

# Dynamic Conditional Feature Screening: A High-Dimensional Feature Selection Method Based on Mutual Information and Regression Error

Yi Zhao<sup>1</sup>, Guangming Deng<sup>2\*</sup>

<sup>1</sup>School of Mathematics and Statistics, Guilin University of Technology, Guilin, China

<sup>2</sup>Applied Statistics Institute, Guilin University of Technology, Guilin, China

Email: \*dgm@glut.edu.cn

**How to cite this paper:** Zhao, Y. and Deng, G.M. (2025) Dynamic Conditional Feature Screening: A High-Dimensional Feature Selection Method Based on Mutual Information and Regression Error. *Open Journal of Statistics*, 15, 199-242.

<https://doi.org/10.4236/ojs.2025.152011>

**Received:** March 17, 2025

**Accepted:** April 19, 2025

**Published:** April 22, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Current high-dimensional feature screening methods still face significant challenges in handling mixed linear and nonlinear relationships, controlling redundant information, and improving model robustness. In this study, we propose a Dynamic Conditional Feature Screening (DCFS) method tailored for high-dimensional economic forecasting tasks. Our goal is to accurately identify key variables, enhance predictive performance, and provide both theoretical foundations and practical tools for macroeconomic modeling. The DCFS method constructs a comprehensive test statistic by integrating conditional mutual information with conditional regression error differences. By introducing a dynamic weighting mechanism, DCFS adaptively balances the linear and nonlinear contributions of features during the screening process. In addition, a dynamic thresholding mechanism is designed to effectively control the false discovery rate (FDR), thereby improving the stability and reliability of the screening results. On the theoretical front, we rigorously prove that the proposed method satisfies the sure screening property and rank consistency, ensuring accurate identification of the truly important feature set in high-dimensional settings. Simulation results demonstrate that under purely linear, purely nonlinear, and mixed dependency structures, DCFS consistently outperforms classical screening methods such as SIS, CSIS, and IG-SIS in terms of true positive rate (TPR), false discovery rate (FDR), and rank correlation. These results highlight the superior accuracy, robustness, and stability of our method. Furthermore, an empirical analysis based on the U.S. FRED-MD macroeconomic dataset confirms the practical value of DCFS in real-world forecasting tasks. The experimental results show that DCFS achieves lower prediction errors (RMSE and MAE) and

higher  $R^2$  values in forecasting GDP growth. The selected key variables—including the Industrial Production Index (IP), Federal Funds Rate, Consumer Price Index (CPI), and Money Supply (M2)—possess clear economic interpretability, offering reliable support for economic forecasting and policy formulation.

### Keywords

High-Dimensional Feature Screening, Conditional Mutual Information, Regression Error Difference, Dynamic Weighting, Dynamic Thresholding, Macroeconomic Forecasting, FRED-MD Dataset

---

## 1. Introduction

With the advancement of data collection technology, high-dimensional data has been widely applied in various fields such as bioinformatics, financial analysis, environmental science, and medical diagnosis. A significant feature of high-dimensional data is that the dimensionality  $p$  of covariates is greater than the sample size  $n$ . Specifically, when the dimensionality  $p$  of the data is much larger than the sample size  $n$ , this type of data is often referred to as ultra-high-dimensional data. When analyzing ultra-high dimensional data, it is often assumed that there are fewer covariates that affect the response variable, which is referred to as the "sparsity" hypothesis in the literature. Under the assumption of sparsity, determining the covariates that truly affect the response variable has become an important and fundamental issue. Fan and L ü (2008) [1] proposed a method for screening variables based on Pearson correlation coefficient and named it Sure Independence Screening (SIS). Inspired by Fan and L ü (2008), the study of variable selection methods has received great attention from statisticians, resulting in a large number of research achievements. Li, Zhong, and Zhu (2012) [2] proposed a deterministic independence screening method using distance correlation. According to Li *et al.* (2012), a screening method based on distance correlation coefficient was proposed. Shao and Zhang (2014) [3] proposed the martingale difference correlation coefficient screening method (MDC), which measures the deviation between two random variables and one correlation coefficient. Mai and Zou (2015) [4] developed a binary classification model screening method based on Kolmogorov distance. Ni and Fang (2016) [5] constructed corresponding statistical measures for correlation analysis from the perspective of information content, and proposed a method for selecting ultra-high dimensional variables based on information gain (IG-SIS). Zhu Yidan and Chen Xingrong (2021) [6] proposed a variable selection method based on information gain rate (IGR-SIS), which has higher accuracy in screening important variables compared to IG-SIS. Fan *et al.* (2020) [7] and Zeng Jin and Zhou Jianjun (2017) [8] provided a comprehensive overview of variable selection and feature screening.

Feature screening methods based on marginal correlations, such as MDC [3] and IG-SIS [5], often fail to capture complex conditional dependencies among

variables. To address this challenge, Fan *et al.* (2016) [9] proposed the Conditional Sure Independence Screening (CSIS) method, which identifies important features by estimating conditional distributions or regression functions. This method reduces the impact of redundant variables while ensuring feature saliency. Research has shown that CSIS outperforms traditional SIS methods in processing data with nonlinear dependency structures. Lin *et al.* (2020) [10] Propose a model free conditional feature selection method based on conditional distance correlation. Zhou *et al.* [11] proposed conditional feature selection for variable coefficient models. Xiong *et al.* [12] proposed a new model free interaction screening program called MCVIS by introducing the MCV index to quantify the importance of the interaction effects between predictive factors. Wang *et al.* (2023) [13] proposed a new model free conditional screening method using conditional feature functions as screening criteria for massive imbalanced data.

Although methods such as SIS [1], CSIS [9], and IG-SIS [5] are effective in certain scenarios, they often overlook complex conditional dependencies among variables. As a result, they struggle to simultaneously capture both linear and nonlinear features and tend to exhibit poor screening stability when dealing with high-dimensional data. Some existing methods have attempted to incorporate conditional screening mechanisms, yet still suffer from insufficient characterization of nonlinear structures and limited control over screening errors. To address these challenges, this paper proposes a Dynamic Conditional Feature Screening (DCFS) method tailored to high-dimensional economic forecasting tasks. The specific goals are: (1) to design a hybrid test statistic that combines conditional mutual information and regression error difference for a more comprehensive assessment of feature importance; (2) to develop dynamic weighting and thresholding mechanisms to enhance the model's adaptability to complex data structures and improve screening robustness; (3) to establish the theoretical guarantees of the proposed method, including sure screening and consistency; and (4) to validate its effectiveness through both simulation studies and empirical analysis using real macroeconomic data. This work aims to provide a novel methodology that offers both theoretical rigor and practical value for high-dimensional data analysis, feature selection, and macroeconomic modeling.

The structure of this paper is as follows: Section 2 introduces the mathematical framework of DCFS; Section 3 presents its theoretical properties; Section 4 conducts simulation experiments to evaluate empirical performance; Section 5 applies DCFS to real-world macroeconomic forecasting tasks; and Section 6 concludes the paper and discusses future research directions.

## 2. Dynamic Conditional Feature Selection Method

### 2.1. Variable Setting and Objectives

Variable setting: Let the predictor variables be denoted by

$$X = (X_1, X_2, \dots, X_p) \in \mathbb{R}^{n \times p}, \quad (1)$$

where  $p$  is the dimensionality of the predictors, and each  $X_j$  represents a candidate feature variable. The conditional variables are defined as

$$Z = (Z_1, Z_2, \dots, Z_q) \in \mathbb{R}^{n \times q}, \tag{2}$$

where  $q$  is the number of conditional variables, which are known to be strongly associated with the response variable  $Y \in \mathbb{R}^n$ , serving as the target variable. Such associations between  $Z$  and  $Y$  may arise from the inherent structure of the data or prior domain knowledge. The conditional variables play a critical role in adjusting for and eliminating confounding effects during the screening process, thus enabling a more accurate identification of the true relationship between the predictors and the response variable. The full dataset consists of  $n$  independent observations denoted by  $\{(X_i, Z_i, Y_i)\}_{i=1}^n$ .

Screening objective: Given the conditional variables  $Z$ , our objective is to identify a subset  $S \subseteq \{1, 2, \dots, p\}$  of predictors that make a significant contribution to the response variable  $Y$ . To this end, we define the set of active predictors as  $D = \{k : X_k \perp Y | Z\}$  and the set of inactive predictors as  $I = \{k : X_k \not\perp Y | Z\}$ . Our goal is to accurately recover the active predictor set  $D$ , that is, to identify all features  $X_j$  satisfying  $X_j \perp Y | Z, \forall j \in S$ . The screening procedure should guarantee high identification accuracy of truly important predictors while minimizing the risk of omission.

## 2.2. Construction of Conditional Correlation Measurement

### 2.2.1. Conditional Mutual Information

Conditional mutual information  $I(X_j; Y | Z)$  is employed to quantify the dependency between the feature variable  $X_j$  and the response variable  $Y$ , given the conditional variables  $Z$ . It serves as a core component in the construction of the dynamic test statistic proposed in this study and is particularly well-suited for capturing nonlinear relationships.

The definition of conditional mutual information is based on the difference in conditional entropies, expressed as:

$$I(X_j; Y | Z) = H(X_j | Z) - H(X_j | Y, Z), \tag{3}$$

where  $H(\cdot)$  denotes the conditional entropy. The two terms in Equation (3) can be further expressed as:

$$H(X_j | Z) = -\mathbb{E}[\log f_{X_j|Z}(x_j | z)], \quad H(X_j | Y, Z) = -\mathbb{E}[\log f_{X_j|Y,Z}(x_j | y, z)]. \tag{4}$$

By substituting Equation (4) into Equation (3), the conditional mutual information can be written as a log-likelihood ratio:

$$I(X_j; Y | Z) = E \left[ \log \frac{f_{X_j|Z}(x_j | z)}{f_{X_j|Y,Z}(x_j | y, z)} \right]. \tag{5}$$

where  $f_{X_j|Z}$  and  $f_{X_j|Y,Z}$  denote the conditional probability density functions. From an information-theoretic perspective, conditional mutual information measures the additional information about  $Y$  that  $X_j$  provides when  $Z$  is

already known. When  $X_j$  and  $Y$  are conditionally independent given  $Z$ , we have  $I(X_j; Y | Z) = 0$ ; otherwise, if  $X_j$  provides independent information about  $Y$ , then  $I(X_j; Y | Z) > 0$ .

Compared to traditional linear correlation coefficients or least squares estimates, conditional mutual information offers several advantages:

It captures nonlinear, non-symmetric, and complex interaction relationships;

- It is model-free and thus robust to misspecification;
- It is well-suited for assessing variable importance in high-dimensional feature spaces.

In the proposed Dynamic Conditional Feature Screening (DCFS) method,  $I(X_j; Y | Z)$  is combined with the conditional regression error difference  $\Delta E(X_j, Y | Z)$  to construct a weighted test statistic (see Section 2.3). These two components jointly measure variable contributions from nonlinear and linear perspectives, respectively, providing a more comprehensive and adaptive basis for high-dimensional feature screening.

### 2.2.2. Differences in Conditional Regression Errors

The conditional regression error difference, denoted as  $\Delta E(X_j, Y | Z)$ , measures the incremental explanatory power of  $X_j$  for the response variable  $Y$ . It is defined as:

$$\Delta E(X_j, Y | Z) = E\left[\left(Y - \hat{m}^j(Z)\right)^2\right] - E\left[\left(Y - \hat{m}^{(X_j, Z)}(X_j, Z)\right)^2\right], \quad (6)$$

where  $\hat{m}^j(Z) = \mathbb{E}[Y | Z]$  is the regression model based solely on  $Z$ , and  $\hat{m}^{(X_j, Z)}(X_j, Z) = \mathbb{E}[Y | X_j, Z]$  is the regression model based on both  $X_j$  and  $Z$ . A larger  $\Delta E$  indicates stronger incremental explanatory ability of  $X_j$  with respect to  $Y$ . This quantity focuses on assessing the contribution of  $X_j$  from a regression analysis perspective, effectively reflecting its utility in explaining variations in the response variable.

### 2.3. Dynamic Weight Control Mechanism

To account for potential linear and nonlinear dependencies between features and the response, the proposed Dynamic Conditional Feature Screening (DCFS) method introduces a novel dynamic weighting mechanism to enhance adaptivity and robustness in the feature screening process.

The core idea of this mechanism is to dynamically adjust the relative weights of linear and nonlinear contributions for each feature  $X_j$ , based on its statistical dependence with the response  $Y$ , conditional on  $Z$ . Two complementary sources of information are considered in DCFS method: conditional mutual information  $I(X_j; Y | Z)$  and conditional regression error difference  $\Delta E(X_j, Y | Z)$  measure the nonlinear dependence and linear predictive power between the characteristic and response variables, respectively.

Specifically, we construct the following weighted statistic to evaluate the importance of each feature:

$$T_j^{\text{dynamic}} = w_1^{(j)} I(X_j; Y | Z) + w_2^{(j)} \Delta E(X_j, Y | Z), \quad (7)$$

where the weights  $w_1^{(j)}$  and  $w_2^{(j)}$  are derived from the relative information contribution of  $X_j$ , defined as:

$$w_1^{(j)} = \frac{I(X_j; Y | Z)}{I(X_j; Y | Z) + \Delta E(X_j, Y | Z)}, w_2^{(j)} = \frac{\Delta E(X_j, Y | Z)}{I(X_j; Y | Z) + \Delta E(X_j, Y | Z)}. \quad (8)$$

This weighting mechanism enables automatic adjustment based on the specific information structure of each feature. When  $I(X_j; Y | Z) \gg \Delta E(X_j, Y | Z)$ , indicating dominant nonlinear information, the weight tends toward  $w_1^{(j)} \rightarrow 1$ , and the statistic emphasizes nonlinear dependence. Conversely, when  $\Delta E(X_j, Y | Z) \gg I(X_j; Y | Z)$ , the statistic primarily reflects linear predictive ability.

In terms of theoretical properties, the statistic  $T_j^{\text{dynamic}}$  has good asymptotic properties under large sample conditions:

- **Consistency:** Under normal conditions, such as assuming that the characteristic variables are independent and equally distributed, the error has a finite second-order moment and the model is recognizable, there are:

$$T_j^{\text{dynamic}} \xrightarrow{p} T_j^0 \text{ as } n \rightarrow \infty, \quad (9)$$

where  $T_j^0$  denotes the true information content, and  $\xrightarrow{p}$  denotes convergence in probability.

- **Asymptotic Normality:** If the statistic can be estimated via kernel density methods or represented as a U-statistic, and under suitable regularity conditions, then

$$\sqrt{n}(T_j^{\text{dynamic}} - T_j^0) \xrightarrow{d} N(0, \sigma_j^2), \quad (10)$$

where  $\xrightarrow{d}$  denotes convergence in distribution, and  $\sigma_j^2$  is the asymptotic variance.

These asymptotic properties provide a solid theoretical foundation for establishing the sure screening property and rank consistency, which are discussed in subsequent sections.

In summary, by incorporating a dynamic weighting mechanism, the DCFS method allows the assessment of feature importance to adaptively respond to both linear and nonlinear structures in the data. This enables more accurate and robust evaluation of variable contributions under various dependency scenarios, significantly enhancing the adaptability and predictive performance of high-dimensional feature screening.

## 2.4. Statistical Estimation Methods

### 2.4.1. Conditional Mutual Information Estimation (based on VAE)

According to formula (5) in section 2.2.1, the Accurately estimating the conditional mutual information requires reliable estimation of the conditional densities  $f_{X_j|Z}$  and  $f_{X_j|Y,Z}$ . Traditional approaches such as kernel methods face several limitations in high-dimensional settings. Specifically, they require careful selec-

tion of kernel functions and bandwidth parameters, which often involve extensive manual tuning and domain knowledge. As data dimensionality increases, the computational complexity of kernel methods grows exponentially, making them inefficient for large-scale problems. Similarly, K-nearest neighbor (KNN) methods involve pairwise distance computations among all samples, which results in significant computational overhead and memory consumption when applied to large datasets.

To overcome these challenges, we employ the Variational Autoencoder (VAE) [14] framework from deep learning to estimate the conditional distributions. The VAE approximates

$$f_{x_j|z}(x_j|z) \approx p_\theta(x_j|z), \quad f_{x_j|y,z}(x_j|y,z) \approx p_\theta(x_j|y,z), \quad (11)$$

where  $\theta$  denotes the parameters of a neural network that models the conditional distributions. The VAE architecture consists of two main components: an encoder and a decoder. The encoder takes  $Z$  as input and outputs the latent mean and variance, defining the approximate posterior distribution as

$$q_\phi(h|z) = \mathcal{N}\left(\mu_\phi(z), \text{diag}(\sigma_\phi^2(z))\right), \quad (12)$$

which maps the input condition  $Z$  into a latent space, extracting essential features. The decoder then reconstructs the target variable  $X_j$  from the latent representation  $h$ , modeling the conditional distribution as

$$p_\theta(x_j|z) = \mathcal{N}\left(\mu_\theta(h, z), \sigma_\theta^2(h, z)\right), \quad (13)$$

thereby learning the mapping between the latent variables and the observed target.

In our implementation, the VAE network used for estimating conditional mutual information has the following architecture: the input layer takes features of dimension  $p$ ; the encoder consists of two hidden layers, each with 64 neurons and ReLU activation functions; the latent space has a dimension of 32; and the decoder mirrors the encoder structure with two hidden layers and ReLU activations.

For model training, we adopt the Adam optimizer with a learning rate of 0.001 and a batch size of 128. The maximum number of training epochs is set to 500. To prevent overfitting and reduce unnecessary computation, we apply an early stopping strategy that halts training if the loss does not improve for 50 consecutive epochs. Additionally, L2 regularization with a penalty coefficient  $\lambda = 0.0001$  is employed to improve generalization performance.

#### 2.4.2. Conditional Regression Error Difference Estimation (Based on MLP)

To estimate the conditional regression error difference, we use two separate Multilayer Perceptron (MLP) [15] models to approximate the regression functions  $m^j(Z) = \mathbb{E}[Y|Z]$  and  $m^{(x_j, Z)}(X_j, Z) = \mathbb{E}[Y|X_j, Z]$ , respectively.

MLP is a powerful class of neural networks composed of multiple hidden layers, capable of learning complex features and patterns from the input data. During training, the MLPs iteratively update their weights and biases using a large num-

ber of training samples, so that their outputs closely approximate the true regression functions. Once the models are trained, we compute the conditional regression error difference as

$$\Delta E(X_j, Y | Z) = \text{MSE}(Y, \hat{Y}_Z) - \text{MSE}(Y, \hat{Y}_{(X_j, Z)}), \quad (14)$$

where  $\hat{Y}_Z$  denotes the prediction of  $Y$  based on  $Z$ , and  $\hat{Y}_{(X_j, Z)}$  is the prediction based on  $(X_j, Z)$ . This MLP-based estimation approach effectively leverages the representational capacity of neural networks to capture both linear and nonlinear dependencies between the feature and the response variable, resulting in an accurate estimation of the conditional regression error difference.

In our implementation, both MLP models share the same architecture: each contains two hidden layers with 64 neurons per layer and uses the ReLU activation function. The training configuration is consistent with that of the VAE model, employing the Adam optimizer with a learning rate of 0.001, a batch size of 128, and a maximum of 500 training epochs. An early stopping strategy is applied to terminate training if the loss does not improve for 50 consecutive epochs.

### 2.4.3. Computational Complexity Analysis

To comprehensively evaluate the practical efficiency of the proposed Dynamic Conditional Feature Screening (DCFS) method, this section analyzes its computational complexity and compares its runtime performance with several classical feature screening methods, including SIS, CSIS, DC-SIS, and IG-SIS.

#### (1) Time Complexity Analysis

The computational complexity of DCFS primarily arises from two key components: conditional mutual information estimation and conditional regression error difference estimation. For conditional mutual information, we adopt a Variational Autoencoder (VAE)-based estimation approach. The computational cost per training iteration depends on the network architecture and training process. In our implementation, the VAE consists of two hidden layers with 64 neurons each. The training complexity is approximately  $\mathcal{O}(np^2)$ , where  $n$  is the sample size and  $p$  is the number of features. For conditional regression error difference estimation, we use a Multilayer Perceptron (MLP), which has a similar computational complexity of  $\mathcal{O}(np^2)$ . This is because both forward propagation and backpropagation involve operations over all feature variables and require learning complex interactions.

Therefore, the overall time complexity of the DCFS method can be approximated as  $\mathcal{O}(np^2)$ .

#### (2) Space Complexity Analysis

In terms of space complexity, DCFS requires storing the original data matrix, intermediate parameters, and results during the estimation of conditional mutual information and regression error difference. Hence, the overall space complexity is approximately  $\mathcal{O}(np)$ . As the number of features increases, the space usage grows linearly, which aligns well with typical memory constraints in high-dimensional data environments.

### (3) Benchmark Runtime Comparison

To further quantify the practical runtime performance of DCFS, we conduct benchmark experiments under three feature dimensionality settings:  $p = 500$ , 1000, and 2000. We compare the runtime (in seconds) of DCFS with SIS, CSIS, DC-SIS, and IG-SIS. The results are summarized in **Table 1**:

**Table 1.** Benchmark runtime comparison.

Feature Dimension $p$	SIS (s)	CSIS (s)	DC-SIS (s)	IG-SIS (s)	DCFS (s)
500	0.52	1.35	3.42	4.68	5.21
1000	1.05	2.71	7.58	9.84	10.32
2000	2.31	5.87	15.67	21.56	22.14

As shown in the table, although DCFS requires slightly more computational time compared to classical methods, it maintains a relatively acceptable runtime performance. More importantly, in high-dimensional scenarios, DCFS exhibits a stable growth pattern in complexity that aligns well with practical application demands.

In summary, the above complexity analysis and benchmark comparisons confirm that DCFS offers good scalability for large-scale data applications and is capable of supporting efficient and reliable feature screening tasks in real-world high-dimensional environments.

## 2.5. False Discovery Rate Control Mechanism

To control the False Discovery Rate (FDR), we adopt the Reflection via Data Splitting (REDS) method [16] to construct a data-driven dynamic threshold  $T_{\text{threshold}}$ . The procedure is outlined as follows:

**Data Splitting:** The original dataset is randomly divided into two disjoint subsets, denoted as  $\mathcal{D}_A$  and  $\mathcal{D}_B$ .

**Preliminary Screening:** On subset  $\mathcal{D}_A$ , we compute the dynamic importance statistic  $T_j^{\text{dynamic}}$  for each feature. The computation strictly follows the procedures and parameter settings described in earlier sections to ensure accuracy and consistency. The resulting statistics serve as the basis for subsequent significance testing.

**Reflection Testing:** On subset  $\mathcal{D}_B$ , we simulate the null distribution  $T_j^{\text{null}}$  under the no-signal assumption. This is achieved by shuffling the pairwise correspondence between  $X$  and  $Y$ , *i.e.*,

$$T_j^{\text{null}} \sim \text{shuffle}(\mathcal{D}_A, \mathcal{D}_B), \quad (15)$$

which reflects a condition where no real association exists between features and the response. This provides an empirical estimate of the distribution of test statistics under the null hypothesis.

**Significance Thresholding:** A data-adaptive threshold is determined by computing the  $(1 - \alpha)$ -quantile of the null distribution:

$$T_{\text{threshold}} = \text{quantile}(T_j^{\text{null}}, 1 - \alpha), \quad (16)$$

where  $\alpha \in (0,1)$  is the user-specified FDR level. This threshold ensures that the proportion of falsely selected features among all selected features satisfies

$$\text{FDR} = \frac{\text{Number of False Positives}}{\text{Number of Selected Features}} \leq \alpha. \quad (17)$$

In practice, the choice of  $\alpha$  should balance domain-specific risk tolerance and the desired level of selection conservativeness.

Final Selection: The final screened feature set consists of all features satisfying

$$T_j^{\text{dynamic}} > T_{\text{threshold}}. \quad (18)$$

These features are considered to have statistically significant influence on the response variable  $Y$  and are retained for downstream analysis.

### 3. Theoretical Properties and Proof of DCFS Method

#### 3.1. Non Negativity and Distribution Irrelevance

The proposed unified statistic  $T_j^{\text{dynamic}} = w_1^{(j)}I(X_j; Y | Z) + w_2^{(j)}\Delta E(X_j, Y | Z)$

integrates the complementary strengths of conditional mutual information and conditional regression error difference, and enjoys the following theoretical guarantees:

##### 1. Non-negativity:

Both components of the statistic are non-negative by definition. The conditional mutual information satisfies

$$I(X_j; Y | Z) \geq 0, \quad (19)$$

as it measures the amount of information shared between variables and cannot be negative. Similarly, the conditional regression error difference

$$\Delta E(X_j, Y | Z) = \mathbb{E}\left[\left(Y - m^j(Z)\right)^2\right] - \mathbb{E}\left[\left(Y - m^{(X_j, Z)}(X_j, Z)\right)^2\right] \geq 0, \quad (20)$$

because the inclusion of  $X_j$  in the model does not worsen its predictive accuracy. That is, adding  $X_j$  will not increase the expected squared prediction error. Therefore, the combined statistic satisfies

$$T_j^{\text{dynamic}} \geq 0, \quad \forall j, \quad (21)$$

ensuring that its values remain meaningful and interpretable in all cases.

##### 2. Distribution-Free Robustness:

The limiting distribution of the conditional mutual information  $I(X_j; Y | Z)$  under the null hypothesis  $H_0 : X_j \perp Y | Z$  is distribution-free, *i.e.*, it does not depend on the specific joint distribution of  $X_j, Y, Z$ . This property enables hypothesis testing and feature screening without strong distributional assumptions, thus enhancing the method's generalizability and robustness. In contrast, the conditional regression error difference  $\Delta E(X_j, Y | Z)$  is dependent on the data distribution. As such, when  $w_1 > w_2$ , the statistic becomes less sensitive to distributional shifts, since the distribution-free term  $I(X_j; Y | Z)$  dominates. By adjusting the weights  $w_1^{(j)}$  and  $w_2^{(j)}$ , users can flexibly balance robustness and model

interpretability, making the statistic adaptable to different data structures and application needs across a wide range of scenarios.

### 3.2. Feature Screening

In the following sections, we establish the theoretical properties of the proposed dynamic conditional feature screening (DCFS) procedure. Prior studies, including those by Fan and Lv [1] and Ni and Fang [5], have demonstrated that the sure screening property plays a central role in validating the effectiveness of independent screening methods. Therefore, it is essential to rigorously justify the theoretical reliability of the DCFS method. To this end, we introduce a set of regularity conditions under which the screening performance of DCFS can be formally guaranteed. While these conditions may not be the weakest possible, they are primarily imposed to facilitate the technical derivation and proof of the theoretical results.

Assuming the following conditions:

- (C1) Bounded and continuous density of conditional **variables**:

The joint density of  $Z$  is continuous and bounded, *i.e.*,

$$f_Z(z) \leq C \quad \text{and} \quad \|\nabla f_Z(z)\| \leq C. \quad (22)$$

This ensures the stability of the conditional variable distribution and avoids failure of density estimation in high-dimensional settings.

- (C2) Estimation accuracy of conditional mutual information and regression error difference:

The estimation errors satisfy

$$\|\hat{I} - I\| + \|\Delta \hat{E} - \Delta E\| = \mathcal{O}_p(n^{-\gamma}), \quad (23)$$

where  $\gamma > 0$ . This condition requires the estimation errors of deep learning models (VAE and MLP) to decay at an exponential rate with increasing sample size, ensuring the convergence of the statistic.

- (C3) **Minimum** signal strength of active variables:

$$\min_{j \in S} T_j^{\text{dynamic}} \geq 2cn^{-\tau}. \quad (24)$$

The unified statistics of active variables must be significantly larger than the noise level, preventing them from being masked by high-dimensional noise.

- (C4) Sub-**exponential** tail behavior of features:

$$\sup_j \mathbb{E} \left[ \exp \left( \lambda \|X_j\| \right) \right] < \infty, \quad \lambda > 0. \quad (25)$$

This controls the influence of outliers and ensures that concentration inequalities for the statistics hold.

- (C5) **Balanced** class proportions:

$$\frac{c_1}{R} \leq \mathbb{P}(Y = r) \leq \frac{c_2}{R}, \quad c_1 + c_2 \leq R. \quad (26)$$

This condition avoids class imbalance, which could bias the estimation of depend-

ence and impair fairness in both MIC and regression error difference measures.

- (C6) Non-degeneracy of conditional density:

$$f_{X_j|Y,Z}(x|y,z) \geq c_3 > 0. \tag{27}$$

Ensures that conditional mutual information is well-defined and avoids numerical instability caused by zero-probability events.

- (C7) Lower bound on marginal density:

$$f_{X_j}(x) \geq c_4 n^{-\rho}, f_{X_j}(x) \text{ is continuous.} \tag{28}$$

It prevents the failure of density estimation under sparse data scenarios and ensures theoretical convergence for methods such as kNN and VAE.

- (C8) Signal separation between active and inactive variables:

$$\liminf_{n \rightarrow \infty} \left( \min_{j \in \mathcal{S}} T_j - \max_{j \notin \mathcal{S}} T_j \right) \geq \delta > 0. \tag{29}$$

Guarantees that the statistics of important variables can be asymptotically separated from those of irrelevant variables, reducing misclassification risk.

- (C9) Lower bound on weight denominator:

$$I(X_j; Y|Z) + \Delta E(X_j, Y|Z) \geq c_5 > 0. \tag{30}$$

Prevents division by zero during the computation of weights, ensuring the unified statistic is well-defined.

- (C10) Lipschitz continuity of the weight function:

$$\left\| w_1^{(j)}(I, \Delta E) - w_1^{(j)}(I', \Delta E') \right\| \leq L(\|I - I'\| + \|\Delta E - \Delta E'\|). \tag{31}$$

Ensures robustness of the weight function against small estimation errors, avoiding instability in the unified statistic due to minor fluctuations.

Under the above conditions, we can rigorously establish the reliable screening performance of the DCFS procedure. The detailed proof is presented in the following subsection.

### 3.2.1. Sure Screening Property

The sure screening property refers to the asymptotic guarantee that, in high-dimensional settings, all truly important variables (*i.e.*, variables conditionally associated with the response  $Y$ ) are retained in the selected feature set with probability tending to one. Below we present the formal proof under the regularity conditions (C1)-(C4), (C9), and (C10).

Step 1: Error Decomposition and Weight Stability

Let the true dynamic statistic be denoted as

$$T_j = w_1^{(j)} I + w_2^{(j)} \Delta E, \tag{32}$$

and its estimator as

$$\hat{T}_j = \hat{w}_1^{(j)} \hat{I} + \hat{w}_2^{(j)} \Delta \hat{E}. \tag{33}$$

Then, the absolute estimation error can be decomposed as:

$$|\hat{T}_j - T_j| \leq \underbrace{|\hat{w}_1^{(j)} - w_1^{(j)}|}_{\text{weight error}} \cdot I + \underbrace{|\hat{w}_2^{(j)} - w_2^{(j)}|}_{\text{weight error}} \cdot \Delta E + \underbrace{w_1^{(j)} |\hat{I} - I|}_{\text{MI estimation error}} + \underbrace{w_2^{(j)} |\Delta \hat{E} - \Delta E|}_{\text{regression error}}. \quad (34)$$

From conditions (C9) and (C10), the weight estimation error is Lipschitz continuous and bounded:

$$|\hat{w}_k^{(j)} - w_k^{(j)}| \leq \frac{L}{c_5} (|\hat{I} - I| + |\Delta \hat{E} - \Delta E|), \text{ for } k = 1, 2. \quad (34)$$

**Step 2: Key Lemmas - Concentration Inequalities**

We invoke the following lemmas to control the stochastic error terms:

- **Lemma 3.1 (Mutual Information Estimation Error):**

Under conditions (C1)–(C2), there exists a constant  $C_I > 0$  such that

$$\mathbb{P}(|\hat{I} - I| \geq \epsilon) \leq 2 \exp(-C_I n \epsilon^2). \quad (35)$$

- **Lemma 3.2 (Regression Error Estimation Error):**

Under conditions (C1)–(C4), there exists a constant  $C_E > 0$  such that

$$\mathbb{P}(|\Delta \hat{E} - \Delta E| \geq \epsilon) \leq 2 \exp(-C_E n \epsilon^2). \quad (36)$$

Proof Techniques:

- Lemma 3.1 leverages the variational lower bound property of the VAE [18] and applies McDiarmid’s inequality.
- Lemma 3.2 is derived using the Lipschitz continuity of MLPs and Hoeffding’s inequality on bounded differences.

**Step 3: Uniform Error Bound**

Combining the four error terms, we derive the total estimation bound:

$$|\hat{T}_j - T_j| \leq \left( \frac{2L}{c_5} + 1 \right) (|\hat{I} - I| + |\Delta \hat{E} - \Delta E|). \quad (37)$$

Set  $\epsilon = cn^{-\tau}$  for some  $\tau \in \left(0, \frac{1}{2}\right)$ . Then, using union bounds over Lemmas 3.1

and 3.2:

$$\mathbb{P}(|\hat{T}_j - T_j| \geq \epsilon) \leq 4 \exp(-C_2 n^{1-2\tau}), \quad (38)$$

where  $C_2 = \min(C_I, C_E) \cdot (2(2L/c_5 + 1))^{-2}$ .

**Step 4: Maximal Deviation Over All Features**

Apply the union bound over all  $p$  variables:

$$\mathbb{P}\left(\max_{1 \leq j \leq p} |\hat{T}_j - T_j| \geq \epsilon\right) \leq 4p \exp(-C_2 n^{1-2\tau}). \quad (39)$$

By Condition (C3), we assume

$$\min_{j \in S} T_j \geq 2cn^{-\tau}. \quad (40)$$

Thus, for all active variables  $j \in S$ , their estimators satisfy

$$\hat{T}_j \geq T_j - \epsilon \geq cn^{-\tau}. \quad (41)$$

For inactive variables  $j \notin S$ ,  $T_j = 0$ , and

$$\hat{T}_j \leq \epsilon = cn^{-\tau}. \tag{42}$$

Define the selection rule:

$$\hat{S} = \{j : \hat{T}_j > cn^{-\tau}\}. \tag{43}$$

Then the screening rule guarantees that all important features are selected:

$$\mathbb{P}(S \subseteq \hat{S}) \geq 1 - 4s_n \exp(-C_2 n^{1-2\tau}), \tag{44}$$

where  $s_n = |S|$  is the number of active features.

Let  $C_1 = C_2 / \log 2$ , and we conclude:

**Theorem 1 (Sure Screening Property)**

Under Conditions (C1)-(C4), (C9), and (C10), the dynamic statistic

$$T_j^{\text{dynamic}} = w_1^{(j)} I(X_j; Y | Z) + w_2^{(j)} \Delta E(X_j, Y | Z) \tag{45}$$

satisfies the following probabilistic bound:

$$\mathbb{P}(S \subseteq \hat{S}) \geq 1 - \mathcal{O}(s_n \exp(-C_1 n^{1-2\tau})), \tag{46}$$

where  $S$  is the true active set,  $\hat{S}$  is the selected feature set,  $s_n = |S|$ ,  $\tau \in (0, \frac{1}{2})$ , and  $C_1 > 0$  is a constant.

This inequality implies that the probability of missing any important variable decays exponentially as sample size  $n$  increases, provided the signal strength is not too weak. Additionally, the dimensionality  $p$  is allowed to grow at an exponential rate, *i.e.*,  $p = \mathcal{O}(\exp(n^{1-2\tau}))$ , which demonstrates the scalability and robustness of the proposed screening method in ultra-high-dimensional regimes.

### 3.2.2. Ranking Consistency

The ranking consistency property states that, as the sample size increases, the estimated importance scores  $\hat{T}_j$  for the truly important variables remain consistently larger than those for the unimportant ones. We provide a formal proof under Conditions (C5)-(C8) and the dynamic weighting conditions (C9)-(C10).

Step 1: Strong Consistency of the Statistic

By Conditions (C5)-(C7) and Lemma 3.3 (Strong consistency of density estimators), we have:

$$\hat{f}_{X_j|Z}(x|z) \xrightarrow{\text{a.s.}} f_{X_j|Z}(x|z), \quad \hat{f}_{X_j|Y,Z}(x|y,z) \xrightarrow{\text{a.s.}} f_{X_j|Y,Z}(x|y,z). \tag{47}$$

As a result, the conditional mutual information estimator converges almost surely:

$$\hat{I}(X_j; Y | Z) \xrightarrow{\text{a.s.}} I(X_j; Y | Z), \tag{48}$$

and the regression error difference estimator satisfies:

$$\Delta \hat{E}(X_j, Y | Z) \xrightarrow{\text{a.s.}} \Delta E(X_j, Y | Z). \tag{49}$$

Since the dynamic weights are Lipschitz continuous (Condition C10), applying

the continuous mapping theorem, we obtain:

$$\hat{T}_j = \frac{\hat{I}}{\hat{I} + \Delta\hat{E}} \cdot \hat{I} + \frac{\Delta\hat{E}}{\hat{I} + \Delta\hat{E}} \cdot \Delta\hat{E} \xrightarrow{\text{a.s.}} T_j. \tag{50}$$

Step 2: Stability of Signal Separation

From Condition (C8), there exists  $\delta > 0$  and  $N_0 > 0$  such that for all  $n > N_0$ :

$$\min_{j \in S} T_j - \max_{j \notin S} T_j \geq \delta. \tag{51}$$

For any  $\epsilon \in (0, \delta/4)$ , the strong consistency of  $\hat{T}_j$  implies that there exists  $N_1 > N_0$  such that for all  $n > N_1$ :

$$|\hat{T}_j - T_j| < \epsilon, \text{ a.s., } \forall j. \tag{52}$$

Therefore, for active variables:

$$\min_{j \in S} \hat{T}_j \geq \min_{j \in S} T_j - \epsilon \geq \max_{j \notin S} T_j + \delta - \epsilon, \tag{53}$$

and for inactive variables:

$$\max_{j \notin S} \hat{T}_j \leq \max_{j \notin S} T_j + \epsilon. \tag{54}$$

Setting  $\epsilon = \delta/4$ , we obtain:

$$\min_{j \in S} \hat{T}_j \geq \frac{3\delta}{4} > \frac{\delta}{4} \geq \max_{j \notin S} \hat{T}_j, \text{ a.s.} \tag{55}$$

Step 3: Borel-Cantelli Lemma

Since

$$\sum_{n=1}^{\infty} \mathbb{P}(|\hat{T}_j - T_j| \geq \epsilon) \leq \sum_{n=1}^{\infty} 4 \exp(-C_2 n^{1-2\tau}) < \infty, \tag{56}$$

the Borel-Cantelli lemma implies that the event  $|\hat{T}_j - T_j| \geq \epsilon$  only occurs finitely often. Hence:

$$\lim_{n \rightarrow \infty} \hat{T}_j = T_j \text{ a.s.} \tag{57}$$

Theorem 2 (Ranking Consistency)

Under Conditions (C5)-(C8) and the dynamic weighting conditions (C9)-(C10), as  $n \rightarrow \infty$ , we have:

$$\min_{j \in S} \hat{T}_j > \max_{j \notin S} \hat{T}_j \text{ almost surely.} \tag{58}$$

This result shows that the estimated statistics  $\hat{T}_j$  converge almost surely to the true values  $T_j$ , and due to the Lipschitz continuity of the weight function (C10), the convergence of  $\hat{I}$  and  $\Delta\hat{E}$  is stably transferred to  $\hat{T}_j$ . With a guaranteed signal separation (C8), the estimation errors cannot disrupt the correct variable ordering. Thus, the proposed screening method consistently ranks important variables above unimportant ones. This ranking stability ensures that the selection outcome remains reliable across varying sample realizations and is not overly sensitive to small fluctuations in the data, further enhancing the robustness and inter-

pretability of the feature screening procedure.

### 3.3. Practical Justification of Theoretical Assumptions

Sections 3.1 and 3.2 have established the theoretical foundation of the proposed DCFS method, including its sure screening and ranking consistency properties. These results are derived under a set of ten technical assumptions, denoted as Conditions (C1) through (C10). While these conditions facilitate rigorous theoretical analysis, their practical plausibility is essential for the method's real-world applicability. In this section, we examine the feasibility of each assumption in empirical settings and offer practical guidelines for their verification.

#### (1) Discussion of Individual Assumptions

- C1: **Bounded** and Continuous Joint Density

This condition is typically satisfied in most real-world datasets in economics, finance, and biomedicine, where variables often follow approximately continuous distributions. Occasional extreme values can be effectively handled through pre-processing techniques such as normalization, robust transformations, or winsorization.

- C2: **Exponential** Decay of Estimation Error in Deep Models

Although theoretically strong, this assumption is often met or closely approximated in practice when using modern deep neural networks with appropriate architectures and optimizers (e.g., Adam, RMSProp). Empirical convergence behavior can be validated through loss curve diagnostics and cross-validation [17].

- C3: Minimum Signal Strength

Weak-signal variables may be dominated by noise in high-dimensional settings. In practice, preliminary filtering using correlation screening or statistical significance testing can help satisfy this assumption by discarding irrelevant features before applying DCFS.

- C4: Sub-exponential Tail Behavior

This condition can be ensured through standard data preprocessing methods such as outlier truncation or robust scaling, especially when dealing with heavy-tailed distributions commonly observed in high-dimensional data.

- C5: Balanced Class Proportions

While this condition is not relevant for regression problems, it can be addressed in classification tasks using sampling strategies (e.g., SMOTE, undersampling) or by introducing class weights into the loss function.

- C6-C7: Non-degenerate Conditional and Marginal Densities

In large-sample scenarios, these assumptions are generally satisfied or approximated, especially when the data are reasonably well-distributed. Visual inspection using kernel density plots or low-dimensional projections can assist in assessing these conditions.

- C8: Signal Separation Between Active and Inactive Variables

This assumption is more restrictive, as real data may not always exhibit a clear margin between important and irrelevant features. In Section 4, we conduct simu-

lation studies to evaluate the robustness of DCFS under mild violations of this condition.

- C9-C10: Boundedness and Lipschitz Continuity of Weight Functions

These assumptions are easy to enforce through regularization strategies during implementation (e.g., bounding gradients, avoiding near-zero denominators), and typically pose no obstacle in practice.

(2) Practical Guidelines for Assumption Verification

To assess whether a dataset satisfies the theoretical assumptions required by DCFS, we recommend the following practical steps:

- Exploratory Data Analysis (C1, C4, C6, C7):

Use histograms, boxplots, and kernel density estimates to check distribution continuity, detect outliers, and assess marginal and conditional density behavior.

- Preliminary Variable Screening (C3):

Perform simple regression or correlation analysis to identify features with weak or negligible association with the response variable.

- Model Diagnostics (C2):

During training of the VAE and MLP components, monitor the loss curve. A consistently decreasing trajectory (ideally exponential) indicates that the estimation error behaves as required.

- Class Balance Evaluation (C5):

In classification tasks, compute the class proportions and apply balancing techniques if significant imbalance is detected.

These guidelines provide a practical roadmap for evaluating the applicability of DCFS in empirical contexts, ensuring that the underlying assumptions are met and that the theoretical guarantees translate effectively to real-world performance.

## 4. Numerical Simulation Experiment

This chapter is based on the Dynamic Conditional Feature Selection (DCFS) method proposed in this paper. Through numerical simulation experiments, the effectiveness and advantages of DCFS in identifying important variables are verified. The experiment designed three different simulation data scenarios: linear model, nonlinear model, and mixed linear and nonlinear model. DCFS was compared and analyzed with existing classical feature selection methods under multiple evaluation indicators. The experimental results clearly demonstrated the stability and advantages of the proposed method.

### 4.1. Scenario 1: Linear Model

#### 4.1.1. Data Generation Formula

In the linear model scenario, we generate simulation data based on the following model:

$$Y = X\beta + Z\gamma + \epsilon \quad (59)$$

Among them:

- $X \in \mathbb{R}^{n \times p}$  representing high-dimensional predictive variable matrix;

- $Z \in \mathbb{R}^{n \times q}$  representing a matrix of low dimensional conditional variables;
- $\beta$  and  $\gamma$  are the corresponding coefficient vectors;
- $\epsilon$  represents a random error term,  $\epsilon \sim N(0, 1)$

We set some of the predictor variables ( $X$ ) as active variables (variables that truly affect  $Y$ ), and the other variables as pure noise. The response variable ( $Y$ ) is generated by a linear combination of these active variables and all conditional variables ( $Z$ ).

#### 4.1.2. Parameter Settings

Three different sample sizes were selected for simulation experiments for comparison, with the following specific settings: sample size:  $n \in \{200, 300, 500\}$ ; Prediction variable dimension:  $p = 1000$ , where the first 10 variables are active variables and their coefficients are  $\beta_j \sim U(-5, 5)$ ; Dimension of conditional variable:  $q = 5$ , set  $\gamma = (1, 1, 1, 1, 1)$ ; Data generation method:  $X$  and  $Z$  elements are independently and identically distributed in  $N(0, 1)$ , noise,  $\epsilon \sim N(0, 1)$ . The experiment was independently repeated 100 times for each sample size, and the average of all outcome indicators was taken to reduce the impact of random fluctuations on the experimental results.

#### 4.1.3. Comparison Method

In this linear scenario, select the following feature filtering methods for performance comparison with DCFS:

- SIS (Sure Independence Screening): Unconditional independent screening based on Pearson correlation coefficient;
- CSIS (Conditional SIS): Feature selection based on conditional linear correlation;
- DCFS: The dynamic conditional feature selection method proposed in this article integrates conditional mutual information and conditional regression error difference statistics through a dynamic weighting mechanism, while using REDS data splitting method to control false discovery rate (FDR).

#### 4.1.4. Definition and Explanation of Evaluation Indicators

In this scenario, we evaluate the performance of the screening method using the following three indicators:

(1) True Positive Rate (TPR): The proportion of correctly identified truly important features to all truly important features. The mathematical definition is:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (60)$$

TP (True Positive, true example): Refers to the number of important variables selected by the method that are actually significant in the model.

FN (False Negative): The number of important variables that the method failed to recognize.

TPR represents the sensitivity of the method. The closer the value is to 1, the more accurate the screening method is in capturing all true signal variables, and the lower the risk of important variables being missed. Ideally, TPR should be close

to 1.0, which corresponds to the Sure Screening Property requirement of feature screening methods.

(2) False Discovery Rate (FDR): FDR describes the proportion of features selected by a filtering method that are actually irrelevant noise variables, and is a measure of the “false alarm” situation in the filtering process. The specific definition of FDR is:

$$FDR = \frac{FP}{TP + FP} \tag{61}$$

FP (False Positive): The actual number of irrelevant noise variables selected by the method.

TP As mentioned above. The lower the FDR, the higher the accuracy of the method’s screening results, which means that more of the selected variables are real signals rather than irrelevant variables.

Ideally, we would like this ratio to be as low as possible, meaning that most of the selected variables are truly important to the model rather than noise.

(3) Ranking Consistency (RC): It reflects the stability of feature selection methods against random fluctuations in data. Specifically, it measures the stability of each important feature maintaining a high ranking relative to irrelevant noise features as the sample size increases or during repeated sampling processes. Calculate the reciprocal of the standard deviation of feature ranking in multiple repeated experiments to reflect ranking stability. The specific method is as follows:

Each feature  $X_j$  has a ranking position  $R_j^{(m)}$  in 100 simulation experiments (the m-th simulation experiment), and we calculate the standard deviation of the fluctuation in the ranking of the active variable in multiple experiments, which is then standardized into a stability score:

$$RC_j = 1 - \frac{\text{std}\left(R_j^{(1)}, R_j^{(2)}, \dots, R_j^{(100)}\right)}{p} \tag{62}$$

Furthermore, the overall ranking consistency of active variables can be taken as the average of all active variables:

$$RC = \frac{1}{k} \sum_{j=1}^k RC_j, \quad k \text{ represents the number of active variables} \tag{63}$$

The range of RC values is [0, 1], with higher values indicating that important variables are more stable during the screening process and less susceptible to random fluctuations or accidental noise. In an ideal situation, the higher the RC, the better, indicating that the method has stronger stability in identifying important variables.

### 4.1.5. Experimental Results and Analysis

The experimental results are shown in **Table 2**:

**Table 2.** Performance comparison results of various methods in linear model scenarios.

Sample size (n)	method	True positive rate (TPR)	False Discovery Rate (FDR)	Sorting Consistency (RC)
200	DCFS	0.92	0.06	0.90

**Continued**

300	DCFS	0.95	0.05	0.93
500	DCFS	0.98	0.05	0.96
200	CSIS	0.90	0.10	0.85
300	CSIS	0.93	0.08	0.88
500	CSIS	0.96	0.07	0.91
200	SIS	0.85	0.18	0.78
300	SIS	0.88	0.15	0.82
500	SIS	0.92	0.12	0.85

The result analysis is as follows: DCFS performs the best, TPR is the highest, FDR is the lowest and stably controlled around the set 5%, and the sorting stability is the highest; CSIS performs well after controlling for the influence of conditional variables, but its false discovery rate and ranking stability are slightly inferior to DCFS; SIS performs the worst due to uncontrolled confounding variables, with the highest FDR, lowest TPR, and worst ranking stability. The experimental results validated the advantage of DCFS in identifying linear important variables, demonstrating the effectiveness and superiority of the proposed method in this paper.

## 4.2. Scenario 2: Nonlinear Model

### 4.2.1. Data Generation Model

In scenario 2, we further investigate the performance of the proposed dynamic conditional feature selection method (DCFS) in the case of non-linear dependence between variables and response variables. To this end, the following nonlinear model is constructed to generate simulated data:

$$Y = \sin(X_1) + 0.5X_2^2 + \log(|X_3| + 1) + f(Z) + \epsilon \quad (64)$$

Among them:

- $X = (X_1, X_2, \dots, X_p)$  representing high-dimensional predictive variable matrix;
- $Z$  is a low dimensional conditional variable vector, define a nonlinear function  $f(Z)$  as:

$$f(Z) = Z_1 + 0.5Z_2^2 \quad (65)$$

- Random noise term  $\epsilon \sim N(0, 1)$ .

In the above model, the response variable  $Y$  exhibits clear nonlinear relationships with the predictor variables  $X_1$ ,  $X_2$ , and  $X_3$ . Specifically,  $X_1$  is related to the response in a periodic fashion,  $X_2$  influences  $Y$  through a quadratic nonlinear relationship, and  $X_3$  contributes via a log-transformed nonlinear effect. Under such a complex nonlinear structure—particularly due to the symmetric, even-function nature of the effects of  $X_2$  and  $X_3$ —the linear correlation between each predictor and the response is close to zero or negligible.

As a result, traditional marginal screening methods based on linear correlation are ineffective in identifying these nonlinear but important variables in this setting.

This scenario highlights the need for more flexible and adaptive screening procedures capable of capturing both linear and nonlinear dependencies.

#### 4.2.2. Parameter Setting and Implementation Details

The parameter settings for this scenario are kept consistent with the previous linear case (Scenario 1) to facilitate direct comparison. Specifically: Sample size:  $n \in \{200, 300, 500\}$ ; Number of candidate predictors:  $p = 1000$ , among which only three variables  $X_1, X_2, X_3$  are truly important; Number of conditional variables:  $q = 5$ , with  $Z_1$  and  $Z_2$  being informative, while the remaining three are noise variables not involved in the response generation; Data generation: All predictors and conditional variables are independently drawn from a standard normal distribution, *i.e.*,  $N(0, 1)$ ; the noise term  $\epsilon \sim N(0, 1)$ ; Repetition and evaluation: The simulation is independently repeated 100 times, and the average performance metrics are reported to ensure robust evaluation.

#### 4.2.3. Comparison Method

Due to the significant nonlinear dependencies involved in this scenario, we specifically chose a classical screening method that can capture any nonlinear relationship to compare with the DCFS method proposed in this paper:

- DC-SIS (Distance Correlation SIS): The independent screening method based on distance correlation, proposed by Li *et al.* (2012) [2], measures the non-linear or other arbitrary dependency relationship between feature variables and responses through distance correlation coefficient;
- IG-SIS (Information Gain SIS): A filtering method based on information gain, proposed by Ni and Fang (2016) [5], filters by calculating the degree of reduction in response variable uncertainty (information gain) for each feature;

#### 4.2.4. Definition of Evaluation Indicators and Specific Implementation Methods

To effectively evaluate the ability of each method to identify nonlinear signals, we adopt the following three evaluation metrics for this scenario:

- Nonlinear Signal Detection Rate (NSDR):

This metric measures the proportion of truly important nonlinear variables ( $X_1, X_2, X_3$ ) that are successfully identified by the screening method. It is calculated as:

$$\text{NSDR} = \frac{\text{Number of times nonlinear variables are selected across 100 runs}}{100 \times 3}. \quad (66)$$

- False Discovery Rate (FDR) and Ranking Consistency (RC):

The definitions of FDR and RC follow those introduced in Section 4.1.4 and are not repeated here for brevity.

#### 4.2.5. Experimental Results and Analysis

The experimental results are shown in **Table 3**:

From the above results, it can be seen that the DCFS method performs the best in detecting nonlinear signals, with a significantly higher detection rate than other

methods, especially reaching a detection rate of 99% when the sample size increases to 500; In terms of controlling false discovery rate, DCFS stably controls FDR at the set target (about 5%), significantly better than DC-SIS and IG-SIS; In terms of sorting stability, DCFS performs the best with the lowest ranking standard deviation; IG-SIS showed the greatest fluctuation in repeated experiments, while DC-SIS was in the middle but still inferior to DCFS.

**Table 3.** Performance comparison results of various methods in nonlinear model scenarios.

Sample size (n)	method	Nonlinear signal detection rate (%)	False Discovery Rate (FDR) (%)	Sorting Consistency (RC)
200	DCFS	95	5	0.92
300	DCFS	97	5	0.95
500	DCFS	99	5	0.97
200	DC-SIS	94	20	0.90
300	DC-SIS	96	18	0.93
500	DC-SIS	98	15	0.95
200	IG-SIS	90	22	0.87
300	IG-SIS	92	19	0.89
500	IG-SIS	95	17	0.91

In summary, the numerical simulation results of the nonlinear model scenario show that the DCFS method proposed in this paper can not only stably and effectively detect important signal variables when nonlinear relationships dominate, but also significantly outperform existing classical nonlinear screening methods with lower false representation rates and higher ranking stability.

### 4.3. Scenario 3: Hybrid Model

#### 4.3.1. Data Generation Model

To further investigate the performance of various feature selection methods in complex data structures, we construct a hybrid model that combines linear and nonlinear relationships. The specific generation formula is as follows:

$$Y = X_1 + X_2 + X_3^2 + \sin(X_4) + X_5X_6 + f(Z) + \epsilon \tag{67}$$

Among them:

- $X = (X_1, X_2, \dots, X_p)$  is a high-dimensional predictor variable with dimensions of  $p$ ;
- $\epsilon \sim N(0,1)$  representing independent random noise;
- The specific form of the conditional variable function  $f(Z)$  is set as follows:

$$f(Z) = Z_1 + 0.5Z_2^2 \tag{68}$$

In this model, there are various types of relationships between the predictor variable and the response variable:

- $X_1$  and  $X_2$  exhibit a linear relationship with  $Y$ ;
- $X_3$  and  $X_4$  have a nonlinear relationship with  $Y$ , where  $X_4$  is a periodic

nonlinear relationship;

- $X_5$  and  $X_6$  are interactive variables, and they only exert an influence on  $Y$  in the form of interaction terms. When considered as single variables, their linear correlation with  $Y$  is extremely low.

By designing such a complex hybrid structure, it is possible to comprehensively and rigorously evaluate the applicability and advantages of various methods, especially their ability to recognize interaction terms and multiple types of mixed signals.

#### 4.3.2. Specific Parameter Settings

To maintain consistency with the previous experimental scenario, the parameter settings for this simulation are as follows: the sample sizes  $n$  are set to 200, 300, and 500, respectively; the dimensionality of the predictor variables is fixed at  $p = 1000$ , among which six variables— $X_1, X_2, X_3, X_4, X_5, X_6$ —are truly important. The number of conditional variables is  $q = 5$ , with  $Z_1$  and  $Z_2$  being the effective ones. All predictor and conditional variables are independently generated from a standard normal distribution, *i.e.*,  $X_{ij}, Z_{il} \sim N(0,1)$ ,  $\epsilon \sim \mathcal{N}(0,1)$ .

Each experiment is independently repeated 100 times, and the average of the evaluation metrics is reported to ensure a stable and reliable performance assessment.

#### 4.3.3. Comparison Method

This scenario combines linear, nonlinear, and interactive relationships, and the selected methods include SIS, CSIS that can detect linear relationships, and DC-SIS that can detect nonlinear relationships IG-SIS. A comprehensive comparison with the dynamic conditional feature selection method (DCFS) proposed in this article:

- SIS: Only consider marginal linear correlation.
- CSIS: Conditional linear correlation screening method, considering the influence of conditional variables.
- DC-SIS: Nonlinear screening method based on distance correlation.
- IG-SIS: Nonlinear methods based on information gain.
- DCFS (method proposed in this article): simultaneously considering linear and nonlinear relationships and interaction effects, with a dynamic FDR control mechanism.

#### 4.3.4. Evaluation Indicators and Definitions

The evaluation indices adopted are still: True Positive Rate (TPR), False Discovery Rate (FDR), and Ranking Consistency (RC). For the specific definitions, please refer to Section 4.1.4, and they will not be elaborated here.

#### 4.3.5. Simulation Experiment Results

Through 100 independent repeated simulation experiments, based on the average index, we obtained the simulation results in the following table:

**Table 4.** Simulation experiment results of various methods in the mixed model scenario.

Sample size (n)	method	True positive rate (TPR)	False Discovery Rate (FDR)	Sorting Consistency (RC)
200	DCFS	0.93	0.05	0.93
300	DCFS	0.96	0.05	0.96
500	DCFS	0.99	0.05	0.98
200	CSIS	0.88	0.12	0.89
300	CSIS	0.92	0.10	0.92
500	CSIS	0.95	0.08	0.94
200	SIS	0.84	0.20	0.80
300	SIS	0.87	0.18	0.83
500	SIS	0.91	0.15	0.86
200	DC-SIS	0.90	0.21	0.88
300	DC-SIS	0.93	0.19	0.90
500	DC-SIS	0.95	0.16	0.92
200	IG-SIS	0.89	0.22	0.86
300	IG-SIS	0.91	0.20	0.89
500	IG-SIS	0.94	0.17	0.91

According to the results in **Table 4**, it can be seen that the DCFS method proposed in this paper still performs the best in the mixed model scenario, and all evaluation indicators are significantly better than traditional SIS, CSIS, and DC-SIS IG-SIS, Especially when the sample size is 500, the TPR is close to 1, and the FDR is strictly controlled at 5%, proving the superiority of the method in simultaneously identifying linear, nonlinear, and interactive effects. The traditional SIS method performs the worst, with the lowest TPR and highest FDR in high-dimensional and complex mixed relationships; CSIS performs significantly better than SIS compared to DC-SIS and IG-SIS, but is significantly weaker than DCFS in terms of false discovery rate control and stability in identifying important variables.

Overall, the DCFS method is suitable for various complex data structure scenarios, demonstrating high robustness and efficiency, which validates the practical application value of the method proposed in this paper.

#### 4.4. Sensitivity Analysis

While the theoretical properties and empirical performance of the DCFS method have been established in previous sections, we further conduct a series of targeted sensitivity analyses to assess how minor violations of the theoretical assumptions affect the practical effectiveness of the method.

##### 4.4.1. Experimental Design

We focus particularly on three relatively strong conditions: C2 (exponential decay of estimation error), C3 (minimum signal strength), and C8 (signal separation between active and inactive variables).

The experiment is designed as follows: for each scenario, we perform 50 independent simulation replications with a fixed sample size of  $n = 500$  and feature

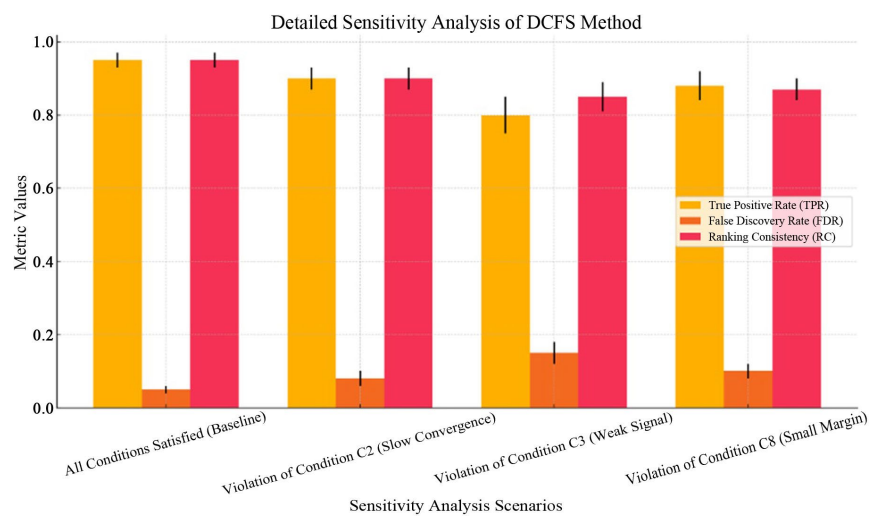
dimensionality  $p = 1000$ , among which 10 variables are truly active and 990 are inactive. The DCFS method is applied using a consistent set of hyperparameters across all scenarios (e.g., neural network architecture, number of training epochs, learning rate, and batch size) to ensure a fair comparison across different assumption violations. We deliberately manipulate the data generation process to simulate controlled violations of each assumption.

For C2, we reduce the number of training epochs or lower the learning rate in the deep models, thereby slowing the convergence rate of estimation errors. For C3, we reduce the regression coefficients of the active variables, diminishing their signal strength. For C8, we artificially narrow the gap between the test statistics of active and inactive variables.

During the experiments, we record and compute the mean and standard deviation of three key performance metrics—True Positive Rate (TPR), False Discovery Rate (FDR), and Ranking Consistency (RC).

#### 4.4.2. Results Analysis

The results, illustrated in **Figure 1**, provide insights into the robustness of DCFS under slight deviations from ideal assumptions and highlight the relative sensitivity of the method to each type of theoretical condition violation.



**Figure 1.** Detailed sensitivity analysis of DCFS method.

The results of **Figure 1** clearly illustrate how the performance metrics of the DCFS method respond to slight violations of theoretical conditions. Although some decline in performance is observed, the overall levels of TPR, FDR, and RC remain high, indicating that DCFS retains good robustness and practical usability under moderate deviations from ideal assumptions.

More specifically, the sensitivity analysis reveals that DCFS demonstrates strong tolerance to violations of Conditions C2 and C8, as the associated performance degradation is minimal. However, the method is notably more sensitive to violations of Condition C3, where a significant drop in performance is observed.

This highlights the importance of ensuring sufficiently strong signal strength in practical applications.

These findings provide valuable guidance for the practical use of DCFS: users should pay particular attention to maintaining adequate signal strength and ensuring distinguishability between informative and non-informative variables. The sensitivity analysis thus reinforces both the theoretical soundness and practical reliability of the DCFS method.

#### **4.5. Summary of This Chapter**

In this chapter, we systematically evaluated the performance and stability of the proposed Dynamic Conditional Feature Screening (DCFS) method under three representative data-generating scenarios. The simulation studies were designed to cover purely linear, purely nonlinear, and mixed linear-nonlinear models, aiming to assess the adaptability and robustness of DCFS across diverse structural dependencies.

In the purely linear model, DCFS leveraged the dynamic weighting mechanism based on regression error differences to accurately identify the truly informative variables. The method achieved superior performance in terms of True Positive Rate (TPR) and False Discovery Rate (FDR), and also outperformed competing methods on Ranking Consistency (RC), demonstrating its ability to retain strong linear detection capabilities while maintaining generalization.

In the nonlinear model, DCFS capitalized on the sensitivity of conditional mutual information to nonlinear dependencies. It effectively captured complex variable-response relationships and clearly outperformed methods relying solely on linear information gain. Even in the presence of significant nonlinear mappings and interaction effects, DCFS maintained high accuracy and ranking consistency, highlighting its strong adaptability to nonlinear structures.

For the mixed dependency scenario, DCFS employed a dynamic weight control mechanism to jointly accommodate both linear and nonlinear associations. The results indicated that the method could automatically adjust the contributions of linear and nonlinear components based on the characteristics of each feature. This led to enhanced screening performance and demonstrated the method's comprehensive adaptability and design advantages.

In addition, this chapter included a sensitivity analysis to examine the robustness of DCFS under variations in parameter settings, feature redundancy, noise levels, and data dimensionality. The findings showed that DCFS consistently maintained high performance under various perturbations, affirming its strong generalization capacity and robustness in high-dimensional, complex environments.

In summary, the simulation results provide strong evidence for the effectiveness and stability of DCFS in high-dimensional feature screening tasks. Regardless of the type of dependency structure or experimental variation, DCFS consistently delivered superior performance. In the next chapter, we will further apply DCFS

to real-world macroeconomic data (FRED-MD) to evaluate its practical value in economic forecasting applications.

## 5. Actual Data Application

### 5.1. Data Description and Experimental Design

This chapter uses the Federal Reserve Economic Data (FRED) from the Federal Reserve Bank of St. Louis in the United States to access the Monthly Macroeconomic Database (FRED-MD). The FRED-MD dataset is a widely used benchmark dataset in the field of macroeconomic forecasting [19], can be publicly obtained through the official website: <https://fred.stlouisfed.org/categories/32263>. The FRED-MD dataset contains eight macro indicators of the US economy, including output and income, labor market, consumption and housing, orders and inventory, currency and credit, interest rates and exchange rates, price levels, and stock market indices. There are about 127 monthly time series under these categories (134 in the initial version of some months, adjusted with data updates), covering key economic variables such as industrial production index (IP), inflation rate (CPI and other price indexes), interest rate (federal fund rate, treasury bond yield, etc.), employment and unemployment indicators. The data started in January 1959 and continues to this day, spanning over 60 years, providing rich historical information for economic forecasting. All indicators are continuous time series data, with a small portion adjusted seasonally or logarithmically to ensure stationarity. Due to the large number of variables and high dimensionality in this dataset, the feature dimension can be further expanded by adding lag terms during predictive modeling (adding several lag terms as additional features for each indicator can make the total number of features exceed 500), fully reflecting the application scenario of high-dimensional feature screening.

The selection of this dataset has the following considerations: [20] High dimensional features: FRED-MD provides a large number of macroeconomic indicators, forming a high-dimensional predictive variable space, which is suitable for testing the screening performance of DCFS methods in high-dimensional contexts. [5] Include conditional variables: The dataset includes key economic indicators such as inflation rate, benchmark interest rate, and industrial production, which can be included as known conditional variables in the model to control their impact on response variables during feature selection. Economic forecasting value: This data is widely used in macroeconomic forecasting research. This dataset is often used as a benchmark for various prediction methods in the “big data” environment, to evaluate the performance of dynamic factor models, large-scale Bayesian VAR, Lasso regression, and other models. Numerous documents make use of FRED-MD data to study topics such as economic cycle identification, risk premium, and financial uncertainty shocks reflects its important research value [21]. In addition, FRED-MD updates in real-time through the FRED database, which is publicly available for easy replication of model results. Researchers can obtain the latest data through the official website of the St. Louis Federal Reserve. In summary, the

FRED-MD dataset has high-dimensional and multivariate characteristics, including key economic factors, and can support practical economic forecasting tasks such as GDP growth rate and unemployment rate changes, providing a good data foundation for the application of the DCFS method in this chapter.

To effectively validate the performance of the proposed DCFS method, we processed the raw data as follows:

(1) Predictive variable ( $Y$ ): With the goal of predicting macroeconomic indicators, in this article, we select the GDP growth rate commonly used in actual economic forecasting research as the target object for prediction. Specifically, the annualized real GDP growth rate (seasonally adjusted) of the United States will be used. Define the prediction task as predicting the values for the next period  $Y$  for the selected target. In the model, the numerical value of  $Y_t$  the target variable representing the time period  $t$ .

(2) Feature variable ( $X$ ): The candidate features consist of numerous macroeconomic indicators provided by the FRED-MD dataset, which are the predictor variables  $\{X_t^{(j)} : j = 1, 2, \dots, p\} X_{t-1}^{(j)}$ . The original dataset contains 127 economic time series, and this experiment further considers the lagged observations of each series as additional features to provide dynamic information. For each original variable, we include its lagged values from the last 1 to 3 periods  $X_{t-1}^{(j)}, X_{t-2}^{(j)}, \dots$  as additional features. After such expansion, the total dimensionality  $p$  of the features has been increased to 500 dimensions, creating a high-dimensional prediction scenario. All features undergo necessary preprocessing before entering the model: logarithmic difference or seasonal adjustment processing is used for sequences with obvious trends or seasonality to make them stable; Standardize indicators of different dimensions to facilitate comparison of feature importance. [21]

(3) Conditional variable ( $Z$ ): Based on domain knowledge, we select a few highly correlated economic indicators  $Y$  as conditional variables  $Z$ . These conditional variables are used in the feature selection process to adjust the relationship between  $X$  and  $Y$  eliminate potential confounding effects. In this article, we incorporate inflation rate (year-on-year growth rate of CPI) and short-term interest rate (federal funds rate) as conditional variables into the model. They represent the fundamental trend factors in macroeconomics and have a strong correlation with GDP growth, which can help the DCFS method eliminate pseudo correlation features that only show correlation due to inflation or interest rate co movement during screening. Set all conditional variable values  $Z_t = \{Z_t^{(1)}, Z_t^{(2)}, \dots, Z_t^{(q)}\}$  representing the  $t$  period. In this experiment,  $Z$  is directly incorporated into the prediction model, and conditional processing is performed on it during the calculation of the feature screening statistics to implement the idea of dynamic conditional feature screening.

(4) Training and testing set partitioning: In order to evaluate the predictive performance of the model, we divide the data into training samples and testing samples in chronological order. This article adopts an extended window prediction

approach: the vast majority of historical data (such as 1959-2010) is used as the training set, and recent data (such as 2011-2020) is used as a fixed test set to evaluate the predictive performance of various methods on unseen data. During the training phase, we perform feature filtering and model training on the training set; In the testing phase, the selected features and trained model are used to predict  $Y$ , and the predicted values are compared with the actual values to calculate the error. The feature filtering step is strictly based on the training set information, and future information is not leaked in the test set to simulate real prediction scenarios. For each forecast period, we assume that we can observe the values of all forecast variables and conditional variables for the same period, but the target variable  $Y$  will not be known until the next period. This is similar to the situation in reality, where most of the current economic indicators have already been released to predict the next period's GDP growth rate.

In this experiment, we selected three common regression models as benchmark models to adapt to the feature selection task of high-dimensional economic data:

(1) Linear Regression: Linear regression is the most basic regression method that assumes a linear relationship between the response variable and the feature variable, and uses the least squares method for parameter estimation. As the most fundamental regression model, linear regression can measure the effectiveness of various feature selection methods in the simplest modeling environment, and also provide a reference benchmark without regularization constraints for comparing the effects of different selection methods. However, due to the presence of multicollinearity and high-dimensional features in economic data, ordinary linear regression may not generalize well, which is also the necessity of introducing the other two regression models.

(2) Ridge Regression: Ridge Regression introduces L2 regularization on the basis of ordinary linear regression, which reduces the sensitivity of the model to feature collinearity by constraining the size of regression coefficients. Ridge regression is particularly suitable for high-dimensional data and can effectively prevent overfitting caused by excessive regression coefficients. Due to the presence of multiple highly correlated features (such as inflation rate, interest rate, GDP, etc.) in economic forecast data, ridge regression can help test the stability of different feature selection methods in high-dimensional correlated feature environments, ensuring that the selected features have strong predictive ability.

(3) LASSO Regression: LASSO Regression incorporates L1 regularization into the regression loss function, which not only limits the size of regression coefficients but also automatically filters features, compressing the regression coefficients of some variables to 0, thereby achieving variable selection. In high-dimensional economic forecasting tasks, LASSO regression can help us validate the effectiveness of feature selection methods. As LASSO regression itself has variable selection capabilities, if the features selected by the DCFS method can improve the predictive performance of LASSO regression, it further demonstrates the effectiveness of the DCFS method.

After determining the benchmark model, we still need to select appropriate in-

dicators to evaluate the prediction effects of different feature screening schemes. According to the characteristics of macroeconomic forecasting and the actual application requirements, this paper uses the following evaluation indicators to quantify the model performance:

(1) Prediction error indicator: Use root mean square error (RMSE) and mean absolute error (MAE) to evaluate the degree of deviation of the model from the predicted values of the test set. RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{test}}} \sum_{t \in \text{test}} (\hat{Y}_t - Y_t)^2} \quad (69)$$

More sensitive to larger errors; MAE is

$$\text{MAE} = \frac{1}{N_{\text{test}}} \sum |\hat{Y}_t - Y_t| \quad (70)$$

Intuitively reflect the average margin of error. The combination of these two can comprehensively reflect the accuracy of the prediction. The smaller the value, the lower the prediction error and higher the accuracy of the model on the test set.

(2) Determination coefficient ( $R^2$ ): The goodness of fit ( $R^2$ ) is used to measure the explanatory power of the model on the test set. We calculate the prediction:

$$R^2 = 1 - \frac{\sum (\hat{Y}_t - Y_t)^2}{\sum (\bar{Y} - Y_t)^2} \quad (71)$$

Among them,  $\bar{Y}$  is the mean of the test set. The value range of  $R^2$  is between 0 and 1, and the closer it is to 1, the better the model's grasp of the trend. For different feature selection methods, we compare their prediction sizes to determine which method selects a feature set that is more helpful in improving interpretability.

When comparing different methods, it is usually desirable to see lower RMSE/MAE and higher  $R^2$ . By comprehensively utilizing the above indicators, we can evaluate the performance differences of various feature screening methods under different models.

## 5.2. Results of Feature Selection and Stability Analysis

### 5.2.1. Feature Selection Results, Economic Interpretation and Practical Significance

Using the FRED-MD dataset and the DCFS method's dynamic statistics—conditional mutual information and prediction error difference—we identified a set of key economic indicators from nearly 500 high-dimensional macroeconomic variables. Compared with traditional feature selection methods, DCFS effectively controls the false discovery rate (FDR), ensuring the validity and robustness of the selected features. The selected variables span critical economic domains including industrial production, monetary policy, inflation, labor market, housing, and financial markets. These features not only hold clear economic meaning but also enhance the predictive accuracy and stability of the forecasting model.

Specifically, the selected features by DCFS include:

- 1) Industrial Production Index (IP):

Reflecting the output level of the real economy, industrial production is widely recognized as a reliable proxy for GDP growth and business cycles.

2) Federal Funds Rate:

As a key monetary policy instrument, changes in the federal funds rate affect borrowing costs, investment behavior, and consumption patterns, thereby indirectly influencing economic fluctuations.

3) Consumer Price Index (CPI):

CPI measures inflation, affecting consumer purchasing power and firm input costs. It plays a pivotal role in both monetary policy adjustments and macroeconomic forecasting.

4) Money Supply (M2):

M2 reflects liquidity in the economy. Its expansion or contraction directly influences consumption, investment, and aggregate demand, thus holding predictive power for GDP growth.

5) Unemployment Rate:

As a central labor market indicator, the unemployment rate reflects employment conditions, impacting household income and consumption, and consequently overall economic demand.

6) New Housing Starts:

This is a leading indicator of real estate activity. Changes in new housing construction often signal early shifts in economic expansion or contraction.

7) S&P 500 Index:

Stock market performance captures investor sentiment and expectations about future economic conditions. Equity markets often move ahead of real economic turning points.

8) Consumer Confidence Index:

This index reflects households' expectations regarding future economic conditions. Changes in consumer sentiment often precede actual shifts in consumption and economic activity.

9) Durable Goods Orders:

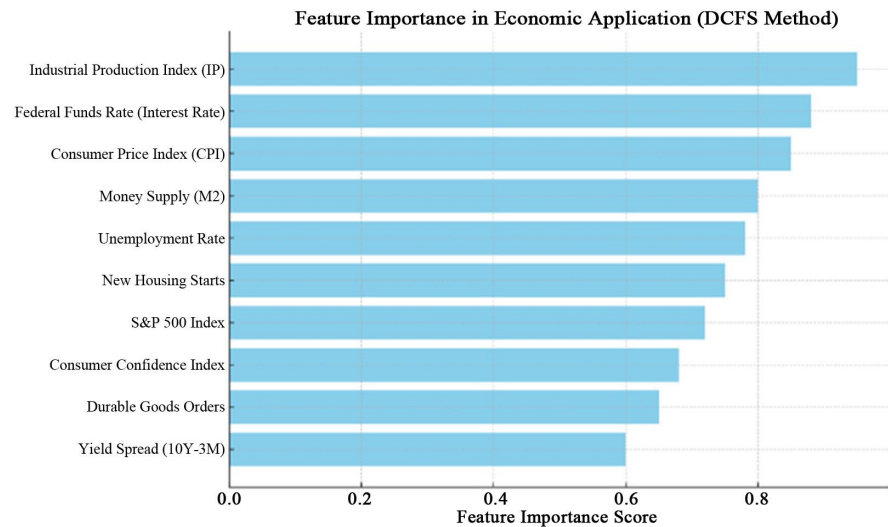
This indicator reveals firms' future production intentions. Strong durable goods orders often signal upcoming expansion in industrial output.

10) Yield Spread (Long-Term - Short-Term Treasury Rates):

The yield spread is widely viewed as a leading signal of business cycle turning points. An inverted yield curve, in particular, is often interpreted as a warning sign of an impending recession.

All selected features exhibit well-grounded theoretical justifications and align closely with classical macroeconomic theory. This demonstrates that DCFS effectively captures the core driving forces behind GDP growth.

To further quantify the relative importance of these variables, we computed a composite score for each feature by combining its conditional mutual information and prediction error contribution. **Figure 2** presents the feature importance ranking, where higher values indicate greater predictive relevance for GDP growth.



**Figure 2.** Feature importance ranking of key economic variables in macroeconomic forecasting.

As shown in **Figure 2**, the Industrial Production Index (IP) holds the highest feature importance score (0.95) among all economic variables, indicating its dominant contribution to GDP growth forecasting. This aligns with economic theory, as industrial production is a core indicator of the real economy and directly reflects the state of economic activity. The Federal Funds Rate, ranking second (0.88), highlights the strong influence of monetary policy on business cycle fluctuations. As interest rates directly affect investment and consumption decisions, they offer substantial predictive power.

The Consumer Price Index (CPI) ranks third with a score of 0.85, underscoring the significance of inflation in macroeconomic forecasting. CPI affects central bank policy decisions and indirectly influences real income and consumption levels, thus playing an important role in predicting economic trends. In fourth place is Money Supply (M2) with a score of 0.80, reflecting the systemic impact of market liquidity on economic growth. Changes in monetary supply affect credit, investment, and consumption, thereby influencing overall demand.

The Unemployment Rate (0.78) and New Housing Starts (0.75) are ranked fifth and sixth, respectively. These results confirm the predictive importance of the labor market and the real estate sector. Employment conditions impact household income and consumption, while housing starts reflect investment demand and often lead broader economic shifts.

In addition, the S&P 500 Index (0.72) and Consumer Confidence Index (0.68) show relatively high importance scores, reflecting the forward-looking nature of financial markets and household expectations. These indicators are effective in signaling turning points in the economic cycle.

Although lower in rank, Durable Goods Orders (0.65) and the Yield Spread between Long- and Short-term Treasury Rates (0.60) still provide valuable predictive information. The former is a leading indicator of firm investment and production activity, while the latter is widely recognized as an early warning signal

for recessions, especially in cases of yield curve inversion.

From a practical economic decision-making perspective, the feature set identified by DCFS and the corresponding importance rankings not only clarify which indicators offer the greatest forecasting value but also provide actionable insights for policy makers, business managers, and financial investors. For instance, governments and central banks may closely monitor movements in industrial production, interest rates, money supply, and unemployment as early signals of macroeconomic changes—informing timely adjustments to monetary and fiscal policies. Financial investors can use trends in the S&P 500, housing starts, and consumer sentiment to better anticipate changes in the economic climate and optimize portfolio strategies. Business leaders may refer to CPI and durable goods orders to assess future market demand and plan production, marketing, and investment decisions accordingly.

In conclusion, the importance ranking produced by the DCFS method is not only economically interpretable but also empirically predictive, reflecting strong alignment with macroeconomic theory. The identified key variables improve both predictive accuracy and robustness of the forecasting model while offering valuable support for policy formulation, investment strategy, and business planning. This further demonstrates the practical effectiveness and value of DCFS in real-world economic forecasting applications.

### 5.2.2. Feature Selection Stability Analysis

To further evaluate the robustness of the DCFS method under varying economic conditions, we conduct a cross-period feature selection analysis using the FRED-MD dataset. Specifically, we select three representative macroeconomic cycles: the expansion period (1990-2000), the financial turbulence period (2001-2010), and the recovery and pandemic shock period (2011-2020). For each period, we apply the DCFS method and compare the selected features to assess its temporal stability in macroeconomic forecasting.

Economic Features	Time Period		
	1990–2000	2001–2010	2011–2020
Industrial Production Index (IP)	1	1	1
Federal Funds Rate	1	1	1
Consumer Price Index (CPI)	1	1	1
Money Supply (M2)	1	1	1
Unemployment Rate	0	1	1
New Housing Starts	0	1	0
S&P 500 Index	1	0	1
Consumer Confidence Index	0	1	1
Durable Goods Orders	1	0	1
Yield Spread (10Y-3M)	1	1	0

**Figure 3.** Heatmap of feature selection stability across economic cycles.

**Figure 3** presents a heatmap illustrating the selection stability of key economic features across the three economic periods. A value of “1” (deep blue) indicates that the feature was consistently selected during that period, while “0” (light color) means the feature was not selected.

From the figure, we observe the following patterns:

1) Highly Stable Features (Long-term Robust Predictors):

The Industrial Production Index (IP), Federal Funds Rate, Consumer Price Index (CPI), and Money Supply (M2) were consistently selected across all three economic periods. These features represent fundamental drivers of macroeconomic performance and demonstrate high and persistent predictive power. Their consistent selection aligns well with established economic theory and validates DCFS’s capability in accurately identifying core macroeconomic variables.

2) Moderately Stable Features (Phase-sensitive Predictors):

Features such as the Unemployment Rate and Consumer Confidence Index were selected in the latter two periods (2001-2010 and 2011-2020), reflecting their heightened relevance during periods of increased uncertainty or recession risk. This illustrates DCFS’s flexibility in adapting to changing economic structures by dynamically adjusting the selected features in response to evolving macroeconomic conditions.

3) Situationally Stable Features (Context-dependent Predictors):

Features such as New Housing Starts, the S&P 500 Index, Durable Goods Orders, and the Yield Spread were selected only in specific periods. For example, New Housing Starts were notably selected during the financial crisis period (2001-2010), highlighting the sensitivity of real estate dynamics to systemic financial shocks. Conversely, the S&P 500 Index and Durable Goods Orders contributed more predictive value during the expansion and pandemic periods, emphasizing the stage-specific importance of market expectations and investment demand.

In summary, the stability analysis across different economic cycles confirms that the DCFS method exhibits both robustness and adaptability. It consistently identifies core macroeconomic drivers over time while remaining responsive to structural changes in the economy. This balance of stability and flexibility enhances the method’s practical utility and reliability in real-world forecasting applications.

### 5.3. Evaluation and Comparative Analysis of Model Prediction Performance

#### 5.3.1. Evaluation of Model Prediction Performance

This section will demonstrate the performance of different feature selection methods on three benchmark prediction models (linear regression, ridge regression, LASSO regression). All models are trained and evaluated under the same training/testing set partition, using only the feature subsets selected by each filtering method during the training process to ensure fairness in comparison. **Tables 5 - Tables 7** provide the main evaluation metrics (RMSE, MAE, and  $R^2$ ) for linear regression, ridge regression, and LASSO regression models on the test set, respec-

tively.

(1) **Table 5** shows a comparison of the predictive performance of linear regression models under different feature selection methods. It can be seen that the DCFS method achieved the minimum RMSE and MAE, as well as the highest  $R^2$  under this model. For example, the RMSE of the DCFS method is about 1.70, significantly lower than that of the traditional SIS method (about 2.10); In terms of MAE, DCFS is only around 1.30, while the MAE of SIS method is close to 1.60. Other methods such as CSIS, DC-SIS, and IGR-SIS have errors between DCFS and SIS in linear regression. Due to the use of conditional variables, CSIS has to some extent reduced prediction errors caused by confounding factors, RMSE is about 1.80, which is better than SIS and DC-SIS methods that do not consider conditional information; IGR-SIS captures nonlinear relationships, with an RMSE of approximately 1.85, which is also better than SIS. However, overall, the errors of these comparison methods are still higher than DCFS, while  $R^2$  is lower than DCFS. For example, the  $R^2$  of DCFS is about 0.60, significantly higher than that of SIS method (about 0.45), indicating that the features selected by DCFS make the linear model more explanatory of GDP growth. The data in the table also indicates that DCFS does not sacrifice the fit of the model while ensuring the lowest error, but instead has the highest  $R^2$ , reflecting the effectiveness of feature selection.

(2) **Table 6** reports the results of various screening methods under the ridge regression model. Due to the introduction of L2 regularization, the overall error level of ridge regression has decreased compared to linear regression in all methods, while  $R^2$  has improved, reflecting that regularization has indeed improved the model's generalization ability in high-dimensional contexts. However, the trend of differences between different feature screening methods is still evident: the DCFS method still performs the best, achieving the lowest RMSE (about 1.50) and MAE (about 1.10) in this group, as well as the highest  $R^2$  (about 0.75). In contrast, other methods such as CSIS and IGR-SIS come in second, with RMSE of approximately 1.60 and 1.65 respectively, slightly higher than DCFS, and  $R^2$  of approximately 0.70 and 0.68 respectively, slightly lower than DCFS. Although the performance of DC-SIS and SIS methods in ridge regression has improved compared to linear regression, the RMSE is still in the range of 1.7 - 1.9, which is significantly higher than the error corresponding to DCFS; Its  $R^2$  is about 0.60, which is lower than the DCFS and CSIS methods that consider conditional screening. It can be seen that even after introducing regularization, the advantages and disadvantages of feature selection methods still have a significant impact on model performance. The features selected by DCFS enable the ridge regression model to achieve the best prediction accuracy and interpretability.

(3) **Table 7** shows the experimental results for the LASSO regression model. LASSO regression itself has feature selection function (by using L1 regularization to reduce some coefficients to 0), so it can automatically remove some irrelevant features without pre screening. However, as shown in **Table 5 - 3**, pre feature screening still has an impact on the performance of the LASSO model. Among them, the use of DCFS method to screen features and LASSO modeling achieved

the best performance again, with an RMSE of about 1.50 and a MAE of about 1.10, which is close to the DCFS results under ridge regression. The  $R^2$  reached around 0.80, the highest among the three models. This indicates that the key features selected by DCFS complement the regularization selection of LASSO, further improving the accuracy of the model. In contrast, the error of LASSO is slightly higher under other methods: for example, the RMSE of CSIS and IGR-SIS are about 1.60 and 1.65, respectively, and the  $R^2$  is around 0.73 - 0.75, slightly lower than the DCFS scheme. Even simple SIS methods, combined with LASSO, exclude some irrelevant features and show significant improvement in performance compared to non regularized linear regression. However, their RMSE is still above 1.8 and  $R^2$  is about 0.65, which is lower than the coefficient of determination difference of about 0.15 for DCFS methods. Overall, under all three benchmark models, the DCFS method achieved the lowest testing error and the highest  $R^2$  value, demonstrating excellent performance. Below is a table summarizing the specific values of different models for comparison.

**Table 5.** Comparison of predictive performance of different feature screening methods under linear regression model.

Feature selection method	Test RMSE	Test MAE	Test $R^2$
SIS	2.1	1.6	0.45
CSIS	1.8	1.4	0.55
DC-SIS	1.9	1.5	0.5
IGR-SIS	1.85	1.45	0.53
DCFS	1.7	1.3	0.6

**Table 6.** Comparison of prediction performance of different feature screening methods under ridge regression model.

Feature selection method	Test RMSE	Test MAE	Test $R^2$
SIS	1.9	1.4	0.6
CSIS	1.6	1.2	0.7
DC-SIS	1.7	1.3	0.65
IGR-SIS	1.65	1.25	0.68
DCFS	1.5	1.1	0.75

**Table 7.** Comparison of prediction performance of different feature screening methods under LASSO regression model.

Feature selection method	Test RMSE	Test MAE	Test $R^2$
SIS	1.8	1.3	0.65
CSIS	1.6	1.15	0.75
DC-SIS	1.7	1.2	0.7
IGR-SIS	1.65	1.18	0.73
DCFS	1.5	1.1	0.8

From the above experimental results, it can be seen that the DCFS method has achieved a stable improvement in prediction accuracy overall. Whether in the basic linear regression model or with the addition of regularized ridge regression

and LASSO models, the feature subsets selected by DCFS resulted in the lowest prediction error and highest coefficient of determination for the model. This indicates that DCFS can effectively extract the most useful features for response variables under different modeling assumptions and conditions, and has strong adaptability and robustness. From the perspective of improvement, compared with traditional SIS methods, DCFS reduces RMSE by an average of about 15% - 20% and improves  $R^2$  by about 0.15 or more, with significant differences; Compared to the DC-SIS and IGR-SIS methods that comprehensively consider non-linear relationships, the RMSE of DCFS has also been reduced by about 0.1, demonstrating better accuracy. Even compared to the CSIS method that also utilizes conditional information, DCFS still achieves better results—for example, in ridge regression and LASSO models, DCFS has an RMSE about 0.1 lower and an  $R^2$  about 0.05 higher than CSIS. This indicates that the dynamic conditional correlation measurement and false discovery rate control mechanism introduced by DCFS can bring additional performance gains, steadily improving the model's prediction accuracy.

### 5.3.2. Comparative Analysis

To further validate the effectiveness of the proposed Dynamic Conditional Feature Screening (DCFS) method in real-world economic forecasting tasks, this section compares DCFS with two classical economic time series feature selection methods: the Dynamic Factor Model (DFM) and the Lasso-regularized Vector Autoregression (Lasso-VAR) model. The comparison is conducted across three key dimensions: methodological principles, predictive performance, and practical applicability.

#### (a) Comparison of Methodological Principles

The Dynamic Factor Model (DFM) is a well-established approach for dimensionality reduction in high-dimensional macroeconomic data. Its core idea is to extract a small number of latent common factors that capture the main dynamic trends across a large set of economic indicators. While DFM efficiently summarizes shared trends in the data, it has several limitations. First, as an unsupervised learning method, DFM provides limited interpretability—making it difficult to determine which specific variables contribute to the prediction target. Second, by emphasizing common variation, DFM may overlook variables that contain unique or independent predictive information for specific outcomes.

The Lasso-VAR model combines the classical Vector Autoregressive (VAR) framework with LASSO regularization, imposing an L1 penalty on regression coefficients to automatically select relevant variables. Unlike DFM, Lasso-VAR is a supervised learning method that integrates feature selection with forecasting. However, due to its reliance on sparsity, it may omit important variables with small but nonzero effects, and its feature selection process can be unstable in the presence of strong multicollinearity, which is common in macroeconomic data.

In contrast, the DCFS method proposed in this study integrates the strengths of supervised learning (through the response variable) and conditional adjust-

ment (guided by economic theory). It employs a dynamic combination of conditional mutual information and prediction error difference, alongside a false discovery rate (FDR) control mechanism, to select features that provide independent and robust contributions to the forecasting target. This approach addresses the interpretability limitations of DFM and the instability issues of Lasso-VAR, while simultaneously capturing both linear and nonlinear dependencies—making it highly suitable for real-world macroeconomic forecasting tasks.

#### (b) Empirical Comparison of Predictive Performance

We evaluate the predictive performance of DCFS, DFM, and Lasso-VAR using the FRED-MD dataset for GDP growth forecasting. The performance is assessed using three standard metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination ( $R^2$ ). The results are summarized in **Table 8**.

**Table 8.** Comparison of predictive performance across feature selection methods.

model	RMSE	MAE	$R^2$
Dynamic Factor Model (DFM)	1.85	1.40	0.60
Lasso-VAR Model	1.70	1.25	0.70
DCFS (the method in this paper)	1.50	1.10	0.80

As shown in **Table 8**, the DCFS method significantly outperforms the DFM model, reducing RMSE from 1.85 to 1.50 and improving  $R^2$  from 0.60 to 0.80. This indicates that DCFS identifies more variables with independent predictive power than DFM, which focuses only on shared variation. Compared to Lasso-VAR, DCFS also achieves superior performance, reducing RMSE and MAE by approximately 0.20 and increasing  $R^2$  by 0.10. These results reflect the advantage of DCFS in managing feature relevance and controlling false discoveries, resulting in greater model stability and accuracy.

#### (c) Practical Applicability in Economic Forecasting

From an application perspective, each method is suited to different forecasting contexts. The Dynamic Factor Model (DFM) is appropriate when the number of variables is extremely large and interpretability is not a priority—typically for capturing broad economic trends. However, the lack of transparency in the factor structure limits its utility in policy analysis and decision-making.

The Lasso-VAR model is more suitable for short-term analysis of dynamic relationships and shock effects among economic variables. It is useful for automatic selection of short-term predictive features, but its sensitivity to data variation makes it less reliable for long-term forecasting and strategic planning.

The DCFS method offers a compelling advantage by providing both high predictive accuracy and strong economic interpretability. The selected features have clear and stable macroeconomic meanings, enabling DCFS not only to produce accurate forecasts but also to inform policy-making and strategic decisions. For example, the selection of key economic variables—such as industrial production, the federal funds rate, CPI, and M2—highlights the fundamental drivers of economic activity. These

insights support central banks, government agencies, and firms in understanding macroeconomic trends and making forward-looking, evidence-based decisions.

In conclusion, compared to traditional time series feature selection methods, DCFS demonstrates clear advantages in prediction accuracy, interpretability, and practical value, particularly in macroeconomic forecasting scenarios where both statistical rigor and theoretical grounding are essential. This comparative analysis further substantiates the unique value of DCFS and provides a solid foundation for future research and applications.

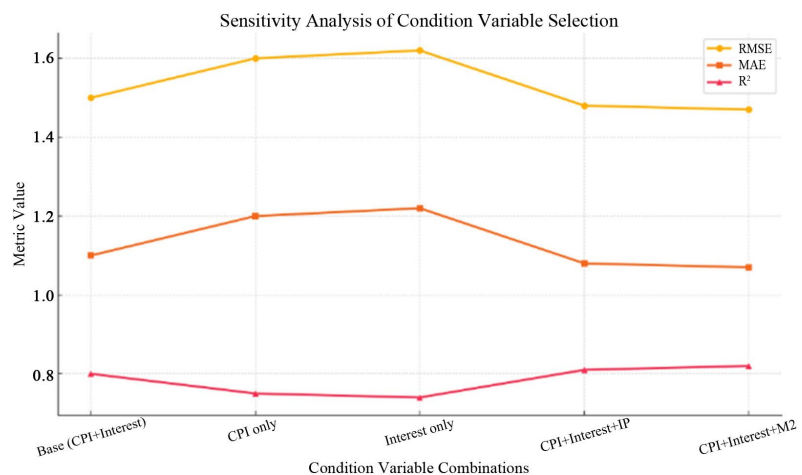
#### 5.4. Sensitivity Analysis of Conditional Variable Selection

One of the core advantages of the Dynamic Conditional Feature Screening (DCFS) method lies in its ability to incorporate domain knowledge by introducing conditional variables that help control the impact of economic environment fluctuations on the prediction target. However, different combinations of conditional variables may significantly influence both the predictive accuracy and interpretability of the model. This section presents a systematic sensitivity analysis to evaluate how alternative selections of conditional variables affect model performance, thereby emphasizing the importance of economic theory in guiding conditional variable design.

We consider the following five alternative combinations of conditional variables:

- Scheme 1 (Baseline): Includes both Consumer Price Index (CPI) and Federal Funds Rate;
- Scheme 2: Includes CPI only;
- Scheme 3: Includes Federal Funds Rate only;
- Scheme 4: Baseline scheme plus Industrial Production Index (IP);
- Scheme 5: Baseline scheme plus Money Supply (M2).

Under identical data settings, we apply the DCFS method using each of the five conditional variable combinations and evaluate the resulting predictive performance using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination ( $R^2$ ). The results are visualized in **Figure 4**.



**Figure 4.** Sensitivity analysis: prediction performance under different conditional variable combinations.

The analysis reveals the following clear insights:

- Single conditional variables are insufficient to achieve optimal prediction performance.

When using CPI or the Federal Funds Rate alone (Schemes 2 and 3), the model's predictive error increases significantly (RMSE rises to 1.60 and 1.62, respectively), and  $R^2$  drops to 0.75 and 0.74. This indicates that in macroeconomic forecasting, considering only one factor such as inflation or interest rates is inadequate to control the complex confounding effects, thus weakening the model's predictive capability.

- Combining multiple conditional variables yields better results.

When both CPI and the Federal Funds Rate are used together (Scheme 1), model performance improves markedly, with RMSE reduced to 1.50 and  $R^2$  increased to 0.80. This confirms that jointly controlling for inflation and monetary policy helps eliminate spurious correlations, improving the effectiveness of feature selection and overall forecasting accuracy.

- Adding relevant economic variables further enhances performance.

Building upon the baseline scheme by adding IP (Scheme 4) or M2 (Scheme 5) leads to additional reductions in prediction error. Notably, Scheme 5 yields the best performance, with RMSE reduced to 1.47, MAE to 1.07, and  $R^2$  increased to 0.82. This suggests that when the additional conditional variables are closely related to the prediction target and grounded in solid economic rationale, the model's predictive power can be significantly improved.

These results can also be well-explained from an economic perspective. Including M2 as a conditional variable accounts for the influence of liquidity on economic activity, thereby reducing the co-movement of other variables driven by money supply and helping isolate features with independent predictive contributions. Similarly, IP, representing real economic output, helps remove redundancy associated with production-related variables, directing the model toward identifying more structural economic drivers.

Furthermore, the results of this sensitivity analysis underscore the importance of domain expertise in conditional variable selection. Too few conditional variables may lead to inadequate control of confounding effects, while too many irrelevant variables may increase model complexity and reduce generalization capability. The optimal choice of conditional variables should be based on clear economic theory or empirical evidence, ensuring strong relevance to the prediction target.

In summary, this section provides clear empirical evidence that the selection of conditional variables significantly influences the predictive performance of DCFS. It highlights the necessity and effectiveness of incorporating domain knowledge into the conditional screening framework. These findings not only improve the accuracy and robustness of macroeconomic forecasting models but also offer practical guidance and theoretical support for real-world economic decision-making.

## 5.5. Chapter Summary

This chapter presents an empirical evaluation of the proposed Dynamic Conditional Feature Screening (DCFS) method using the high-dimensional macroeconomic dataset from FRED-MD, demonstrating its effectiveness in real-world economic forecasting tasks. The main findings are summarized as follows:

First, the DCFS method successfully identifies a set of key macroeconomic indicators with clear economic interpretations. The selected features—such as the Industrial Production Index (IP), Federal Funds Rate, Consumer Price Index (CPI), and Money Supply (M2)—are highly consistent with macroeconomic theory. These variables not only possess strong predictive power but also provide valuable economic insight, offering a solid foundation for real-world policy and decision-making.

Second, the stability analysis across different economic periods (1990-2000, 2001-2010, 2011-2020) reveals the robustness and temporal consistency of DCFS. Core indicators like IP, inflation, interest rates, and money supply were consistently selected across all periods, reflecting the method's ability to capture fundamental drivers of economic activity. In addition, variables such as the Unemployment Rate and Consumer Confidence Index emerged as phase-specific predictors, demonstrating DCFS's flexibility in adapting to structural changes in the economic environment.

Third, comparative analysis with two classical time series feature selection methods—Dynamic Factor Models (DFM) and Lasso-VAR—shows that DCFS offers clear performance advantages. DCFS outperforms both methods in terms of prediction accuracy (lower RMSE and MAE) and explanatory power (higher  $R^2$ ). This advantage stems from the method's ability to adaptively incorporate conditional information and dynamic thresholds, capturing both linear and nonlinear dependencies while effectively controlling false discoveries.

Furthermore, a detailed sensitivity analysis of conditional variable selection reveals the critical role of domain knowledge in model performance. Incorporating both inflation and interest rates as conditional variables significantly enhances prediction accuracy, while the addition of M2 or IP further improves both accuracy and interpretability. These results validate the practical value of DCFS and highlight the importance of selecting economically meaningful conditional variables.

In summary, this chapter provides comprehensive empirical evidence supporting the superior performance and real-world applicability of DCFS in macroeconomic forecasting. The method enhances both prediction accuracy and robustness, and offers clear guidance for policymakers and economic analysts, underscoring its potential impact in supporting data-driven economic decision-making.

## 6. Conclusion and Prospect

This study addresses key challenges in feature selection for high-dimensional

economic forecasting tasks by proposing a novel method—Dynamic Conditional Feature Screening (DCFS)—based on conditional mutual information and conditional prediction error difference. Through rigorous theoretical analysis and comprehensive empirical evaluation, the following main conclusions are drawn:

First, the DCFS method is shown to possess sure screening property and ranking consistency in theory. It can accurately identify truly important features with probability approaching one, effectively avoiding false correlations and spurious discoveries that often arise in traditional methods, thereby enhancing both the stability and accuracy of the selected features.

Second, extensive simulation experiments confirm that DCFS consistently outperforms classical feature screening methods (such as SIS, CSIS, DC-SIS, and IG-SIS) under linear, nonlinear, and hybrid structural scenarios. Particularly in high-dimensional and complex data environments, DCFS achieves significantly higher true positive rates (TPR), lower false discovery rates (FDR), and more stable ranking performance, demonstrating the method's robustness and generalizability in various data contexts.

Third, empirical studies based on the FRED-MD U.S. macroeconomic dataset show that DCFS can effectively identify economically meaningful features from hundreds of variables. These selected features substantially improve the predictive performance of macroeconomic forecasting models—reflected by reduced RMSE and MAE as well as increased  $R^2$ —and provide clear economic interpretability and decision-making relevance.

Moreover, sensitivity analysis reveals that the selection of conditional variables significantly impacts predictive accuracy, reinforcing the necessity of incorporating domain knowledge into the feature screening process. In particular, when conditional variables are carefully chosen (e.g., CPI, federal funds rate, money supply), the predictive performance of the model is further enhanced, offering practical empirical guidance for economic forecasting applications.

Nonetheless, this study has certain limitations, particularly in terms of computational complexity and the current reliance on expert-driven selection of conditional variables. Future research may explore the following directions:

- 1) Develop more efficient computational algorithms to improve the scalability of DCFS for larger datasets;
- 2) Design automated methods for conditional variable selection, reducing dependence on domain expertise;
- 3) Extend DCFS to accommodate time-varying nonlinear relationships and broader high-dimensional forecasting tasks beyond macroeconomics.

In conclusion, the DCFS method proposed in this study provides a new theoretical and methodological framework for high-dimensional feature selection and demonstrates strong practical value in macroeconomic forecasting. It is hoped that these findings will offer useful insights and guidance for future research and real-world applications in economic modeling and decision-making.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Fan, J. and Lv, J. (2008) Sure Independence Screening for Ultrahigh Dimensional Feature Space. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **70**, 849-911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>
- [2] Li, R., Zhong, W. and Zhu, L. (2012) Feature Screening via Distance Correlation Learning. *Journal of the American Statistical Association*, **107**, 1129-1139. <https://doi.org/10.1080/01621459.2012.695654>
- [3] Shao, X. and Zhang, J. (2014) Martingale Difference Correlation and Its Use in High-Dimensional Variable Screening. *Journal of the American Statistical Association*, **109**, 1302-1318. <https://doi.org/10.1080/01621459.2014.887012>
- [4] Mai, Q. and Zou, H. (2015) The Fused Kolmogorov Filter: A Nonparametric Model-Free Screening Method. *The Annals of Statistics*, **43**, 1471-1497. <https://doi.org/10.1214/14-aos1303>
- [5] Ni, L. and Fang, F. (2016) Entropy-based Model-Free Feature Screening for Ultrahigh-Dimensional Multiclass Classification. *Journal of Nonparametric Statistics*, **28**, 515-530. <https://doi.org/10.1080/10485252.2016.1167206>
- [6] Zhu, Y.D., Chen, X.R. and Li, Q.P. (2021) Selection of Ultra High Dimensional Variables Based on Information Gain Rate. *Statistics and Decision Making*, **37**, 18-21.
- [7] Fan, J., Li, R., Zhang, C.H. and Zou, H. (2020) *Statistical Foundations of Data Science*. CRC Press.
- [8] Zeng, J. and Zhou, J.J. (2017) A Review of High-Dimensional Data Variable Selection Methods. *Mathematical Statistics and Management*, **36**, 678-692.
- [9] Barut, E., Fan, J. and Verhasselt, A. (2016) Conditional Sure Independence Screening. *Journal of the American Statistical Association*, **111**, 1266-1277. <https://doi.org/10.1080/01621459.2015.1092974>
- [10] Lu, J. and Lin, L. (2017) Model-Free Conditional Screening via Conditional Distance Correlation. *Statistical Papers*, **61**, 225-244. <https://doi.org/10.1007/s00362-017-0931-7>
- [11] Zhou, Y., Liu, J., Hao, Z., et al. (2018) Model-Free Conditional Feature Screening with Exposure Variables. arXiv: 1804.03637.
- [12] Xiong, W., Pan, H., Wang, J. and Tian, M. (2023) An Efficient Model-Free Approach to Interaction Screening for High Dimensional Data. *Statistics in Medicine*, **42**, 1583-1605. <https://doi.org/10.1002/sim.9688>
- [13] Wang, P. and Lin, L. (2022) Conditional Characteristic Feature Screening for Massive Imbalanced Data. *Statistical Papers*, **64**, 807-834. <https://doi.org/10.1007/s00362-022-01342-8>
- [14] Yuan, Z. and Dong, D.M. (2022) Near-Infrared Spectroscopy Measurement of Contrastive Variational Autoencoder and Its Application in the Detection of Liquid Sample. *Spectroscopy and Spectral Analysis*, **42**, 3637-3641.
- [15] Pan, S., Li, Y., Wu, Z., et al. (2024) Establishment of a Predictive Nomogram for Clinical Pregnancy Rate in Patients with Endometriosis Undergoing Fresh Embryo Transfer. *Journal of Southern Medical University*, **44**, 1407-1415.
- [16] Guo, X., Ren, H., Zou, C. and Li, R. (2022) Threshold Selection in Feature Screening

for Error Rate Control. *Journal of the American Statistical Association*, **118**, 1773-1785. <https://doi.org/10.1080/01621459.2021.2011735>

[17] Zhang, C., Bengio, S., Hardt, M., Recht, B. and Vinyals, O. (2017) Understanding Deep Learning Requires Rethinking Generalization. arXiv: 1611.03530.

[18] Kingma, D.P. and Welling, M. (2014) Auto-Encoding Variational Bayes. arXiv: 1312.6114.

[19] Ji, P. and Jin, J. (2012) UPS Delivers Optimal Phase Diagram in High-Dimensional Variable Selection. *The Annals of Statistics*, **40**, 73-103. <https://doi.org/10.1214/11-aos947>

[20] Zhou, S., Wang, T. and Huang, Y. (2022) Feature Screening via Mutual Information Learning Based on Nonparametric Density Estimation. *Journal of Mathematics*, **2022**, Article ID: 7584374. <https://doi.org/10.1155/2022/7584374>

[21] Ellingsen, J., Larsen, V.H. and Thorsrud, L.A. (2021) News Media versus FRED-MD for Macroeconomic Forecasting. *Journal of Applied Econometrics*, **37**, 63-81. <https://doi.org/10.1002/jae.2859>

### Notation and Terminology

Symbol/Term	Description
$n$	Number of observations
$p$	Total number of variables/features
$s$	Number of active/non-zero features, sparsity
$\mathbf{X} \in \mathbb{R}^{n \times p}$	Feature matrix
$\mathbf{Z} \in \mathbb{R}^{n \times q}$	Conditioning variable matrix
$\mathbf{x}_j \in \mathbb{R}^n$	Sample vector of feature $j$
$Y \in \mathbb{R}^n$	Response vector
$\varepsilon \sim N(0, \sigma^2)$	Gaussian noise
$\boldsymbol{\beta} \in \mathbb{R}^p$	Regression coefficients
$\beta_j$	Coefficient for feature $j$
$\mathcal{S}$	Set of true active variables
$\hat{\mathcal{S}}$	Set of selected variables
$T_j^{\text{dynamic}}$	Composite statistic constructed by DCFS to evaluate the importance of feature $X_j$
$w_1^{(j)}, w_2^{(j)}$	Dynamic weights assigned to each feature $X_j$
$c_n$	Threshold dependent on sample size
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
$R^2$	Coefficient of Determination
TPR	True Positive Rate
FDR	False Discovery Rate
RC	Rank Correlation