RESEARCH STATEMENT

Zhigen Zhao

Data science research is evolving rapidly alongside the advancement of AI and the increasing complexity of data. Traditional statistical approaches, often grounded in the central limit theorem (CLT) and strong model assumptions, are increasingly insufficient to address the intricacies of today's Big Data landscape. The proliferation of AI-driven technologies has led to high-dimensional datasets that capture a growing number of features per subject. The "3Vs" of Big Data—volume, velocity, and variety—pose unique challenges to both statistical and machine learning methodologies:

Volume: Classical asymptotic theories, such as consistency and limiting distributions, often fail to scale effectively to the massive datasets routinely processed by modern AI systems.

Velocity: The demand for real-time or near-real-time insights necessitates fast and scalable algorithms.

Variety: Data heterogeneity from sources such as structured datasets, unstructured images, and text introduces challenges in integration and modeling.

These issues are especially pressing in AI-driven domains such as genomics and precision medicine, brain connectivity analysis, consumer behavior modeling, and beyond. The integration of AI with statistical frameworks offers promising pathways to develop robust, scalable, and interpretable methodologies. Despite extensive progress in statistical methodology over the past two decades, several fundamental challenges persist:

- **Dimension Reduction**: How can dimensionality be reduced effectively in the presence of numerous predictors without losing valuable information?
- **Sparse and Weak Signals**: How do we identify significant variables when signals are weak and rare, while controlling error rates such as the false discovery rate (FDR)?
- **Beyond Linearity**: How can we construct methods that operate beyond the assumption of linearity and avoid reliance on rigid model specifications?
- **Post-Selection Inference**: After variable selection, how do we account for the selection process itself in downstream inference?

Addressing these challenges has the potential to significantly impact both scientific discovery and technological innovation. My research has focused on theoretical and methodological development to tackle these problems, with contributions organized into three core areas:

Statistical Inference Based on Model-Lean Assumptions

My representative works in this category include Lin, Zhao, and Liu (2018, 2019); Xing, Zhao, and Liu (2021); Zhao and Xing (2024) and Liang, Liu, Wang, and Zhao (2024). We established the phase transition threshold for sliced inverse regression (SIR), demonstrating that consistency is achieved if and only if the ratio $p/n \to 0$, and introduced Lasso-SIR to address scenarios where $p \gg n$. We further proposed the Gaussian Mirror framework for controlling FDR in high-dimensional settings and extended this to model-free settings via integration with SIR.

In another line of work, we developed adaptive, model-free tests for independence and exchangeability using a binary expansion framework (Zhang et al. (2021)). This led to the BREVITY framework, which represents exchangeable distributions in terms of their binary digits, offering an equivalence condition for exchangeability and enabling novel testing procedures.

Statistical Inference for Heterogeneity

Heterogeneity is a central concern in data analysis. My research has addressed this through multiple lines of work. Liu, Sarkar, and Zhao (2016), and Zhao et al. (2024) proposed FDR-controlling procedures that incorporate group structure. Other work, such as Hwang et al. (2009) and Kwon and Zhao (2023), addressed the challenge of inference under unknown and unequal variances, using empirical Bayes methods to jointly shrink means and variances.

Statistical Inference Post Selection

Modern data analysis often involves a variable selection stage prior to model fitting. My work (e.g., Zhao and Hwang (2012); Hwang and Zhao (2013); Zhao and Sarkar (2015) has developed empirical Bayes confidence intervals for selected parameters, accounting for the bias introduced by selection. In Kwon and Zhao (2023), we proposed a nonparametric empirical Bayes variance estimator, relying on the empirical distribution of sample variances, and demonstrated its superiority in selected settings. In Zhao et al. (2024), I was invited to contribute a discussion on "data fission" (Leiner et al. (2023)), where I highlighted its novel connection to empirical Bayes methodology.

In summary, my research is dedicated to developing practical and robust solutions to the complex challenges posed by big data analysis. These efforts not only advance statistical theory and methodology but also have tangible applications across a wide range of scientific and technological fields. In the following sections of my research statement, I will detail my past, ongoing, and future research endeavors across the following thematic areas: Section 1 will summarize my funded research and funding plans; Section 2 highlights my contributions to sufficient dimension reduction; Section 3 explores my work in multiple hypothesis testing; Section 4 focuses on my investigations in Bayes and empirical Bayes inference; and Section 5 examines model-free inference on independence and exchangeability.

In recognition of my contributions to the field, I was recently invited to serve as a **Co-Editor-in-Chief** of *Communications in Statistics: Theory and Methods (CSTM)*, a well-established international journal that receives over 1,000 submissions annually and covers all areas of statistical research. I am honored to have been offered this role and have officially accepted the appointment. This editorial position reflects the visibility of my research and the confidence the broader statistical community has placed in my judgment. The journal's wide scope and high submission volume offer a unique opportunity to engage with emerging trends across diverse subfields and to help guide the development and dissemination of rigorous, innovative, and impactful work. This leadership role underscores both my commitment to advancing the discipline and the global reach of my scholarly contributions. As a Co-Editor of CSTM, I have initiated a special issue entitled "Celebrating 30 Years of False Discovery Rate: Insights, Methods, and Frontiers," which has already attracted considerable interest from renowned statistical researchers. To further strengthen the journal's reach and influence, I have also initiated the creation of a new section, Editorial Invitation Articles, which will feature invited contributions from leading experts to highlight emerging research directions and promote the visibility of the journal. These initiatives underscore my commitment to advancing scholarship while positioning Communications in Statistics as a premier venue for cutting-edge statistical research.

1 Funding

My research has been supported by several NSF grants for which I served as Principal Investigator, including *DMS-1208735* (2012–2015, \$174,976), *IIS-1633283* (2016–2020 \$250,000), and *DMS-2311216* (2023–2026, \$197,007).

The NSF grant *DMS-1208735*, titled "Bayesian Decision Theoretic Methods for Some High-Dimensional Multiple Inference Problems," focused on developing Bayesian methods for high-dimensional data with sparse signals. It addressed two key goals: (i) constructing multiple testing procedures that balance type II errors and false discoveries, and (ii) constructing confidence intervals for selected parameters using zero-inflated mixture priors.

The NSF grant *IIS-1633283* tackled Big Data challenges by overcoming the curse of dimensionality, where the number of variables (p) grows exponentially relative to the number of observations (n). This project introduced a novel packing-based framework and addressed three problems: (i) spurious correlation theory, (ii) detection of low-rank structures, and (iii) optimal testing for rare and weak signals. The project also contributed to developing scalable methods and student training.

The most recent grant, *DMS-2311216*, targets multiple testing under dependent data. It focuses on two aims: (i) developing a model-free FDR method for general multiple-index models and (ii) exploring optimal testing procedures in regression models with rare and weak signals.

2 Model-free Dimension Reduction

In the realm of regression analysis, our primary goal is to understand the relationship between the response variable Y and a set of p predictors denoted as $\boldsymbol{X}$. Specifically, we aim to uncover insights about $Y|\boldsymbol{X}$, representing the conditional distribution of Y given $\boldsymbol{X}$. Within the extensive literature on sufficient dimension reduction (SDR), the focus centers on identifying a minimal subspace $\mathcal{S}_{Y|\boldsymbol{X}}$ within the column space of $\boldsymbol{X}$, known as the central subspace such that

$$Y \perp\!\!\!\perp \boldsymbol{X} | \mathbf{P}_{\mathcal{S}} \boldsymbol{X}. \tag{1}$$

A prevalent method to estimate the central space $\mathcal{S}$ is sliced inverse regression (SIR), proposed in Li (1991). From a series of works, we have filled the theoretical gap of SIR under various theoretical and methodological settings.

In Lin, Zhao, and Liu (2018), we bridge a significant gap in the literature concerning sufficient dimension reduction (SDR) by establishing a crucial phase transition. Specifically, we demonstrate that the sliced inverse regression (SIR) method achieves consistency if and only if the ratio $\rho = \frac{p}{n} \to 0$. When ρ does not tend to zero, achieving consistent estimation necessitates imposing a sparsity condition. This result serves as a foundational theoretical framework for various regularized methods within the realm of sufficient dimension reduction. In scenarios where the number of predictors p is comparable to or exceeds the sample size n, we introduced the Lasso-SIR method, as detailed in Lin, Zhao, and Liu (2019). This innovative approach utilizes SIR within a Lasso regression framework to obtain an estimate of the sufficient dimension reduction (SDR) space.

Notably, Lasso-SIR has demonstrated consistency and attains the optimal convergence rate under specific sparsity conditions, particularly when p is of the order $o(n^2\lambda^2)$. These two studies provide significant and foundational contributions to the theory of sufficient dimension reduction and have collectively been cited in over 230 publications. In addition, I developed the R package *Lasso-SIR*, which has been downloaded more than 26,000 times.

Furthermore, in Lin, Zhao, and Liu (2021), we make significant contributions to the detection boundary of the single index model, represented as $y = f(\boldsymbol{\beta X}, \epsilon)$. In this context, traditional measures of signal strength, such as the norm of the parameter vector, become non-identifiable and no longer applicable. To address this challenge, we introduce the concept of the "generalized signal-to-noise ratio (gSNR)" denoted as λ. We establish the unique non-zero eigenvalue of $\mathrm{var}[(\mathbb{E}(\boldsymbol{x}|y)]$ as λ and present pioneering results on the detection boundary of the single index model, framed in terms of λ.

Despite the substantial body of research focusing on the consistency of sliced inverse regression (SIR), there remains a notable gap in the literature regarding the limiting distribution of SIR when the number of predictors p grows with the sample size n. Existing distributional results typically assume a fixed value for p or a fixed number of signals. In a recent manuscript (Zhao and Xing, 2024), leveraging recent advancements in high-dimensional Gaussian approximation theory, we have made significant strides in deriving the limiting distribution of SIR while allowing for p to diverge to infinity. This breakthrough result is agnostic to conditions related to the sparsity of "uniform signal strength," thereby opening new avenues for statistical inference in the context of SIR. These theoretical development paves the road for model free multiple testing methods for matrix valued tensor structured data (Yan, Zhang, and Zhao, 2025).

3 High Dimensional Inference via Multiple Comparison

The field of multiple hypothesis testing plays a pivotal role in contemporary scientific research, particularly when dealing with a vast number of hypotheses of interest. Over the past three decades, there has been a significant surge in research dedicated to this domain. Since becoming a part of Temple University, I have actively contributed to this field, developing a series of critical methods aimed at addressing fundamental challenges within it.

In a sole-authored paper (Zhao, 2022), I conducted a theoretical investigation comparing the strengths and weaknesses of two types of testing methods: the p-value based approach and the local false discovery rate (FDR) based method. In collaboration with colleagues He, Sarkar, and Zhao (2015), we explored the development of an optimal multiple testing procedure that incorporates a severity function to account for unequal penalties associated with type II errors. When dealing with hypotheses organized into grouping structures, we introduced an optimal method in Liu, Sarkar, and Zhao (2016) designed to control both overall and within-group FDR. This contribution garnered significant attention, as evidenced by its inclusion in the list of the "**most downloaded papers in the last 90 days**" on the journal's website. The corresponding R-package *GroupTest* has been downloaded more than 33,000 times.

Our research journey continued in Sarkar and Zhao (2022), where we presented an innovative framework for multiple testing of hypotheses grouped in a one-way classification format. This approach utilized hypothesis-specific local FDRs, effectively capturing the dependence structure arising from the grouping of hypotheses. Notably, our work in this area has earned recognition. In a discussion paper featured in JRSSB (Cai et al., 2019), Dr. Tony Cai and his collaborators acknowledged the significance of our contributions by commenting that "important recent progress for exploiting the grouping and hierarchical structures among hypotheses under more generic settings has been made in Liu, Sarkar, and Zhao (2016) ... The logical correlation (2.2) can be conceptualized as a hierarchical constraint and exploited more efficiently (Sarkar and Zhao, 2022)."

Nowadays, it is important to control false discovery rate for high-dimensional regression models. In Ji and Zhao (2014), we tackled the problem of optimal multiple testing procedures in high-dimensional regression models where signals are so weak and rare that achieving "selection consistency" is unattainable. This work represents the pioneering achievement in establishing the rate-optimality of testing procedures for high-dimensional regression models. Its significance is underscored by the recognition it received in an article by Emmanuel J. Candès (Su and Candès, 2016), who noted that "Having said this, we are aware of recent significant progress on this problem, including the development of the knockoff filter and other innovative ideas [31, 38]," with reference [38] pointing to this particular work. In Xing, Zhao, and Liu (2021), we introduced the Gaussian Mirror method, which generates pairs of mirror variables for each predictor by adding and subtracting Gaussian perturbations. These mirror variables naturally yield test statistics symmetric around zero under the null hypothesis, enabling us to estimate false discovery proportions and control the FDR. This work has been extended in various directions. Since its publication in 2023, this paper has received over 60 citations from researchers in the field.

In Zhao and Xing (2024), we investigated the model-free multiple testing problem within the context of sufficient dimension reduction, where we aim to test the relevance of each covariate to the central subspace $\mathcal{S}$ without prior knowledge of the link function. We estimated β_i using Sliced Inverse Regression (SIR) and constructed Angular Balanced Statistics (ABS). Leveraging the limiting distribution of SIR, as discussed in Section 2, we demonstrated that these statistics exhibit symmetry around zero under the null hypothesis, leading to the proposal of a model-free multiple testing approach using ABS. In Yu, Zu, and Zhao (2024); Sanni, Yu, and Zhao (2024), we introduced, for the first time in the literature, a multiple testing procedure to control the False Discovery Rate (FDR) in high-dimensional quantile regression analysis. In Yan et al. (2025), we developed model-free multiple testing method for matrix-valued tensor data.

4 Post Selection Inferece

In my dissertation, I focused on constructing empirical Bayes confidence intervals for multiple parameters, particularly those selected under the assumption of unequal and unknown variances. Building on this foundation, I have continued to explore and deepen this line of research during my tenure at Temple University, resulting in a series of published articles. Below, I summarize my key contributions in this area.

In a series of works (Hwang, Qiu, and Zhao, 2009; Zhao and Hwang, 2012; Hwang and Zhao, 2013), we have examined empirical Bayes confidence intervals for means, particularly when assuming unequal and unknown variances. In the latter two papers, our focus shifts to parameters post-selection. While Bayesian methods are generally robust to selection bias, the estimation error introduced by estimating the prior can significantly impact the final inference, especially for the selected parameters. In these papers, we conducted a comprehensive investigation into second-order corrections to account for such estimation errors, ensuring the attainment of desired coverage probabilities for the selected parameters. In Zhao (2010), I explored the concept of double shrinkage estimators for mean parameters. Interestingly, this methodology has been applied in Dharmarathne et al. (2023) to aggregate expert judgments on uncertain quantities, providing reliable weights for experts' opinions. This represents a novel application of my work that I had not originally anticipated. In Zhao (2024), I was invited to write a discussion on the concept of data fission introduced in Leiner et al. (2023), where I illustrated its intrinsic connection to empirical Bayes methods in a novel way. We will continue exploring this direction further in collaboration with my current student.

In all these aforementioned works, we relied on a parametric assumption for the prior distribution of the variances. However, in an article with my PhD student (Kwon and Zhao, 2023), we took a different approach and introduced an arbitrary prior for σ_i^2 as $\sigma_i^2 \sim g(\sigma_i^2)$. From this non-parametric prior, we derived a Bayesian estimator of the variance as

$$\widehat{\sigma}_{i,B}^2 = \frac{k}{2}\left\{ \frac{\int_{s_i^2}^{\infty} (s^2)^{-(\frac{k}{2}-2)}\, dF(s^2)}{\int_{s_i^2}^{\infty} (s^2)^{-(\frac{k}{2}-1)}\, dF(s^2)} - s_i^2 \right\}. \tag{2}$$

A distinctive feature of this estimator is its reliance solely on $F(s^2)$, the cumulative distribution function of the sample variances. We refer to it as an "F-modeling" based estimator, which contrasts with the "f-modeling" based estimator that depends on the data through its marginal probability density function. To enhance its applicability, we developed a data-driven version by replacing the cumulative distribution function with its empirical counterpart. This data-driven approach has demonstrated superior performance, particularly when applied to parameters selected after the selection process.

5 Model-free Testing of Independence and Exchangebility

In this line of research, we focused on developing adaptive procedures to test independence and exchangeability—problems that have been studied for over a century. Mathematically, the characteristic function is one of the fundamental tools for investigating random variables. Practically, binary digits form the foundation of modern computation. Within this framework, we bridge these two foundational concepts to develop **model-free** inferential methods for testing independence and exchangeability.

Let $\mathbf{U} = (U_1, U_2, \ldots, U_p)$ be a p-dimensional continuous vector within the range $[-1, 1]^p$. According to Lemma 2.1 of Zhang, Zhao, and Zhou (2021), there exists a sequence of random variables $\{A_{j,d}\}$, where $j = 1, 2, \ldots, p$ and $d = 1, 2, \ldots, D$, which only take values -1 and 1, such that

$$\max_{1\le j\le p} |U_j - U_{j,D}| \to 0 \text{ uniformly as } D \to \infty,$$

where $U_{j,D} = \sum_{d=1}^{D} \frac{A_{j,d}}{2^d}$. The variables $\{A_{j,d}\}$ are referred as bits. Let $\mathbf{U}_D = (U_{1,D}, U_{2,D}, \ldots, U_{p,D})$ be the binary expansion of $\mathbf{U}$ upto a depth of D. For a given depth D, let Λ be a $p \times D$ matrix with each entry being either 0 or 1. These matrices are referred to as binary interactions. Let $\mathbf{B}^{p\times D}$ be the set which consists of all possible binary interactions including Λ_0, a unique binary interaction in which all entries are zero.

For any binary interaction $\Lambda \in \mathbf{B}^{p\times D}$, define the interaction of bits as

$$A_\Lambda = \prod_{j=1}^{p}\prod_{d=1}^{D} (A_{j,d})^{\lambda_{jd}}, \tag{3}$$

where λ_{jd} is the entry of Λ_j at the j-th row and d-th column. For any $\mathbf{t} \in \mathcal{R}^p$ and $\Lambda \in \mathbb{B}^{p\times D}$, define

$$\Psi_\Lambda(t) = \prod_{j=1}^{p}\prod_{d=1}^{D}\left\{\cos\left(\frac{t_j}{2^d}\right)\right\}^{1-\lambda_{jd}}\left\{i\sin\left(\frac{t_j}{2^d}\right)\right\}^{\lambda_{jd}}.$$

We then have the following theorem

Theorem 5.1 (Binary Expansion Approximation of Uniformity, BEAUTY)**.** *Let* $\mathbf{U}$ *be a* p*-dimensional random vector such that* $U_j \in [-1,1], \forall j$*. Let* $\phi_{\mathbf{U}}(\mathbf{t})$ *be the characteristic function of* $\mathbf{U}$ *for any* $\mathbf{t} = (t_1,\ldots,t_p)^T \in \mathbb{R}^p$*. We have*

$$e^{i\mathbf{t}^T\mathbf{U}_D} = \sum\nolimits_{\Lambda\in\mathbf{B}^{p\times D}} A_\Lambda \Psi_\Lambda(\mathbf{t}) \tag{4}$$

and

$$\phi_{\mathbf{U}}(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^T\mathbf{U})] = \lim_{D\to\infty}\sum\nolimits_{\Lambda\in\mathbf{B}^{p\times D}} \Psi_\Lambda(\mathbf{t})\mathbb{E}[A_\Lambda], \tag{5}$$

where $\Psi_\Lambda(\mathbf{t}) = \prod_{j=1}^{p}\prod_{d=1}^{D}\{\cos(t_j/2^d)\}^{1-\Lambda_{jd}}\{i\sin(t_j/2^d)\}^{\Lambda_{jd}}$*.*

In the set of all binary interactions $\mathbb{B}^{p\times D}$, we define an equivalent relation $\sim$. For any $\Lambda^1, \Lambda^2 \in \mathbb{B}^{p\times D}$, we say that Λ^1 is equivalent to Λ^2, denoted as $\Lambda^1 \sim \Lambda^2$, if Λ^1 can be obtained from Λ^2 through row permutations. Specifically, there exists a $\sigma \in \mathrm{Sym}(M)$ such that

$$\Lambda^1_{j,d} = \Lambda^2_{\sigma(j),d}, \quad d = 1,2,\ldots,D.$$

For any $\Lambda \in \mathbb{B}^{p\times D}$, let $[\Lambda] = \{\Lambda' \in \mathbb{B}^{p\times D}, \Lambda' \sim \Lambda\}$ be the equivalent class of Λ which consists of all the binary interactions which are equivalent to Λ. Let $\mathbb{B}^{p\times D}/\sim$ be the quotient space. We then have the following theorem.

Theorem 5.2. *If* $\mathbf{U_D}$ *is exchangeable, then for any* $\Lambda^1 \sim \Lambda^2 \in \mathbb{B}^{p\times D}$*,*

$$\mathbb{E}[A_{\Lambda^1}] = \mathbb{E}[A_{\Lambda^2}]. \tag{6}$$

The converse is also true.

Based on this result, we then have the following Binary REpresentation of Variables with exchangeabilITY (BREVITY).

Theorem 5.3. *Let* $\mathbf{U_D}$ *be the binary expansion up to the depth* D *for an exchangeable random vector* $\mathbf{U}$*, then*

$$\Phi_{\mathbf{U}}(\mathbf{t}) = \lim_{D\to\infty}\Phi_{\mathbf{U_D}}(\mathbf{t}) = \lim_{D\to\infty}\sum_{[\Lambda]\in\mathbb{B}^{p\times D}/\sim}\mathbb{E}(A_\Lambda)\sum_{\Lambda'\sim\Lambda}\Psi_{\Lambda'}(\mathbf{t}).$$

Note that Theorems 5.1 to 5.3 make no distributional assumptions on $\mathbf{U}$, paving the way for **model-free** statistical inference on independence and exchangeability using the binary interactions A_Λ (Zhang, Zhao, and Zhou (2021); Zhao, Tian, Zhang, and Zhou (2024)). This framework can be further extended to flexibly study other distributional properties through binary expansion.

6 Summary

I derive great satisfaction from leveraging insights gained through the analysis of novel data types to deepen our understanding of traditional challenges in statistical inference. My research is driven by a passion for seamlessly integrating theory and methodology to tackle fundamental issues such as heterogeneity and dependence. Throughout my career, I have had the privilege of addressing theoretical problems that have led to the development of practical methods with significant real-world applications.

References

Cai, T., W. Sun, and W. Wang (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. Journal of the Royal Statistical Society Series B: Statistical Methodology 81(2), 187–234.

Dharmarathne, G., G. F. Nane, A. Robinson, and A. M. Hanea (2023). Shrinking the variance in experts' "classical" weights used in expert judgment aggregation. Forecasting 5(3), 522–535.

He, L., S. K. Sarkar, and Z. Zhao (2015). Capturing the severity of type II errors in high-dimensional multiple testing. Journal of Multivariate Analysis 142, 106–116.

Hwang, J. T., J. Qiu, and Z. Zhao (2009). Empirical Bayes confidence intervals shrinking both means and variances. Journal of the Royal Statistical Society. Series B 71(1), 265–285.

Hwang, J. T. and Z. Zhao (2013). Empirical Bayes confidence intervals for selected parameters in high dimension with application to microarray data analysis. Journal of the American Statistical Association 108(502), 607–618.

Ji, P. and Z. Zhao (2014). Rate optimal multiple testing procedure in high-dimensional regression.

Kwon, Y. and Z. Zhao (2023). On f-modelling-based empirical bayes estimation of variances. Biometrika 110(1), 69–81.

Leiner, J., B. Duan, L. Wasserman, and A. Ramdas (2023). Data fission: splitting a single data point. Journal of the American Statistical Association, 1–12.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. Journal of the American Statistical Association 86(414), 316–327.

Liang, Z., J. Liu, X. Wang, and Z. Zhao (2024). Bayesian analysis of multiple index additive models. Submitted.

Lin, Q., Z. Zhao, and J. S. Liu (2018). On consistency and sparsity for sliced inverse regression in high dimensions. The Annals of Statistics 46(2), 580 – 610.

Lin, Q., Z. Zhao, and J. S. Liu (2019). Sparse sliced inverse regression via lasso. Journal of the American Statistical Association 114(528), 1726–1739.

Lin, Q., Z. Zhao, and J. S. Liu (2021). Testing model utility for single index models under high dimension. Festschrift in Honor of R. Dennis Cook: Fifty Years of Contribution to Statistical Science, 65.

Liu, Y., S. K. Sarkar, and Z. Zhao (2016). A new approach to multiple testing of grouped hypotheses. Journal of Statistical Planning and Inference 179, 1–14.

Sanni, S., Y. Yu, and Z. Zhao (2024). Ultra-high dimensional threshold selection for quantile mirror feature screening with fdr error rate control: A case study on high blood pressure analysis. Submitted.

Sarkar, S. K. and Z. Zhao (2022). Local false discovery rate based methods for multiple testing of one-way classified hypotheses. Electronic Journal of Statistics 16(2), 6043–6085.

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

Xing, X., Z. Zhao, and J. S. Liu (2021). Controlling false discovery rate using gaussian mirrors. Journal of the American Statistical Association, 1–20.

Yan, L., X. Zhang, and Z. Zhao (2025). Model-free multiple testing for matrix-valued predictors with false discovery control.

Yu, Y., T. Zu, and Z. Zhao (2024). Fdr control for high dimensional quantile variable selection. Submitted.

Zhang, K., Z. Zhao, and W. Zhou (2021). Beauty powered beast. arXiv preprint arXiv:2103.00674.

Zhao, Z. (2010). Double shrinkage empirical Bayesian estimation for unknown and unequal variances. Statistics and Its Interface 3, 533–541.

Zhao, Z. (2022). Where to find needles in a haystack? TEST 31(1), 148–174.

Zhao, Z. (2024). Discussion on "data fission: Splitting a single data point". Journal of the American Statistical Association. To appear.

Zhao, Z. and J. T. Hwang (2012). Empirical Bayes false coverate rate controlling confidence interval. Journal of the Royal Statistical Society. Series B 74(5), 871–891.

Zhao, Z. and S. K. Sarkar (2015). A Bayesian approach to constructing multiple confidence intervals of selected parameters with sparse signals. Statistica Sinica, 725–741.

Zhao, Z., Y. Tian, K. Zhang, and W. Zhou (2024). Test of exchangeability under binary expansion. Submitted.

Zhao, Z. and X. Xing (2024). On the testing of multiple hypothesis in sliced inverse regression. Electronic journal of statistcs. Revision submitted.