped with the lockdown during the coronavirus pandemic in 2020. Figure 1 provides an overview of the recession phases and also shows their relationship to GDP growth. One advantage of this definition of recession is that they provide the exact start and end month of a recession.

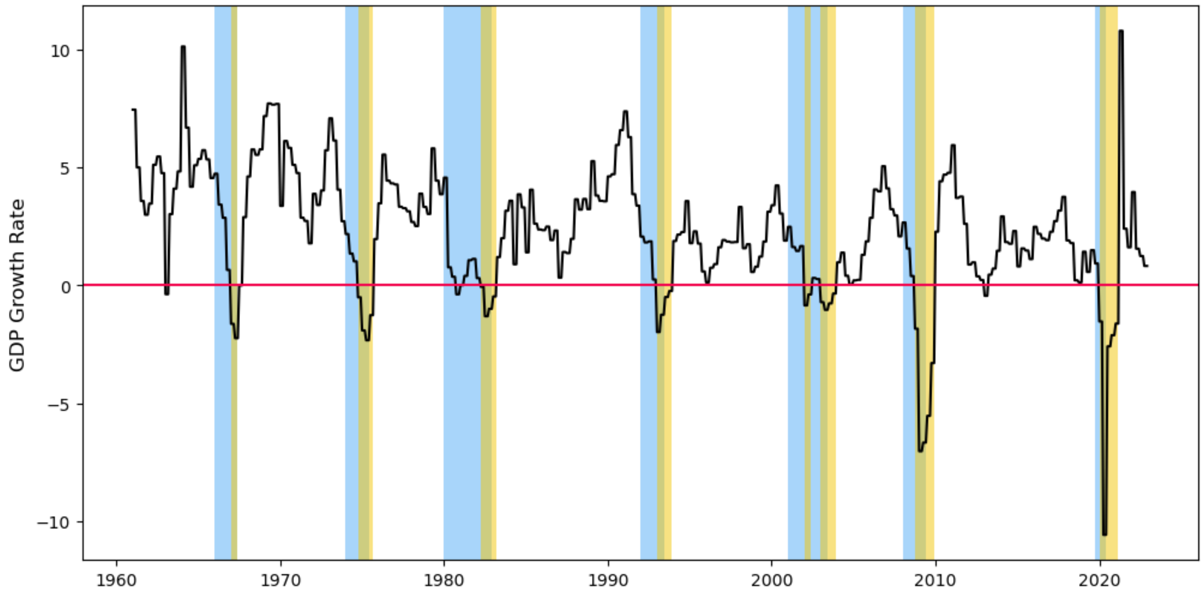


Figure 1: The periods highlighted in blue indicate recessions as classified by the German Council of Economic Experts, whereas the periods highlighted in orange indicate periods with negative GDP growth rates in two subsequent quarters. The black line shows the GDP growth rates compared to the same period in the previous year.

The recession definitions provided by expert boards, such as the GCEE, are not strictly rule-based and therefore contain an element of subjectivity. An alternative concept is the technical recession, which occurs when GDP decreases in two subsequent quarters. Both concepts of recession overlap to some extent, as Figure 1 shows. The blue periods indicate GCEE recessions, while the orange periods represent technical recessions. It is noticeable that the GCEE recessions start earlier than the decline in GDP, namely immediately after the peak of the growth rates, and end at their trough. In contrast, technical recessions are much shorter. They start later but last a little longer, as the recovery of the growth rates is still included as long as they are negative. This observation implies the following: Models that can predict a recession twelve months in advance are also able to predict the GDP decline, but much earlier than the recession.

3.2 Indicators

To forecast recessions, this paper uses a large set of predictors, predominantly from the OECD Main Economic Indicators Database. Ideally, indicators should fulfill the following requirements: They should be (1) released monthly in real time, (2) not subject to major revisions, and (3) available for a long period of time. (See also: Döpke 2015, p. 43)

The last requirement is applied to ensure that the dataset has as many past recessions as possible. In this paper, the starting year was set at 1973 to ensure that the consumer confidence indicator, a well-known recession indicator, is available for the whole study period.[1]

The other requirements aim to prevent the use of future information for forecasting (look-ahead-bias). Specifically, the restriction on monthly data excludes artificially interpolated monthly values derived from quarterly data. Interpolation shifts future information into the past, where it was not originally known. When indicators are revised or not published in real time, a similar problem can arise. In these cases, the forecasts would be based on figures that had not even been published in this form at the time of the forecast. For those reasons, GDP growth rates are excluded, as such data is published quarterly and subject to subsequent revisions. Industrial production has also been removed because revisions are frequently carried out here. To identify indicators that are published with delay, the initially uploaded data from 2021-2022 was analyzed. In order to keep many indicators in the dataset, real-time publication is defined generously. Only indicators with a publication delay exceeding, on average, two months were excluded from the sample.

Further indicators were added to the study sample due to their frequent use in the literature. These include two stock indices, namely the DAX and the SP 500. Although the latter primarily reflects US

[1] Indicator values before 1990 refer to West Germany.

economic trends, some contributions have observed a predictive power for recessions in other countries. (e.g., Nissilae 2020) As the DAX was first launched in 1988, retrospectively calculated values were used for the previous period. (Stehle et al., 1996, p. 24)

Field	Examples
Financial	Long-term interest rates, Interbank Rates, Term Spread
International	Real Effective Exchange Rates, Export Orders
Labor Market	Unemployment Rate, Job vacancies
Prices	Oil Prices, Housing-Prices, Energy Prices
Real Economy	Orders, Producer Price Index
Stock Market	Standard & Poors 500, DAX
Survey-Based	Business Situation, Orders Inflow, Consumer Confidence

Table 1: Thematic overview of indicators and selected examples.

Furthermore, the growth rate of the M1 and M3 monetary aggregates in the eurozone, oil prices, the US term spread, the difference between the German and US term spread and some US bond spreads were added to the study sample. The US term spread has frequently been used to successfully forecast recessions in other countries, and the difference between the US and domestic term spread has proven to be a reliable indicator to predict recessions in Germany (Nyberg 2010). Table 1 provides an overview of the indicators' thematic coverage; Appendix A1 contains a complete list of indicators with data sources and applied transformations.

In general, indicators are included as differences or percentage changes from the previous year. Exceptions are indicators whose values are stationary over time, such as interest rates or survey-based indicators with a fixed scale. Indicators that were indexed to a certain year are only included as percentage changes in the study sample.

4. Classifiers

Predicting recessions is a classification task, as the recession indicator y_t is a binary variable with the values:

$$y_t = \begin{cases} 1, & \text{if recession in period } t \\ 0, & \text{if expansion in period } t \end{cases} \tag{1}$$

All classification models estimate the probability of a recession occurring within one year in the future, as a function $F(..)$ of some indicators X_{t-h} at the time of the forecasts. The models are calibrated with a threshold of 50%, i.e. probabilities higher than this value indicate a future recession.

$$\Pr(y_t = 1|X_{t-h}) = F(X_{t-h}) \tag{2}$$
$$\text{if } \Pr(y_t > 0.5|X_{t-h}) \text{ then } \hat{y}_t = 1$$

Two groups of classifiers will be used to forecast recessions.

On the one hand, probit and regularized logistic regression are from the group of generalized linear models, and on the other hand, Random Forest and Logit Boost are ensemble classifiers that are based on decision trees. While linear models have the advantage of being easy to interpret due to their coefficients, they are restricted by their parametric structure. Interaction effects and non-linearities must be specified ex ante. Tree-based methods, on the other hand, can model non-linear relationships and interaction effects between indicators easily due to their non-parametric structure.

4.1 Logistic and Probit Regression

Both, logistic and probit regression, parameterize the relationship between the indicators and the recession variable in a linear way with coefficients ($\hat{y}_t = \beta_0 + \beta^T x_{t-h}$). The main difference to the linear model is that both binary regression models use a sigmoid function, ensuring the predictions are probabilities bounded between 0 and 1. As a result, both regression models lack a closed form and therefore the estimation is done iteratively with maximum likelihood. The logistic regression uses the cumulative distribution function of the logistic distribution as the sigmoid function, while the probit model uses the same of the standard normal distribution.

This study uses probit regression only for the benchmark model, with the term spread as a single predictor. For the challenger models, built with sequential forward selection, two variants of logistic regression are applied. Since a logistic regression with several indicators can quickly overfit, both variants add a penalty to the maximization problem (formula 3), which ensures that coefficients of less relevant indicators are closer to zero. Two penalties are used, namely the L1 penalty (also called Least Absolute Shrinkage, LASSO), and the L2 penalty (Ridge). The LASSO model minimizes the sum of the absolute values of the coefficients (formula 4, Tibshirani, 1996), while the Ridge model minimizes the sum of the squared coefficients (formula 5, Hoerl & Kennard, 1970).

$$\max_{\beta_0, \beta} \sum_{i=1}^{N} \left[y_t(\beta_0 + \beta' x_{t-h}) - \log\left(1 + e^{\beta_0 + \beta' x_{t-h}}\right) \right] \tag{3}$$

$$-\lambda \sum_{i=1}^{p} |\beta_j| \tag{4}$$

$$-\lambda \sum_{i=1}^{p} {\beta_j}^2 \tag{5}$$

4.2 Random Forest

Random Forest (Breimann, 2001) consists of numerous randomized decision trees (S). The randomness results from the fact, that each of the trees was trained on a different random subsample of the training data. Every subsample only contained a fixed number of observations and indicators, whereby observations could be included in the sample more than once. The random sampling leads to unique trees, that are largely independent of each other.

Random Forest

For $s = 1,2, \dots S$:

a) Sample observations with replacement from the training data.

b) Train a decision tree, where each split considers a random subset of all available features.

Classify new observations by taking the majority class vote of the S trees.

Random Forest is an ensemble method that combines single decision trees with majority voting: Each decision tree gives a prediction and if the majority (e.g. 60%) of the decision trees predict a recession, the final recession probability would be 60%. Combining many random decision trees with the principle of majority voting, makes Random Forest more resilient to overfitting (Hastie et al., 2009, pp. 587-588).

4.3 Logit Boost

The algorithm was developed by Friedman et al. (2000) as a statistical view of Adaboost. They examined that Adaboost can be understood as an additive logistic regression model. The loss function was the main difference between Adaboost and additive logistic regression, as Adaboost minimizes an exponential loss function. With their Logit Boost algorithm, they bring boosting even closer to logistic regression by implementing a logistic loss function.

Logit Boost (Friedman et al. 2000)

1. Start with equal weights for all periods $w_t = \frac{1}{T} \; \forall \; t = 1 \dots T$ and an initial guess for the recession probability $p(x_{t-h}) = 0.5$ and $F(x) = 0$.
2. Repeat for $m = 1,2, \dots M$:

   a.) Compute the differences between the predicted probabilities and the observed outcome (z_t) and the resulting weights.

$$z_t = \frac{y_t - p(x_{t-h})}{p(x_{t-h})(1 - p(x_{t-h}))}$$

$$w_t = p(x_{t-h})(1 - p(x_{t-h}))$$

   b.) Fit the function $f_m(x_{t-h})$ on z_t using the weights w_t. In the implementation f_m is a regression tree stump.

c.) Update $F(x_{t-h}) \leftarrow F(x_{t-h}) + lf_m(x_{t-h})$ with the learning rate l. Update probabilities from the new $F(x_{t-h})$:

$$p(x_{t-h}) \leftarrow \frac{e^{F(x_{t-h})}}{e^{F(x_{t-h})} + e^{-F(x_{t-h})}}$$

The basic idea of boosting is to combine several weak learners (M) in a sequential way to a strong classifier: Starting with an initial prediction that is the same for all periods, i.e. a recession probability of 50%, a regression tree with only one node is used to explain the difference between the initial prediction and the observed states. However, as the tree is a stump, it can only slightly improve the initial predictions. The new predictions are calculated by adding the predicted difference of the first tree multiplied with the learning rate to the initial prediction. More trees will be added to further reduce the difference between the predictions and the observed outcome. This process continues until there is no difference anymore or a fixed number of trees is reached. The final prediction is calculated by adding all the trees predictions multiplied with the learning rate to initial prediction.

5. Empirical Design

In the first step, the time series is divided into a train and a hold-out sample. The train sample spans from 1973 to December 2005 and includes four recession phases, while the hold-out sample begins in January 2006 and includes two recession phases. The hold-out sample is only used to evaluate the model's predictive power.

The train sample is used to select relevant indicators, tune hyperparameters, and train the final models. An expanding window with three passes is applied to evaluate the selected features and hyperparameters. Therefore, the models are trained on a subsample of the train sample and then tested on a subsequent subsample of the training dataset (which are called the training and validation fold). In the first pass, the training fold covers the period from 1974 to 1983, and the validation fold covers the period until 1992. In the second pass, the training fold is extended by the validation fold from the first pass, and a new subsample is used for validation (period between 1992 and 2001). In the last run, the validation fold covers the period from 2001 to 2005. Figure 2 illustrates the training and validation folds, which were built in such a way that two conditions were fulfilled: First, the initial training fold contains one complete business cycle, and second, the validation folds of the first and second passes contain the year before the subsequent recession (but not the recession itself).

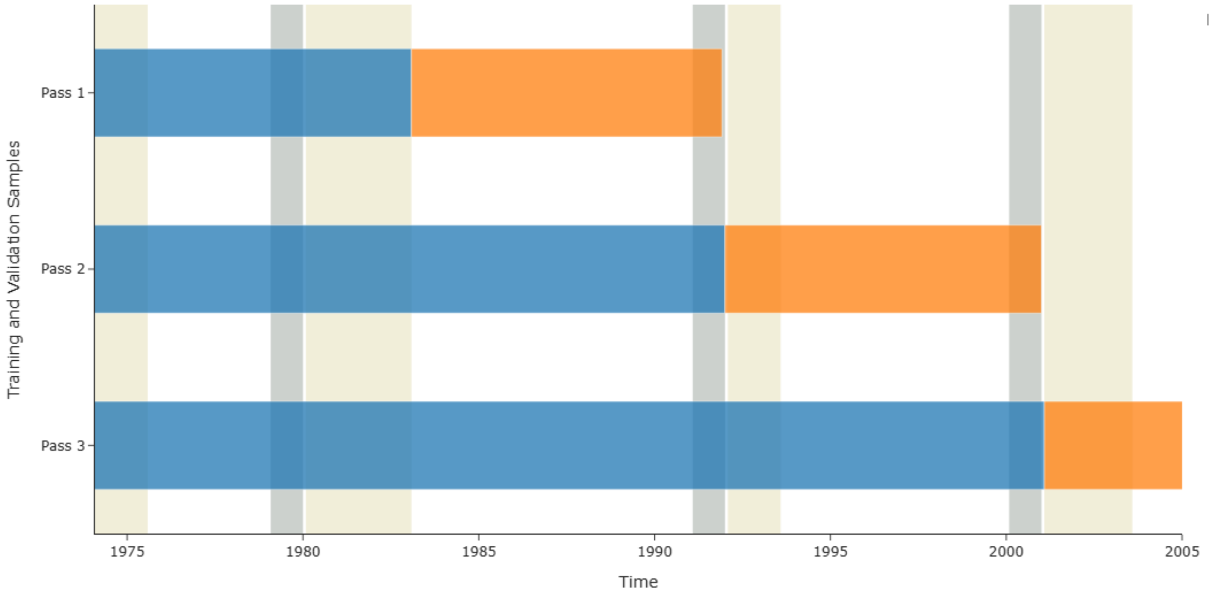


Figure 2: Implementation of the expanding window: Blue periods indicate the training folds, and orange areas represent the validation folds. The validation folds in the 1st and 2nd passes include the twelve months (dark grey) before the subsequent recession period (light grey).

This way, each combination of indicators from the sequential floating forward selection and the hyperparameters are validated. Finally, the performance metrics of the different validation folds are averaged, and the hyperparameter or indicator set with the highest mean performance is selected.

The model's performance is measured using the negative Brier score (eq. 6). The Brier score is the mean squared deviation between the model prediction $\hat{y}_t$ and the actual state y_t for all observations in the validation fold T. If the predictions are always close to the actual state, the model performance is high, and the Brier score is small. The negative Brier score is therefore maximized.

$$BS = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \tag{6}$$

The expanding window approach ensures that the chronology of the observations is always retained, which allows the calculation of in-sample metrics for the different validation folds, i.e., the business cycles between 1983 and 2006. The metrics there tend to be higher than in the hold-out sample, as the validation folds were used to optimize the models. The metrics therefore primarily reflect the model's ability to adapt to historical data and do not allow any conclusions about their predictive power in the future.

To assess the model's predictive power, only data after 2006 is used. Due to this strict separation, the out-of-sample results are worse than in related contributions that use a hybrid setting (e.g., Vrontos et al., 2021) or evaluate the performance entirely in the validation samples of the rolling window (e.g., Nevasalmi, 2022).

5.1 Feature Selection

In this study, the embedded ability of the ML models to focus on relevant predictors is augmented by a wrapper method, which is called Sequential Forward Floating Selection (SFFS). (Pudil et. al 1994) This selection method is often used for classifiers that lack regularization and therefore cannot perform embedded feature selection. (Raschka & Mirjalili, 2017) The method requires two components to select relevant indicators: a classifier (such as logistic regression or Random Forest) and data to train and evaluate the predictive power of the models.

The algorithm is initiated by training and evaluating classification models using a single indicator. The indicator with the best performance is selected. After that, each remaining indicator is added to the selected indicator, and models with two indicators are trained and validated. The indicator that has the highest performance, in conjunction with the indicator from the initial step, is added to the feature set. This step is repeated until the model performance no longer improves. The whole process is known as sequential forward selection. In the floating variant used here, after each iteration it is checked whether a feature can be removed from the feature set without a decrease in performance.

The procedure is computationally complex and is therefore often referred to as a "brute-force" or "greedy" method. The selected variables are also nested, i.e. the method develops a strong path dependency and builds local-optimal models, as the best feature is selected conditionally on the already selected variables. However, it has two advantages: First, it selects variables based on their performance within the selected classifier. Therefore, the selected indicators differ between the classifiers. Non-parametric classifiers can rely on different variables than parametric ones. Second, the indicators are selected based on their performance in the validation fold. During the selection process, an indicator therefore has to prove its forecasting power on new data, which reduces the risk of overfitting. (Guyon & Elisseeff, 2003) This is primarily intended to prevent the selection of indicators that only work well for one business cycle but do not have a general predictive power.

5.2 Hyperparameter Tuning

The ML-methods have various hyperparameters that must be defined prior to model training. In this study, for Random Forest, the depth of the trees, the size of the samples, and the number of features are tuned. For Logit Boost, the learning rate and the number of trees, and for LASSO/ Ridge, the strength of the penalty are defined within a tuning. The hyperparameter tuning is conducted using the hyperparameter optimization framework Optuna (Akiba et al. 2019).

5.3 Out-of-Sample Evaluation

Various metrics are used on the out-of-sample data set to assess the forecasting power of the final models. Most of the metrics can be calculated using four numbers that describe the model accuracy:

- True Positives (TP): Number of correctly predicted recessions
- False Positives (FP): Number of incorrectly predicted recessions
- True Negatives (TN): Number of correctly predicted expansion periods
- False negatives (FN): number of incorrectly predicted expansion periods

Recall (eq. 7) shows the percentage of recessions that the model was able to predict correctly. Conversely, a 70% recall indicates that the model incorrectly classified 30% of the recession months as expansions.

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

Precision (eq. 8) is used to measure the reliability of a recession forecast. It shows the proportion of true recessions to the total number of predicted recessions.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

The F1 score (eq. 9) is a composite measure that combines precision and recall using the harmonic mean. This is helpful as there is a trade-off between recall and precision. When a model predicts all recessions, it often results in a loss of precision, as it needs to label more periods as recessions in order to achieve a high recall.

$$F1 = 2\frac{Precision * Recall}{Precision + Recall} \tag{9}$$

The area under the curve (AUC) is the only metric that is not derived from the number of correctly and incorrectly classified periods but from the ROC curve. An AUC of 1 implies that the model has relevant explanatory power, whereas 0.5 indicates that the model is comparable to random chance. The AUC primarily illustrates the alignment of recession probabilities with actual recessions, i.e., whether they increase before a recession and decrease before an expansion period. However, the AUC does not evaluate the final predictions based on the cut-off value. A model with a high AUC and low values of the other metrics implies that the model would be better if we change the cut-off value. The difficulty is that the models are trained with a cut-off-value of 50% and when forecasting future recessions, we would assume that this cut-off-value is optimal. Changing the cut-off value to improve predictions on the hold-out sample would not be practical, as we would not be able to determine the ideal cut-off value ex ante (look ahead bias). For this reason, the performance evaluation focuses on the other metrics.

5.4 SHAP

SHAP values are calculated to analyze the indicators' impact on recession forecasts. The theory of SHapley Additive exPlanations (SHAP) originates from cooperative game theory and was used to distribute the outcome of a game fairly among a coalition of players (Shapley, 1997).

The Shapley value for a player i (eq. 10) is the sum of the weighted marginal contributions of player i to all possible coalitions S of other players which the player can join. The marginal contribution is the outcome of the coalition including player i $F_{S\subset\{i\}}(x_{S\subset\{i\}})$ minus the outcome without player i $F_S(X_S)$. The individual marginal contributions are weighted by a fraction. $|F|!$ represents the number of coalitions that can be formed with all players. $|S|!$ is the number of players in the coalition and $(|F| - |S| - 1)!$ the number of possibilities in which further players can join the coalition after play i has already joined.

$$\Phi_i = \sum_{S\subseteq F\{I\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[F_{S\subset\{I\}}(X_{S\subset\{I\}}) - F_S(X_S)\right] \tag{10}$$

This theoretical framework was transferred to machine learning as a model-agnostic explainer. (Lundberg & Lee, 2017). In Machine Learning, SHAP values have two important properties that are desirable: (1.) The SHAP values of all indicators always add up to the model prediction. (2.) SHAP values are consistent across various models, allowing for a comparison of the impact of indicators across models such as LASSO and Logit Boost, and business cycles such as the recessions of 1992 and 2008.

6. Results

6.1 Sequential Floating Forward Selection (SFFS)

In SFFS, the classification models reach their optimum very quickly. After that, their performance either does not improve anymore or decreases. The performance of Random Forest in the validation samples already starts to decrease after the inclusion of the fifth indicator. Ridge, LASSO, and Logit Boost processed a larger number of indicators (namely 7, 10, and 11) before their performance reached an optimum. Basically, the results indicate that there are a few strong indicators that drive up performance at the beginning, but then there are no more that can further improve the predictive power in later iterations.

Figure 2 illustrates how the model's performance improves or deteriorates with the inclusion of more indicators. The shaded areas show the optimal number of indicators for the different classifiers. As a (negative) Brier score closer to zero indicates better model performance, LASSO slightly outperforms Ridge and Logit Boost. Random Forest has the lowest performance in the validation samples.

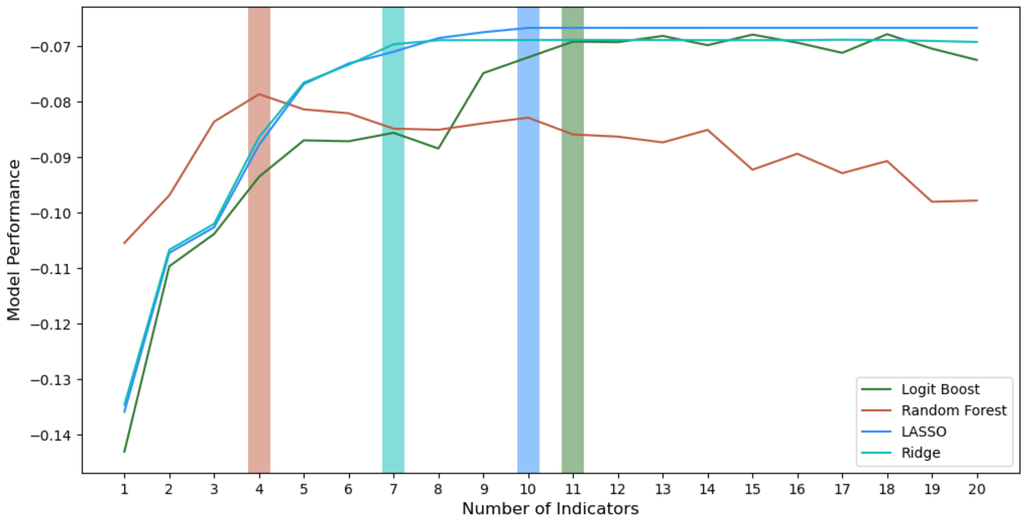


Figure 3: Model performance after each iteration of the SFFS, measured as negative Brier-Score.

Random Forest	Logit Boost	RIDGE	LASSO
Call Money/Interbank Rate	Business Situation (DF)	Business Confidence Indicator (DF)	Consumer Confidence
Long-term interest rates (10 years)	Finnished Goods Stocks (DF)	Consumer Confidence	Export Orders
Term Spread	Orders Inflow	Selling Prices (DF)	Order Books (DF)
Unemployment rate (SA)	Orders Inflow (DF)	BAA vs. 10-Year Treasury (DF)	Orders Inflow (DF)
	Selling Prices (DF)	Term Spread Germany – Term Spread USA (DF)	Selling Prices (DF)
	AAA vs. 10-Year Treasury (DF)	Term Spread USA (DF)	Term Spread USA (DF)
	Monetary Aggregate M1 (%)	Energy Prices (%)	Term Spread Germany – Term Spread USA (DF)
	Term Spread Germany – Term Spread USA (DF)		BAA vs. 10-Year Treasury (DF)
	Term Spread USA (DF)		Call Money/Interbank Rate
	Energy Prices (%)		Energy Prices (%)
	Housing Prices (%)		

Table 2: Indicators selected by SFFS. Ordered by category and its frequency: ● Financial Markets, ● Prices, ● Business Tendency / Consumer Confidence Surveys, ● Labor Market. Indicators with "%" denote percentage changes, and with DF differences to the same month in the previous year.

Table 2 shows which indicators were selected in the SFFS. The indicators are sorted according to their category, which is color-coded. The category with the most indicators in the model is placed at the top. The Random Forest model primarily uses indicators from the financial market that are well-known from the literature. These include the term spread, a long-term interest rate (10 years), and a short-term interest rate (interbank interest rate, less than 24 hours). The sole indicator from a different thematic field is the unemployment rate in a seasonal adjusted form.

In contrast, the other model consists of between 7 and 11 indicators. The three models can be summarized as follows: They contain indicators from three thematic fields. All models include energy prices (Logit Boost also includes housing prices). In addition, there is a roughly equal number of survey-based indicators and financial market indicators. These differ to some extent. Selling prices appear in all three models, and in 2 out of 3 models, the consumer confidence indicator, orders inflow, the US term spread, the difference between the US and German term spread, and a bond spread are included. It is very noticeable that the German term spread was not selected as an indicator, but either the US term spread or the difference to it. The Random Forest model differs from the other models by not using dynamic indicators, meaning it does not select any changes or differences from the previous year.

In summary, the models are very diverse regarding their number of indicators included and the indicators themselves. Random Forest employs a relatively small indicator set, predominantly from the financial markets. Conversely, the other models rely on approximately twice as many indicators, with an additional focus on expectations and (energy) prices.

6.2 Model Performance

The evaluation of the models' predictive power is conducted in two steps. First a visual analysis of the plotted recession probabilities is applied, and then a more mathematical approach with some performance measures. The plots in Figure 3 show the probabilities for a recession occurring in twelve months. A probability exceeding the red line, which indicates a 50% threshold, leads to a recession prediction. The yellow areas correspond to recessions shifted 12 months backwards. These regions denote the periods during which a model should ideally predict a recession.

The primary focus of the model evaluation lies on the forecasts generated for the hold-out sample, starting from the black line in 2006, as this data is unknown to the models. For the sake of completeness, the plots in Figure 3 also show the recession probabilities from the validation periods obtained in the expanding windows. However, the recession probabilities prior to 2006 may exhibit an overly optimistic bias, as feature selection and hyperparameter tuning were conducted here.

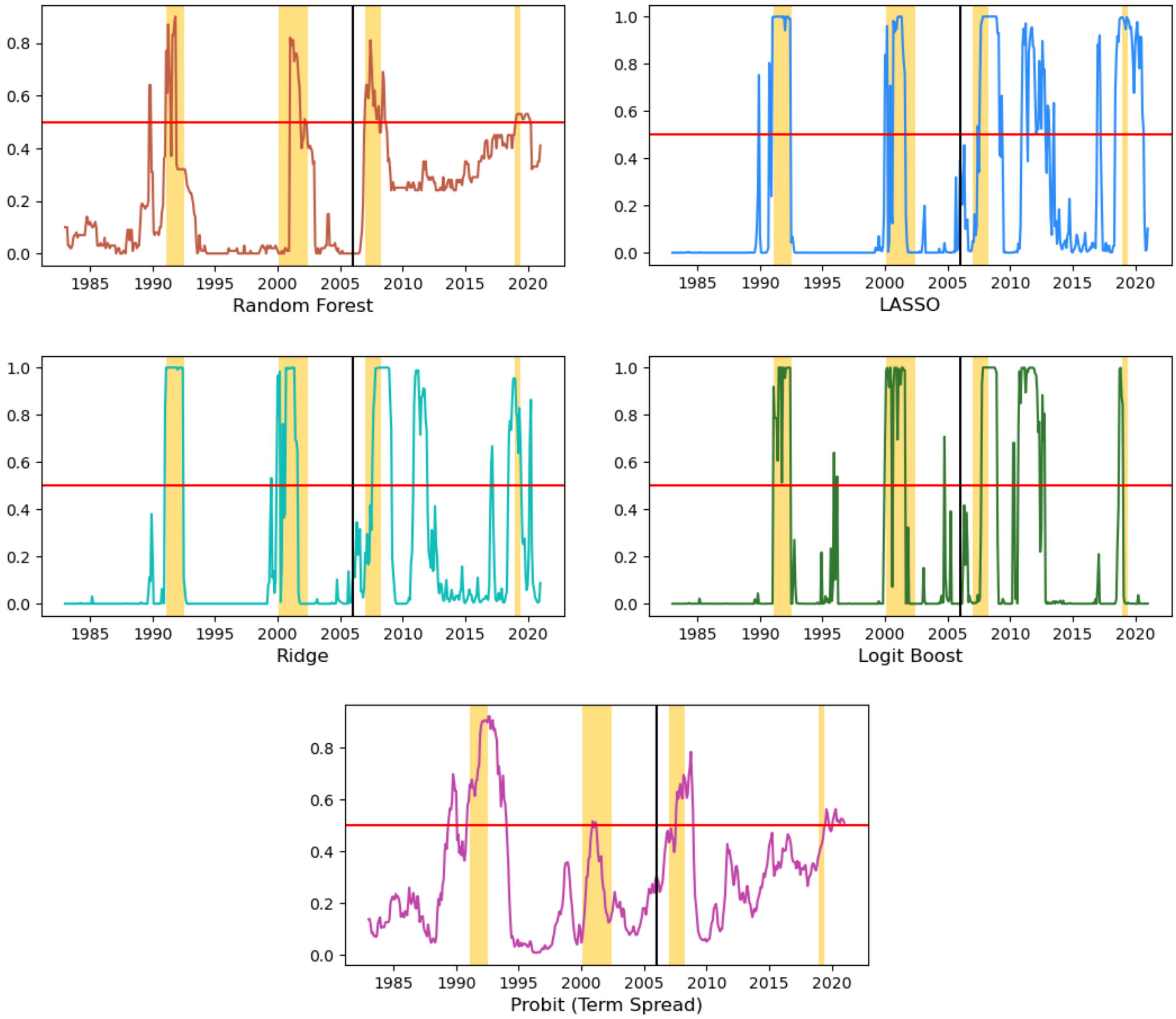


Figure 4: Recession Probability Plots

When looking at the plots, one can see how well the model predictions match with the actual state of the economy. All four ML-models predict the true recessions during the periods highlighted in yellow; only the Probit model fails to predict the 2019 recession. During most of the expansion phases, the models predict expansions. There is one major exception: LASSO, Ridge, and Logit Boost predict a completely fictitious recession between 2010 and 2012. In addition, all models give recession signals during the coronavirus pandemic, even though the official recession is already over in the second half of 2020.

The timeliness of the forecasts also plays a role: For the recession starting in January 2008, Random Forest made its initial prediction in January 2007, while Ridge predicted the recession for the first time in June 2007. In August, Probit and LASSO sent the first recession signals, followed by Logit Boost in October. For the recession starting in October 2019 the forecasts of LASSO and Logit Boost are timely, predicting a recession in October 2018 or earlier. In contrast, Random Forest predicted the recession in February 2019, and the Probit model in June 2019 for the first time. In terms of timeliness, the ML-methods outperformed the benchmark model before the recession in 2019.

The performance metrics in Table 3 show that the ML models can explain the business cycles in the validation folds (1983 - 2005) significantly better than the probit model. However, as expected, the out-of-sample performance is lower than in the validation folds. In the out-of-sample period, Random Forest, LASSO, and Ridge outperform the probit model.

Classifier	**In-Sample** 1983/01 – 2005/12				**Out-Of-Sample** 2006/01 – 2020/12			
	AUC	**F1**	**PRE**	**REC**	**AUC**	**F1**	**PRE**	**REC**
Random Forest	**0.81**	**0.59**	**0.91**	**0.44**	**0.96**	**0.71**	**0.66**	**0.78**
LASSO	**0.95**	**0.81**	**0.94**	**0.71**	**0.84**	**0.40**	0.27	**0.78**
Ridge	**0.94**	**0.81**	**0.94**	**0.71**	**0.87**	**0.53**	**0.42**	**0.74**
Logit Boost	**0.89**	**0.83**	**0.92**	**0.75**	0.76	0.36	0.29	0.48
Probit	0.78	0.42	0.42	0.42	0.84	0.37	0.33	0.41

Table 3: In- and Out-Of-Sample Performance

Table 4 shows the out-of-sample performance separated by recession phases, namely 2008 and 2019. It can be seen that the 2008 recession can be forecasted very well by the probit model. Only Random Forest can improve the term spread prediction. There are two reasons for this: First, the 2008 recession was caused by the financial crisis. Models that rely exclusively on financial market indicators have a clear advantage. The recession forecasts of LASSO, Ridge, and Logit Boost are worse, as financial market indicators only contribute partly to their forecasts. Secondly, the poor performance of these models stems not only from their pre-recession predictions in 2008, but also from their prediction of a fictitious recession in 2010. Although Ridge recognizes just as many months of the 2008 recession as the probit model (both models have the same recall), it has a lower precision, F1 and AUC due to the misclassified expansion phase in 2010-2011.

For the 2019 recession, however, all ML-models performed better than the benchmark model. The latter loses a lot of predictive power due to quantitative easing. Probit and Random Forest, which rely solely on financial market indicators, now have a weaker performance than the more diversified models.

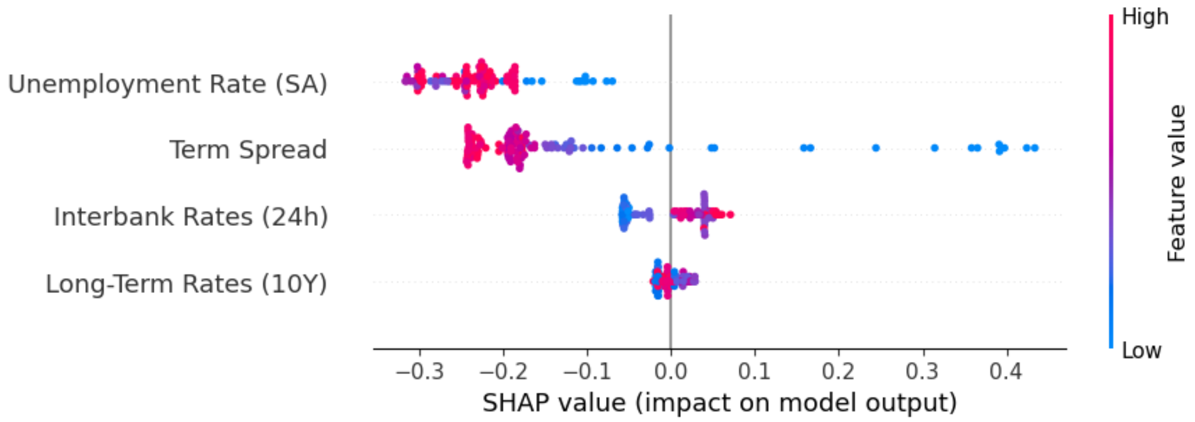
Classifier	Out-Of-Sample							
	2008 2006/01 – 2014/12				**2019** 2015/01 – 2020/12			
	AUC	**F1**	**PRE**	**REC**	**AUC**	**F1**	**PRE**	**REC**
Random Forest	**0.99**	**0.89**	**0.89**	**0.89**	**0.85**	**0.43**	**0.36**	**0.56**
LASSO	0.75	0.38	0.26	**0.67**	**0.99**	**0.45**	**0.29**	**1.00**
Ridge	0.81	0.44	0.34	0.61	**0.97**	**0.72**	**0.56**	**1.00**
Logit Boost	0.68	0.31	0.22	0.50	**0.89**	**0.62**	**1.00**	**0.44**
Probit	0.94	0.63	0.65	0.61	0.51	0.00	0.00	0.00

Table 4: Out-Of-Sample Performance by recession period.

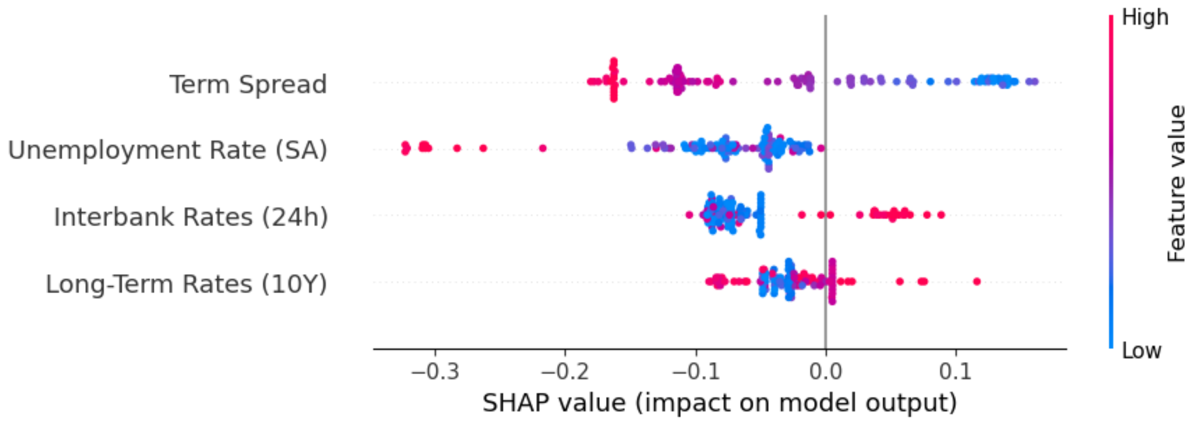
6.3 Forecasting Power of Indicators

After analyzing the model's performance, the focus is now on the indicators and their impact on the model's forecasts. For this purpose, SHAP-values are calculated and visualized with the help of Beeswarm plots. These plots display the SHAP values for each indicator across different observations in the dataset. Each point on the plot corresponds to a specific month, whereby the color illustrates the related indicator value: e.g., blue for low values and red for high values. The horizontal axis shows the size of the SHAP value: A negative SHAP value implies that the indicator reduces the probability of a recession in a given month, while a positive SHAP value implies an increase. The indicators are listed according to their mean absolute SHAP value.

A. Recession 1992



B. Recession 2008



C. Recession 2019

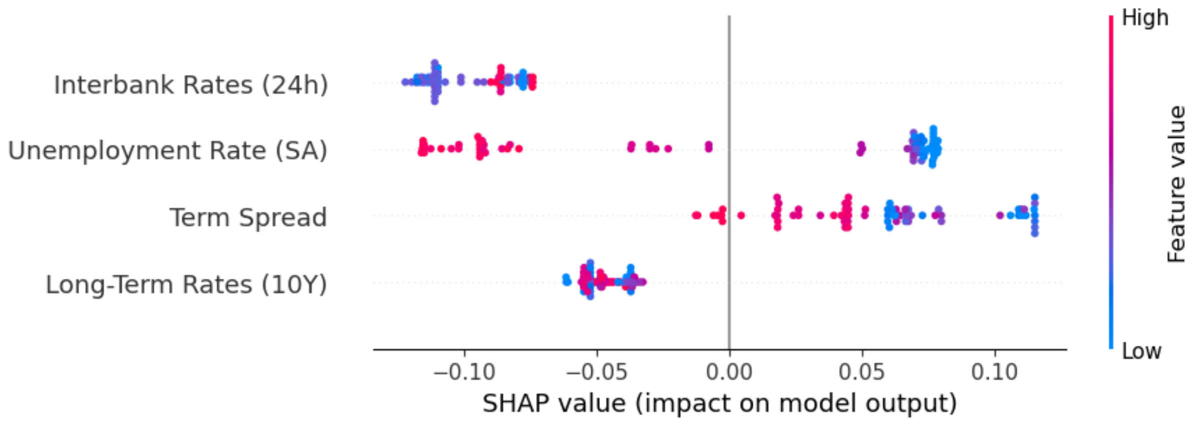


Figure 5: Beeswarm-Plots for Random Forest

In the Random Forest model (Figure 5, Appendix A2), the unemployment rate plays an important role during the business cycles of the validation sample. The model interprets high unemployment rates during recessions as a signal for an upcoming expansion and low rates as a signal for an impending recession. Since 2005, however, the unemployment rate has been falling continuously and has lost much of its forecasting power.

The term spread is another important indicator for the model. The colors in the plot, suggest the well-known linear relationship between the SHAP values and the term spread. The recession risk increases when the term spread is small or negative and decreases when the term spread is large and positive. For the recession forecasts of 2008, the term spread is the strongest indicator. However, this indicator also lost some of its predictive power between 1992 and 2019. This becomes evident when comparing the range of positive SHAP values. In 1992, the term spread had SHAP values between -0.25 and 0.4, depending on the month. In 2019, the SHAP values range only between -0.025 and 0.15.

This loss of importance is related to quantitative easing, which started in 2015. Given the persistently low interest rates, the term spread remains near zero throughout the entire period (Figure 6). There are no large and positive term spreads anymore, which helped the model to distinguish expansions from recessions. As a result, the term spread has permanently increased the risk of recession. Since 2015, there have only been a few months in which the term spread has reduced the recession risk.

Quantitative easing also changes the predictive power of interest rates, albeit in the opposite direction. High interbank rates, which served as a recession warning signal until 2015, are suppressed by QE. The low short-term rates consequently reduce the recession risk between 2015 and 2021. Both effects are responsible for the persistently high recession probabilities (roughly 30%) in the last business cycle, and the fact that we could only predict the 2019 recession with probabilities slightly above the threshold.

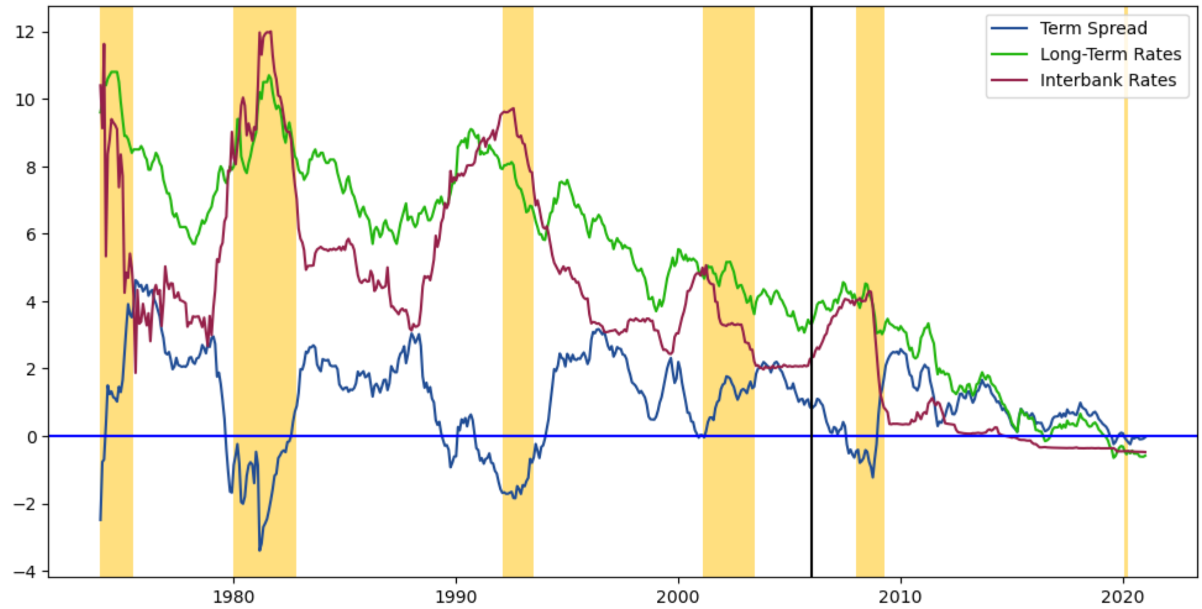


Figure 6: Financial Indicators over Time

While Random Forest depends heavily on the term spread, the other models are much more diversified in their indicators. In the other three models, energy prices are the most important early warning indicator. All models interpret high price increases as a recession signal. In Logit Boost, energy prices always have the highest mean absolute SHAP value and in LASSO and Ridge the indicator is in every recession period among the three most important indicators.

Apart from the fact that all models use energy prices, every model uses roughly the same proportion of survey-based and financial market indicators to improve model predictions. The Beeswarm plots for the individual models can be found in the appendices (A3-A5). Figure 7 shows the most important indicators of the models, as well as how their forecasting power changed over time. The forecasting power is measured by the mean absolute shap value. For the purpose of comprehensiveness, indicators with a permanent mean absolute SHAP value of less than 0.05 are omitted. For Random Forest, the changes have already been discussed in detail using the Beeswarm plots. Figure 7 displays the main effects, namely the loss of forecasting power of the term spread and unemployment rate.

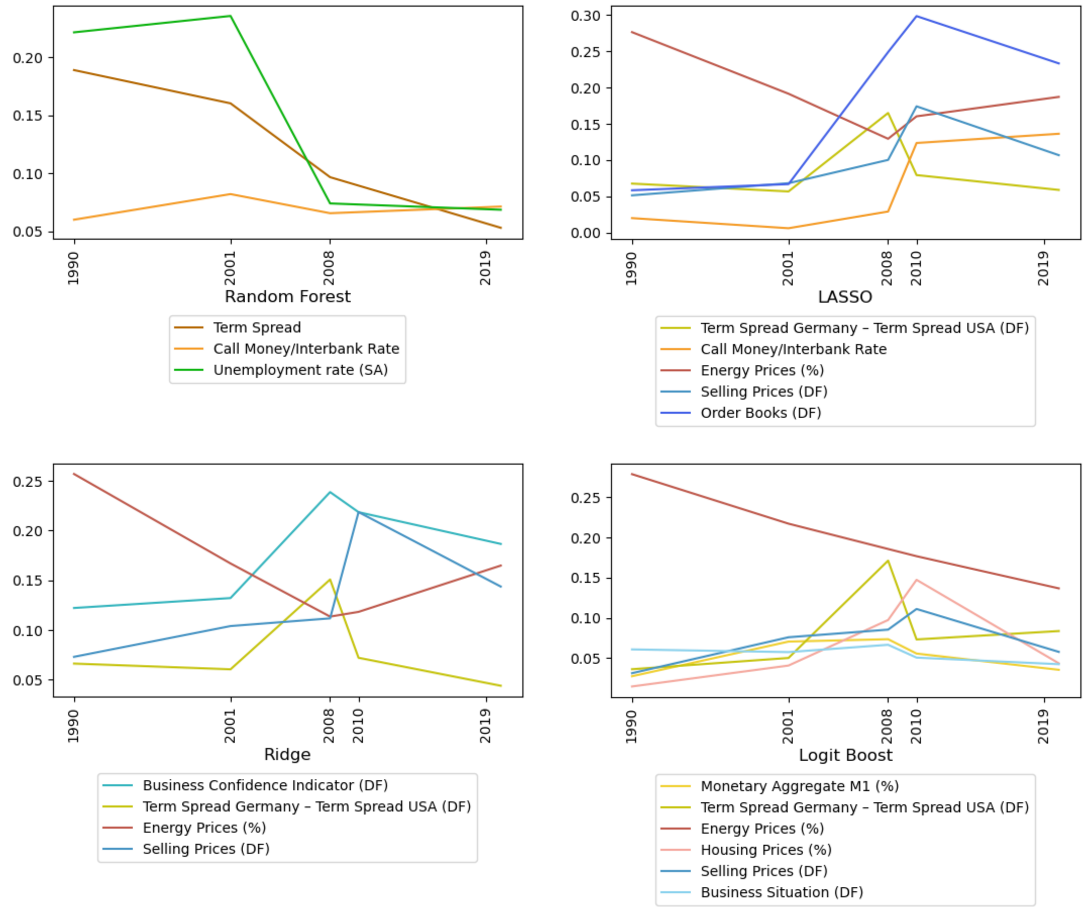


Figure 7: Mean absolute SHAP values for indicators over time. The figure is restricted to indicators with a minimum mean absolute SHAP value of 0.05. The calculations were carried out for the following periods: 1983-1991, 1992-2001, 2006-2015 (only Random Forest), 2006-2009, 2009-2015, 2015-2020.

In Logit Boost, energy prices are the most important indicator but their contribution to the model forecasts decreases over time. Before the 2008 recession, the term spread gained significant relevance. Housing prices also became important for the prediction of the recession in 2008 and the (fictious) recession in 2010. The other indicators fluctuate slightly without a discernible trend. With the exception of the three indicators mentioned, great stability can be detected.

A similar trend can be seen in Ridge and LASSO with regard to energy prices. Over the years, the importance of energy prices has decreased. Both models heavily rely on a financial market indicator for the 2008 recession, such as the term spread, or the term spread difference. From 2008 onwards, we can recognize an increase in the importance of survey-based indicators like selling prices, order books, or the business confidence indicator. The models interpret high increases in selling prices, a decline in order books and a loss of confidence in the business situation as recession signals. In LASSO, the influence of short-term interest rates on the model forecast increases. However, this effect is solely due to quantitative easing and the permanently low interest rates. It leads to a permanent reduction in the recession risk, which is not meaningful at all.

The figures show that – in contrast to Random Forest – the other models have at least one strong indicator in every study period. This is a positive characteristic of these models, which yields to forecasts that are always very clear and do not fluctuate close to the cut-off value of 50 %, which was the case in the Random Forest model. However, one disadvantage of these models is that they predict a fictitious recession.

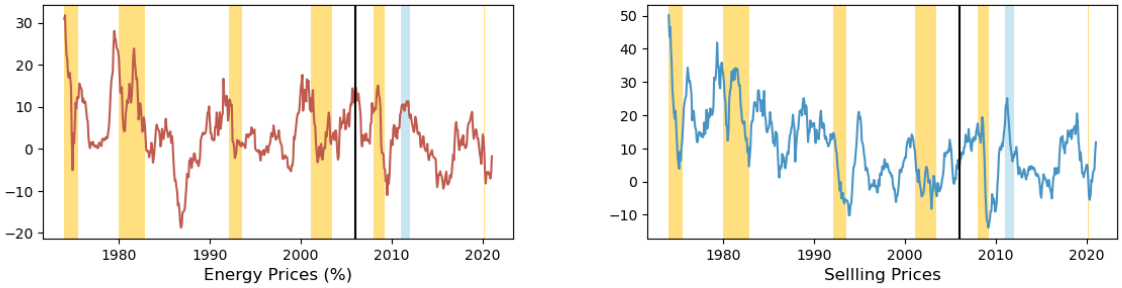


Figure 8: Energy Prices and Business Confidence over Time.

Between 2010 and 2012 all three models gave false recession signals, which mainly resulted from the increasing selling and energy prices (Figure 8). It should be noted that one year after the recession forecasts, GDP growth rates were very low (below 1%) and even negative in the first quarter of 2013. Although this period is not defined as a recession by the GCEE, it is close to a technical recession. Heilemann (2019) discusses Germany's various recession phases and also refers to the period from April 2012 to April 2013 as recession. This fictional recession is therefore not necessarily a bad prediction. Moreover, it shows that the models are very sensitive and already give recession signals when the GDP growth rates are small and negative, but not yet long enough for a recession.

In a final assessment of the models, it can be argued that LASSO and Ridge are best suited to forecast recessions. Although they performed not as well as Random Forest in the 2008 recession, in most recessions they performed better due to the diversified selection of indicators. This is also evident when considering the validation fold metrics. They are less suitable for predicting the recession in 2008 than Random Forest, which is due to the special character of this recession as a direct consequence of the financial crisis. However, they are no worse than the benchmark probit model (measured by recall). The models' strength lies, in particular, in their ability to provide recession risks that clearly exceed or lie below the cut-off value, even during quantitative easing.

7. CONCLUSION

The results from Germany are consistent with many findings that have already been obtained from research on US recessions. The indicators selected in the SFFS have already been used frequently in the literature (c.f. Nissilä 2020, Vrontos et al. 2021). Random Forest in particular has chosen very conventional indicators, e.g. the term spread and short-term interest rates. The other models also used many survey-based indicators, the importance of which has often been emphasized. (Christiansen et al. 2014). The small number of indicators selected, supports the results from Ng (2014) and Puglia & Tucker (2020), who have already pointed out that a few strong indicators are sufficient for forecasting.

The SFFS has found many strong indicators that can make good recession forecasts over a long period of time. The research setting aimed to avoid overoptimistic results and the so-called look-ahead bias. To achieve this, the selection of indicators, hyperparameters, and model training were conducted in an expanding window on data from before 2006, while the performance was evaluated out of sample (2006 – 2021). The out of sample results show that many indicators still have predictive power for the later recessions of 2008 and 2019. Only indicators based on interest rates lost forecasting power during the period of quantitative easing. In particular, energy prices and selling prices had a high predictive power. It was noticeable, that many models included indicators related to the US term spread but not the German term spread.

In summary, Random Forest, LASSO, and Ridge outperformed the benchmark model over the whole out-of-sample period. All ML models were able to predict the 2019 recession more precisely than the probit model. For the 2008 recession, the term spread is the best indicator. The probit model and Random Forest, which rely heavily on the term spread, provide the best predictions. The evaluation of LASSO and Ridge was complicated by the fact that a period of weak and negative GDP growth rates was classified as a recession, which was not defined in the GCEE chronology. Nevertheless, LASSO and Ridge are much better suited to forecast recessions during periods of unconventional monetary policy due to their diversified selection of indicators.

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2623-2631.

Berge, T. J. (2015). Predicting recessions with leading indicators: Model averaging and selection over the business cycle. Journal of Forecasting, 34(6), 455–471. https://doi.org/10.1002/for.2345

Breiman, L. (2001). Random Forest. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Cadahia Delgado, P., Congregado, E., Golpe, A., Vides, J. C. (2022). The Yield Curve as a Recession Leading Indicator. An Application for Gradient Boosting and Random Forest, International Journal of Interactive Multimedia and Artificial Intelligence, Special Issue, 7-19. https://doi.org/10.9781/ijimai.2022.02.006

Chauvet, M., & Potter, S. (2005). Forecasting recessions using the yield curve. Journal of Forecasting, 24(2), 77-103. https://doi.org/10.1002/for.932

Chen, Z., Iqbal, A., & Lai, H. (2011). Forecasting the probability of US recessions: a Probit and dynamic factor modelling approach. Canadian Journal of Economics, 44(2), 651-672. https://doi.org/10.1111/j.1540-5982.2011.01648.x

Chinn, M., & Kucko, K. (2015). The predictive power of the yield curve across countries and time. International Finance, 18(2), 129-156. https://doi.org/10.1111/infi.12064

Choi, J., Ge, D., Kang, K. H., & Sohn, S. (2023). Yield spread selection in predicting recession probabilities. Journal of Forecasting, 42(7), 1772–1785. https://doi.org/10.1002/for.2980

Christiansen, C., Eriksen, J. N., & Møller, S. V. (2014). Forecasting US recessions: The role of sentiment. Journal of Banking & Finance, 49, 459-468. https://doi.org/10.1016/j.jbankfin.2014.06.017

Döpke, J., Fritsche, U., & Pierdzioch, C. (2015). Predicting Recessions in Germany With Boosted Regression Trees, DEP (Socioeconomics) Discussion Papers Macroeconomics and Finance Series, 5/2015.

Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. International Journal of Forecasting, 33(4), 745-759. https://doi.org/10.1016/j.ijforecast.2017.02.003

Estrella, A. & Mishkin, F. S. (1998). Predicting U.S. Recessions: Financial Variables as Leading Indicators. The Review of Economics and Statistics, 80 (1), 45–61. https://doi.org/10.1162/003465398557320

Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. The Journal of Finance, Vol. 46, Nr. 2, S. 555-576. https://doi.org/10.1111/j.1540-6261.1991.tb02674.x

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Special invited paper. additive logistic regression: A statistical view of boosting. Annals of statistics, 337-374.

GCEC (2023), Konjunkturzyklus-Datierung, https://www.sachverstaendigenrat-wirtschaft.de/themen/konjunktur-und-wachstum/konjunkturzyklus-datierung.html

Gilchrist, S., & Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. American Economic Review, 102(4), 1692-1720. https://doi.org/10.1257/aer.102.4.1692

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3, 1157-1182.

Hasse, J. B., & Lajaunie, Q. (2020). Does the yield curve signal recessions? New evidence from an international panel data analysis. AMSE Working Paper, Nr. 13.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. New York, Springer.

Heilemann, U. (2019). Rezessionen in der Bundesrepublik Deutschland von 1966 bis 2013. Wirtschaftsdienst, 99(8), 546-552.

Hoerl, A. & Kennard, R. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 12 (1), 55-67. https://doi.org/10.1080/00401706.1970.10488634

Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.

NBER (2008). Business Cycle Dating Committee Announcement January 7, 2008. https://www.nber.org/news/business-cycle-dating-committee-announcement-january-7-2008

Nevasalmi, L. (2022). Recession forecasting with high-dimensional data. Journal of Forecasting, 31(4), 752-764. https://doi.org/10.1002/for.2823

Ng, S. (2014). Boosting recessions. Canadian Journal of Economics, 47(1), 1-34. http://www.jstor.org/stable/42705937

Nissilä, W. (2020). Probit based time series models in recession forecasting–A survey with an empirical illustration for Finland. Bank of Finland Economic Review, 7/2020.

Nyberg, H. (2010). Dynamic probit models and financial variables in recession forecasting. Journal of Forecasting, 29(1), 215-230. https://doi.org/10.1002/for.1161

Plakandaras, V., Cunado, J., Gupta, R., & Wohar, M. E. (2017). Do leading indicators forecast US recessions? A nonlinear re-evaluation using historical data. International Finance, 20(3), 289-316. https://doi.org/10.1111/infi.12111

Pönka, H. (2017). The role of credit in predicting US recessions. Journal of Forecasting, 36(5), 469-482. https://doi.org/10.1002/for.2448

Psimopoulos, A. (2020). Forecasting economic recessions using machine learning: an empirical study in six countries. South-Eastern Europe Journal of Economics, 18(1), 40-99.

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. Pattern recognition letters, 15(11), 1119-1125. https://doi.org/10.1016/0167-8655(94)90127-9

Puglia, M. & Tucker, A. (2020). Machine Learning, the Treasury Yield Curve and Recession Forecasting, Finance and Economics Discussion Series 2020-038, Board of Governors of the Federal Reserve System (U.S.).

Puglia, M., & Tucker, A. (2021). Neural Networks, the Treasury Yield Curve, and Recession Forecasting. The Journal of Financial Data Science, 3(2), 149-175. https://doi.org/10.3905/jfds.2021.1.061

Raschka, S., & Mirjalili, V. (2017). Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow (Second edition). Birmingham, Packt.

Shapley, L. (1997). 7. A Value for n-Person Games. Contributions to the Theory of Games II (1953) 307-317.. In: H. Kuhn (Ed.), Classics in Game Theory (pp. 69-79). Princeton: Princeton University Press. https://doi.org/10.1515/9781400829156-012

Stehle, R., Maier, J., & Huber, R. (1996). Rückberechnung des DAX für die Jahre 1955 bis 1987. SFB 373 Discussion Paper 7.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso, Journal of the Royal Statistical Society: Series B (Methodological), 58(1) 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vrontos, S. D., Galakis, J., & Vrontos, I. D. (2021). Modeling and predicting US recessions using machine learning techniques. International Journal of Forecasting, 37(2), 647-671. https://doi.org/10.1016/j.ijforecast.2020.08.005

Yazdani, A. (2020). Machine Learning Prediction of Recessions: An Imbalanced Classification Approach. The Journal of Financial Data Science, 2(4), 21-32. https://doi.org/10.3905/jfds.2020.1.040

Zyatkov, N., & Krivorotko, O. (2021). Forecasting Recessions in the US Economy Using Machine Learning Methods. 2021 17th International Asian School-Seminar" Optimization Problems of Complex Systems (OPCS), 139-146. https://doi.org/10.1109/OPCS53376.2021.9588678

Appendix

A1: List of Indicators

Subject	L	%	Δ	Source	OECD / FRED Code
Financial Markets					
Long-term interest rates (10-year) (EURO)	1	0	1	OECD	EA19_IRLTLT01_ST
Monetary Aggregate M3	0	1	0	OECD	EA19_MABMM301_IXOB
Monetary Aggregate M3 (SA)	0	1	0	OECD	EA19_MABMM301_GYSA
Monetary Aggregate (M1)	0	1	0	OECD	EA19_MANMM101_IXOB
Monetary Aggregate (M1) (SA)	0	1	0	OECD	EA19_MANMM101_IXOBSA
3-month or 90-day rates and yields	1	0	1	OECD	IR3TIB01_ST
Long-term interest rates (10-years)	1	0	1	OECD	IRLTLT01_ST
Call money/interbank rate (< 24 hrs.)	1	0	1	OECD	IRSTCI01_ST
Term Spread	1	0	1	OECD	LOCOSIOR_ST
Moody's Seasoned Aaa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity	1	0	1	FRED	AAA10YM
Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	1	0	1	FRED	AAAFFM
Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity	1	0	1	FRED	BAA10YM
Moody's Seasoned Baa Corporate Bond Minus Federal Funds Rate	1	0	1	FRED	BAAFFM
Term Spread USA	1	0	1	FRED	TB6SMFFM
Term Spread USA (Version 2)	1	0	1	FRED	T10YFFM
Term Spread Germany - Term Spread USA	1	0	1		own calculation

Subject	L	%	Δ	Source	OECD / FRED Code
International					
Real Effective Exchange Rates	0	1	0	OECD	CCRETT01_IXOB
Real Effective Exchange Rates (EURO)	0	1	0	OECD	EA19_CCRETT01_IXOB
US$ exchange rate (National currency:USD)	1	0	1	OECD	CCUSMA02_ST
US$ exchange rate (USD:national currency)	1	0	1	OECD	CCUSSP01_ST
Export Orders	1	0	1	OECD	LOCOBXOR_STSA
Labor Market					
Job vacancies	0	1	0	OECD	LMJVTTUV_ST
Job vacancies (SA)	0	1	0	OECD	LMJVTTUV_STSA
Registered unemployment	0	0	1	OECD	LMUNRLTT_ST
Registered unemployment (SA)	0	0	1	OECD	LMUNRLTT_STSA
Unemployment rate	1	0	0	OECD	LMUNRRTT_ST
Unemployment rate (SA)	1	0	0	OECD	LMUNRRTT_STSA
Prices					
Food and non-Alcoholic beverages	0	1	0	OECD	CP010000_GY
Electricity, gas and other fuels	0	1	0	OECD	CP040500_GY
Overall	0	1	0	OECD	CPALTT01_GY
Energy	0	1	0	OECD	CPGREN01_GY
Housing	0	1	0	OECD	CPGRHO01_GY
All items non-food non-energy	0	1	0	OECD	CPGRLE01_GY
Oil Prices	0	1	0	FRED	WTISPLC

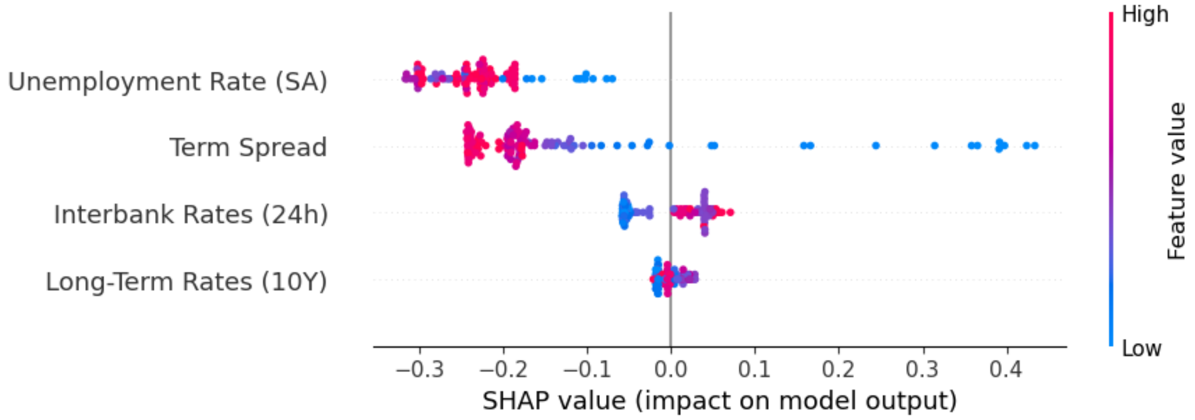
Subject	L	%	Δ	Source	OECD / FRED Code
Real Economy					
Orders	0	1	0	OECD	LOCOODOR_IXOBSA
Orders (Manufacturing)	0	1	0	OECD	ODMNTO01_IXOB
Producer Price Index (Manufacturing)	0	1	0	OECD	PIEAMP02_GY
Stock Market					
DAX	0	1	0	Yahoo Finance / Stehle	
Standard & Poors 500	0	1	0	Shiller	
OECD Share Prices Index	0	1	0	OECD	SPASTT01_GY
Business Tendency / Consumer Confidence Surveys					
Business Situation	1	0	1	OECD	BSBUCT02_STSA
Business Confidence indicator	1	0	1	OECD	BSCICP02_STSA
Finished Goods Stocks	1	0	1	OECD	BSFGLV02_STSA
Order Books	1	0	1	OECD	BSOBLV02_STSA
Orders Inflow	1	0	1	OECD	BSOITE02_STSA
Production Future Tendency	1	0	1	OECD	BSPRFT02_STSA
Production Tendency	1	0	1	OECD	BSPRTE02_STSA
Selling Prices	1	0	1	OECD	BSSPFT02_STSA
Consumer Confidence Indicator	1	0	1	OECD	CSCICP02_STSA

The "OECD / FRED Code" column shows the official abbreviation of the indicator, which is used in the database.
The columns "L", "%", " Δ " show whether an indicator is used in the analysis in level, as a percentage change or difference to the previous year. If the OECD code ends in "GY", the original variable was already available as a percentage change.

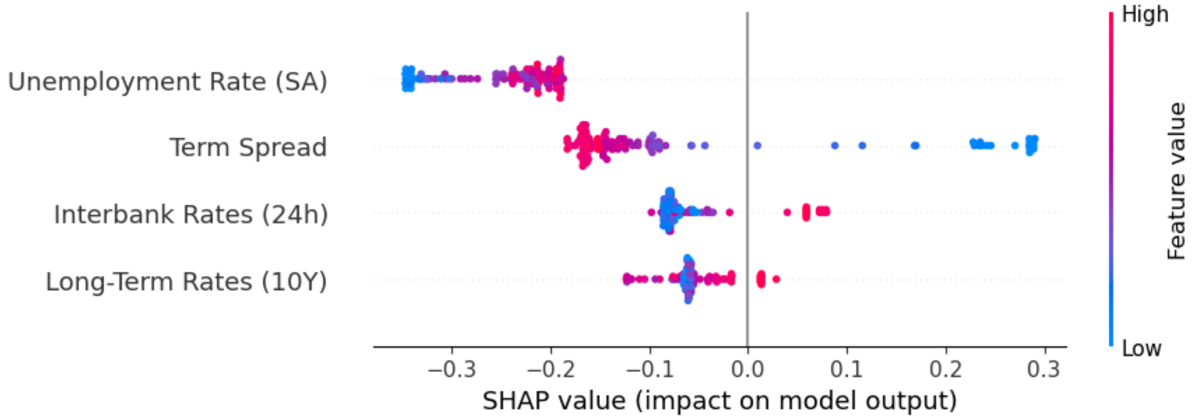
Data Sources:
OECD: https://www.oecd-ilibrary.org/economics/data/main-economic-indicators_mei-data-en
FRED: https://fred.stlouisfed.org/
Yahoo Finance: https://de.finance.yahoo.com/quote/%5EGDAXI/
Stehle: https://www.econstor.eu/handle/10419/66277
Shiller: http://www.econ.yale.edu/~shiller/data.htm

A2: Random Forest - Beeswarm-Plots

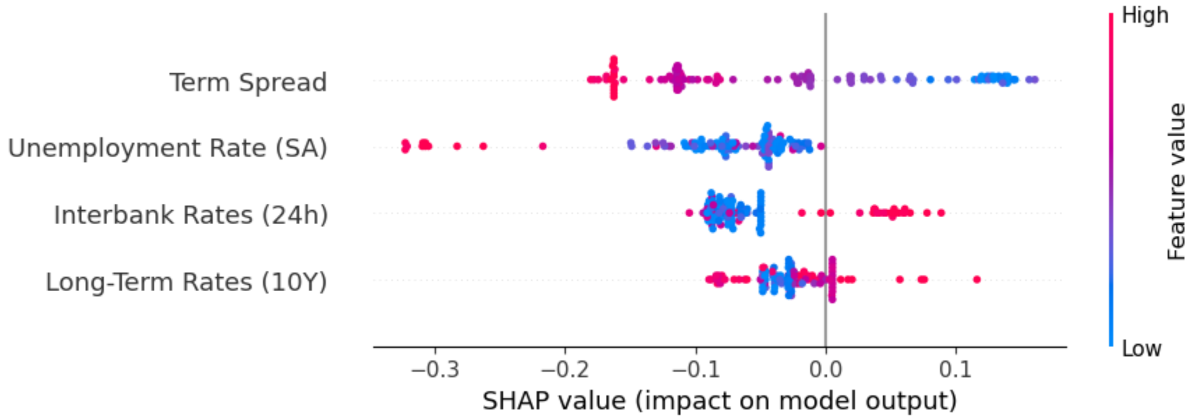
A. Recession 1992



B. Recession 2001



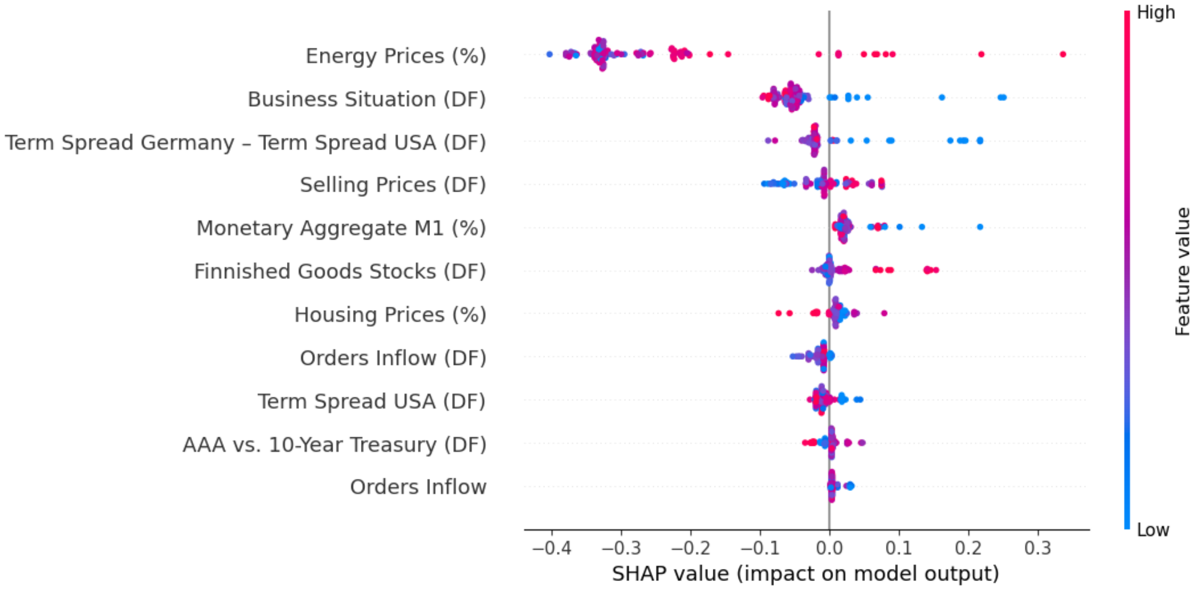
C. Recession 2008



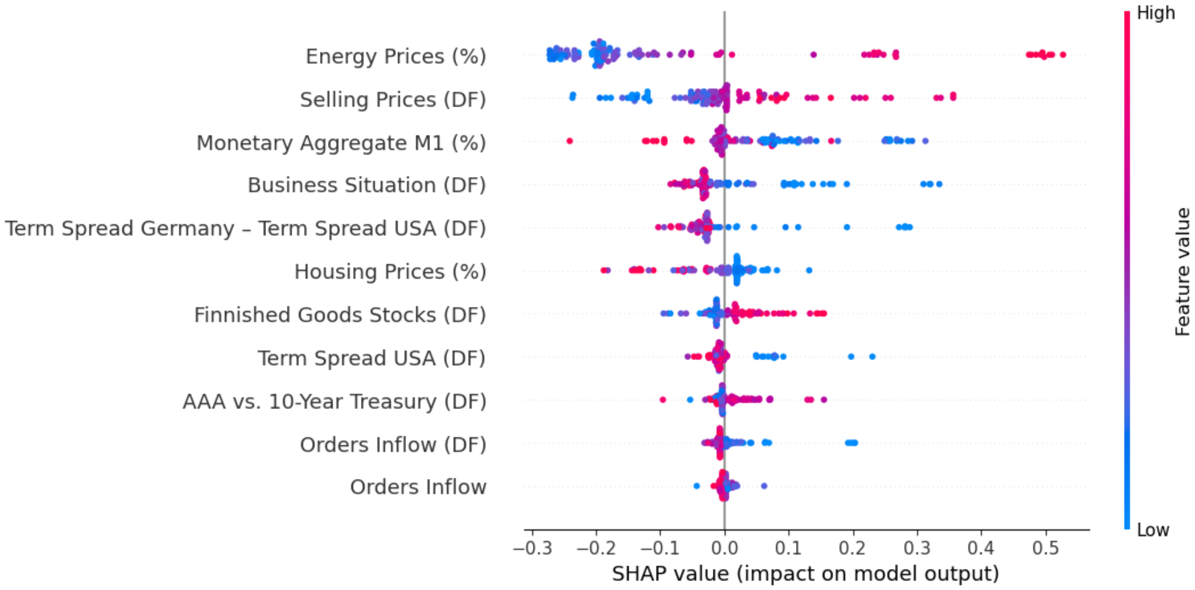
D. Recession 2019

A3: Logit Boost - Beeswarm-Plots

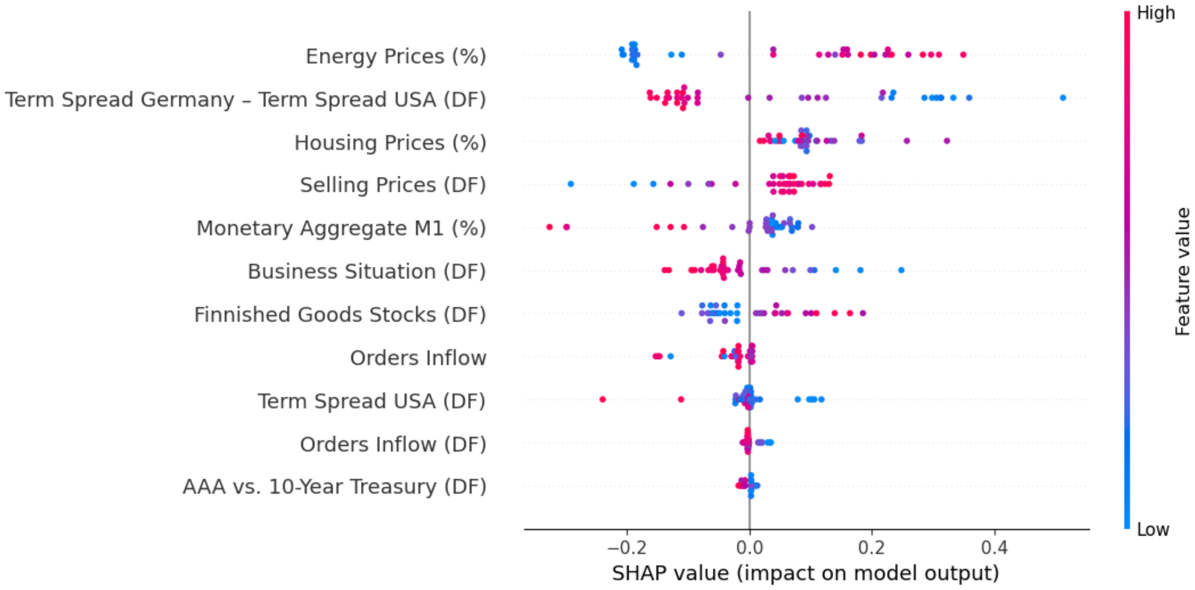
A. Recession 1992



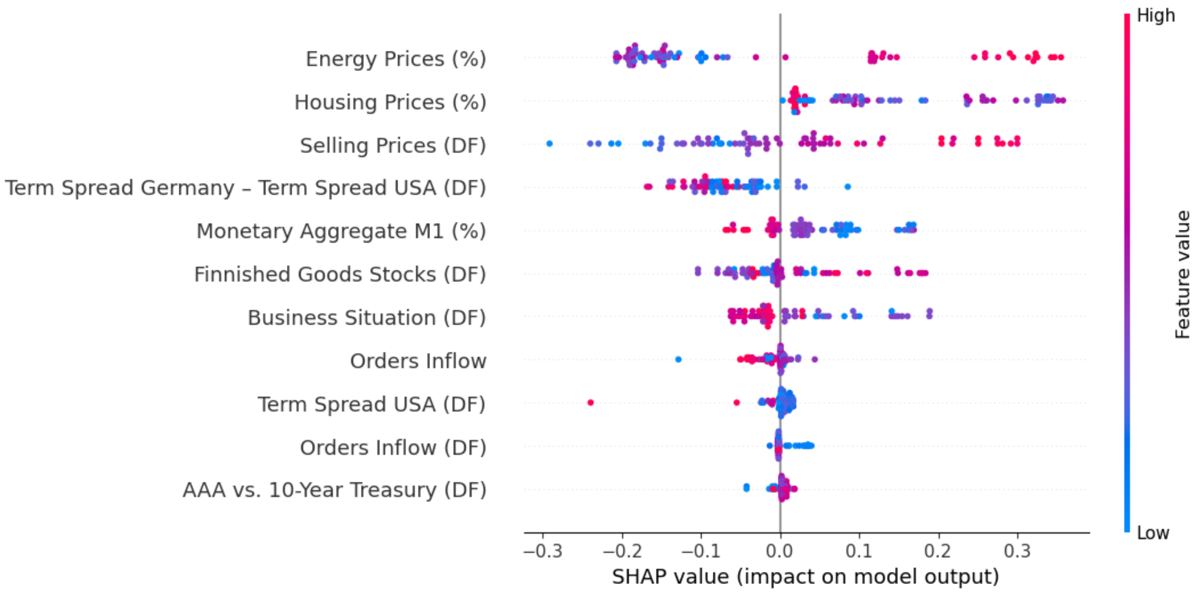
B. Recession 2001



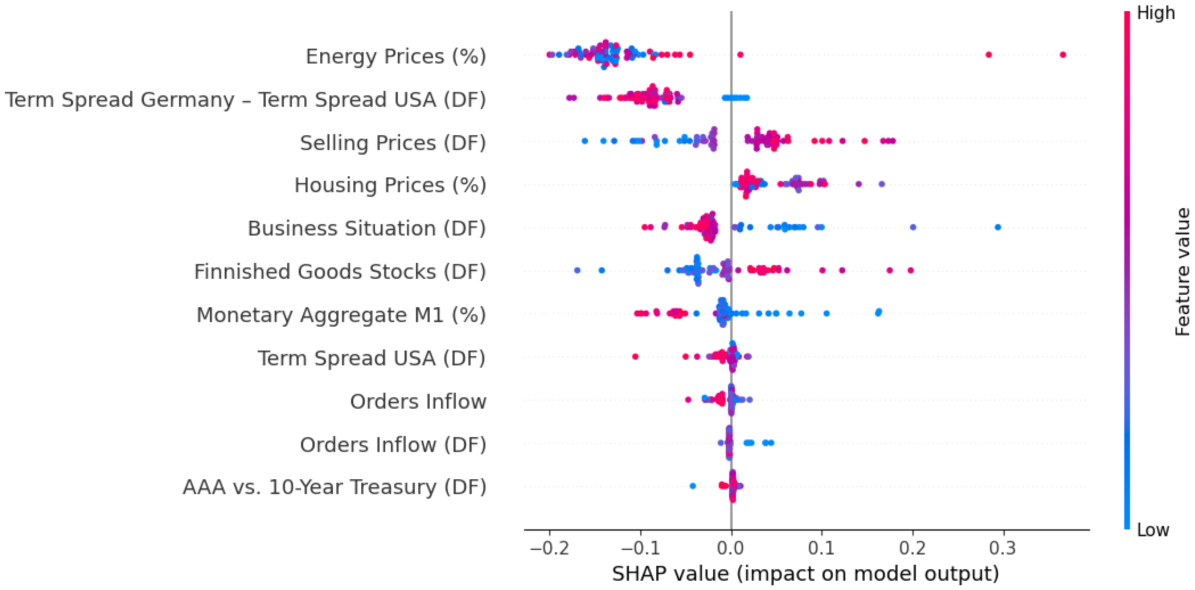
C. Recession 2008



D. Fictious Recession 2010

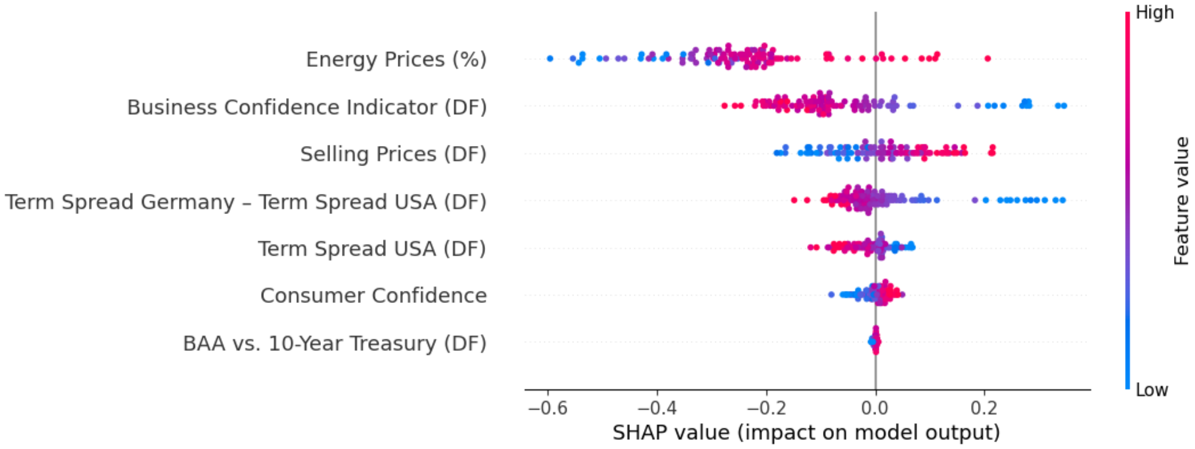


E. Recession 2019

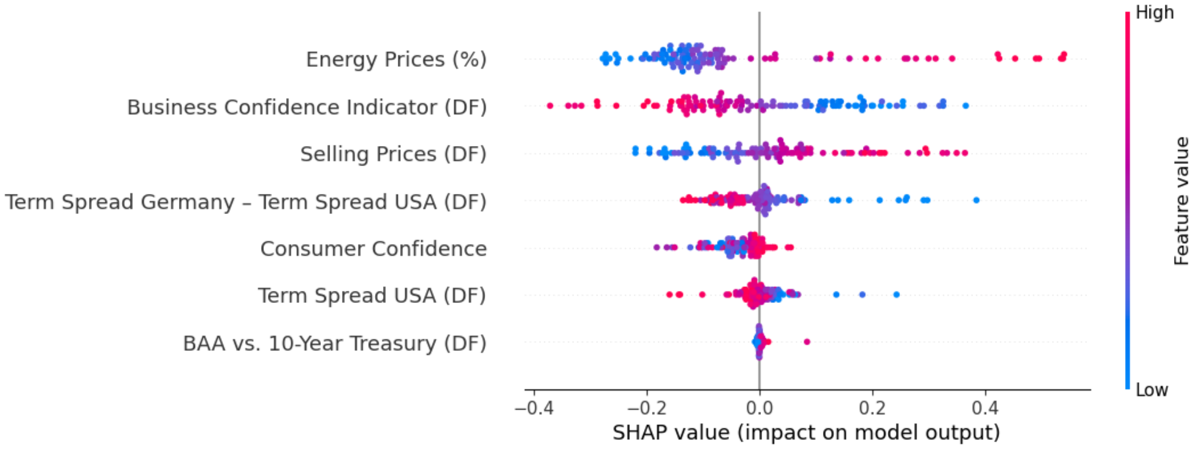


A4: Ridge - Beeswarm-Plots

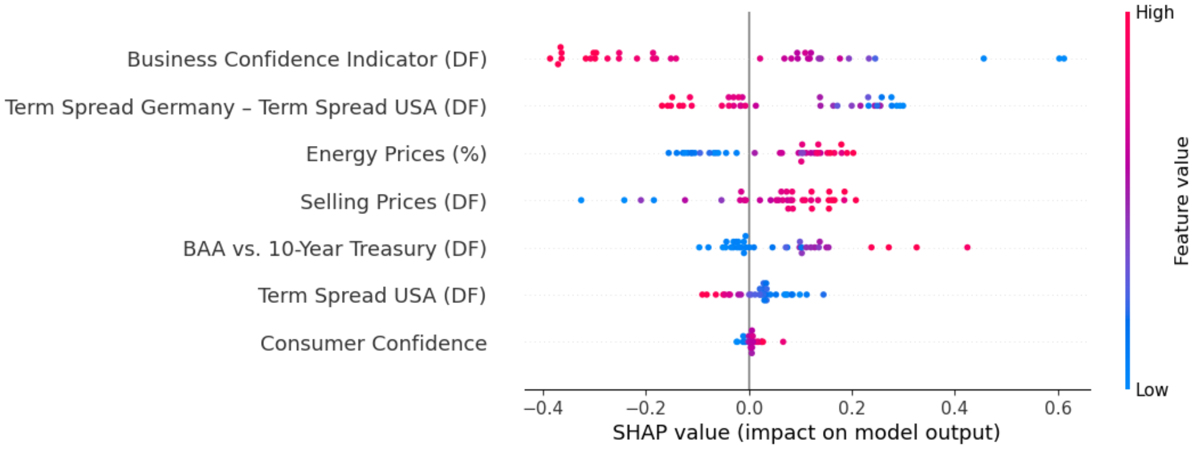
A. Recession 1992



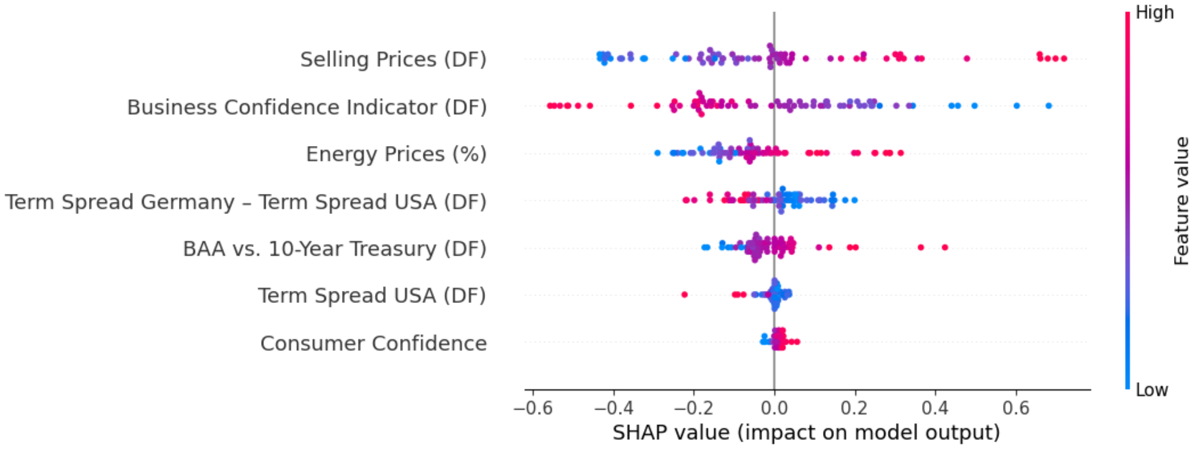
B. Recession 2001



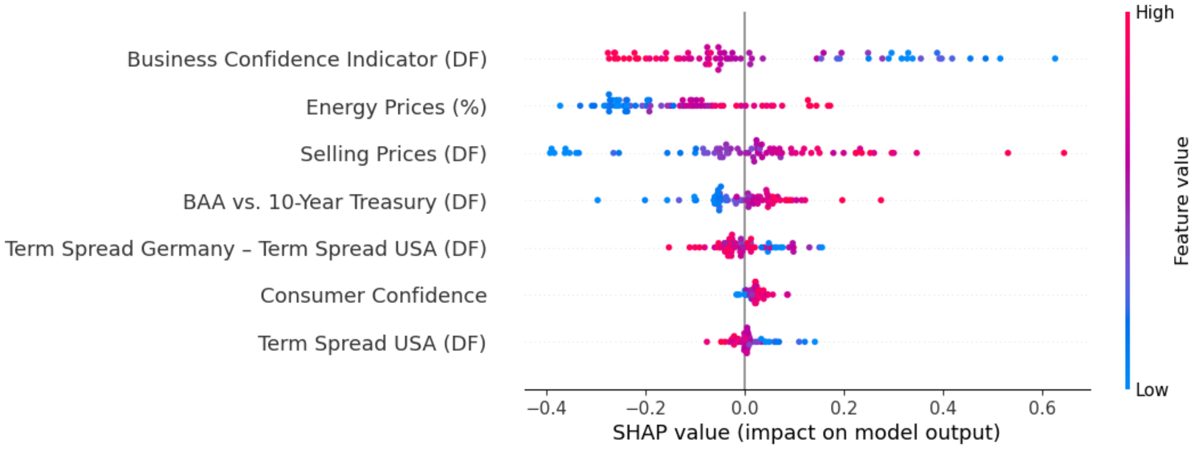
C. Recession 2008



D. Fictious Recession 2010

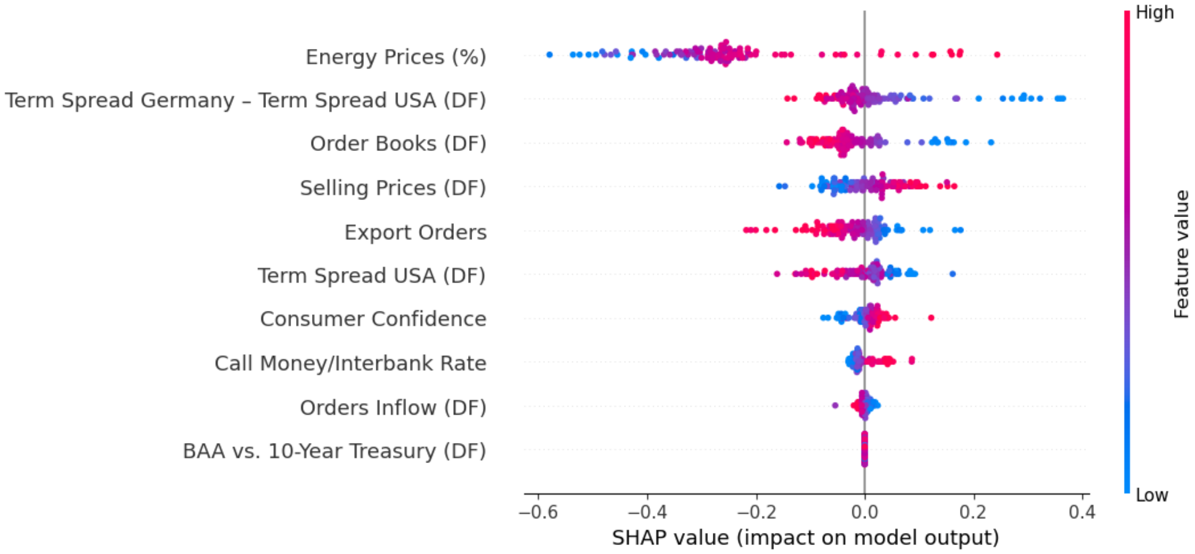


E. Recession 2019

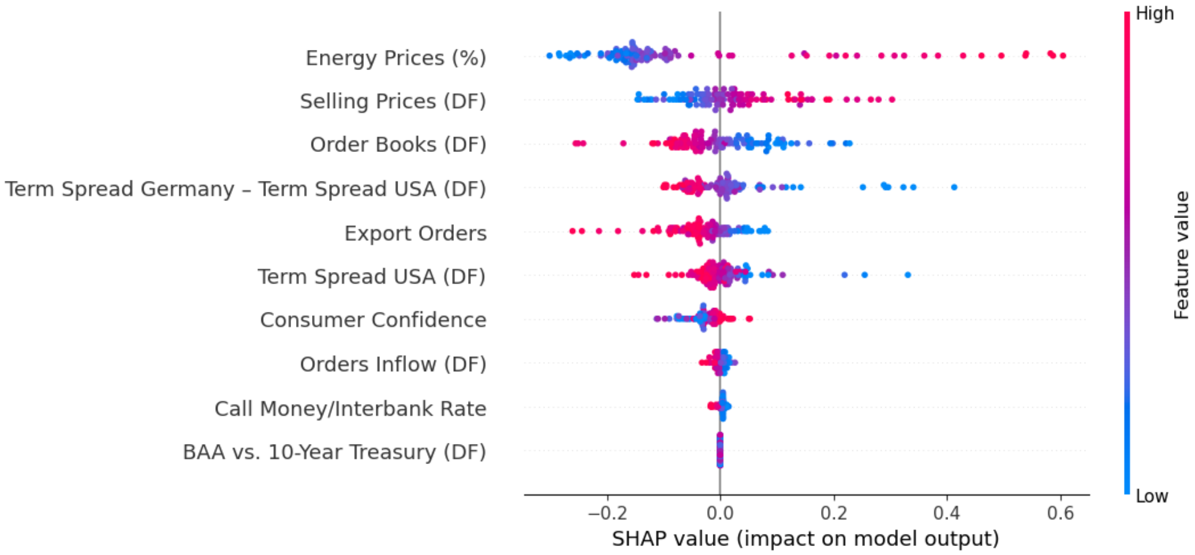


A5: LASSO - Beeswarm-Plots

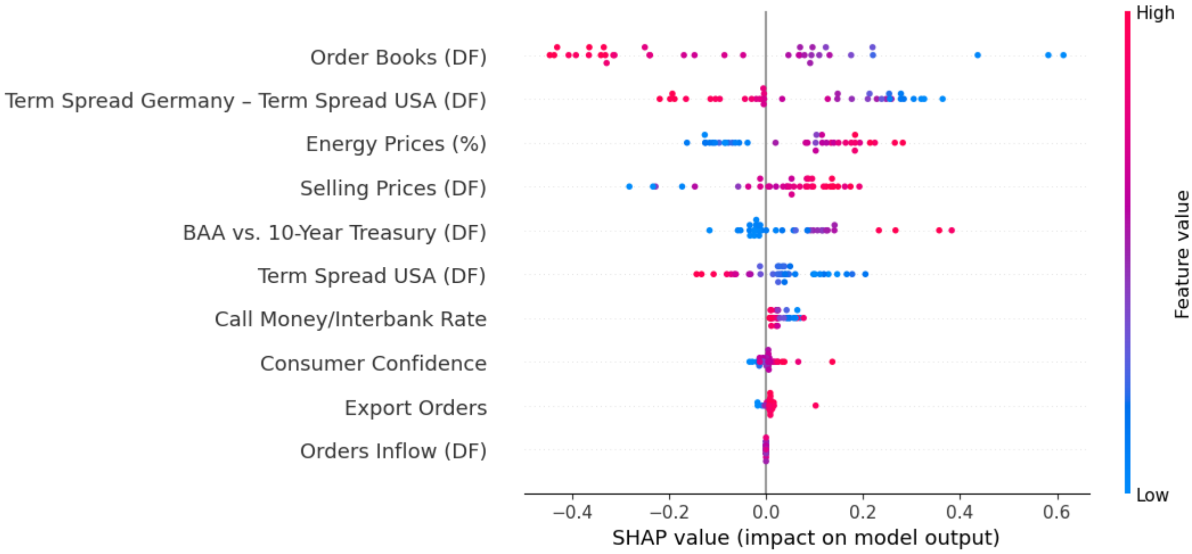
A. Recession 1992



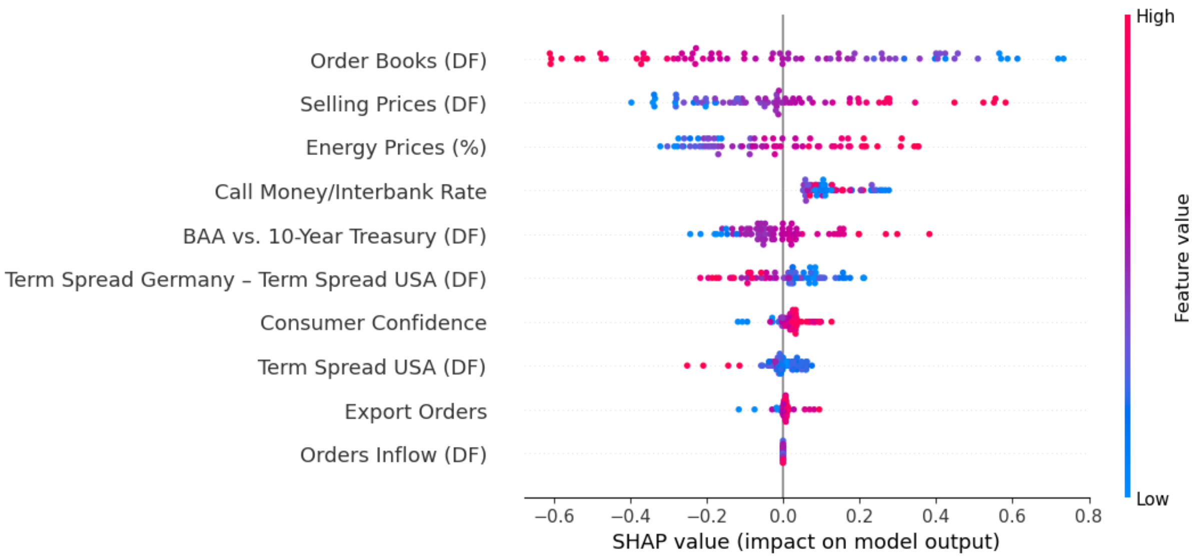
B. Recession 2001



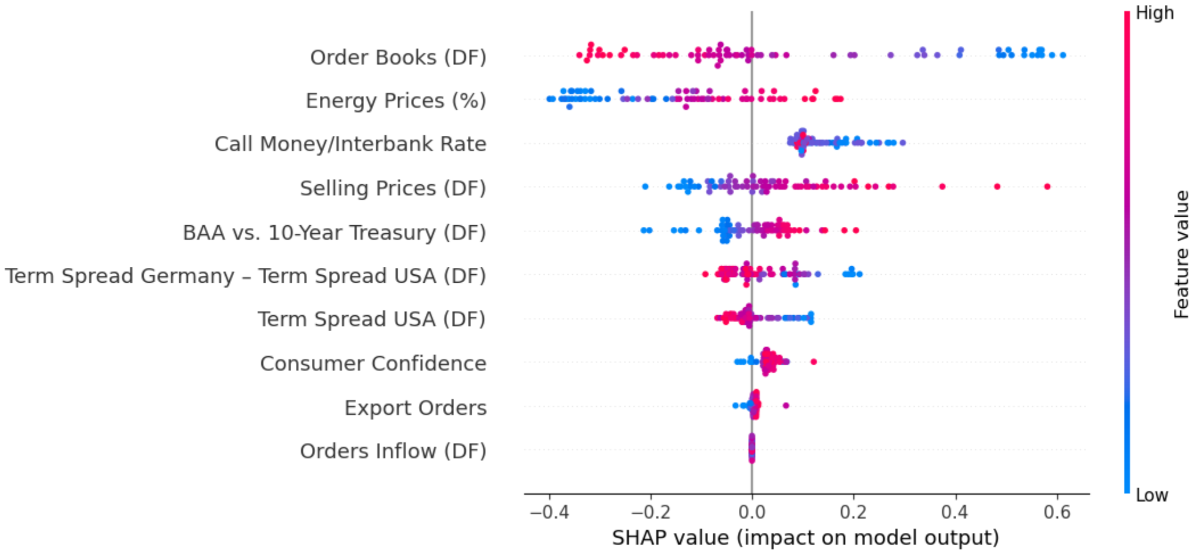
C. Recession 2008



D. Fictious Recession 2010



E. Recession 2019



A6: List of Python Packages

Python Package	Webpage
Logitboost	https://logitboost.readthedocs.io/
Matplotlib	https://matplotlib.org/
Mlxtend	https://rasbt.github.io/mlxtend/
Numpy	https://numpy.org/
Optuna	https://optuna.org/
Pandas	https://pandas.pydata.org/
Scikit-Learn (Sklearn)	https://scikit-learn.org/stable/
SHAP	https://shap.readthedocs.io/en/latest/

PREVIOUS DISCUSSION PAPERS

416 Rademacher, Philip, Forecasting Recessions in Germany with Machine Learning, September 2024.

415 Erhardt, Katharina and Gupta, Apoorva, Go Wide or Go Deep: Margins of New Trade Flows, August 2024.

414 Gupta, Apoorva and Stiebale, Joel, Gains from Patent Protection: Innovation, Market Power and Cost Savings in India, May 2024.

413 Cobb-Clark, Deborah A., Dahmann, Sarah C., Kamhöfer, Daniel A., and Schildberg-Hörisch, Hannah, Schooling and Self-Control, March 2024.

412 Huelden, Tobias, Jascisens, Vitalijs, Roemheld, Lars and Werner, Tobias, Human-Machine Interactions in Pricing: Evidence from Two Large-Scale Field Experiments, March 2024.

411 Hermes, Henning, Lergetporer, Philipp, Mierisch, Fabian, Schwerdt, Guido and Wiederhold, Simon, Does Information about Inequality and Discrimination in Early Child Care Affect Policy Preferences? January 2024.

410 Hunold, Matthias and Werner, Tobias, Algorithmic Price Recommendations and Collusion: Experimental Evidence, December 2023.

409 Herzog, Sabrina, Schildberg-Hörisch, Hannah, Trieu, Chi and Willrodt, Jana, Who is in Favor of Affirmative Action? Representative Evidence from an Experiment and a Survey, November 2023.

408 Stadelmann, David, Thomas, Tobias and Zakharov, Nikita, Too Hot to Play it Cool? Temperature and Media Bias, November 2023.

407 Dertwinkel-Kalt, Markus and Wey, Christian, Why “Energy Price Brakes” Encourage Moral Hazard, Raise Energy Prices, and Reinforce Energy Savings, September 2023.
Forthcoming in: RAND Journal of Economics.

406 Bertermann, Alexander, Kamhöfer, Daniel A. and Schildberg-Hörisch, Hannah, More Education Does Make You Happier – Unless You Are Unemployed, September 2023.

405 Trieu, Chi, Who’s Who: How Uncertainty About the Favored Group Affects Outcomes of Affirmative Action, August 2023.
Published in: Journal of the Economic Science Association, 9 (2023), pp. 252-292.

404 Fischer, Kai, Martin, Simon and Schmidt-Dengler, Philipp, The Heterogeneous Effects of Entry on Prices, July 2023.

403 García-Vega, María, Gupta, Apoorva and Kneller, Richard, Is Acquisition-FDI During an Economic Crisis Detrimental for Domestic Innovation?, July 2023.

402 Haucap, Justus and Stiebale, Joel, Non-price Effects of Mergers and Acquisitions, July 2023.

401 Simion, Ștefania and Sulka, Tomasz, Multidimensional Cognitive Ability, Intermediate Channels, and Financial Outcomes, May 2023.

400 Bartling, Björn, Cappelen, Alexander W., Hermes, Henning, Skivenes, Marit and Tungodden, Bertil, Free to Fail? Paternalistic Preferences in the United States, May 2023.

399 Kandelhardt, Johannes, Flexible Estimation of Random Coefficient Logit Models of Differentiated Product Demand, May 2023.

398 Hermes, Henning, Lergetporer, Philipp, Mierisch, Fabian, Peter, Frauke and Wiederhold, Simon, Discrimination on the Child Care Market: A Nationwide Field Experiment, April 2023.

397 Schmal, W. Benedikt, Haucap, Justus and Knoke, Leon, The Role of Gender and Coauthors in Academic Publication Behavior, March 2023.
Published in: Research Policy, 52 (2023), 104874.

396 Magin, Jana Anjali, Neyer, Ulrike and Stempel, Daniel, The Macroeconomic Effects of Different CBDC Regimes in an Economy with a Heterogeneous Household Sector, March 2023.

395 Dertwinkel-Kalt, Markus and Wey, Christian, Resale Price Maintenance in a Successive Monopoly Model, February 2023.
Forthcoming in: Journal of Industrial Economics.

394 Hermes, Henning, Krauß, Marina, Lergetporer, Philipp, Peter, Frauke and Wiederhold, Simon, Early Child Care and Labor Supply of Lower-SES Mothers: A Randomized Controlled Trial, December 2022.

393 Chowdbury, Shyamal, Schildberg-Hörisch, Hannah, Schneider, Sebastian O., and Sutter, Matthias, Information Provision Over the Phone Saves Lives: An RCT to Contain COVID-19 in Rural Bangladesh at the Pandemic's Onset, November 2022.

392 Normann, Hans-Theo and Sternberg, Martin, Human-Algorithm Interaction: Algorithmic Pricing in Hybrid Laboratory Markets, October 2022.
Published in: European Economic Review, 152 (2023), 104347.

391 Hunold, Matthias and Petrishcheva, Vasilisa, Foreclosure and Tunneling with Partial Vertical Ownership, September 2022.

390 Haucap, Justus and Heldman, Christina, The Sociology of Cartels, August 2022.
Published in: European Journal of Law and Economics, 56 (2023), pp. 289-323.

389 Döpper, Hendrik, Sapi, Geza and Wey, Christian, A Bargaining Perspective on Vertical Integration, May 2022.
Published in: Canadian Journal of Economics, 57 (2024), pp. 199-224.

388 Bachmann, Ronald, Gonschor, Myrielle, Lewandowski, Piotr and Madoń, Karol, The Impact of Robots on Labour Market Transitions in Europe, May 2022.

387 Fremerey, Melinda, Hörnig, Lukas and Schaffner, Sandra, Becoming Neighbors with Refugees and Voting for the Far-Right? The Impact of Refugee Inflows at the Small-Scale Level, April 2022.
Published in: Labour Economics, 86 (2024), 102467.

386 Fischer, Kai, Alcohol Prohibition and Pricing at the Pump, March 2022.

385 Cobb-Clark, Deborah A., Dahmann, Sarah C., Kamhöfer, Daniel A. and Schildberg-Hörisch, Hannah, The Determinants of Population Self-Control, March 2022.
Forthcoming in: Economic Journal under the title "Surveillance and Self-Control".

384 Sulka, Tomasz, Planning and Saving for Retirement, March 2022.
Published in: European Economic Review, 160 (2023), 104609.

383 Cattan, Sarah, Kamhöfer, Daniel A., Karlsson, Martin and Nilsson, Therese, The Long-term Effects of Student Absence: Evidence from Sweden, March 2022.
Published in: Economic Journal, 133 (2023), pp. 888-903.

382 Martin, Simon and Rasch, Alexander, Collusion by Algorithm: The Role of Unobserved Actions, March 2022.
Published in: International Journal of Industrial Organization, 92 (2024), 103036 under the title "Demand Forecasting, Signal Precision, and Collusion with Hidden Actions".

381 Haucap, Justus, Nedic, Radivoje and Şimşek, Talha, An Empirical Analysis of German Casino Locations, March 2022.
Published in: European Journal of Law and Economics, 55 (2023) pp. 291-311.

380 Haucap, Justus, Heldman, Christina and Rau, Holger A., Gender and Collusion, March 2022.

379 Schain, Jan Philip, Foreign Institutional Investors and the Great Productivity Slowdown, November 2022 (First Version February 2022).

378 Neyer, Ulrike and Stempel, Daniel, How Should Central Banks React to Household Inflation Heterogeneity?, January 2022.

Older discussion papers can be found online at:
http://ideas.repec.org/s/zbw/dicedp.html

Heinrich-Heine-Universität Düsseldorf

Düsseldorfer Institut für Wettbewerbsökonomie (DICE)

Universitätsstraße 1, 40225 Düsseldorf

ISSN 2190-992X (online)
ISBN 978-3-86304-415-2