The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Integral-based Knockoffs Inference for Partially Linear Models

Hao Wang[1,*], Biqin Song[1,*], Rushi Lan[2,†], Hong Chen[1]

[1]Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education, and College of Informatics, Huazhong Agricultural University, Wuhan 430062, China

[2]Guangxi Key Laboratory of Image and Graphic Intelligent Processing, Guilin University of Electronic Technology, Guilin 541004, China

hao_wang@webmail.hzau.edu.cn, biqin.song@mail.hzau.edu.cn, rslan@guet.edu.cn

Abstract

Partial linear models (PLM) have attracted much attention for regression estimation and variable selection due to their feasibility on utilizing linear and nonlinear approximations jointly. However, theoretical understanding of how they control the false discovery rate (FDR) during variable selection remains limited. To address this issue, we formulate a new integral-based knockoffs (IKO) inference scheme for controlled variable selection in PLM, where integral-based knockoff statistics are used to measure the variable importance and B-splines (or random Fourier features) are employed for approximating nonlinear components. In theory, FDR control is guaranteed for both linear and nonlinear parts, and the statistical analysis for its power is established. Empirical evaluations validate the effectiveness of our proposed approach.

Introduction

Linear and non-parametric models are two essential classes of statistical modeling tools, each offering distinct advantages. Linear models are straightforward and provide excellent interpretability but are constrained by their assumption of linear relationships. In contrast, non-parametric models are highly flexible, and capable of accommodating a wide range of functional forms in the data, but they often lack interpretability. Consequently, a fundamental trade-off between accuracy and interpretability must be considered. Semi-parametric models bridge this gap by combining the interpretability of linear models with the flexibility of non-parametric models. These models allow covariates to be modeled either linearly or nonlinearly, making them suitable for capturing complex data structures in various scientific domains, including econometrics, social sciences, information sciences, and biomedicine (Ruppert, Wand, and Carroll 2003; Härdle et al. 2004; Fuller 2009). A classic example of a semi-parametric model is the partially linear model (PLM), which has been extensively studied and applied in diverse fields (Engle et al. 1986; Härdle, Liang, and Gao 2000). Given the response $y \in \mathcal{R}$ and input variables $(\mathbf{z}, \mathbf{x}) = (z_1, \ldots, z_q, x_1, \ldots, x_p) \in \mathcal{R}^{q+p}$, the general partial linear models is given by

$$y = \mu + \mathbf{z}^\top \boldsymbol{\beta} + f(\mathbf{x}) + \varepsilon, \tag{1}$$

where μ represents the intercept, $\boldsymbol{\beta}$ is the vector of coefficients corresponding to the linear terms, f is a nonlinear mapping from $\mathcal{R}^p$ to $\mathcal{R}$, and ε denotes the error term satisfying a normal distribution with mean 0 and variance σ^2.

Variable selection for PLM has garnered significant attention over the past two decades. It is often approached using specially designed penalty functions under various smooth regression conditions. Examples include the component selection and smoothing operator penalty (COSSO) (Lin and Zhang 2006), the smoothly clipped absolute deviation (SCAD) penalty (Xie and Huang 2009), and the doubly penalized procedure (Wang et al. 2014a). Additional variable selection methods for PLM have been proposed in studies such as Wang et al. (2014b); Su and Candès (2016); Lian, Zhao, and Lv (2019); Lv and Lian (2022). While these methods have advanced variable selection techniques, they generally focus on selecting relevant variables and don't address controlling selection errors, such as false discovery rate (FDR).

A recent and effective approach for FDR control is the knockoff framework, first introduced by Barber and Candès (2015). The core concept of knockoffs involves three main steps. First, generate synthetic knockoffs variables that replicate the structure of the original variables but have no causal influence on the response. Second, compute an importance score for each original and knockoff variable. Finally, select variables by applying a threshold: those with scores exceeding the corresponding knockoff scores are chosen. This method achieved effective FDR control in Gaussian linear models where the dimensionality d does not exceed the sample size n. The knockoff filter was later extended to high-dimensional linear models with the aid of data splitting and feature screening techniques (Barber and Candès 2019). To handle general high-dimensional linear models, the model-X knockoff framework (Candès et al. 2018) was proposed. This framework accommodates arbitrary dependence structures between the response variable and input variables and eliminates the need for precise p-value calculations. Yingying Fan and Lv (2020) expanded the model-X knockoff framework to accommodate the unknown covariate distribution, and Xiaowu Dai and Li (2023) advanced it

[*]These authors contributed equally.

[†]Corresponding author.

further by proposing a kernel knockoff method, enabling the knockoff framework to handle more flexible, nonlinear settings beyond the original linear model assumptions. Su et al. (2024) introduced the generalized knockoffs framework for variable selection in PLM but was inherently limited to linear model assumptions. Simultaneously, only a few studies, such as Yingying Fan and Lv (2020); Weinstein et al. (2023), have conducted comprehensive theoretical analyses of power. Notably, their power analyses were restricted to linear model settings. Developing a broader theoretical understanding of power across different models remains a significant and open challenge in the field.

To the best of our knowledge, few studies have investigated knockoff inference for PLM, where the **key challenge** is how to measure the importance of variables in nonlinear part. Most existing knockoff frameworks are designed for linear setting, with various statistics used to assess the variable importance, such as Lasso Coefficient Difference (LCD) statistic (Candès et al. 2018) and Lasso Signed Max (LSM) statistic (Barber and Candès 2015). In nonlinear setting, Selection Probability Difference (SPD) statistic (Xiaowu Dai and Li 2023) assesses variable importance but requires repeated experiments for reliable results.

To address the aforementioned gaps, we propose an integral-based knockoff statistic capable of measuring the importance of variables in both linear and nonlinear parts. Then we propose a variable selection procedure called *Integral-based Knockoff* (IKO), where integral-based knockoff statistics are used to measure the variable importance and B-splines (or random Fourier features) are employed for approximating nonlinear components. Our method ensures FDR control in both the linear and nonlinear parts of PLM. Theoretical guarantees for FDR control and Power are provided, and our method is validated through experimental evaluations, demonstrating its effectiveness. This work significantly contributes to the methodologies of variable selection for PLM and knockoff frameworks as follows:

i) *PLM-based knockoff inference with Integral-based Knockoff*. A new integral-based knockoff statistic is proposed to measure variable importance. Meanwhile, the knockoff framework is extended to PLM, ensuring FDR control for both linear and nonlinear parts.

ii) *Theoretical guarantees and empirical effectiveness*. Theoretical analyses for both FDR control and power are established, and experimental results validate the effectiveness of the proposed method.

We compared integral-based knockoff statistic with the above statistics in both linear (L) and nonlinear (N) settings in terms of ability to measure variable importance, conduct FDR and power analysis, and whether requiring repeated experiments to calculate the statistics in Table 1.

Properties	LCD	LSM	SPD	Ours
Variable importance (L)	✓	✓	✓	✓
Variable importance (N)	×	×	✓	✓
FDR (L)	✓	✓	✓	✓
FDR (N)	×	×	✓	✓
Power (L)	✓	✓	✓	✓
Power (N)	×	×	✓	✓
No repeatability	✓	✓	×	✓

Table 1: Comparison of different knockoff statistics (✓-has the given information, ×-hasn't the given information)

Preliminaries

Without loss of generality, this paper assumes that all covariates are scaled to $[0,1]$. Consider that $(\mathbf{z},\mathbf{x}) = (z_1,\dots,z_q,x_1,...,x_p)$ are covariates and y is response variable. we observe the i.i.d. (independent and identically distributed) samples $\{\mathbf{z}_i,\mathbf{x}_i,y_i\}_{i=1}^n$, where y_i's the response, $(\mathbf{z}_i,\mathbf{x}_i) = (z_{i1},...,z_{iq},x_{i1},...,x_{ip})^\top \in \mathcal{R}^{q+p}$ are the linear and nonlinear inputs respectively. The intrinsic relationship between the input and its output is characterized by

$$y_i = \mu + \sum_{k=1}^{q} z_{ik}\beta_k + \sum_{j=1}^{p} f_j(x_{ij}) + \varepsilon_i. \tag{2}$$

Here μ is an intercept, $\beta_1,\dots,\beta_q$ are the coefficient for linear term, $f_1,...,f_p$ are consisting of some unknown unary smooth functions with $\mathbf{E}(f_j) = 0$, and $\varepsilon_i \sim \mathcal{N}(0,\sigma^2)$ is random noise. By the above definition, we are able to select variables by the following criteria:

$$\mathcal{L} = \{z_k : k\in[q], \beta_k \neq 0\}, \mathcal{N} = \{x_j : j\in[p], f_j \not\equiv 0\},$$
$$\mathcal{O} = \{z_k, x_j : k\in[q], j\in[p], \beta_k = 0, f_j \equiv 0\},$$

where $[q]=\{1,...,q\}$, $[p]=\{1,...,p\}$(The same symbol will be used in the following text). And $\mathcal{L}\cap\mathcal{O}=\emptyset, \mathcal{N}\cap\mathcal{O}=\emptyset$.

B-spline and Random Fourier Features estimation

In this paper we use two methods, B-spline (Schumaker 2007) and Random Fourier Features(RFF) (Băzăvan, Li, and Sminchisescu 2012; Rahimi and Recht 2007), to approximate the nonlinear part $\{f_j\}_{j=1}^p$.

B-spline: By B-spline, f_j satisfies the hypothesis that $f_j \in \mathcal{F}^\alpha[0,1]$ for $\alpha \geq 1$, where $\mathcal{F}^\alpha[0,1] = \{f : [0,1] \to \mathcal{R}, f^{(t)}$ is Lipschitz continuous of order $s\}$ with $s = \alpha - t$. Let $I = \{0 = t_0 <,\dots,t_{D^*} = 1\}$ be a partition of $[0,1]$ with D^* interior knots. Then polynomial spline of order ϱ is globally $\varrho - 2$ continuously differentiable on $[0,1]$ and is a polynomial of degree $\varrho - 1$ in each subinterval $[t_d, t_{d+1})$. Further more, according to de Boor (1978); Huang (2003); Schumaker (2007), when restricting $f_j \in \mathcal{F}^\alpha[0,1], \alpha \geq 1$, there is a collection of B-spline basis functions $\mathbf{B}_j(\cdot) = (B_{j1}(\cdot),\dots,B_{jD}(\cdot))^\top$ with $D = D^* + \varrho$. Then we have

$$f_j(\cdot) = \mathbf{B}_j(\cdot)^\top \boldsymbol{\gamma}_j = \sum_{d=1}^{D} \gamma_{jd} B_{jd}(\cdot), \tag{3}$$

where $\boldsymbol{\gamma}_j = (\gamma_{j1},\dots,\gamma_{jD})^\top \in \mathcal{R}^D$ is a vector of coefficients. Therefore, the model (2) can be rewritten as

$$y_i = \mu + \sum_{k=1}^{q} z_{ik}\beta_k + \sum_{j=1}^{p}\sum_{d=1}^{D} \gamma_{jd} B_{jd}(x_{ij}) + \varepsilon_i. \tag{4}$$

Random Fourier Features (RFF): By RFF, if the kernel functions $K(\cdot,\cdot)$ is shift-invariant, i.e., $K(x,x') = K(x-x')$, and integrates to 1, i.e., $\int_{\mathcal{X}} K(x-x')d(x-x') = 1$. Then $K(\cdot,\cdot)$ satisfies the Fourier expansion by the Bochner's theorem (Bochner 1934):

$$K(x,x') = \int_{\mathcal{R}} p(w)e^{\sqrt{-1}(x-x')}dw,$$

where $p(w) = \int_{\mathcal{X}} K(x)e^{-2\pi\sqrt{-1}wx}dx$ is probability density function. Then, we assume that $f_j \in \mathcal{F}$, where $\mathcal{F}$ is a reproducing kernel Hilbert space (RKHS). By the representer theorem (Schölkopf, Herbrich, and Smola 2001) and the approximate theory of RFF (Rahimi and Recht 2007; Băzăvan, Li, and Sminchisescu 2012), we have

$$f_j(\cdot) = \sum_{i=1}^{n} \alpha_{ji}K(\cdot,x_{ij}) \approx \sum_{v=1}^{V} c_{jv}R_{jv}(\cdot) = \mathbf{R}_j(\cdot)^\top \mathbf{c}_j, \tag{5}$$

where $\mathbf{c}_j = (c_{j1},\dots,c_{jV})^\top \in \mathcal{R}^V$ are coefficients. $\mathbf{R}_j(\cdot) = (R_{j1}(\cdot),...,R_{jV}(\cdot))^\top$ are V Fourier basis functions defined as follow,

$$\begin{aligned} &\omega_{jv} \overset{i.i.d.}{\sim} p(\omega), \quad b_{jv} \overset{i.i.d.}{\sim} Uniform[0,2\pi], \\ &R_{jv}(\cdot) = \sqrt{\frac{2}{V}}\cos\left((\cdot)\omega_{jv} + b_{jv}\right), j \in [p], v \in [V]. \end{aligned} \tag{6}$$

Random Fourier features can achieve computational complexity reduction (Rudi and Rosasco 2017). It means that the computation complexity of the nonlinear part in the (5) is only $O(nV^2)$ instead of $O(n^3)$ by the kernel estimator. The computational savings are considerable if $V/n \to 0$ with $n \to \infty$. Then, the model (2) can be rewritten as

$$y_i = \mu + \sum_{k=1}^{q} z_{ik}\beta_k + \sum_{j=1}^{p}\sum_{v=1}^{V} c_{jv}R_{jv}(x_{ij}) + \varepsilon_i. \tag{7}$$

Knockoff variable construction

Now we introduce how to construct knockoff random variables. According to Candès et al. (2018), the knockoff random variables $(\tilde{\mathbf{z}},\tilde{\mathbf{x}}) = (\tilde{z}_1,\dots,\tilde{z}_q,\tilde{x}_1,\dots,\tilde{x}_p) \in \mathcal{R}^{q+p}$ of the original random variable $(\mathbf{z},\mathbf{x}) \in \mathcal{R}^{q+p}$ are constructed with the following two properties (pairwise exchangeability and independence):

i) $(\mathbf{z},\mathbf{x},\tilde{\mathbf{z}},\tilde{\mathbf{x}}) \overset{d}{=} (\mathbf{z},\mathbf{x},\tilde{\mathbf{z}},\tilde{\mathbf{x}})_{swap(\mathcal{S})}, k \in [q], j \in [p]$,

ii) $y \perp\!\!\!\perp (\tilde{\mathbf{z}},\tilde{\mathbf{x}}) \mid (\mathbf{z},\mathbf{x})$;

here $\overset{d}{=}$ denotes identically distributed. $(\cdot)_{swap(\mathcal{S})}$ is an operator swapping $z_k \in \mathcal{S}$ with $\tilde{z}_k$, $x_j \in \mathcal{S}$ with $\tilde{x}_j$. for instance, with $q=3, p=3$ and $\mathcal{S} = \{z_2, x_3\}$,

$$\begin{aligned} &(z_1,z_2,z_3,x_1,x_2,x_3,\tilde{z}_1,\tilde{z}_2,\tilde{z}_3,\tilde{x}_1,\tilde{x}_2,\tilde{x}_3)_{swap(\mathcal{S})} \overset{d}{=} \\ &(z_1,\tilde{z}_2,z_3,x_1,x_2,\tilde{x}_3,\tilde{z}_1,z_2,\tilde{z}_3,\tilde{x}_1,\tilde{x}_2,x_3). \end{aligned}$$

There are two classic data-based methods to generate knockoff variables. The first method is the model-X knockoffs (Candès et al. 2018), which constructs knockoffs by using the expectation and covariance, like the follows:

$$\mathbf{E}[\mathbf{z},\mathbf{x}] = \mathbf{E}[\tilde{\mathbf{z}},\tilde{\mathbf{x}}],$$

$$\mathbf{cov}[(\mathbf{z},\mathbf{x},\tilde{\mathbf{z}},\tilde{\mathbf{x}})] = \begin{bmatrix} \mathbf{\Phi} & \mathbf{\Phi} - \text{diag}(\boldsymbol{s}) \\ \mathbf{\Phi} - \text{diag}(\boldsymbol{s}) & \mathbf{\Phi} \end{bmatrix},$$

where $\mathbf{\Phi}$ is covariance matrix of $[\mathbf{Z},\mathbf{X}] = (\mathbf{z}_1^\top,\dots,\mathbf{z}_n^\top,\mathbf{x}_1^\top,\dots,\mathbf{x}_n^\top)^\top \in \mathcal{R}^{n\times(q+p)}$, $diag(\boldsymbol{s})$ is a diagonal matrix with all components of s being positive and such that $\mathbf{cov}[(\mathbf{z},\mathbf{x},\tilde{\mathbf{z}},\tilde{\mathbf{x}})]$ is positive definite. To obtain a good selection power, diag$(\boldsymbol{s})$ should be constructed as large as possible (Candès et al. 2018).

The second method is deep knockoffs (Yaniv Romano and Candès 2020), a sampling machine based on deep generative models, which approximates model-X knockoffs for unspecified data distributions. The key idea is to optimize the criterion evaluating the validity of the generated knockoffs by refining the knockoff sampling mechanism iteratively. This method leverages higher-order moments, enabling a better approximation of exchangeability.

In this article, we construct knockoffs with model-X knockoff when the input data approximately follows a multivariate normal distribution, and with deep knockoffs otherwise. In the next section, we extend model-X knockoffs framework to PLM (2), aiming to select true variables while keeping FDR at the target level. For a true variables index set $\mathcal{S}$, selected set $\hat{\mathcal{S}}$, true nulls $\mathcal{S}_0$ and the target FDR=$\tau \in (0,1)$, FDR and mFDR are defined as,

$$\mathbf{FDR}(\hat{\mathcal{S}}) = \mathbf{E}\left[\frac{|\hat{\mathcal{S}} \cap \mathcal{S}_0|}{|\hat{\mathcal{S}}| \vee 1}\right], \mathbf{mFDR}(\hat{\mathcal{S}}) = \mathbf{E}\left[\frac{|\hat{\mathcal{S}} \cap \mathcal{S}_0|}{|\hat{\mathcal{S}}| + 1/\tau}\right].$$

And another important indicator is Power defined as,

$$\mathbf{Power}(\hat{\mathcal{S}}) = \mathbf{E}\left[\frac{|\hat{\mathcal{S}} \cap \mathcal{S}|}{|\mathcal{S}| \vee 1}\right].$$

Knockoff Inference for PLM

Inspired by Huang, Horowitz, and Wei (2010), we implement the following regularization strategy,

$$\begin{aligned} &\min_{\boldsymbol{\beta},\mathbf{f}} \frac{1}{n}\sum_{i=1}^{n}\{y_i - \mu - \sum_{k=1}^{q}(z_{ik}\beta_k + \tilde{z}_{ik}\beta_{k+q}) \\ &-\sum_{j=1}^{p}[f_j(x_{ij}) + f_{j+p}(\tilde{x}_{ij})]\}^2 + \sum_{k=1}^{2q}\lambda_1|\beta_k| + \sum_{j=1}^{2p}\lambda_2\left\|\mathcal{P}_{f_j}\right\|_2. \end{aligned} \tag{8}$$

where $\boldsymbol{\beta} = (\beta_1,\dots,\beta_{2q})^\top \in \mathcal{R}^{2q}$, $\mathbf{f} = (f_1,...,f_{2p})^\top$. $\lambda_1,\lambda_2 > 0$ are tuning parameter. The penalty function $\|\mathcal{P}_{f_j}\|_2$ can be $\|\boldsymbol{\gamma}_j\|_2$ with B-spline estimation in (3), or $\|\mathbf{c}_j\|_2$ with Random Fourier Features estimation in (5). These two types of penalties are group Lasso penalties which have been similarly used in Li, Wang, and Nettleton (2019); Xiaowu Dai and Li (2023) for punishing the nonlinear part. The Lasso penalty $|\beta_k|$ is for punishing the linear part.

Supposing given input $\{\mathbf{z}_i,\mathbf{x}_i\}_{i=1}^n$ and by the representation (3) and (5), we can get the solution of (8) as follows,

$$\hat{y} = \hat{\mu} + \sum_{k=1}^{q}(z_k\hat{\beta}_k + \tilde{z}_k\hat{\beta}_{k+q}) + \sum_{j=1}^{p}[\hat{f}_j(x_j) + \hat{f}_{j+p}(\tilde{x}_j)], \tag{9}$$

where

$$\hat{f}_j(x_j) + \hat{f}_{j+p}(\tilde{x}_j) \overset{\text{B-spline}}{=} \mathbf{B}_j(x_j)^\top \boldsymbol{\gamma}_j + \mathbf{B}_{j+p}(\tilde{x}_j)^\top \boldsymbol{\gamma}_{j+p}$$
$$\overset{\text{RFF}}{=} \mathbf{R}_j(x_j)^\top \mathbf{c}_j + \mathbf{R}_{j+p}(\tilde{x}_j)^\top \mathbf{c}_{j+p}. \quad (10)$$

Here $\overset{\text{B-spline}}{=}$ and $\overset{\text{RFF}}{=}$ denotes estimations by B-spline and Random Fourier Features respectively. The expression (9) indicates that whether we choose the linear part depends on whether $\hat{\beta}_k$ is 0 or not; Meanwhile, whether the nonlinear part can be selected depends on whether the vector $\boldsymbol{\gamma}_j$ or $\mathbf{c}_j$ is 0 or not by (10). Therefore, two parameters are used to select linear and nonlinear parts in (9). It's clearly that $f_j \equiv 0$ when $\boldsymbol{\gamma}_j$ or $\mathbf{c}_j = \mathbf{0}$. Through the above analysis, we can define the following index sets:

$$\hat{\mathcal{L}} = \{z_k : k \in [q], \hat{\beta}_k \neq 0\}, \hat{\mathcal{N}} = \{x_j : j \in [p], \hat{\mathbf{c}}_j \neq 0\},$$
$$\hat{\mathcal{O}} = \{z_k, x_j : k \in [q], j \in [p], \hat{\beta}_k = 0, \hat{\boldsymbol{\gamma}}_j \text{ or } \hat{\mathbf{c}}_j = \mathbf{0}\}. \quad (11)$$

Integral-based knockoff statistics

In this subsection, we will show how to construct the important score and knockoff statistic which help us select important variables and control the false discovery rate (FDR).

In this paper, an integral-based knockoff statistic is proposed. The L-1 norm of functions $\{z_k\beta_k, \tilde{z}_k\beta_{k+q}\}_{k=1}^q$ and $\{f_j(x_j), f_{j+p}(\tilde{x}_j)\}_{j=1}^p$ on $[0,1]$ is defined as important scores that reflect the importance of original and knockoff variables, respectively. Specifically, the larger the L-1 norm value, the more important the corresponding variable is.

$$\Pi_k^{\mathcal{L}} = 2 \cdot \|z\beta_k\|_1 = 2 \cdot \int_0^1 |z\beta_k| dz = |\beta_k|, k \in [2q],$$
$$\Pi_j^{\mathcal{N}} = \|f_j(x)\|_1 = \int_0^1 |f_j(x)| dx, j \in [2p]. \quad (12)$$

It is clear that $2 \cdot \|z\beta_k\|_1 = |\beta_k|$, which is the same to the Lasso Coefficient Difference (LCD) statistic (Candès et al. 2018). Based on the L-1 norm of functions, we can define the knockoff statistics for original variables $(\mathbf{z}, \mathbf{x}) = (z_1, \ldots, z_q, x_1, \ldots, x_p) \in \mathcal{R}^{q+p}$.

$$\Delta_k^{\mathcal{L}} = \Pi_k^{\mathcal{L}} - \Pi_{k+q}^{\mathcal{L}}, k \in [q], \Delta_j^{\mathcal{N}} = \Pi_j^{\mathcal{N}} - \Pi_{j+p}^{\mathcal{N}}, j \in [p]. \quad (13)$$

The higher the value of knockoff statistic $\Delta_k^{\mathcal{L}}$ or $\Delta_j^{\mathcal{N}}$, the more important the kth or jth original variables is. We compare our statistics with the SPD in Figure 1. It demonstrates that our statistic distinguishes between important variables and nulls well.

Select data-dependent thresholds

The key step is to control FDR of variable selection. Given the target nominal FDR $\tau \in (0,1)$, we need to choose data-dependent threshold values $\mathcal{T}^{\mathcal{L}}$ and $\mathcal{T}^{\mathcal{N}}$ for linear and nonlinear variable selection respectively. There are two ways to construct thresholds, one is the classic knockoff filter (Barber and Candès 2015; Candès et al. 2018) to choose thresholds as follows:

$$\mathcal{T} = \min\left\{t \in \Delta_+ : \frac{|\{x_j : \Delta_j \leq -t\}|}{|\{x_j : \Delta_j \geq t\}|} \leq \tau\right\}. \quad (14)$$

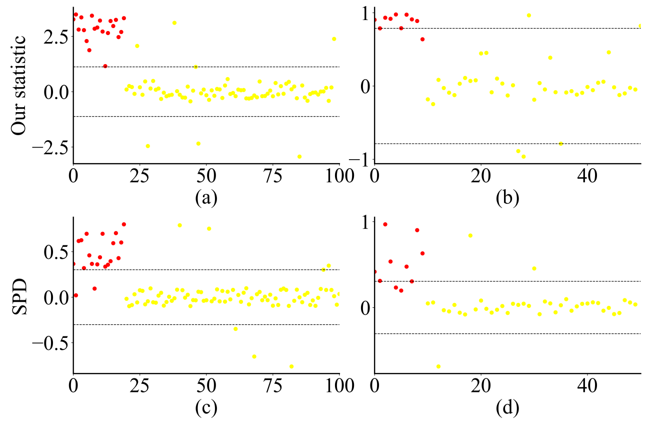

Figure 1: (a) and (b) are the integral-based statistics values of linear and nonlinear parts estimated by IKO-RFF. (c) and (d) are the SPD values of linear and nonlinear parts estimated by IKO-RFF. Where $n = 1000, \rho = 0.2, (q,p) = (100, 50), (|\mathcal{L}|, |\mathcal{N}|) = (20, 10)$. Red points are true variables. Yellow points are true nulls. The horizontal axis represents the variable index.

Algorithm 1: Integral-Based Knockoff Inference Procedure

Input: training data $\{(\mathbf{z}_i, \mathbf{x}_i, y_i)\}_{i=1}^n$, the nominal FDR level $\tau \in (0,1)$.

1: construct knockoffs $\{\tilde{\mathbf{z}}_\mathbf{i}, \tilde{\mathbf{x}}_\mathbf{i}\}_{i=1}^n$ via model-X knockoff or deep knockoffs.
2: estimate $\boldsymbol{\beta}$ and $\mathbf{f}$ by the model (8).
3: compute important scores $\Pi_k^{\mathcal{L}}$ and $\Pi_j^{\mathcal{N}}$ by (12).
4: compute integral-based knockoff statistics $\Delta_k^{\mathcal{L}}$ and $\Delta_j^{\mathcal{N}}$ based on the important scores $\Pi_k^{\mathcal{L}}$ and $\Pi_j^{\mathcal{N}}$ via (13).
5: select thresholds $\mathcal{T}^{\mathcal{L}}$ and $\mathcal{T}^{\mathcal{N}}$ by (14) or (15).
6: choose variables via thresholds in (16).

Output: $\hat{\mathcal{L}}, \hat{\mathcal{N}}, \hat{\mathcal{O}}$.

Here $\Delta_+ = \{|\Delta_j| : |\Delta_j| > 0\}$ is the set of nonzero values. Δ_j denotes the knockoff statistic and we set $\mathcal{T} = \infty$ if the above set is null set. Another one is a more conservative knockoff filter but being used commonly as follows:

$$\mathcal{T}_+ = \min\left\{t \in \Delta_+ : \frac{|\{x_j : \Delta_j \leq -t\}| + 1}{|\{x_j : \Delta_j \geq t\}|} \leq \tau\right\}. \quad (15)$$

After threshold values $\mathcal{T}^{\mathcal{L}}$ and $\mathcal{T}^{\mathcal{N}}$ are constructed via (14) or (15). The variable selection index set (11) can be redefined as in (16) . We call this procedure Integral-based Knockoff (IKO) which can control FDR of $\hat{\mathcal{L}}$ and $\hat{\mathcal{N}}$. It is summarized in Algorithm 1.

$$\hat{\mathcal{L}} = \{z_k : k \in [q], \Delta_k^{\mathcal{L}} \geq \mathcal{T}^{\mathcal{L}}\}, \hat{\mathcal{N}} = \{x_j : j \in [p], \Delta_j^{\mathcal{N}} \geq \mathcal{T}^{\mathcal{N}}\},$$
$$\hat{\mathcal{O}} = \{z_k, x_j : k \in [q], j \in [p], k \notin \hat{\mathcal{L}}, j \notin \hat{\mathcal{N}}\}. \quad (16)$$

Theoretical Analysis

This section builds theoretical guarantees for the IKO procedure to control FDR and power asymptotic property. The supplementary material provides all proofs.

FDR analysis

In this subsection, theoretical analysis shows that the IKO procedure can control FDR at any given nominal level τ and sample size. In subsection , we employ B-spline and Random Fourier Features methods to learn $\{f_j\}_{j=1}^{p}$ in (2). By knockoff construction, the B-spline basis functions and Random Fourier Features of the knockoff variables are crafted to replicate the structure of the B-spline basis functions and Random Fourier Features of the original variables. This design ensures that, although the knockoff variables are not associated with the response when conditioned on the original variables, they still closely mirror the properties of the original variables (Xiaowu Dai and Li 2023). Immediately, we use integral-based knockoff statistics to measure variable importance. Finally, IKO controls the finite-sample FDR, the same as the existing knockoff methods (Barber and Candès 2015; Candès et al. 2018; Xiaowu Dai and Li 2023).

Assumption 1 (PLM decomposition) *For variables $(\mathbf{z},\mathbf{x}) = (z_1,\ldots,z_q,x_1,\ldots,x_p) \in \mathcal{R}^{q+p}$, the response y can be represented as follow decomposition,*

$$y = y^{\mathcal{L}} + y^{\mathcal{N}} + \varepsilon^{\mathcal{L}} + \varepsilon^{\mathcal{N}} = \sum_{k=1}^{q} \beta_k z_k + \sum_{j=1}^{p} f_j(x_j) + \varepsilon^{\mathcal{L}} + \varepsilon^{\mathcal{N}}.$$

Remark 1 *Assumption 1 decomposes PLM into linear and nonlinear parts, facilitating theoretical analysis.*

Assumption 2 (Irrepresentable condition)
i) $\forall k \in [q]$ and $\beta_\kappa \in \mathcal{R}$, $z_k \neq \sum_{\kappa=1;\kappa\neq k}^{q} \beta_\kappa z_\kappa$;
ii) $\forall j \in [p]$ and $f_\iota \in \mathcal{F}^\alpha$ or $\mathcal{F}$, $f_j \neq \sum_{\iota=1;\iota\neq j}^{p} f_\iota$.

Proposition 1 *Suppose Assumption 1 and 2 hold.*
i) $\forall k \in [q], z_k \in \mathcal{O}$ if and only if $\beta_k = 0$;
ii) $\forall j \in [p], x_j \in \mathcal{O}$ if and only if $f_j \equiv 0$.

Lemma 1 (Sign-flip property for the nulls) *Suppose Assumptions (1) and (2) hold.*
i) Let $(\zeta_1,\ldots,\zeta_q)$ be a set of independent random variables, such that $\zeta_k = 1$ if $z_k \in \mathcal{L}$, and $\zeta_k = \pm 1$ with equal probability 1/2 if $z_k \in \mathcal{L}^\perp$. Then, $\left(\Delta_1^{\mathcal{L}},\ldots,\Delta_q^{\mathcal{L}}\right) \overset{d}{=} \left(\Delta_1^{\mathcal{L}}\cdot\zeta_1,\ldots,\Delta_q^{\mathcal{L}}\cdot\zeta_q\right)$.
ii) Let $(\epsilon_1,\ldots,\epsilon_p)$ be a set of independent random variables, such that $\epsilon_j = 1$ if $x_j \in \mathcal{N}$, and $\epsilon_j = \pm 1$ with equal probability 1/2 if $x_j \in \mathcal{N}^\perp$. Then, $\left(\Delta_1^{\mathcal{N}},\ldots,\Delta_p^{\mathcal{N}}\right) \overset{d}{=} \left(\Delta_1^{\mathcal{N}}\cdot\epsilon_1,\ldots,\Delta_p^{\mathcal{N}}\cdot\epsilon_p\right)$.

Remark 2 *Lemma 1 demonstrates that the knockoff statistics, $\Delta_k^{\mathcal{L}}$ and $\Delta_j^{\mathcal{N}}$, exhibit symmetric properties for null variables, where $z_k \notin \mathcal{L}$ or $x_j \notin \mathcal{N}$. The symmetry of these null variables plays a critical role in the IKO procedure, enabling the selection of a data-dependent threshold while effectively controlling FDR (Barber and Candès 2015).*

The following part demonstrates that the IKO procedure effectively controls FDR for any sample size. The results, derived without requiring prior knowledge of the noise level, remain robust across different distributions and varying numbers of variables.

Theorem 1 (FDR control of IKO) *For any $\tau \in (0,1)$ and any sample size n, choose thresholds $\mathcal{T}^{\mathcal{L}} > 0$ and $\mathcal{T}^{\mathcal{N}} > 0$ via $\mathcal{T}$ (14), then $\mathbf{mFDR}(\hat{\mathcal{L}}) \leq \tau$ and $\mathbf{mFDR}(\hat{\mathcal{N}}) \leq \tau$; Meanwhile, choose the threshold $\mathcal{T}^{\mathcal{L}} > 0$ and $\mathcal{T}^{\mathcal{N}} > 0$ via $\mathcal{T}_+$ (15), then $\mathbf{FDR}(\hat{\mathcal{L}}) \leq \tau$ and $\mathbf{FDR}(\hat{\mathcal{N}}) \leq \tau$.*

Remark 3 *Theorem 1 guarantees valid FDR control without imposing any restrictions on the variable dimension $q+p$ or the sample size n.*

Power analysis

To the best of our knowledge, the existing literature lacks comprehensive theoretical analysis of power for knockoff, with the exception of Yingying Fan and Lv (2020); Weinstein et al. (2023), which examined power for linear regressions within the model-X knockoff framework and for thresholded Lasso knockoffs, respectively, and Xiaowu Dai and Li (2023), which investigated power for nonlinear additive models. This section demonstrates that the IKO procedure achieves full power as $n \to \infty$. To establish the theoretical results, we introduce some basic regularity conditions.

Assumption 3 (Minimum signal) *For slowly diverging sequence $\vartheta_n \to \infty$, as $n \to \infty$, let $\eta \equiv C_\eta\{n^{-\beta/(2\beta+1)} + [(\log p)/n]^{1/2}\}$. For some constant $C_\eta > 0$, such that, $\min_{z_k\in\mathcal{L}} |\beta_k| \geq \vartheta_n\{(\log q)/n\}^{1/2}, \min_{x_j\in\mathcal{N}} \|f_j(x_j)\|_1 \geq \vartheta_n\eta$, where the RKHS $\mathcal{F}$ is embedded to a βth order Sobolev space with $\beta > 1$.*

In order to introduce the following conditions, some notations are given. $\Psi = [\mathbf{Z}, \tilde{\mathbf{Z}}], \mathbf{\Sigma} = [\mathbf{R}, \tilde{\mathbf{R}}]$. $\mathbf{\Sigma}_{\mathcal{N}}$ is the design matrix consisted of the jth and $j+p$th columns of $\mathbf{\Sigma}$ for $x_j \in \mathcal{N}$. $\mathbf{\Sigma}_j$ is the design matrix consisted of the jth and $j+p$th columns of $\mathbf{\Sigma}$. It is the same as other matrices.

Assumption 4 (Minimal eigenvalue) *Suppose there is a constant $C_{\min} > 0$, such that the minimal eigenvalue λ_{min} of matrix $\mathbf{E}[n^{-1}\Psi_{\mathcal{L}}^\top\Psi_{\mathcal{L}}]$ and $\mathbf{E}[n^{-1}\mathbf{\Sigma}_{\mathcal{N}}^\top\mathbf{\Sigma}_{\mathcal{N}}]$ satisfies that, $\lambda_{\min}(\mathbf{E}[n^{-1}\Psi_{\mathcal{L}}^\top\Psi_{\mathcal{L}}]) \geq C_{\min}, \lambda_{\min}(\mathbf{E}[n^{-1}\mathbf{\Sigma}_{\mathcal{N}}^\top\mathbf{\Sigma}_{\mathcal{N}}]) \geq C_{\min}$.*

Assumption 5 (Mutual incoherence) *Suppose there exists a constant $\xi \in (0,1]$, such that,*

$$\max_{z_k\notin\mathcal{L}} \|\Psi_k^\top\Psi_{\mathcal{L}}(\Psi_{\mathcal{L}}^\top\Psi_{\mathcal{L}})^{-1}\|_2 \leq 1-\xi.$$

Assumption 6 (Mutual incoherence) *Suppose $p < e^n$ and there exists a constant $0 \leq \xi_{\mathbf{\Sigma}} < 1$, such that,*

$$\max_{x_j\notin\mathcal{N}} \|\mathbf{\Sigma}_j^\top\mathbf{\Sigma}_{\mathcal{N}}[\mathbf{\Sigma}_{\mathcal{N}}^\top\mathbf{\Sigma}_{\mathcal{N}}]^{-1}\|_2 \leq \xi_{\mathbf{\Sigma}},$$

$$\frac{\xi_{\mathbf{\Sigma}}\sqrt{|\mathcal{N}|}+1}{\lambda_2}\eta + \xi_{\mathbf{\Sigma}}\sqrt{|\mathcal{N}|} < 1,$$

Remark 4 *All of these assumptions are reasonable and commonly employed in the current literature. Specifically, Assumption 3 represents a minimal regularization condition that ensures the solution of the PLM model (8) does not omit a significant portion of important variables. This assumption can also be found in Zhao and Yu (2006); Ravikumar, Wainwright, and Lafferty (2010); Xiaowu Dai and Li (2023).*

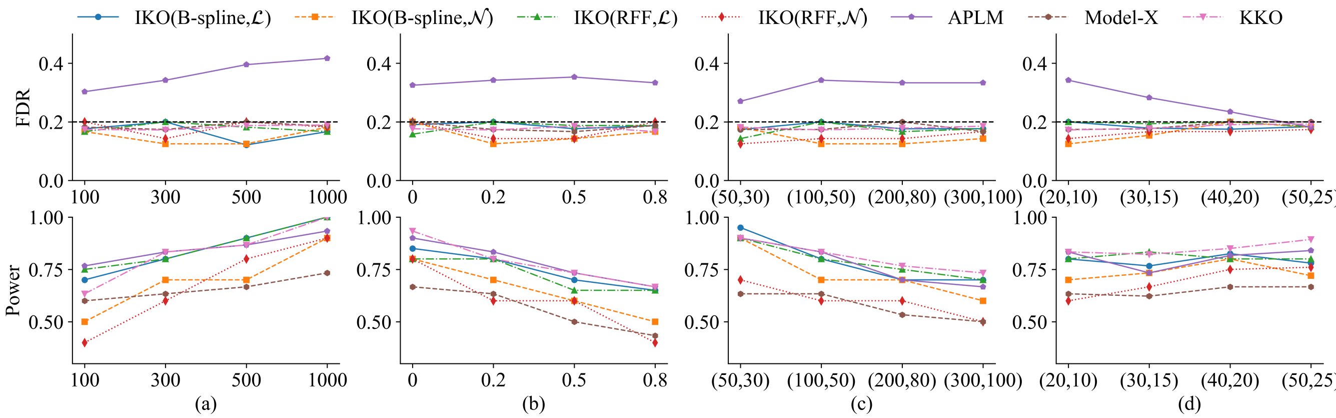

Figure 2: Comparisons in terms of FDR and power with the different (a) sample sizes n, (b) correlations ρ, (c) variable dimensions (q, p), (d) sparsity $(|\mathcal{L}|, |\mathcal{N}|)$. $\mathcal{L}$ and $\mathcal{N}$ denote the results on the set $\hat{\mathcal{L}}$ and $\hat{\mathcal{N}}$, respectively.

		IKO								APLM		Model-X		KKO	
		RFF				B-spline									
		$\mathcal{L}$		$\mathcal{N}$		$\mathcal{L}$		$\mathcal{N}$							
		FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
n	100	0.1667	0.7500	0.2000	0.4000	0.1765	0.7000	0.1667	0.5000	0.3030	0.7667	0.1818	0.6000	0.1739	0.6333
	300	0.2000	0.8000	0.1429	0.6000	0.2000	0.8000	0.1250	0.7000	0.3421	0.8333	0.1739	0.6333	0.1724	0.8333
	500	0.1818	0.9000	0.2000	0.8000	0.1212	0.9000	0.1250	0.7000	0.3953	0.8667	0.2000	0.6667	0.1875	0.8667
	1000	0.1667	1.0000	0.1818	0.8000	0.1667	1.0000	0.1818	0.9000	0.4167	0.9333	0.1852	0.7333	0.1892	1.0000
ρ	0	0.1579	0.8000	0.2000	0.8000	0.1905	0.8500	0.2000	0.8000	0.3250	0.9000	0.2000	0.6667	0.1765	0.9333
	0.2	0.2000	0.8000	0.1429	0.6000	0.2000	0.8000	0.1250	0.7000	0.3421	0.8333	0.1739	0.6333	0.1724	0.8000
	0.5	0.1875	0.6500	0.1429	0.6000	0.1765	0.7000	0.1429	0.6000	0.3529	0.7333	0.1667	0.5000	0.1852	0.7333
	0.8	0.1875	0.6500	0.2000	0.4000	0.1875	0.6500	0.1667	0.5000	0.3333	0.6667	0.1875	0.4333	0.1667	0.6667
q,p	50,30	0.1429	0.9000	0.1250	0.7000	0.1739	0.9500	0.1818	0.9000	0.2703	0.9000	0.1739	0.6333	0.1818	0.9000
	100,50	0.2000	0.8000	0.1429	0.6000	0.2000	0.8000	0.1250	0.7000	0.3421	0.8333	0.1739	0.6333	0.1724	0.8333
	200,80	0.1667	0.7500	0.1429	0.6000	0.1765	0.7000	0.1250	0.7000	0.3333	0.7000	0.2000	0.5333	0.1786	0.7667
	300,100	0.1765	0.7000	0.1667	0.5000	0.1765	0.7000	0.1429	0.6000	0.3333	0.6667	0.1667	0.5000	0.1852	0.7333
$\lvert\mathcal{L}\rvert, \lvert\mathcal{N}\rvert$	20,10	0.2000	0.8000	0.1429	0.6000	0.2000	0.8000	0.1250	0.7000	0.3421	0.8333	0.1739	0.6333	0.1724	0.8333
	30,15	0.1935	0.8333	0.1667	0.6667	0.1786	0.7667	0.1538	0.7333	0.2826	0.7333	0.1765	0.6222	0.1778	0.8222
	40,20	0.2000	0.8000	0.1667	0.7500	0.1750	0.8250	0.2000	0.8000	0.2344	0.8167	0.2000	0.6667	0.1905	0.8500
	50,25	0.1837	0.8000	0.1739	0.7600	0.1837	0.7800	0.1818	0.7200	0.1818	0.8400	0.2000	0.6667	0.1928	0.8933

Table 2: FDR and Power with different sample sizes n, correlations ρ, variable dimensions (q, p), sparsity $(|\mathcal{L}|, |\mathcal{N}|)$.

Assumption 4 pertains to the minimal eigenvalue condition, which guarantees that the Gram matrix of the important set in the augmented design matrix is invertible. Similar conditions have been imposed in Lasso regressions (Ravikumar, Wainwright, and Lafferty 2010; Raskutti, J Wainwright, and Yu 2012; Yingying Fan and Lv 2020). Assumptions 5 and 6 imply that the correlation between true signals and null variables should not have a strong influence. Similar assumptions appear in Zhao and Yu (2006); Ravikumar, Wainwright, and Lafferty (2010); Wainwright (2019).

With these regularity conditions established, we analyze the statistical power of the IKO procedure.

Theorem 2 (Power) *Suppose Assumptions 3-6 hold, Then the oracle IKO procedure has an asymptotic property that* $\mathbf{Power}(\hat{\mathcal{L}}) \to 1$ *and* $\mathbf{Power}(\hat{\mathcal{N}}) \to 1$*, as* $n \to \infty$.

Remark 5 *Theorems 1 and 2 demonstrate that the IKO procedure can achieve both effective FDR control and strong statistical power under certain regularity conditions. Unlike Theorem 1, which imposes no restrictions on the relationship between variable dimension and sample size, Theorem 2 specifically applies to cases where* $p < n$ *or* $n < p < e^n$.

Experimental Analysis

This section evaluates the proposed IKO's empirical performance through experiments in both simulated and real data. In all cases, when using B-splines, we set $D^* = 4$ and employ cubic splines ($\varrho = 4$). When using Random Fourier Features (RFF), we set $V = 4$ and use the Laplacian kernel $K(x, x') = \frac{1}{2}e^{-|x-x'|}$. the parameter (λ_1, λ_2) is selected by cross-validation. And the target FDR of $\hat{\mathcal{L}}$ and $\hat{\mathcal{N}}$ are 0.2.

Simulations

The following model was used to generate simulation data.

$$y = \sum\nolimits_{k=1}^{q} 3.5 z_k + \sum\nolimits_{j=1}^{p} \frac{\sin(\gamma_{1j} x_j)}{2 - \sin(\gamma_{2j} x_j)} + \varepsilon$$

where $\varepsilon \sim \mathcal{N}(0,1)$, $\gamma_{1j}, \gamma_{2j} \sim U(1,10)$, the designed matrix $[\mathbf{Z}, \mathbf{X}] \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$, where $\mathbf{\Omega} = (\rho^{|i-j|})_{1<i,j<q+p}$.

To further demonstrate the effectiveness of IKO, we compare it with APLM (Li, Wang, and Nettleton 2019), Model-X (Candès et al. 2018) via R package knockoff, and KKO (Xiaowu Dai and Li 2023) via R package kko. APLM is the regularization model (8) that does not incorporate knockoffs, while Model-X and KKO represent linear and nonlinear additive knockoff frameworks, respectively. In Figure 2 and Table 2, simulation experiments were conducted under various data settings, including different correlations ρ, sample sizes n, variable dimensions (q,p), and sparsity $(|\mathcal{L}|, |\mathcal{N}|)$.

(a) $n = \{100, 300, 500, 1000\}$,
$\rho = 0.2, (q,p) = (100, 50), (|\mathcal{L}|, |\mathcal{N}|) = (20, 10)$.

(b) $\rho = \{0, 0.2, 0.5, 0.8\}$,
$n = 300, (q,p) = (100, 50), (|\mathcal{L}|, |\mathcal{N}|) = (20, 10)$.

(c) $(q,p) = \{(50,30), (100,50), (200,80), (300,100)\}$,
$(|\mathcal{L}|, |\mathcal{N}|) = (20, 10), n = 300, \rho = 0.2$.

(d) $(|\mathcal{L}|, |\mathcal{N}|) = \{(20,10), (30,15), (40,20), (50,25)\}$,
$n = 300, \rho = 0.2, (q,p) = (100, 50)$.

Figure 2 and Table 2 demonstrate the performance of IKO. It can be seen that the performance of the selection of RFF and B-spline in IKO is about the same. As the sample size increases, the power of both the linear and nonlinear parts tends to approach 1. In contrast, power decreases when the variable dimension is increased or when correlations increase. Although power shows a slight upward trend as sparsity changes, the effect is not obvious. For example, the power performance of IKO is not consistently strong compared to KKO. A possible reason for the decline in power is the adoption of the double-penalty strategy to fine-grained distinguish key variables (linear or non-linear). While APLM exhibits good power performance, it fails to effectively control FDR. Due to incorporation of nonlinear data, power of Model-X consistently remains relatively low.

Real data experiments

In real-world setting, we conducted a toy example using IKO to analyze the Boston housing dataset . It includes 506 samples and 14 variables, comprising 13 input features and 1 target variable (house price: MEDV). The features are as follows: CRIM: Per capita crime rate by town. ZN: Proportion of residential land zoned for lots over 25,000 sq.ft. INDUS: Proportion of non-retail business acres per town. CHAS: Charles River dummy variable (= 1 if tract bounds river, 0 otherwise). NOX: Nitric oxide concentration (parts per 10 million). RM: Average number of rooms per dwelling. AGE: Proportion of owner-occupied units built before 1940. DIS: Weighted distances to five Boston employment centers. RAD: Index of accessibility to radial highways. TAX: Full-value property-tax rate per \$10,000. PTRATIO: Pupil-teacher ratio by town. B: $1000(Bk - 0.63)^2$, where Bk is the proportion of blacks by town LSTAT: Percentage of lower status of the population. MEDV: Median value of owner-occupied homes in \$1000's. We scale all variables to $[0,1]$.

Variables	APLM	Model-X	KKO	IKO(B-spline)	IKO(RFF)
CRIM	✓		✓	✓	✓
ZN					
INDUS	✓				
CHAS					
NOX	✓		✓		✓
RM	✓	✓	✓	✓	✓
AGE					
DIS	✓		✓	✓	✓
RAD	✓		✓		
TAX	✓		✓	✓	✓
PTRATIO	✓	✓	✓	✓	✓
B					
LSTAT	✓	✓	✓	✓	✓

Table 3: Real data experiments, ✓denotes that variable is selected.

$q+p$	$\mathcal{L}$(B-spline)	$\mathcal{N}$(B-spline)	$\mathcal{L}$(RFF)	$\mathcal{N}$(RFF)
13	3	3	3	4
75(62)	3	4	3	4
100(87)	4(1)	4	4(1)	4
125(112)	6(2)	4	5(1)	4

Table 4: IKO variable selection results with noise variables added and 100 data points from the Boston housing dataset. The number inside the parentheses indicates how many additional noise variables are included.

Inspired by Hao Helen Zhang and Liu (2011), we assume that PTRATIO, RAD, B, LSTAT, CRIM, ZN, and CHAS belong to the linear part; NOX, RM, DIS, TAX, AGE, and INDUS belong to nonlinear part. For comparison, we also fit the data by APLM, Model-X, and KKO in Table 3. Besides, we sample 100 data points from the Boston housing dataset and use a Gaussian distribution to generate random noise variables. We then evaluate the performance of IKO under the conditions $n > q+p$ and $n < q+p$, as shown in Table 4. The similar strategy was used for Butcher and Smith (2020).

Conclusion

We introduce the knockoff framework to PLM scenarios. Specifically, a new integral-based knockoff statistic is proposed which measures variable importance not only for linear part but also for nonlinear part. The traditional statistics LCD Candès et al. (2018) is a part of our proposed statistics, specifically for the linear part. This method ensures FDR control across both linear and nonlinear parts. Compared to closely related works (Candès et al. 2018; Li, Wang, and Nettleton 2019; Xiaowu Dai and Li 2023), experimental results demonstrate that IKO achieves competitive performance. Additionally, we have established theoretical guarantees for both FDR control and power performance. In the future, it would be interesting to extend this work to multivariate collaborative selection.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (NSFC) (Nos. 62376104, 62172120, 12426512).

References

Barber, R. F.; and Candès, E. J. 2015. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5): 2055 – 2085.

Barber, R. F.; and Candès, E. J. 2019. A knockoff filter for high-dimensional selective inference. *The Annals of Statistics*, 47(5): 2504 – 2537.

Băzăvan, E. G.; Li, F.; and Sminchisescu, C. 2012. Fourier Kernel Learning. In Fitzgibbon, A.; Lazebnik, S.; Perona, P.; Sato, Y.; and Schmid, C., eds., *Computer Vision – ECCV 2012*, 459–473. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-33709-3.

Bochner, S. 1934. A theorem on Fourier-Stieltjes integrals. *Bulletin of the American Mathematical Society*, 40: 271–276.

Butcher, B.; and Smith, B. J. 2020. Feature Engineering and Selection: A Practical Approach for Predictive Models. *The American Statistician*, 74(3): 308–309.

Candès, E.; Fan, Y.; Janson, L.; and Lv, J. 2018. Panning for Gold: 'Model-X' Knockoffs for High Dimensional Controlled Variable Selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(3): 551–577.

de Boor, C. 1978. A Practical Guide to Splines. In *Applied Mathematical Sciences*.

Engle, R. F.; Granger, C. W.; Rice, J.; and Weiss, A. 1986. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, 81(394): 310–320.

Fuller, W. A. 2009. *Measurement error models*. John Wiley & Sons.

Hao Helen Zhang, G. C.; and Liu, Y. 2011. Linear or Nonlinear? Automatic Structure Discovery for Partially Linear Models. *Journal of the American Statistical Association*, 106(495): 1099–1112. PMID: 22121305.

Härdle, W.; Liang, H.; and Gao, J. 2000. *Partially linear models*. Springer Science & Business Media.

Härdle, W.; Müller, M.; Sperlich, S.; Werwatz, A.; et al. 2004. *Nonparametric and semiparametric models*, volume 1. Springer.

Huang, J.; Horowitz, J. L.; and Wei, F. 2010. Variable selection in nonparametric additive models. *The Annals of Statistics*, 38(4): 2282 – 2313.

Huang, J. Z. 2003. Local asymptotics for polynomial spline regression. *The Annals of Statistics*, 31(5): 1600 – 1635.

Li, X.; Wang, L.; and Nettleton, D. 2019. Additive partially linear models for ultra-high-dimensional regression. *Stat*, 8(1): e223. E223 sta4.223.

Lian, H.; Zhao, K.; and Lv, S. 2019. Projected spline estimation of the nonparametric function in high-dimensional partially linear models for massive data. *The Annals of Statistics*, 47(5): 2922–2949.

Lin, Y.; and Zhang, H. H. 2006. Component selection and smoothing in multivariate nonparametric regression. *The Annals of Statistics*, 34(5): 2272 – 2297.

Lv, S.; and Lian, H. 2022. Debiased distributed learning for sparse partial linear models in high dimensions. *Journal of Machine Learning Research*, 23(2): 1–32.

Rahimi, A.; and Recht, B. 2007. Random features for large-scale kernel machines. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, NIPS'07, 1177–1184. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781605603520.

Raskutti, G.; J Wainwright, M.; and Yu, B. 2012. Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *Journal of machine learning research*, 13(2).

Ravikumar, P.; Wainwright, M. J.; and Lafferty, J. D. 2010. High-dimensional Ising model selection using l1-regularized logistic regression. *The Annals of Statistics*, 38(3): 1287 – 1319.

Rudi, A.; and Rosasco, L. 2017. Generalization properties of learning with random features. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 3218–3228. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510860964.

Ruppert, D.; Wand, M. P.; and Carroll, R. J. 2003. *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Schölkopf, B.; Herbrich, R.; and Smola, A. J. 2001. A Generalized Representer Theorem. In Helmbold, D.; and Williamson, B., eds., *Computational Learning Theory*, 416–426. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-44581-4.

Schumaker, L. 2007. *Spline Functions: Basic Theory*. Cambridge Mathematical Library. Cambridge University Press, 3 edition.

Su, H.; Sun, Q.; Yi, M.; Li, G.; and Yuan, P. 2024. StabGKnock: controlled variable selection for partially linear models using generalized knockoffs. *Statistics and Computing*, 34.

Su, W.; and Candès, E. 2016. SLOPE is adaptive to unknown sparsity and asymptotically minimax. *The Annals of Statistics*, 44(3): 1038 – 1068.

Wainwright, M. J. 2019. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press.

Wang, L.; Xue, L.; Qu, A.; and Liang, H. 2014a. Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates. *The Annals of Statistics*, 42(2): 592 – 624.

Wang, L.; Xue, L.; Qu, A.; and Liang, H. 2014b. Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates. *The Annals of Statistics*, 42(2): 592 – 624.

Weinstein, A.; Su, W. J.; Bogdan, M.; Barber, R. F.; and Candès, E. J. 2023. A power analysis for model-X knockoffs with -regularized statistics. *The Annals of Statistics*, 51(3): 1005 – 1029.

Xiaowu Dai, X. L.; and Li, L. 2023. Kernel Knockoffs Selection for Nonparametric Additive Models. *Journal of the American Statistical Association*, 118(543): 2158–2170. PMID: 38143786.

Xie, H.; and Huang, J. 2009. SCAD-penalized regression in high-dimensional partially linear models. *Annals of Statistics*, 37: 673–696.

Yaniv Romano, M. S.; and Candès, E. 2020. Deep Knockoffs. *Journal of the American Statistical Association*, 115(532): 1861–1872.

Yingying Fan, G. L., Emre Demirkaya; and Lv, J. 2020. RANK: Large-Scale Inference With Graphical Nonlinear Knockoffs. *Journal of the American Statistical Association*, 115(529): 362–379. PMID: 32742045.

Zhao, P.; and Yu, B. 2006. On Model Selection Consistency of Lasso. *Journal of Machine Learning Research*, 7(90): 2541–2563.