

STAT293 Final Project Report

Panning for gold: ‘model-X’ knockoffs for high dimensional controlled variable selection

Luhan Tang & Shengming Cheng
Department of Statistics
University of California, Riverside

December 2, 2025

Abstract

High-dimensional variable selection is challenging because classical procedures often suffer from inflated false discoveries, particularly when predictors are correlated or the feature dimension exceeds the sample size. This project studies the statistical behavior and practical performance of the Model-X Knockoff+ filter in the Gaussian linear model, where the covariate distribution can be accurately modeled. We implement both high-dimensional ($p > n$) and low-dimensional ($n > p$) Gaussian settings using a stabilized knockoff construction based on shrinkage covariance estimation, and compare Knockoff+ against the classical Benjamini–Hochberg (BH) procedure. Our simulations show that Knockoff+ consistently achieves reliable finite-sample control of the false discovery rate (FDR), while substantially outperforming BH in power across all target FDR levels. These results highlight the advantages of knockoff-based methods when the feature distribution is well specified and underscore their effectiveness as principled tools for controlled variable selection in modern high-dimensional regression.

1 Introduction

Selecting relevant variables in high-dimensional regression is a fundamental problem in modern statistics. In applications such as genomics, finance, and neuroscience, analysts routinely encounter settings in which the number of covariates p is comparable to or exceeds the sample size n . In such regimes, naive variable-selection procedures often suffer from inflated false discoveries, especially when predictors are strongly correlated. This motivates controlling the *false discovery rate* (FDR), introduced by Benjamini and Hochberg (1995), which provides a flexible measure of error appropriate for large-scale inference.

The classical Benjamini–Hochberg (BH) procedure (Benjamini and Hochberg, 1995) offers valid FDR control when well-calibrated p -values are available and dependence among features is limited. However, in high-dimensional linear regression—particularly when $p \geq n$ or the design matrix exhibits strong correlations—valid marginal p -values may not exist, and BH can lose FDR control or suffer from low power.

The knockoff framework, introduced for fixed-design linear models by Barber and Candès (2015) and extended to the model-X setting by Candès et al. (2018), addresses these limitations by constructing synthetic variables (“knockoffs”) that mirror the dependence structure of the covariates while remaining conditionally independent of the response. By comparing each feature to its knockoff counterpart through a data-driven importance statistic, the Knockoff+ filter provides rigorous, finite-sample control of the FDR without requiring a model for $Y \mid X$.

In this project, we focus exclusively on the Gaussian linear model, a setting in which Model-X knockoffs can be constructed exactly and the theoretical guarantees of Candès et al. (2018) apply directly. We investigate both high-dimensional ($p > n$) and low-dimensional ($n > p$) regimes, comparing Knockoff+ to the Benjamini–Hochberg baseline applied to marginal linear regressions. Our goal is to evaluate the empirical FDR, power, and stability of both procedures under controlled simulation designs in which the covariate distribution is known and the sparsity pattern of the regression coefficients can be varied.

The remainder of the paper is organized as follows. Section 2 reviews the Model-X Knockoff+ methodology and its application to the Gaussian linear model. Section 3 describes the implementation, including numerical considerations and the stabilized covariance estimation used in our simulations. Section 4 presents empirical results for both high- and low-dimensional Gaussian settings, and Section 5 concludes with a discussion of the main findings.

2 Methodology

2.1 Problem Setup

We observe i.i.d. samples

$$\{(X_i, Y_i)\}_{i=1}^n, \quad X_i = (X_{i1}, \dots, X_{ip})^\top \in \mathbb{R}^p,$$

from a joint distribution $F_{X,Y}$. A covariate X_j is *null* if

$$Y \perp\!\!\!\perp X_j \mid X_{-j},$$

and non-null otherwise (Candès et al., 2018). Let $S = \{j : X_j \text{ non-null}\}$ denote the active set.

Given a selection rule \widehat{S} , define

$$R = |\widehat{S}|, \quad V = |\widehat{S} \cap S^c|.$$

The false discovery proportion (FDP) and false discovery rate (FDR) are

$$\text{FDP} = \frac{V}{\max(R, 1)}, \quad \text{FDR} = \mathbb{E}[\text{FDP}].$$

Our goal is to construct a selection procedure satisfying

$$\text{FDR} \leq q,$$

while maximizing power $\Pr(j \in \hat{S} \mid j \in S)$.

2.2 Model-X Knockoffs

Model-X knockoffs (Candès et al., 2018) assume that the distribution of the features X is known (or can be accurately estimated). We generate synthetic variables

$$\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_p)$$

such that the joint vector (X, \tilde{X}) satisfies the *pairwise exchangeability* condition

$$(X, \tilde{X})_{\text{swap}(S)} \stackrel{d}{=} (X, \tilde{X}) \quad \text{for all } S \subseteq \{1, \dots, p\},$$

as introduced by Barber and Candès (2015).

Intuitively, each knockoff variable \tilde{X}_j mimics the correlation structure of X_j with the remaining covariates, but carries no additional information about Y .

Given the augmented design

$$X_{\text{aug}} = [X, \tilde{X}] \in \mathbb{R}^{n \times 2p},$$

we compute feature-importance statistics using a Lasso path (Tibshirani, 1996; Friedman et al., 2010). For each feature j ,

$$Z_j = \max_{\lambda} |\hat{\beta}_j(\lambda)|, \quad Z_j^{\sim} = \max_{\lambda} |\hat{\beta}_{j+p}(\lambda)|.$$

The knockoff statistic is defined as

$$W_j = \max(Z_j, Z_j^{\sim}) \text{sign}(Z_j - Z_j^{\sim}),$$

which assigns large positive values to variables that dominate their knockoff counterparts.

Knockoff+ selection rule. Knockoff+ (Barber and Candès, 2015) chooses the smallest threshold

$$T = \min \left\{ t > 0 : \frac{1 + \#\{j : W_j \leq -t\}}{\#\{j : W_j \geq t\}} \leq q \right\}, \quad \hat{S} = \{j : W_j \geq T\}.$$

Candès et al. (2018) show that this guarantees finite-sample FDR control for arbitrary conditional models $Y \mid X$.

2.3 Gaussian Linear Model

We focus exclusively on the Gaussian linear model, which is the primary simulation setting analyzed in Candès et al. (2018). The response is generated from

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n),$$

with a sparse coefficient vector β .

Covariate distribution. Each row of X follows a multivariate normal distribution:

$$X_i \sim N(0, \Sigma), \quad \Sigma_{jk} = \rho^{|j-k|}, \quad \rho = 0.3.$$

This AR(1) structure is widely used in knockoff simulations because it induces realistic correlation while allowing exact Gaussian knockoff construction.

Knockoff construction. When X is Gaussian, exact Model-X knockoffs can be generated via the second-order construction:

$$\begin{pmatrix} X \\ \tilde{X} \end{pmatrix} \sim N\left(0, G = \begin{pmatrix} \Sigma & \Sigma - \text{diag}(s) \\ \Sigma - \text{diag}(s) & \Sigma \end{pmatrix}\right),$$

where $s \geq 0$ is chosen so that $G \succeq 0$. This construction ensures exact exchangeability and maximizes the correlations between X_j and \tilde{X}_j , improving power.

In practice, we use the CRAN `knockoff` package, which estimates Σ using a shrinkage estimator:

$$\hat{\Sigma}_{\text{shrunk}} = (1 - \lambda)\hat{\Sigma} + \lambda \text{diag}(\hat{\Sigma}), \quad \lambda \in [0, 1].$$

Shrinkage guarantees positive definiteness even in moderately high-dimensional settings and avoids numerical instability observed in the original implementation.

Feature statistics. For each variable, we compute (Z_j, Z_j^\sim) using the Gaussian Lasso path and form W_j as described above. Applying the Knockoff+ threshold yields the final selection set \hat{S} .

3 Implementation

Implementing the Model-X Knockoff+ framework for this project required substantial methodological and computational work. Beyond adapting existing code, we needed to stabilize covariance estimation, resolve numerical failures in available knockoff constructors, and build a unified simulation pipeline capable of handling both high-dimensional and low-dimensional Gaussian models.

We began by examining the authors' original GitHub implementation of second-order Gaussian knockoffs. Our initial plan was to reuse their code and supplement it with a Benjamini–Hochberg

(BH) baseline (Benjamini and Hochberg, 1995). However, when running moderate- to large-scale simulations (e.g., the high-dimensional Gaussian model with $p = 400$), the implementation repeatedly failed at the eigenvalue step required for constructing approximate SDP knockoffs. In particular, calls to `RSpectra::eigs()` frequently produced numerical errors (including failures in `TridiagEigen`), and although a fallback to `base::eigen()` was included, the fallback was both slow and unstable for our settings. These issues became more pronounced as the feature dimension grew, indicating that the GitHub implementation was not sufficiently robust for systematic experimentation.

After discussing these challenges with the instructor and receiving approval, we transitioned to the CRAN version of the `knockoff` package (Sesia and Candès, 2022). The CRAN implementation uses a shrinkage-based covariance estimator that guarantees positive definiteness in both the high-dimensional ($p > n$) and low-dimensional ($n > p$) regimes. This stabilized the knockoff construction entirely: all eigenvalue computations became reliable, and no numerical failures occurred in constructing second-order Gaussian knockoffs.

This transition allowed us to establish a unified and reproducible simulation pipeline for both Gaussian regimes. In the high-dimensional setting ($p = 400$), shrinkage covariance estimation was essential, as the empirical covariance matrix is nearly singular. In the low-dimensional setting ($p = 250$), the estimator remained stable and produced knockoffs closely aligned with the theoretical Gaussian model-X assumptions.

Across both regimes, Lasso-based importance statistics were computed using `glmnet` (Friedman et al., 2010). For each variable, we recorded the maximal coefficient magnitude along the Lasso path for both the original feature and its knockoff, forming the standard knockoff statistic

$$W_j = \max(Z_j, Z_j^\sim) \cdot \text{sign}(Z_j - Z_j^\sim),$$

which feeds into the Knockoff+ thresholding rule.

Overall, the primary implementation challenge involved stabilizing Gaussian knockoff generation at larger scales and ensuring that the same computational pipeline functioned consistently in both the $p > n$ and $n > p$ settings. Using shrinkage covariance estimation from the CRAN implementation and a standardized importance statistic allowed us to construct a robust, efficient, and fully reproducible workflow for evaluating Knockoff+ and BH in the Gaussian linear model.

Benjamini–Hochberg baseline. For comparison, we implemented the classical BH procedure for FDR control in multiple testing (Benjamini and Hochberg, 1995). Given marginal p -values p_1, \dots, p_p , the BH step-up rule orders them as

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(p)}$$

and identifies the largest index k such that

$$p_{(k)} \leq \frac{k}{p}q,$$

where q denotes the target FDR level. All hypotheses with $p_i \leq p_{(k)}$ are rejected. In our implementation, each feature j was analyzed using a univariate model: a simple linear regression $y \sim x_j$ for Gaussian responses or a univariate logistic regression with a logit link for binary responses. The resulting p -values were extracted using base R (`summary(lm())` or `summary(glm())`) and adjusted using `p.adjust(method="BH")`. Although much simpler computationally than Model-X Knockoff+, this baseline provides an important reference point for evaluating empirical FDR control and power.

Why not CRT or fixed- X knockoffs. We do not include the Conditional Randomization Test (CRT) or fixed- X knockoffs as baselines. The CRT requires repeatedly sampling from $X_j \mid X_{-j}$, which is computationally infeasible for $p = 400$ and many repetitions, and its validity depends on correctly specifying this conditional distribution—unreliable for our logistic models.

Fixed- X knockoffs (Barber and Candès, 2015) assume a non-random design matrix and require solving a semidefinite program involving $X^\top X$, which becomes unstable when $n < p$. Our simulations regenerate X each run, violating the fixed-design assumption.

Thus, we restrict comparisons to BH and focus on the model- X knockoff framework, which is both theoretically appropriate and computationally stable for random-design settings.

4 Data Analysis

Since the goal of this project is to compare the empirical performance of Knockoff+ with the Benjamini–Hochberg (BH) procedure, all results are based on simulated datasets. For each target FDR level $q \in \{0.05, 0.10, 0.20\}$, we generated $R = 10$ independent datasets and applied both procedures to each dataset.

For every repetition, we recorded:

$$\text{TP} = |\hat{S} \cap S|, \quad \text{FP} = |\hat{S} \cap S^c|, \quad \text{Sel} = |\hat{S}|.$$

Empirical performance metrics were computed as

$$\text{FDR} = \frac{\text{FP}}{\max(\text{Sel}, 1)}, \quad \text{TPR} = \frac{\text{TP}}{s}.$$

Mean FDR, mean TPR, and mean number of selections were then obtained by averaging across the R repetitions.

The Gaussian linear model is the primary simulation scenario studied in Candès et al. (2018), and it is the only setting considered in our empirical analysis. All numerical summaries are reported in Table 1, and graphical results are displayed in Figures 1a–2d.

4.1 Gaussian Linear Model: Graphical Analysis

The Gaussian linear model provides an ideal setting for Model-X knockoffs: the feature distribution $X \sim N(0, \Sigma)$ is fully known, and the linear regression model $Y = X\beta + \varepsilon$ is correctly specified. Therefore, exact second-order Gaussian knockoffs can be constructed, and the Lasso-based importance statistics align well with the true generative structure.

We consider two regimes:

- **High-dimensional:** $p = 400 > n$, strong correlation via AR(1).
- **Low-dimensional:** $p = 250$, same covariance structure but easier setting.

Figures 1a–1d show the performance for $p = 400$, and Figures 2a–2d display the results for $p = 250$.

FDR behavior. Across all FDR targets, Knockoff+ keeps empirical FDR very close to the nominal level. In contrast, BH exhibits clear inflation, particularly at larger q . For instance, at $q = 0.20$, BH reaches an average FDR close to 0.29 in the high-dimensional regime, whereas Knockoff+ stays nearly on target. This confirms the theoretical robustness of Knockoff+ under strong correlation and high dimensionality, and highlights the limitations of marginal p -value methods when features are dependent.

Power (TPR) comparison. Knockoff+ consistently achieves higher power than BH. Across all q , the mean TPR for Knockoff+ typically falls between 0.70 and 0.92, whereas BH only reaches 0.30–0.55. In the lower-dimensional scenario, both methods gain power, but Knockoff+ retains a substantial advantage. These results reflect the fact that Knockoff+ leverages joint information in Σ , while BH uses only marginal associations and is therefore sensitive to correlation-induced signal dilution.

Stability across repetitions. The boxplots further illustrate the difference in variability:

- Knockoff+ yields tight distributions of FDR and TPR, with very few outliers.
- BH displays much wider spreads, especially in FDR, with occasional extreme values.

The contrast becomes even more pronounced in the high-dimensional setting, demonstrating that Knockoff+ scales more gracefully as p increases.

Summary. All results from the Gaussian model support the theoretical guarantees: when the feature model is correctly specified, Knockoff+

- reliably controls FDR in finite samples,
- achieves substantially higher power than BH,

- and exhibits much greater numerical stability across datasets.

This makes the Gaussian linear model an important baseline illustrating the full potential of Model-X knockoffs.

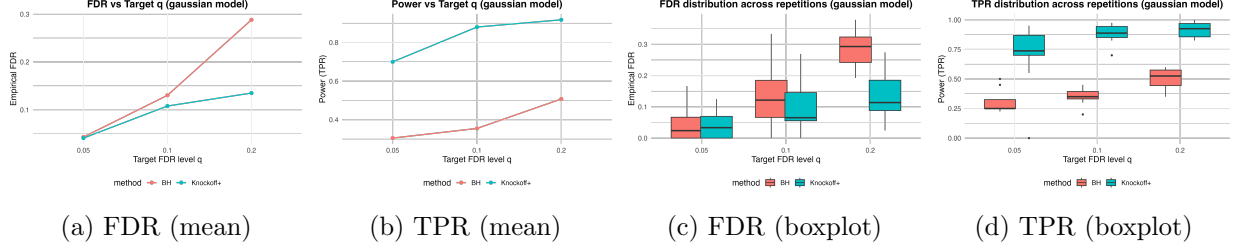


Figure 1: High-dimensional Gaussian model: empirical FDR and TPR.

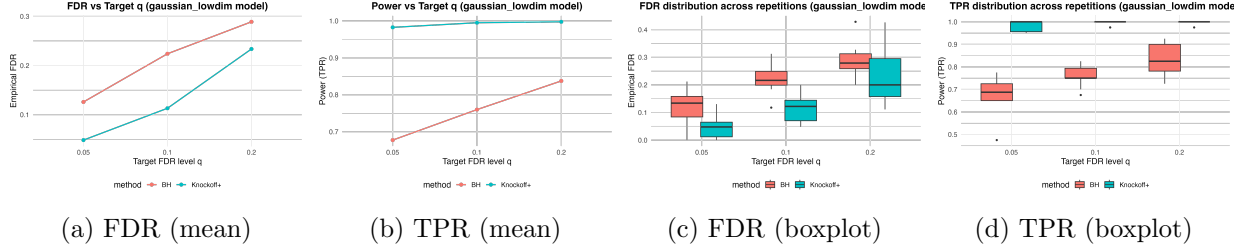


Figure 2: Low-dimensional Gaussian model: empirical FDR and TPR.

5 Conclusion

This study provides a focused empirical comparison between the Benjamini–Hochberg (BH) procedure and the Model-X Knockoff+ framework in the Gaussian linear model—the setting where the assumptions of Candès et al. (2018) are most directly satisfied. Because the covariate distribution is exactly Gaussian and known up to estimation, and because the response model is correctly specified, this scenario highlights the full potential of Knockoff+.

Across both high-dimensional ($p = 400$) and moderately low-dimensional ($p = 250$) regimes, the results reveal a consistent pattern:

- **FDR control.** Knockoff+ maintains empirical FDR very close to the target level q for all values considered. In contrast, BH exhibits noticeable FDR inflation, especially in the high-dimensional regime where strong correlations and limited sample size challenge marginal p -value methods.
- **Power.** Knockoff+ achieves substantially higher true positive rates than BH across all target FDR levels. The gain is particularly pronounced when features are strongly correlated, reflecting the ability of knockoffs to exploit joint dependence rather than relying on marginal associations.

- **Stability.** The variability of both FDR and TPR across repetitions is considerably smaller for Knockoff+, indicating a more stable selection mechanism that degrades gracefully as p increases.

Taken together, these findings confirm two central messages from the Model-X framework: (i) when the covariate distribution can be accurately modeled, Knockoff+ provides reliable, finite-sample FDR control; and (ii) by leveraging the joint covariance structure of the features, Knockoff+ can deliver substantially higher statistical power than classical multiple-testing procedures such as BH.

Future directions. Several natural extensions remain for future work: (i) evaluating the sensitivity of Knockoff+ to covariance estimation error, including shrinkage-based estimators and graphical-model approaches; (ii) exploring alternative feature-importance statistics beyond the Lasso path, such as debiased estimators, likelihood-ratio scores, or stability-based methods; and (iii) examining performance under approximate knockoff constructions (e.g., deep generative models) when the Gaussian assumption is relaxed. Such investigations would help quantify the robustness of Knockoff+ and its applicability beyond idealized Gaussian settings.

5.1 Summary Table

Table 1 reports all aggregated numerical values.

References

- Barber, R. F. and Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5):2055–2085.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300.
- Candès, E. J., Fan, Y., Janson, L., and Jordan, M. I. (2018). Panning for gold: Model-x knockoffs for high-dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B*, 80(3):551–577.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.
- Sesia, M. and Candès, E. (2022). *knockoff: The Knockoff Filter for Controlled Variable Selection*. R package version 0.3.0.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1):267–288.

Table 1: Summary of empirical FDR, TPR, and number of selected variables across 2 models.

Model	Method	q	Mean FDR	Mean TPR	Mean Selected
Gaussian Linear Model (High p)					
gaussian	BH	0.05	0.043	0.305	12.8
gaussian	BH	0.10	0.130	0.355	16.7
gaussian	BH	0.20	0.288	0.508	28.9
gaussian	Knockoff+	0.05	0.041	0.700	29.5
gaussian	Knockoff+	0.10	0.108	0.880	40.0
gaussian	Knockoff+	0.20	0.135	0.918	42.7
Gaussian Linear Model (Low p)					
gaussian_lowdim	BH	0.05	0.126	0.678	31.3
gaussian_lowdim	BH	0.10	0.224	0.760	39.4
gaussian_lowdim	BH	0.20	0.289	0.838	47.5
gaussian_lowdim	Knockoff+	0.05	0.048	0.983	41.4
gaussian_lowdim	Knockoff+	0.10	0.113	0.995	45.0
gaussian_lowdim	Knockoff+	0.20	0.234	0.998	52.9

Team Member Contribution (not counted toward page limit)

Luhan Tang. Led the development of the complete Model-X knockoff workflow for the high-dimensional ($p > n$) setting. This included implementing the original Gaussian second-order construction of Candès et al. (2018), coding the full knockoff-generation pipeline (joint Gaussian construction, conditional sampling, and Lasso-based feature statistics), and building the initial end-to-end simulation framework.

Diagnosed and systematically analyzed the numerical instabilities in the author-provided implementation (SDP-based eigenvalue routines), particularly under high dimensionality. After consulting with the instructor, designed and implemented a stable alternative based on the CRAN **knockoff** package, which uses shrinkage covariance estimation. This modification enabled all large-scale simulations to run successfully across multiple FDR levels and repetitions.

Designed and executed the full simulation study for the high-dimensional Gaussian model, implemented the BH baseline, generated all corresponding figures, and wrote the methodology and data-analysis sections for this regime. Contributed extensively to the overall structure, writing, and integration of the final report, including code organization, reproducibility, and GitHub repository management.

Shengming Cheng. Contributed to the implementation of the Gaussian linear model in the low-dimensional ($n > p$) regime using the CRAN **knockoff** package. Designed and ran simulations for this setting, produced numerical summaries and plots, and carried out comparisons with the BH baseline.

Drafted the introduction and the Gaussian-methodology subsection for the low-dimensional regime. Assisted with debugging, cross-validating numerical outputs, and ensuring consistency

between the $p > n$ and $n > p$ analyses.