*The Annals of Statistics*
2023, Vol. 51, No. 6, 2318–2341
https://doi.org/10.1214/23-AOS2318


# CARVING MODEL-FREE INFERENCE

BY SNIGDHA PANIGRAHI[a]

*Department of Statistics, University of Michigan,* [a]*psnigdha@umich.edu*

Complex studies involve many steps. Selecting promising findings based on pilot data is a first step. As more observations are collected, the investigator must decide how to combine the new data with the pilot data to construct valid selective inference. Carving, introduced by Fithian, Sun and Taylor (2014), enables the reuse of pilot data during selective inference and accounts for overoptimism from the selection process. However, currently, carving is only justified for parametric models such as the commonly used Gaussian model. In this paper, we develop the asymptotic theory to substantiate the use of carving beyond Gaussian models. Our results indicate that carving produces valid and tight confidence intervals within a model-free setting, as demonstrated through simulated and real instances.

**1. Introduction.** Conducting inference for a selected set of findings, also known as selective inference, is a common problem in complex studies. Usually, the investigator starts with pilot data to select a set of promising findings. As additional observations are collected, the investigator faces the question of how to augment the new data with the existing pilot data for drawing valid selective inference. On the one hand, a direct augmentation of the two data sets ignores overoptimism from the selection process. For example, a recent article by [1] highlights replicability concerns with standard inferential methods that do not account for the selection process. On the other hand, valid selective inference, which relies only on the new data, fails to utilize observations from the pilot data. This practice is popularly known as data splitting.

*Carving*, introduced by [5], is an efficient alternative to data splitting. It permits the reuse of pilot data by basing valid selective inference on a conditional distribution. This distribution accounts for overoptimism from the selection process by conditioning on the selection outcome seen in the pilot data. Previous work by [9, 24, 26] gives a recipe to construct pivots from such conditional distributions. Applying the same recipe yields us a pivot for carving, which we call a *carved pivot*.

When data is generated by a Gaussian model, the carved pivot provides exactly-valid selective inference. However, what happens when we drift away from Gaussian data? In model-free settings, can we still use the carved pivot for drawing asymptotically-valid selective inference? Moreover, can we trust selective inference when rare selection outcomes are observed in our pilot data?

This paper demonstrates that a carved pivot produces asymptotically-valid selective inference, even if our data is not from a Gaussian model. Our theory suggests that this is true for a wide range of distributions, and that selective inference using a carved pivot remains valid even for rare selection outcomes.

1.1. *Notation.* We list some basic notation for our paper. For $d \in \mathbb{N}$, let $[d] = \{1, 2, \ldots, d\}$. Let $|E|$ be the cardinality of set $E$ and let $E^c$ be its complement set. Let $V^{(j)}$ be

Received March 2022; revised July 2023.
*MSC2020 subject classifications.* Primary 62E20; secondary 62G15, 62G20.
*Key words and phrases.* Carving, conditional inference, model-free, post-selection inference, randomization, selective inference.

the $j$th component of the vector $V \in \mathbb{R}^d$. Let $V^{(-j)}$ be the subvector of $V$ after we exclude the $j$th component of the original vector and let $V^{(E)}$ be the subvector of $V$ that collects the components in $E \subseteq [d]$. The symbol $V'$ denotes the transpose of the vector $V$. Unless mentioned otherwise, $\|V\|$ is understood as the $\ell_2$-norm of $V$. $\mathcal{P}_E$ denotes a permutation matrix: $\mathcal{P}_E V$ reorders the components of $V$ and returns the vector $(V^{(E)'} V^{(E^c)'})'$. For a positive definite matrix $M$, let $M^{1/2}$ be its principal square root. For any matrix $M \in \mathbb{R}^{d_1 \times d_2}$, $E_1 \subseteq [d_1]$, $E_2 \subseteq [d_2]$, let $M_{E_1,E_2}$ be the submatrix of $M$ that contains rows and columns in the sets $E_1$ and $E_2$, respectively. Also, let $M_{E_1}$ be the submatrix of $M$, which collects its columns in the set $E_1$. We use the notation $I_{d,d}$ and $0_{d_1,d_2}$ for the identity matrix with $d$ rows and columns and the matrix of all zeros with $d_1$ rows and $d_2$ columns, respectively. We use the notation $0_d$ and $1_d$ to denote a vector with all $d$ components equal to zero and a vector with all $d$ components equal to one, respectively. For a positive semidefinite matrix $\Sigma \in \mathbb{R}^{d \times d}$ and $x \in \mathbb{R}^d$, let $\operatorname{Exp}(x, \Sigma) = \exp(-\frac{1}{2} x' \Sigma x)$. Let $\mathbf{1}_{\mathcal{E}}$ represent the indicator function, where $\mathcal{E}$ is a fixed set. At last, let $\Phi$ be the CDF of a standard Gaussian distribution with the density function $\phi$, and let $\bar{\Phi}(x) = 1 - \Phi(x)$ be the corresponding survival function at $x$.

1.2. *Organization.* In Section 2, we present a carved pivot that ensures exactly-valid selective inference with Gaussian data. We introduce a running example in this section that helps us develop the main ideas behind the asymptotic theory. We demonstrate in Section 3 that asymptotically-valid selective inference is dependent on the convergence of specific relative differences. In Section 4, we build the asymptotic theory for $\mathbb{R}^d$-valued random variables with the identity covariance matrix. We then extend the asymptotic theory to variables with a general covariance matrix in Section 5. We study the empirical behavior of the carved pivot on both synthetic and real data in Section 6. Lastly, we conclude our paper with a brief discussion in Section 7. Proofs and supporting results are collected in the Supplementary Material [13].

## 2. Exactly-valid selective inference with carving.

2.1. *Our running example.* Suppose that we observe a triangular array of independent and identically distributed $\mathbb{R}^d$-valued observations

$$\zeta_{i,n} = \begin{pmatrix} \zeta_{i,n}^{(1)} & \zeta_{i,n}^{(2)} & \cdots & \zeta_{i,n}^{(d)} \end{pmatrix}' \overset{\text{i.i.d.}}{\sim} \mathbb{P}_n \quad \text{for } i \in [n]. \tag{2.1}$$

Let

$$\beta_n = \mathbb{E}_{\mathbb{P}_n}[\zeta_{1,n}] \in \mathbb{R}^d$$

be the unknown mean parameter. Let

$$\Sigma = \mathbb{E}_{\mathbb{P}_n}\big[(\zeta_{1,n} - \beta_n)(\zeta_{1,n} - \beta_n)'\big]$$

be the $d \times d$ covariance matrix, which we assume is fixed and invertible. Define

$$V_n = \sqrt{n}\bar{\zeta}_n,$$

where $\bar{\zeta}_n = \frac{1}{n}\sum_{i=1}^n \zeta_{i,n}$. Additionally, let

$$\Sigma_{-j,j} = \operatorname{Cov}_{\mathbb{P}_n}\big(V_n^{(-j)}, V_n^{(j)}\big), \qquad \sigma_j^2 = \operatorname{Var}_{\mathbb{P}_n}\big(V_n^{(j)}\big), \quad \text{for } j \in [d].$$

Throughout, we will assume that the distribution of $V_n$ admits a Lebesgue density.

For a fixed constant $\rho \in (0, 1)$, we consider a Gaussian variable

$$W_n \sim \mathcal{N}\big(0_d, \rho^2 \Sigma\big),$$

which is independent of $V_n$. Then, using $V_n$ and $W_n$, we infer for $\beta_n^{(j)}$, the $j$th component of $\beta_n$, only if

$$V_n^{(j)} + W_n^{(j)} > \Lambda^{(j)}, \tag{2.2}$$

where $\Lambda$ is a fixed vector in $\mathbb{R}^d$. Borrowing the term *randomization* from [24], we call $W_n$ a *randomization variable*. The rule used for selection, in (2.2), is called a *randomized selection* rule. As shown afterwards, there is an asymptotic correspondence between (2.2) and a similar selection on pilot data. We use the symbol $E_n$ to represent the indices of our selected means. Let

$$E_{\text{obs}} \subseteq [d]$$

be the observed value of the random variable $E_n$. For brevity sake, let $|E_{\text{obs}}| = p$.

2.2. *Exactly-valid selective inference with Gaussian data.* Suppose that our data is drawn from

$$\mathbb{P}_n = \mathcal{N}(\beta_n, \Sigma).$$

In this case, $V_n$ is distributed as Gaussian variable with mean vector $\sqrt{n}\beta_n$ and covariance matrix $\Sigma$. Consider $j \in E_{\text{obs}}$. We obtain a conditional distribution for $V_n^{(j)}$, which accounts for the selection process by conditioning on a subset of the selection outcome

$$\{E_n = E_{\text{obs}}\}.$$

We present a pivot for for $\beta_n^{(j)}$ based on this conditional distribution.

First, we introduce some more statistics. Define

$$A_n = V_n^{(E_n^c)} + W_n^{(E_n^c)}, \qquad U_n^{(j)} = V_n^{(-j)} - \frac{1}{\sigma_j^2}\Sigma_{-j,j} V_n^{(j)}.$$

Let $A_{\text{obs}}$ be the observed value of $A_n$. To draw valid selective inference, we construct a pivot by using the conditional distribution of $V_n^{(j)}$ when conditioned on

$$\{E_n = E_{\text{obs}}, A_n = A_{\text{obs}}\} \tag{2.3}$$

and the observed value of $U_n^{(j)}$.

Note that we condition on a subset of the selection outcome by further conditioning on some additional information $A_n$. By adding extra conditioning, the conditional distribution of $V_n^{(j)}$ becomes simpler since the outcome can be described as a set of straightforward sign constraints. Additionally, we condition on $U_n^{(j)}$ to eliminate all parameters except $\beta_n^{(j)}$.

Proposition 1 introduces this pivot for Gaussian data. To state this result, we consider the following matrices:

$$R^{(j)} = \mathcal{P}_{E_{\text{obs}}}\begin{bmatrix} 1 & 0 \\ \frac{1}{\sigma_j^2}\Sigma_{-j,j} & I_{d-1,d-1} \end{bmatrix}, \qquad Q = \begin{bmatrix} I_{p,p} \\ 0_{d-p,p} \end{bmatrix}, \qquad r = \begin{pmatrix} \Lambda^{(E_{\text{obs}})} \\ A_{\text{obs}} \end{pmatrix}.$$

Define $\mathrm{F} : \mathbb{R}^d \to \mathbb{R}$ as

$$\mathrm{F}(V) = \int \mathrm{Exp}\Big(Qt - V + r, \frac{1}{\rho^2}\Sigma^{-1}\Big) \cdot \mathbf{1}_{t \in \mathbb{R}^{p+}}\, dt,$$

and define

$$\mathrm{D}(U; \sqrt{n}\beta_n^{(j)}) = \int_{-\infty}^{\infty} \phi\Big(\frac{1}{\sigma_j}(v - \sqrt{n}\beta_n^{(j)})\Big) \cdot \mathrm{F}\left(R^{(j)} \begin{pmatrix} v & U' \end{pmatrix}'\right) dv.$$

PROPOSITION 1 (Pivot). *Let* $\text{Pivot}^{(j)}((V_n^{(j)}\ (U_n^{(j)})')')$ *be equal to*

$$(\mathrm{D}(U_n^{(j)}; \sqrt{n}\beta_n^{(j)}))^{-1} \cdot \int_{V_n^{(j)}}^{\infty} \phi\Big(\frac{1}{\sigma_j}(v - \sqrt{n}\beta_n^{(j)})\Big) \cdot \mathrm{F}\Big(R^{(j)}\Big(v \quad (U_n^{(j)})'\Big)'\Big)\, dv.$$

*Conditional on the outcome in* (2.3), $\text{Pivot}^{(j)}((V_n^{(j)}\ (U_n^{(j)})')')$ *is distributed as a* Unif(0, 1) *variable*.

The pivot in Proposition 1 applies broadly to several instances of selective inference. We provide more examples in Section 6, which includes inference after variable selection. In each instance, we construct a pivot with a similar representation as Proposition 1.

Before proceeding further, we turn to a special case when $\Sigma = I_{d,d}$. Note that the components of $\beta_n$ do not have any relationship with each other. We obtain a reduced form for our pivot in this special case.

To simplify further, we fix $\Lambda = 0_d$. Our pivot turns out to be a univariate function in the statistic $V_n^{(j)}$, that is, it is free of $U_n^{(j)}$.

COROLLARY 1 (Univariate pivot). *The pivot for* $\beta_n^{(j)}$ *in Proposition* 1 *simplifies as*

$$\text{Pivot}^{(j)}(V_n^{(j)}) = \frac{\int_{V_n^{(j)} - \sqrt{n}\beta_n^{(j)}}^{\infty} \phi(v) \cdot \bar{\Phi}(-\frac{1}{\rho}(v + \sqrt{n}\beta_n^{(j)}))\, dv}{\int_{-\infty}^{\infty} \phi(v) \cdot \bar{\Phi}(-\frac{1}{\rho}(v + \sqrt{n}\beta_n^{(j)}))\, dv}.$$

2.3. *Contributions and related developments.* Consider an array of observations from $\mathbb{P}_n$ in $\mathbb{R}^d$, as described earlier. Now, suppose a similar selection rule is applied only to a random subsample of size $n_1 (< n)$. This subsample plays the role of pilot data in our setup. Let $n_2 = n - n_1$ and let

$$\rho^2 = \frac{n_2}{n_1}$$

be the ratio of the number of observations in the new data to the number of observations in the pilot data. We infer for $\beta_n^{(j)}$ only if the corresponding $Z$-test statistic exceeds a threshold $\tau^{(j)}$ in the pilot data, that is,

$$V_{n_1}^{(j)} > \tau^{(j)}.$$

The selection rule on the subsample can be transformed into the randomized selection rule in (2.2), in an asymptotic sense. To see this, we define

$$W_n^{(j)} = \sqrt{1 + \rho^2} \cdot V_{n_1}^{(j)} - V_n^{(j)} \quad \text{for } j \in [d], \tag{2.4}$$

and also let

$$\Lambda^{(j)} = \sqrt{1 + \rho^2} \cdot \tau^{(j)}.$$

Asymptotically, $W_n$ is distributed as a Gaussian variable with mean $0_d$ and covariance $\rho^2\Sigma$ and is independent of $V_n$.

Specifically, when $\mathbb{P}_n = \mathcal{N}(\beta_n, \Sigma)$, we easily verify that

$$W_n \sim \mathcal{N}(0_d, \rho^2\Sigma),$$

and that $W_n$ is independent of $V_n$. That is, for Gaussian data, the selection on our subsample of size $n_1$ coincides exactly with the randomized selection in (2.2). This example was provided in [18] for $d = 1$. In this situation, we note that Proposition 1 reuses the pilot data to produce a carved pivot for exactly-valid selective inference.

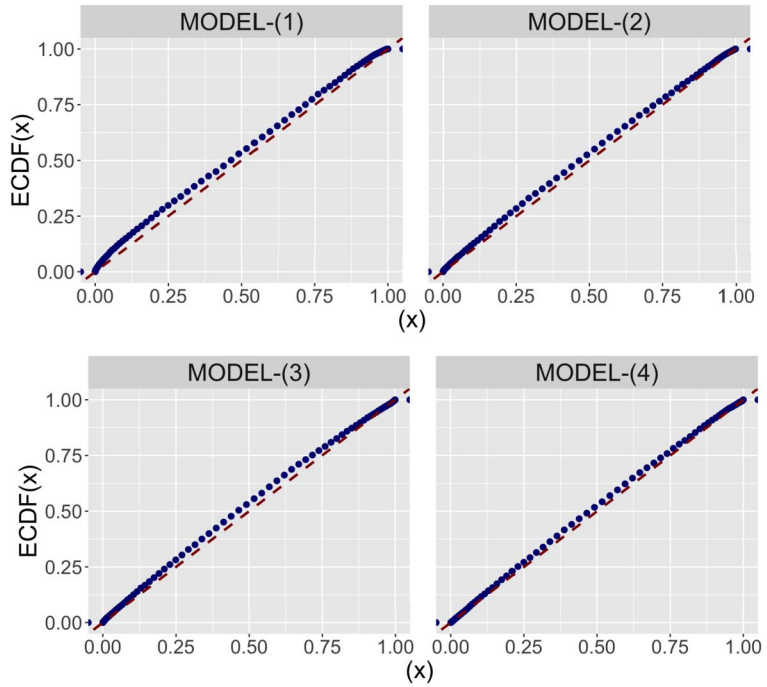


FIG. 1. *The four panels plot the ECDF of the carved pivot when data is generated according to MODELS* (1)–(4). *The red dashed line represents the reference* $y = x$ *curve.*

What happens when we drift away from Gaussian data? We begin with a simple simulation. We draw our data from four different models with non-Gaussian errors and conduct 10,000 rounds of simulation from each model. Figure 1 depicts the empirical cumulative distribution function (ECDF) of the carved pivot. A strong alignment with the $y = x$ line indicates that the carved pivot is well approximated by a Unif(0, 1) variable.

Let $n_1 = n_2 = 25$, that is, $\rho^2 = 1$. Fix $\Sigma = I_{d,d}$ and fix $\tau = 0_d$. Let

$$\sqrt{n}\beta_n^{(j)} = -a_n\bar{\beta} \quad \text{for} j \in [d],$$

where $a_n = n^{1/6-\delta}$, $\delta = 1\text{e}{-3}$ and $\bar{\beta} = 1.5$. We draw

$$\zeta_{i,n} = \beta_n + \text{e}_{i,n}.$$

First, each component of the error vector $\text{e}_{i,n}$ is drawn independently from E, a distribution supported on the real line. Then we standardize each such observation to have mean 0 and variance 1. MODELS (1)–(4) are based on four different choices of E:

1. MODEL-(1) E = Exponential(1) with rate parameter equal to 1 and density equal to $\text{p}(x) = \exp(-x) \cdot \mathbf{1}_{x>0}$.
2. MODEL-(2) E = Exponentially Modified Gaussian distribution(0, 1, 1) with the mean and variance of the Gaussian component equal to 0 and 1 respectively, and with the rate parameter of the exponential component equal to 1, and density equal to

$$\text{p}(x) = \frac{1}{\sqrt{\pi}} \exp(0.5 - x) \int_{\frac{1-x}{\sqrt{2}}}^{\infty} e^{-t^2}\, dt$$

3. MODEL-(3) $\text{E} = 0.8 \cdot \mathcal{N}(0, 0.25) + 0.2 \cdot \mathcal{N}(0, 3)$, which is a mixture of two Gaussian distributions with mixing weights 0.8 and 0.2.
4. MODEL-(4) E = Laplace(0, 1) with location and scale parameters equal to 0 and 1, respectively, and density equal to $\text{p}(x) = (2)^{-1} \exp(-|x|)$.

We make a few interesting observations from Figure 1. First, our plot suggests that the carved pivot produces valid selective inference well beyond Gaussian data. Previously, [24] showed that randomized selection rules with heavy-tailed variables produced asymptotically-valid selective inference. In contrast, randomization variables, based on carving, resemble Gaussian variables in the limit. Second, our plot shows that selective inference remains valid for rare selection outcomes that have vanishing probabilities in the limit. Prior asymptotic work such as those conducted by [11, 24, 25] have only focused on selection outcomes with nonvanishing probabilities. The asymptotic theory in our paper extends the use of carving beyond Gaussian models and confirms the validity of selective inference for rare selection outcomes observed on pilot data.

Our paper is connected with the fast-growing literature on selective inference with randomization. In recent work, [8] showed that randomized rules on Gaussian variables yield bounded confidence intervals for selective inference and [20] have utilized similar rules to construct confidence intervals for the effects of selected genetic variants. [22] proposed repeated carving for more stable inference in high-dimensional settings. [28] investigated the potential of randomization from an algorithmic stability perspective. [19] applied carving to pool summary statistics from prior studies and constructed unbiased estimators for shared parameters. Work by [16, 18] introduced Bayesian methods to construct inference after solving randomized variable selection algorithms. [21] utilized a Gaussian randomization variable to split a data set into two parts. One part is utilized for selection while the other is kept aside for inference. Differently from the previous reference, the results in our paper support the reuse of the first part when moving away from Gaussian data.

**3. Dependence on relative differences.** Our main finding in this section is that asymptotic validity of the carved pivot can be shown to depend on the convergence of specific relative differences. We first discuss some preliminaries.

3.1. *Some preliminaries.* We start from the randomized selection rule in (2.2), where (i) $W_n$ is distributed as a Gaussian random variable

$$W_n \sim \mathcal{N}(0_d, \rho^2 \Sigma),$$

and (ii) $W_n$ is independent of $V_n$. Later, we show that asymptotic guarantees with a Gaussian randomization variable are transferable to carving under some conditions. We come back to this topic in Section 5.

Fixing some more notation, let

$$\mathrm{e}_{i,n} = \Sigma^{-1/2}(\zeta_{i,n} - \beta_n),$$

and let $Z_{i,n} = \frac{1}{\sqrt{n}}\mathrm{e}_{i,n}$. We assume that the components of $V_n$ in the set $E_{\mathrm{obs}}$ are stacked before the ones in its complement set. Hereafter, we find it convenient to work with a standardized version for $V_n$,

$$\mathcal{Z}_n = \Sigma^{-1/2}(V_n - \sqrt{n}\beta_n),$$

which can be written as

$$\mathcal{Z}_n = \sum_{i=1}^{n} \frac{1}{\sqrt{n}}\mathrm{e}_{i,n} = \sum_{i=1}^{n} Z_{i,n}. \tag{3.1}$$

Our pivot, in terms of the standardized variable, is now denoted by

$$\mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n) = \mathrm{Pivot}^{(j)}\big((R^{(j)})^{-1}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n)\big). \tag{3.2}$$

This is based on noting the following equality:

$$\left(V_n^{(j)} \quad U_n^{(j)\prime}\right)' = \left(R^{(j)}\right)^{-1}\left(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n\right).$$

Our next result computes the ratio between the conditional and unconditional likelihood functions, after and before we apply the randomized selection rule. We use the symbol

$$\mathrm{LR}_{\mathbb{P}_n}(\mathcal{Z}_n; \sqrt{n}\beta_n)$$

to denote this ratio at $\beta_n$.

REMARK 1. We stress that $\mathrm{LR}_{\mathbb{P}_n}(\mathcal{Z}_n; \sqrt{n}\beta_n)$ represents how the selection process affects the unconditional likelihood. It is worth noting that this ratio is different from a ratio of the same likelihood function at two distinct values of $\beta_n$.

In the rest of the paper, we use $\mathbb{E}_{\mathbb{P}_n}[Z]$ to denote the expectation of the standardized variable $Z$ when based on the distribution $\mathbb{P}_n$. We use the more specific symbol $\mathbb{E}_{\mathcal{N}}[Z]$ to represent the expectation of $Z$ when $Z \sim \mathcal{N}(0_d, I_{d,d})$.

PROPOSITION 2 (Ratio of likelihood functions after and before selection). *Let* $\mathrm{F}: \mathbb{R}^d \to \mathbb{R}$ *be defined according to Proposition* 1. *Then the ratio between the conditional likelihood and unconditional likelihood functions is equal to*

$$\mathrm{LR}_{\mathbb{P}_n}(\mathcal{Z}_{n;\mathrm{obs}}; \sqrt{n}\beta_n) = \frac{\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_{n;\mathrm{obs}} + \sqrt{n}\beta_n)}{\mathbb{E}_{\mathbb{P}_n}[\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n)]},$$

*where* $\mathcal{Z}_{n;\mathrm{obs}}$ *is the observed value of* $\mathcal{Z}_n$.

REMARK 2. Suppose that $\mathbb{P}_n = \mathcal{N}(\beta_n, \Sigma)$. Equivalently, $\mathcal{Z}_n$ is distributed as $\mathcal{N}(0_d, I_{d,d})$ variable. In this case, we utilize the subscript $\mathcal{N}$ to indicate that our likelihoods are based on Gaussian data, and the above ratio is denoted by

$$\mathrm{LR}_{\mathcal{N}}(\mathcal{Z}_{n;\mathrm{obs}}; \sqrt{n}\beta_n) = \frac{\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_{n;\mathrm{obs}} + \sqrt{n}\beta_n)}{\mathbb{E}_{\mathcal{N}}[\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n)]}.$$

Suppose $\mathcal{Q}$ is a real-valued measurable mapping. Through the ratio in Proposition 2, we define

$$\widetilde{\mathbb{E}}_{\mathbb{P}_n}\left[\mathcal{Q}(\mathcal{Z}_n)\right] = \mathbb{E}_{\mathbb{P}_n}\left[\mathcal{Q}(\mathcal{Z}_n) \cdot \mathrm{LR}_{\mathbb{P}_n}(\mathcal{Z}_n; \sqrt{n}\beta_n)\right]. \tag{3.3}$$

The expectation on the left-hand side of (3.3) is taken with respect to the conditional distribution of $\mathcal{Z}_n$, and the expectation on the right-hand side is taken with respect to the unconditional distribution of $\mathcal{Z}_n$. Once again, for Gaussian data, we use more specific notation with the subscript $\mathcal{N}$ and define

$$\widetilde{\mathbb{E}}_{\mathcal{N}}\left[\mathcal{Q}(\mathcal{Z}_n)\right] = \mathbb{E}_{\mathcal{N}}\left[\mathcal{Q}(\mathcal{Z}_n) \cdot \mathrm{LR}_{\mathcal{N}}(\mathcal{Z}_n; \sqrt{n}\beta_n)\right].$$

We are now ready to formally state our main goal in the paper. Let $\mathrm{H} \in \mathbb{C}^3(\mathbb{R}, \mathbb{R})$ be an arbitrary function with bounded derivatives up to the third order. Let $\mathcal{C}_n$ be a suitable collection of distributions $\mathbb{P}_n$ that we specify later. Using our notation, we prove weak convergence of our pivot by showing that

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{C}_n} \left|\widetilde{\mathbb{E}}_{\mathbb{P}_n}\left[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)\right] - \widetilde{\mathbb{E}}_{\mathcal{N}}\left[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)\right]\right| = 0. \tag{3.4}$$

The above weak convergence statement indicates that our pivot generates asymptotically-valid conditional inference even as we depart from Gaussian data. It is important to mention that this statement assures the validity of selective inference across all distributions in the collection $\mathcal{C}_n$.

3.2. *Relative differences.* Proposition 3 recognizes that weak convergence of our pivot depends on the convergence of specific relative differences. Before we do so, define

$$\begin{aligned} \mathrm{G}_1(\mathcal{Z}_n; \sqrt{n}\beta_n) &= \mathrm{F}\big(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n\big), \\ \mathrm{G}_2(\mathcal{Z}_n; \sqrt{n}\beta_n) &= \mathrm{F}\big(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n\big) \cdot \mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n). \end{aligned} \tag{3.5}$$

Let $\sup f$ denote the supremum of a bounded, real-valued function $f$.

PROPOSITION 3 (Relative differences). *Suppose that*

$$\mathrm{R}_n^{(l)} = \big(\mathbb{E}_{\mathcal{N}}\big[\mathrm{F}\big(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n\big)\big]\big)^{-1} \cdot \big|\mathbb{E}_{\mathbb{P}_n}\big[\mathrm{G}_l(\mathcal{Z}_n; \sqrt{n}\beta_n)\big] - \mathbb{E}_{\mathcal{N}}\big[\mathrm{G}_l(\mathcal{Z}_n; \sqrt{n}\beta_n)\big]\big|,$$

*for* $l \in [2]$. *Let* $\sup|\mathrm{H}| = \mathrm{K} < \infty$. *Then it holds that*

$$\big|\widetilde{\mathbb{E}}_{\mathbb{P}_n}\big[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)\big] - \widetilde{\mathbb{E}}_{\mathcal{N}}\big[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)\big]\big| \le \big(\mathrm{K} \cdot \mathrm{R}_n^{(1)} + \mathrm{R}_n^{(2)}\big).$$

REMARK 3. We note that the relative differences $\mathrm{R}_n^{(l)}$ involve expectations that are computed with respect to the unconditional distribution of $\mathcal{Z}_n$.

As a result of Proposition 3, the weak convergence statement in (3.4) follows immediately once we show that

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{C}_n} \mathrm{R}_n^{(l)} = 0 \quad \text{for } l \in [2].$$

To close this section, we have a simplified expression for the common denominator in our relative differences. Define

$$\bar{\Sigma} = \big(Q'\Sigma^{-1}Q\big)^{-1}, \qquad \bar{\mu}_n = \bar{\Sigma}\,Q'\Sigma^{-1}(\sqrt{n}\beta_n - r), \qquad \Lambda = \Sigma^{-1} - \Sigma^{-1}Q\bar{\Sigma}\,Q'\Sigma^{-1}.$$

PROPOSITION 4. *We have*

$$\mathbb{E}_{\mathcal{N}}\big[\mathrm{F}\big(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n\big)\big] = C_0 \cdot \mathrm{Exp}\bigg(\sqrt{n}\beta_n - r, \frac{1}{(1+\rho^2)} \cdot \Lambda\bigg) \cdot \mathbb{P}_{\mathcal{N}}[T_n > 0_p],$$

*where* $T_n \sim \mathcal{N}(\bar{\mu}_n, (1+\rho^2)\bar{\Sigma})$ *and* $C_0$ *is a constant, which does not depend on* $n$.

We note that $\mathbb{P}_{\mathcal{N}}[T_n > 0_p]$ is the probability of our selection outcome when $\mathbb{P}_n = \mathcal{N}(\beta_n, \Sigma)$. Put another way, Proposition 4 states how the common denominator of our relative differences depends on this probability.

3.3. *Revisiting the univariate pivot.* We revisit our univariate pivot in Corollary 1. Recall that $\Sigma = I_{d,d}$ and $\Lambda = 0_d$. Consistent with our earlier notation, we represent the univariate pivot using the standardized variable through

$$\mathrm{P}^{(j)}\big(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)}\big) = \mathrm{Pivot}^{(j)}\big(\mathcal{Z}_n^{(j)} + \sqrt{n}\beta_n^{(j)}\big). \tag{3.6}$$

For this special case, we define

$$\begin{aligned} \widetilde{\mathrm{G}}_1\big(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)}\big) &= \bar{\Phi}\bigg(-\frac{1}{\rho}\big(\mathcal{Z}_n^{(j)} + \sqrt{n}\beta_n^{(j)}\big)\bigg), \\ \widetilde{\mathrm{G}}_2\big(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)}\big) &= \bar{\Phi}\bigg(-\frac{1}{\rho}\big(\mathcal{Z}_n^{(j)} + \sqrt{n}\beta_n^{(j)}\big)\bigg) \cdot \mathrm{H} \circ \mathrm{P}^{(j)}\big(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)}\big). \end{aligned} \tag{3.7}$$

Letting $\widetilde{D}_n = \mathbb{E}_{\mathcal{N}}[\bar{\Phi}(-\frac{1}{\rho}(\mathcal{Z}_n^{(j)} + \sqrt{n}\beta_n^{(j)}))]$, we now define the relevant relative differences as

$$\widetilde{\mathrm{R}}_n^{(l)} = \widetilde{D}_n^{-1} \cdot \big|\mathbb{E}_{\mathbb{P}_n}\big[\widetilde{\mathrm{G}}_l\big(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)}\big)\big] - \mathbb{E}_{\mathcal{N}}\big[\widetilde{\mathrm{G}}_l\big(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)}\big)\big]\big|, \tag{3.8}$$

for $l \in [2]$. Note that the relative differences are determined solely by the expectations of functions that involve the univariate variable $\mathcal{Z}_n^{(j)}$.

Suppose that $\sup|\mathrm{H}| = \mathrm{K} < \infty$. Once again, we can show that

$$|\widetilde{\mathbb{E}}_{\mathbb{P}_n}[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)})] - \widetilde{\mathbb{E}}_{\mathcal{N}}[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)})]| \leq (\mathrm{K} \cdot \widetilde{\mathrm{R}}_n^{(1)} + \widetilde{\mathrm{R}}_n^{(2)}).$$

We use two facts here. First, the pivot is a function of the univariate variable $\mathcal{Z}_n^{(j)}$. Second, the likelihood ratio, in Proposition 2, is proportional to

$$\prod_{j \in E_{\mathrm{obs}}} \mathrm{LR}_{\mathbb{P}_n}(\mathcal{Z}_{n;\mathrm{obs}}^{(j)}; \sqrt{n}\beta_n^{(j)}),$$

where

$$\mathrm{LR}_{\mathbb{P}_n}(\mathcal{Z}_{n;\mathrm{obs}}^{(j)}; \sqrt{n}\beta_n^{(j)}) = \left\{\mathbb{E}_{\mathbb{P}_n}\left[\bar{\Phi}\left(-\frac{1}{\rho}(\mathcal{Z}_n^{(j)} + \sqrt{n}\beta_n^{(j)})\right)\right]\right\}^{-1} \bar{\Phi}\left(-\frac{1}{\rho}(\mathcal{Z}_{n;\mathrm{obs}}^{(j)} + \sqrt{n}\beta_n^{(j)})\right).$$

It is important to note that this ratio depends on $\beta_n^{(j)}$ only through the univariate variable $\mathcal{Z}_n^{(j)}$. For Gaussian data, the specific ratio is given by

$$\mathrm{LR}_{\mathcal{N}}(\mathcal{Z}_{n;\mathrm{obs}}^{(j)}; \sqrt{n}\beta_n^{(j)}) = \frac{\bar{\Phi}(-\frac{1}{\rho}(\mathcal{Z}_{n;\mathrm{obs}}^{(j)} + \sqrt{n}\beta_n^{(j)}))}{\mathbb{E}_{\mathcal{N}}[\bar{\Phi}(-\frac{1}{\rho}(\mathcal{Z}_n^{(j)} + \sqrt{n}\beta_n^{(j)}))]}.$$

The steps in the proof of Proposition 3 directly lead us to the bound in (3.8) based on our relative differences.

At last, we note that the common denominator in our relative differences is equal to

$$\mathbb{E}_{\mathcal{N}}\left[\bar{\Phi}\left(-\frac{1}{\rho}(\mathcal{Z}_n^{(j)} + \sqrt{n}\beta_n^{(j)})\right)\right] = \mathbb{P}_{\mathcal{N}}[j \in E_{\mathrm{obs}}] = \bar{\Phi}\left(-\frac{\sqrt{n}\beta_n^{(j)}}{\sqrt{(1+\rho^2)}}\right),$$

which is the probability of the selection outcome on Gaussian data.

**4. Weak convergence of univariate pivot.** To better understand the behavior of the multivariate pivot for a general $\Sigma$, we first analyze the simpler univariate pivot.

4.1. *Main results.* In this section, we state our main results in Theorem 1 and Theorem 2. These results establish that our univariate pivot yields asymptotically-valid selective inference for two types of selection outcomes, namely bounded outcomes and rare outcomes. We describe both types of outcomes below.

Suppose that the components of our mean vector are bounded, that is,

$$|\sqrt{n}\beta_n^{(j)}| < R \quad \text{for each } n \in \mathbb{N} \text{ and } j \in [d]. \tag{4.1}$$

Consider the limiting case when $\mathbb{P}_n = \mathcal{N}(\beta_n, I_{d,d})$. Recall that the probability of the selection outcome on Gaussian data is equal to

$$\mathbb{P}_{\mathcal{N}}[j \in E_{\mathrm{obs}}] = \bar{\Phi}\left(-\frac{\sqrt{n}\beta_n^{(j)}}{\sqrt{(1+\rho^2)}}\right).$$

Clearly, this probability is bounded away from 0 whenever the mean satisfies (4.1). This selection outcome is referred to as a bounded outcome.

From now on, fix $j \in E_{\mathrm{obs}}$ and consider the relative differences $\widetilde{\mathrm{R}}_n^{(l)}$ in (3.8).

ASSUMPTION 1. Consider a collection of distributions $\mathcal{P}_{b,n}$ such that the mean of each distribution $\mathbb{P}_n$ in this collection satisfies (4.1). Assume that $\mathcal{P}_{b,n}$ has uniformly bounded third moments in the following sense:

$$\sup_n \sup_{\mathbb{P}_n \in \mathcal{P}_{b,n}} \mathbb{E}_{\mathbb{P}_n}\big[|\mathrm{e}_{1,n}^{(j)}|^3\big] < \infty,$$

where $\mathrm{e}_{1,n}$ is the standardized variable, which was defined in (3.1).

THEOREM 1 (Weak convergence of under bounded outcomes). *Under Assumption 1, we have*

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{P}_{b,n}} \widetilde{\mathrm{R}}_n^{(l)} = 0 \quad \textit{for } l \in [2],$$

*and as a result*,

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{P}_{b,n}} \big|\widetilde{\mathbb{E}}_{\mathbb{P}_n}\big[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)})\big] - \widetilde{\mathbb{E}}_{\mathcal{N}}\big[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)})\big]\big| = 0.$$

Now suppose we consider the case where the mean of our distribution $\mathbb{P}_n$ grows with increasing sample size and

$$\lim_n \sqrt{n}\beta_n^{(j)} = -\infty \quad \text{for all } j \in [d].$$

As the sample size grows bigger, the probability of the selection outcome on Gaussian data approaches 0. This selection outcome is referred to as a rare outcome.

From now on, we focus on a subset of these parameters that result in large deviation-type probabilities. Fix $\bar{\beta} > 0$. Suppose that each component of the mean vector is parameterized as

$$\sqrt{n}\beta_n^{(j)} = -a_n\bar{\beta}, \tag{4.2}$$

where $a_n \to \infty$ as $n \to \infty$ and $a_n = o(n^{1/2})$. Using the Mills ratio for Gaussian tail probabilities, it is easy to see that the probability of the rare outcome vanishes to 0 as

$$\mathbb{P}_{\mathcal{N}}[j \in E_{\mathrm{obs}}] = \bar{\Phi}\bigg(-\frac{\sqrt{n}\beta_n^{(j)}}{\sqrt{(1+\rho^2)}}\bigg) = C_0(a_n\bar{\beta})^{-1} \cdot \phi\bigg(\frac{a_n\bar{\beta}}{\sqrt{1+\rho^2}}\bigg).$$

ASSUMPTION 2. Consider a collection of distributions $\mathcal{P}_{r,n}$ that have means parameterized as per (4.2). Assume that the collection $\mathcal{P}_{r,n}$ has uniformly bounded exponential moments as follows:

$$\sup_n \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \mathbb{E}_{\mathbb{P}_n}\big[\exp\big(\chi\,|\mathrm{e}_{1,n}^{(j)}|\big)\big] < \infty$$

for some $\chi \in \mathbb{R}^+$.

Let $\Psi : \mathcal{K} \to \mathbb{R}$ be a continuous, bounded function. Under the moment condition in Assumption 2, the variable $\mathcal{Z}_n$ obeys Varadhan's principle of large deviations in the following sense:

$$\frac{1}{a_n^2}\log \mathbb{E}_{\mathbb{P}_n}\bigg[\exp\bigg(-a_n^2\Psi\bigg(\frac{1}{a_n}\mathcal{Z}_n\bigg)\bigg) \cdot \mathbf{1}_{\frac{1}{a_n}\mathcal{Z}_n \in \mathcal{K}}\bigg] = \mathrm{r}_{\Psi,n} - \inf_{z \in \mathcal{K}}\bigg\{\frac{1}{2}z^2 + \Psi(z)\bigg\},$$

where $\mathrm{r}_{\Psi,n} = o(1)$. For example, please see [4].

ASSUMPTION 3. Consider $\Psi \equiv \Psi_t$ for $t \in \{0, 1\}$ where $\Psi_1(z) = \frac{1}{\rho^2}(z-\bar{\beta})^2$ and $\Psi_0(z) = 0$. Fix $\mathcal{K} \equiv \mathcal{K}_t$ for $t \in \{0, 1\}$ where $\mathcal{K}_1 = [-c_0, c_0]$ for $c_0 > 0$, and $\mathcal{K}_0 = \mathcal{K}_1^c$. We assume that

$$\sup_n \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \sup_{t\in\{0,1\}} a_n^2 \mathrm{r}_{\Psi_t,n} < \infty.$$

The conditions in Assumptions 2 and 3 imply that

$$\mathbb{E}_{\mathbb{P}_n}\left[\exp\left(-a_n^2 \Psi_t\left(\frac{1}{a_n}\mathcal{Z}_n\right)\right) \cdot \mathbf{1}_{a_n^{-1}\mathcal{Z}_n \in \mathcal{K}_t}\right] \leq \mathrm{K}_0 \exp\left(-a_n^2 \inf_{z\in\mathcal{K}_t}\left\{\frac{1}{2}z^2 + \Psi_t(z)\right\}\right),$$

where $\mathrm{K}_0$ is a constant. As a result, we obtain the rate of decay for large-deviations type probabilities and exponentially vanishing moments.

THEOREM 2 (Weak convergence under rare outcomes). *Suppose that the conditions in Assumptions 2 and 3 are met. Then we have*

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \widetilde{\mathrm{R}}_n^{(l)} = 0 \quad \textit{for } l \in [2],$$

*and*

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \left|\widetilde{\mathbb{E}}_{\mathbb{P}_n}[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)})] - \widetilde{\mathbb{E}}_{\mathcal{N}}[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n^{(j)}; \sqrt{n}\beta_n^{(j)})]\right| = 0.$$

Relative to the conditions in Assumption 1, we impose stronger moment conditions to handle rare outcomes. In return, we can guarantee asymptotically-valid inference through our pivot, even when we condition on rare outcomes with large deviation-type probabilities.

REMARK 4. We exclude the uninteresting case when

$$\sqrt{n}\beta_n^{(j)} \to \infty.$$

This is because selection does not have an impact in large samples and standard inferences do not require an adjustment for selection.

4.2. *Main tool for weak convergence theory.* We present the Stein bound for Gaussian approximations, which is the primary tool in our asymptotic theory. We then provide a brief outline of how it applies to our problem.

Fixing some more notation, we denote by

$$\mathcal{Z}_n[-i] = \mathcal{Z}_n - Z_{i,n} = \sum_{k\in[n]\setminus i} Z_{k,n}$$

the $i$th leave-one out variable. This variable is obtained by dropping $Z_{i,n}$ from the sum defined in (3.1). Let $\mathcal{Z}_n^{(j)}[-i]$ be the $j$th entry of this $i$th leave-one out variable. Consider a real-valued mapping g that is Lebesgue-almost surely differentiable and satisfies $\mathbb{E}_{\mathcal{N}}[|\mathrm{g}(Z)|] < \infty$ for $Z \sim \mathcal{N}(0, 1)$. Define

$$\mathcal{S}_{\mathrm{g}}(z) := \exp\left(\frac{1}{2}z^2\right) \cdot \int_{-\infty}^{z} \{\mathrm{g}(t) - \mathbb{E}_{\mathcal{N}}(\mathrm{g}(Z))\} \cdot \mathrm{Exp}(t, 1)\, dt, \tag{4.3}$$

which is also called the Stein function for g. For $i \in [n]$, we let

$$\mathrm{M}_i(t) = \mathbb{E}_{\mathbb{P}_n}[Z_{i,n}^{(j)}(\mathbf{1}_{[t,\infty)}(Z_{i,n}^{(j)})\mathbf{1}_{[0,\infty)}(t) - \mathbf{1}_{(-\infty,t]}(Z_{i,n}^{(j)})\mathbf{1}_{(-\infty,0)}(t))].$$

Lemma 1 provides a bound to measure the difference between the expectations of a Gaussian variable and its non-Gaussian counterpart. For related literature, we point out to [3]. In this paper, we use the symbol $\mathcal{D}^k f(x_0)$ to denote the $k$th derivative of a differentiable function $f$ at $x_0$ and simply use $\mathcal{D}f(x_0)$ to denote the first derivative of $f$ at $x_0$.

LEMMA 1 (Univariate Stein bound). *We have*

$$\left|\mathbb{E}_{\mathbb{P}_n}\left[\mathrm{g}\left(\mathcal{Z}_n^{(j)}\right)\right] - \mathbb{E}_{\mathcal{N}}\left[\mathrm{g}\left(\mathcal{Z}_n^{(j)}\right)\right]\right| \leq \mathrm{SB}_{\mathbb{P}_n}(\mathrm{g}),$$

*where*

$$\mathrm{SB}_{\mathbb{P}_n}(\mathrm{g}) = n \cdot \int_{-\infty}^{\infty} \sup_{\alpha\in[0,1]} \mathbb{E}_{\mathbb{P}_n}\left[\left(|t| + \frac{1}{\sqrt{n}}|\mathrm{e}_{1,n}^{(j)}|\right) \times \left|\mathcal{D}^2\mathcal{S}_{\mathrm{g}}\left(\alpha t + (1-\alpha)\frac{1}{\sqrt{n}}\mathrm{e}_{1,n}^{(j)} + \mathcal{Z}_n^{(j)}[-1]\right)\right|\right]\mathrm{M}_1(t)\,dt.$$

Equipped with the above bound, we review the relative differences defined in (3.8). We use the Stein bound to write the following inequality:

$$\widetilde{\mathrm{R}}_n^{(l)} \leq (\widetilde{D}_n)^{-1} \cdot \mathrm{SB}_{\mathbb{P}_n}(\widetilde{\mathrm{G}}_l)$$

for $l \in [2]$.

First, we consider bounded outcomes. The probability of a bounded outcome, which is also the common denominator of our relative differences $\widetilde{D}_n$, is bounded away from 0. To prove weak convergence of our pivot, we need to prove that the univariate Stein bound $\mathrm{SB}_{\mathbb{P}_n}(\widetilde{\mathrm{G}}_l)$ uniformly converges to 0 as $n$ tends to infinity.

When dealing with rare outcomes, the uniform convergence of the Stein bound is not enough to guarantee weak convergence of our pivot. This is because the probability of the selection outcome also converges to 0 at an exponentially fast rate. To ensure weak convergence of our pivot, it is necessary for the related Stein bound to converge at a faster rate than the probability of the selection outcome. As a result, proving the asymptotic validity of our pivot requires stronger conditions compared to bounded outcomes.

For both types of outcomes mentioned, we investigate the large-sample behavior of the commensurate Stein bound to prove Theorems 1 and 2. Detailed proofs are deferred to the Supplementary Material.

## 5. Weak convergence of multivariate pivot.

We turn to the multivariate pivot in Proposition 1. Throughout the section, we will use $C_1, C_2, \ldots$ to denote constants that are free of $n$.

### 5.1. *Main results*.

In line with the preceding section, we develop our theory for bounded and rare outcomes.

We start by considering mean parameters, which satisfy

$$\|\sqrt{n}\beta_n - r\| \leq R. \tag{5.1}$$

Suppose that $\mathbb{P}_n = \mathcal{N}(\beta_n, \Sigma)$. Recall that the probability of the selection outcome is equal to

$$\mathbb{P}_{\mathcal{N}}[T_n > 0_p],$$

where $T_n$ is a Gaussian variable as stated in Proposition 4. It is easy to see that the probability of the selection outcome is bounded away from 0, which gives rise to bounded outcomes.

ASSUMPTION 4. We consider a collection of distributions $\mathcal{P}_{b,n}$ with bounded mean parameters as stated in (5.1). Suppose that the collection $\mathcal{P}_{b,n}$ has uniformly bounded moments as follows:

$$\sup_n \sup_{\mathbb{P}_n\in\mathcal{P}_{b,n}} \mathbb{E}_{\mathbb{P}_n}\left[\|\mathrm{e}_{1,n}\|^6\right] < \infty.$$

Let $\mathrm{R}_n^{(l)}$ be defined according to Proposition 3. Theorem 3 assures that our pivot generates asymptotically-valid selective inference for bounded outcomes.

THEOREM 3 (Weak convergence under bounded outcomes). *Under Assumption 4, we have*

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{P}_{b,n}} \mathrm{R}_n^{(l)} = 0 \quad \textit{for } l \in [2],$$

*and as a result*,

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{P}_{b,n}} \left| \widetilde{\mathbb{E}}_{\mathbb{P}_n}\big[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)\big] - \widetilde{\mathbb{E}}_{\mathcal{N}}\big[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)\big] \right| = 0.$$

Now we turn to rare outcomes. Fix $\bar{\beta} \in \mathbb{R}^d$ such that $\bar{\Sigma} Q' \Sigma^{-1} \bar{\beta} \notin (-\infty, 0]^d$. Let the mean for our generating distribution $\mathbb{P}_n$ be parameterized as

$$\sqrt{n}\beta_n - r = -a_n \bar{\beta}, \tag{5.2}$$

where $a_n \to \infty$ as $n \to \infty$ and $a_n = o(n^{1/6})$.

For each $\beta_n$, we consider the matching parameter

$$\bar{\mu}_n = \bar{\Sigma} Q' \Sigma^{-1} (\sqrt{n}\beta_n - r).$$

Based on our parameterization, note that we can write

$$\bar{\mu}_n = -a_n \bar{\mu},$$

where $\bar{\mu} = \bar{\Sigma} Q' \Sigma^{-1} \bar{\beta}$. Formalized next, we first see that the probability of the selection outcome vanishes to zero at an exponentially fast rate.

PROPOSITION 5 (Probability of a rare outcome). *Consider the optimization problem*

$$t_\star = \operatorname*{arg\,min}_{t \geq \bar{\mu}} t' \bar{\Sigma}^{-1} t.$$

*Then there exists a unique (nonempty) set $\mathcal{I} \subseteq [d]$ such that the following assertions are simultaneously true*:

(i) $t_\star^{(\mathcal{I})} = \bar{\mu}^{(\mathcal{I})} \neq 0_{|\mathcal{I}|}$;
(ii) *for* $\mathcal{J} = \mathcal{I}^c$, $t_\star^{(\mathcal{J})} = \bar{\Sigma}_{\mathcal{J},\mathcal{I}} \bar{\Sigma}_{\mathcal{I},\mathcal{I}}^{-1} \bar{\mu}^{(\mathcal{I})} \geq \bar{\mu}^{(\mathcal{J})}$ *whenever* $\mathcal{J} \neq \varnothing$;
(iii) $(\bar{\Sigma}_{\mathcal{I},\mathcal{I}}^{-1} \bar{\mu}^{(\mathcal{I})})^{(j)} > 0$ *for all* $j \in \mathcal{I}$ *and* $t_\star' \bar{\Sigma}^{-1} t_\star = (\bar{\mu}^{(\mathcal{I})})' \bar{\Sigma}_{\mathcal{I},\mathcal{I}}^{-1} \bar{\mu}^{(\mathcal{I})} > 0$.

*Further, we have*

$$\mathbb{P}_{\mathcal{N}}[T_n > 0_p] = \frac{C_3}{(a_n)^{|\mathcal{I}|}} \cdot \mathrm{Exp}\bigg(a_n \bar{\mu}^{(\mathcal{I})}, \frac{1}{(1+\rho^2)} (\bar{\Sigma}_{\mathcal{I},\mathcal{I}})^{-1}\bigg)$$

*for sufficiently large n*.

REMARK 5. The proof for the above result closely follows Proposition 2.1 and Corollary 4.1 in [7]. Therefore, we omit further details of the proof here.

As a corollary, we observe the following.

COROLLARY 2. *Let* $\Delta = \Sigma^{-1} Q \bar{\Sigma}_{\mathcal{I}} \bar{\Sigma}_{\mathcal{I},\mathcal{I}}^{-1} \bar{\Sigma}_{\mathcal{I}}' Q' \Sigma^{-1}$. *It holds that the common denominator of our relative differences is equal to*

$$\mathbb{E}_{\mathcal{N}}\big[\mathrm{F}(\Sigma^{1/2} \mathcal{Z}_n + \sqrt{n}\beta_n)\big] = \frac{C_4}{(a_n)^{|\mathcal{I}|}} \cdot \mathrm{Exp}\bigg(\sqrt{n}\beta_n - r, \frac{1}{(1+\rho^2)} \cdot (\Lambda + \Delta)\bigg).$$

The proof of Corollary 2 follows directly from the claims in Propositions 4 and 5.

Below, we state the assumptions to guarantee weak convergence of the multivariate pivot under rare outcomes.

ASSUMPTION 5. Consider a collection of distributions $\mathcal{P}_{r,n}$ such that the mean grows with $n$ as per (5.2). Assume that the collection $\mathcal{P}_{r,n}$ has a uniformly bounded exponential moment near the origin as follows:

$$\sup_n \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \mathbb{E}_{\mathbb{P}_n}\big[\exp(\chi \|\mathrm{e}_{1,n}\|)\big] < \infty$$

for some $\chi \in \mathbb{R}^+$.

Let $\Psi : \mathcal{K} \to \mathbb{R}$ be a continuous and bounded function. Under Assumption 5, Varadhan's principle of large deviations for $\mathcal{Z}_n$ implies that

$$\frac{1}{a_n^2} \log \mathbb{E}_{\mathbb{P}_n}\left[\exp\left(-a_n^2 \Psi\left(\frac{1}{a_n}\mathcal{Z}_n\right)\right) \cdot \mathbf{1}_{\frac{1}{a_n}\mathcal{Z}_n \in \mathcal{K}}\right] = \mathrm{r}_{\Psi,n} - \inf_{z \in \mathcal{K}}\left\{\frac{1}{2} z'z + \Psi(z)\right\},$$

where $\mathrm{r}_{\Psi,n} = o(1)$.

ASSUMPTION 6. Consider $\Psi \equiv \Psi_t$ where $\Psi_t(z) = \frac{1}{1-t+\rho^2}(\sqrt{t}\Sigma^{1/2}z - \bar{\beta})'(\Lambda + \Delta)(\sqrt{t}\Sigma^{1/2}z - \bar{\beta})$ for $t \in (0,1]$ and $\Psi_0(z) = 0$. Fix $\mathcal{K} \equiv \mathcal{K}_t$ where $\mathcal{K}_t = [-c_0 \cdot 1_d, c_0 \cdot 1_d]$ for $c_0 > 0$ and $t \in (0,1]$, and $\mathcal{K}_0 = \mathcal{K}_1^c$. We impose the condition that

$$\sup_n \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \sup_{t \in [0,1]} a_n^2 \mathrm{r}_{\Psi_t,n} < \infty.$$

ASSUMPTION 7. Additionally, we assume that

$$\sup_n \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \frac{\mathbb{E}_{\mathbb{P}_n}[\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n) \cdot \mathbf{1}_{\mathcal{Z}_n \in \mathcal{K}_n}]}{\mathbb{E}_{\mathcal{N}}[\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n) \cdot \mathbf{1}_{\mathcal{Z}_n \in \mathcal{K}_n}]} < \infty$$

whenever

$$\lim_n \frac{\mathbb{E}_{\mathcal{N}}[\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n) \cdot \mathbf{1}_{\mathcal{Z}_n \in \mathcal{K}_n}]}{\mathbb{E}_{\mathcal{N}}[\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n)]} = 0.$$

In particular, we note the following.

REMARK 6. Similar to our univariate analysis, the conditions in Assumptions 5 and 6 provide a uniform bound on a set of large-deviations type probabilities and exponentially vanishing moments. The condition in Assumption 7 controls the probability of selection outcomes that are rarer than the observed outcome on Gaussian data by imposing the restriction that these probabilities decay at an equal or faster rate than the limiting Gaussian counterpart More specifically, this condition allows us to establish convergence of our relative differences on a set of high probability while controlling their behavior on the complement set.

Theorem 4 proves that our pivot offers asymptotically-valid selective inference, even when rare outcomes are observed.

THEOREM 4 (Weak convergence under rare outcomes). *Suppose that the conditions in Assumptions 5, 6 and 7 are met. Then we have that*

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \mathrm{R}_n^{(l)} = 0,$$

*and that*

$$\lim_{n} \sup_{\mathbb{P}_n \in \mathcal{P}_{r,n}} \big|\widetilde{\mathbb{E}}_{\mathbb{P}_n}\big[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)\big] - \widetilde{\mathbb{E}}_{\mathcal{N}}\big[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)\big]\big| = 0$$

*for* $l \in [2]$.

5.2. *Main tool for weak convergence theory.* To prove our main results in Theorem 3 and Theorem 4, we use a multivariate version of the Stein bound.

Lemma 2 presents this bound for a Lebesgue almost surely three times differentiable mapping $\mathrm{g} : \mathbb{R}^d \to \mathbb{R}$, which is adopted from [2]. Suppose that $\mathbb{E}_{\mathcal{N}}[|\mathrm{g}(Z)|] < \infty$. Let $Z \sim N(0_d, I_{d,d})$. The Stein bound is defined through partial derivatives of

$$\mathcal{S}_{\mathrm{g}}(z) = \int_0^1 \frac{1}{2t}\big(\mathbb{E}_{\mathcal{N}}\big[\mathrm{g}(\sqrt{t}z + \sqrt{1-t}Z)\big] - \mathbb{E}_{\mathcal{N}}\big[\mathrm{g}(Z)\big]\big)\,dt,$$

also called the Stein function for g. Before stating the bound, recall that

$$\mathcal{Z}_n[-i] = \mathcal{Z}_n - Z_{i,n}$$

denotes the $i$th leave-one out variable. Let

$$\mathcal{D}^k f(x_0)[i_1, i_2, \ldots, i_k] = \frac{\partial^k f(x_0)}{\partial x^{(i_1)} \partial x^{(i_2)} \ldots \partial x^{(i_k)}}$$

denote the $k$th order partial derivative of $f$ at $x_0$, for $i_1, i_2, \ldots, i_k \in [d]$, and let $\mathrm{e}^{\star}_{i,n}$ be an independent copy of $\mathrm{e}_{i,n}$, for $i \in [d]$.

LEMMA 2 (Multivariate Stein bound). *We have that*

$$\big|\mathbb{E}_{\mathbb{P}_n}\big[\mathrm{g}(\mathcal{Z}_n)\big] - \mathbb{E}_{\mathcal{N}}\big[\mathrm{g}(\mathcal{Z}_n)\big]\big| \le \mathrm{SB}_{\mathbb{P}_n}(\mathrm{g}),$$

*where*

$$\mathrm{SB}_{\mathbb{P}_n}(\mathrm{g}) = \frac{C_1}{\sqrt{n}} \sum_{\lambda,\gamma \in \{0\} \cup [3] : \lambda+\gamma \le 3} \sum_{j,k,l} \mathbb{E}_{\mathbb{P}_n}\Bigg[ \|\mathrm{e}_{1,n}\|^{\lambda} \|\mathrm{e}^{\star}_{1,n}\|^{\gamma} \sup_{\alpha,\kappa \in [0,1]} \Bigg| \mathcal{D}^3 \mathcal{S}_{\mathrm{g}}\Bigg( \mathcal{Z}_n[-1] + \frac{\alpha}{\sqrt{n}}\mathrm{e}_{1,n} + \frac{\kappa}{\sqrt{n}}\mathrm{e}^{\star}_{1,n} \Bigg)[j,k,l] \Bigg| \Bigg].$$

As before, we revisit our relative differences and use the Stein bound to note that

$$R_n^{(l)} \le \big(\mathbb{E}_{\mathcal{N}}\big[\mathrm{F}\big(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n\big)\big]\big)^{-1} \cdot \mathrm{SB}_{\mathbb{P}_n}(\mathrm{G}_l).$$

In order to establish the weak convergence of our pivot, we analyze how the Stein bound behaves in large samples, similar to what we did for the univariate pivot. However, unlike the univariate theory, the multivariate version of the Stein bound involves higher-order derivatives of the Stein function. As a result, we investigate higher-order smoothness properties of our multivariate pivot. Proofs for Theorem 3 and Theorem 4 are collected in the Supplementary Material.

5.3. *Transfer of asymptotic guarantees to the carved pivot.* Having established weak convergence of our pivot for randomized rules with Gaussian variables, we come back to the selection described in (2.4).

Following the same convention as before, we evaluate the the likelihood ratio after and before we apply the selection rule on the pilot samples. At $(v' \; w')'$, let the joint density for $V_n$ and $W_n$ factorize as

$$\mathrm{p}_n(v, w) = \mathrm{p}_n(v) \cdot \bar{\mathrm{p}}_n\big(w|v\big),$$

where $\mathrm{p}_n$ is the marginal density for $V_n$ and $\bar{\mathrm{p}}_n(\cdot|v)$ is the conditional density of $W_n$ given $V_n = v$. Let $\bar{\mathrm{F}}_n : \mathbb{R}^d \to \mathbb{R}$ assume the value

$$\bar{\mathrm{F}}_n(v) = \int \bar{\mathrm{p}}_n(Qt - v + r|v) \cdot \mathbf{1}_{t \in \mathbb{R}^{p+}}\, dt.$$

PROPOSITION 6. *Under the randomized selection rule in* (2.4)*, the ratio of the conditional and unconditional likelihood functions is*

$$\overline{\mathrm{LR}}_{\mathbb{P}_n}(\mathcal{Z}_{n;\mathrm{obs}}; \sqrt{n}\beta_n) = \frac{\bar{\mathrm{F}}_n(\Sigma^{1/2}\mathcal{Z}_{n;\mathrm{obs}} + \sqrt{n}\beta_n)}{\mathbb{E}_{\mathbb{P}_n}[\bar{\mathrm{F}}_n(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n)]}.$$

Define

$$\overline{\mathbb{E}}_{\mathbb{P}_n}[\mathcal{Q}(\mathcal{Z}_n)] = \mathbb{E}_{\mathbb{P}_n}[\mathcal{Q}(\mathcal{Z}_n) \cdot \overline{\mathrm{LR}}_{\mathbb{P}_n}(\mathcal{Z}_n; \sqrt{n}\beta_n)].$$

The expectation on the left-hand side is taken with respect to the conditional law after selection on pilot data and is expressed as an unconditional expectation on the right-hand side through the above-stated likelihood ratio.

Consider a collection of distributions $\mathcal{C}_n$. The weak convergence of our pivot follows by proving

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{C}_n} \left|\overline{\mathbb{E}}_{\mathbb{P}_n}[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)] - \widetilde{\mathbb{E}}_{\mathcal{N}}[\mathrm{H} \circ \mathrm{P}^{(j)}(\mathcal{Z}_n; \sqrt{n}\beta_n)]\right| = 0 \tag{5.3}$$

for any $\mathrm{H} \in \mathbb{C}^3(\mathbb{R}, \mathbb{R})$ with bounded derivatives up to the third order. We replaced the first term in (3.4) with a conditional expectation given the selection outcome observed in the pilot data.

Our next result establishes that asymptotically-valid selective inference with Gaussian randomized rules transfers to the carved pivot. This result holds as long as the probability of the selection outcome converges to its counterpart with Gaussian randomization.

PROPOSITION 7 (Transfer of asymptotic guarantees). *Suppose that the conditional weak convergence statement in* (3.4) *holds over a collection of distributions in* $\mathcal{C}_n$*. Assume that*

$$\lim_n \sup_{\mathbb{P}_n \in \mathcal{C}_n} \frac{\mathbb{E}_{\mathbb{P}_n}[|\bar{\mathrm{F}}_n(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n) - \mathrm{F}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n)|]}{\mathbb{E}_{\mathbb{P}_n}[\mathrm{F}(\Sigma^{1/2}\mathcal{Z}_n + \sqrt{n}\beta_n)]} = 0.$$

*Then the convergence result of* (5.3) *holds*.

**6. Empirical analysis.** We illustrate how our theory translates to practice in various instances of selective inference.

EXAMPLE 6.1. Selectively inferring for a difference in means. We selectively infer for a difference in means through the two-sample test statistic. In alignment with the running example in our paper, we use the following scheme to draw $n$ independent and identically distributed observations with identity covariance. For $d = 2$, we draw

$$\zeta_{i,n} = \beta_n + \mathrm{e}_{i,n} \quad \text{for } i \in [n].$$

Each component of $\mathrm{e}_{i,n}$ is drawn independently as

$$\mathrm{e}_{i,n}^{(j)} \overset{\text{i.i.d.}}{\sim} \mathrm{E}$$

and standardized such that

$$\mathbb{E}[\mathrm{e}_{i,n}^{(j)}] = 0; \qquad \mathbb{E}[(\mathrm{e}_{i,n}^{(j)})^2] = 1.$$

TABLE 1
*Comparison of inference between carving and data splitting*

| | Gaussian | | Model-1 | | Model-2 | | Model-3 | | Model-4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho^2 = 1/2$ | Cov | Len | Cov | Len | Cov | Len | Cov | Len | Cov | Len |
| $\bar{\beta} = 2$ | | | | | | | | | | |
| Carve | 89% | 0.73 | 94% | 0.74 | 88% | 0.72 | 90% | 0.67 | 90.5% | 0.68 |
| Split | 89.5% | 0.95 | 95% | 0.95 | 87% | 0.95 | 91.5% | 0.95 | 90% | 0.95 |
| $\bar{\beta} = 1$ | | | | | | | | | | |
| Carve | 91.5% | 0.72 | 88.5% | 0.72 | 92% | 0.74 | 88% | 0.72 | 90.5% | 0.72 |
| Split | 88% | 0.95 | 90% | 0.95 | 93% | 0.96 | 91% | 0.95 | 91% | 0.95 |
| $\bar{\beta} = 0$ | | | | | | | | | | |
| Carve | 90% | 0.59 | 90% | 0.58 | 87% | 0.59 | 88.5% | 0.60 | 91.5% | 0.58 |
| Split | 91.5% | 0.95 | 90% | 0.95 | 87.5% | 0.95 | 91% | 0.95 | 91% | 0.95 |

Note that the distribution E is based on five different models, which include Models (1)–(4) described in Section 2 and the baseline Gaussian model. We provide selective inference for $\bar{\beta}_n = \beta_n^{(1)} - \beta_n^{(2)}$ whenever the two-sample statistic

$$V_{n_1} = \frac{\sqrt{n_1}}{\sqrt{2}}(\bar{\zeta}_{n_1}^{(1)} - \bar{\zeta}_{n_1}^{(2)}),$$

exceeds a prefixed threshold of significance. We investigate the performance of our carved pivot for $\bar{\beta}_n$.

For our simulations, the difference of means is parameterized according to $\sqrt{n}\bar{\beta}_n = -a_n\bar{\beta}$ for $a_n = n^{1/6-\delta}$ and $\delta = 1\text{e}{-3}$. We fix $n = 50$. We set our split proportion value at

$$\rho^2 = \frac{n - n_1}{n_1} = 1/2,$$

that is, two-thirds of our data is used to decide whether to pursue inference in the second stage. We vary $\bar{\beta}$ in the set $\{2, 1, 0\}$. For comparison, we consider asymptotic intervals based on the widely used data splitting. The latter procedure simply uses the $n_2$ samples that were held out for inference.

We compare the 90%-confidence intervals from inverting the carved pivot with the 90%-confidence intervals from data splitting and summarize our findings in Table 1. Our method is noted as "Carve" and data splitting is noted as "Split." The cells in this table report the empirical coverage rate "Cov" of the asymptotic confidence intervals and their lengths "Len" when averaged over all our simulations. The first column in the table notes the performance of the exact confidence intervals under the baseline Gaussian model.

As expected, both procedures approximately achieve the target coverage rate. However, carving produces tighter intervals than data splitting across all models and all values of $\bar{\beta}$.

EXAMPLE 6.2. Selectively inferring for the $p$ largest effects. We consider selective inference for the effects of the $p$ largest mean statistics in our pilot data [6]. Let $[V_{n_1}]^{(p)}$ be the $p^{\text{th}}$ largest mean statistic using the components of $V_{n_1}$. We note that our selection rule in this example can be written as

$$\begin{aligned} V_{n_1}^{(j)} &> [V_{n_1}]^{(p+1)} \quad \text{for } j \in E_n, \\ V_{n_1}^{(j)} &\leq [V_{n_1}]^{(p+1)} \quad \text{for } j \in E_n^c. \end{aligned} \tag{6.1}$$

TABLE 2
*Comparison of inference between carving and data splitting*

| | Gaussian | | Model-1 | | Model-2 | | Model-3 | | Model-4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho^2 = 1/2$ | Cov | Len | Cov | Len | Cov | Len | Cov | Len | Cov | Len |
| $\bar{\beta} = -2.5$ | | | | | | | | | | |
| Carve | 93% | 0.60 | 88% | 0.64 | 91.5% | 0.62 | 91% | 0.61 | 87% | 0.59 |
| Split | 90% | 0.95 | 90.5% | 0.95 | 91.5% | 0.95 | 93.5% | 0.95 | 90.5% | 0.95 |
| $\bar{\beta} = -1.5$ | | | | | | | | | | |
| Carve | 92% | 0.68 | 89% | 0.69 | 90% | 0.68 | 87.5% | 0.66 | 90% | 0.66 |
| Split | 91.5% | 0.95 | 93% | 0.95 | 92.5% | 0.95 | 90.5% | 0.95 | 92.5% | 0.95 |
| $\bar{\beta} = 0$ | | | | | | | | | | |
| Carve | 91% | 0.55 | 87.5% | 0.56 | 87.5% | 0.52 | 87% | 0.56 | 90.5% | 0.55 |
| Split | 92% | 0.95 | 88.5% | 0.95 | 88% | 0.95 | 90% | 0.95 | 88.5% | 0.95 |

Suppose that $\mathbb{P}_n = \mathcal{N}(\beta_n, \Sigma)$. Lemma 8 gives a carved pivot after conditioning on the event

$$\{E_n = E_{\text{obs}}, A_n = A_{\text{obs}}\},$$

where

$$A_{\text{obs}} = \left( \left( \sqrt{1+\rho^2}[V_{n_1}]^{(p+1)} \cdot 1_p \right)' \quad \left( V_n^{(E_n^c)} + W_n^{(E_n^c)} \right)' \right)' = \begin{pmatrix} A'_{1,n} & A'_{2,n} \end{pmatrix}'.$$

To state the pivot, define the matrices

$$R^{(j)} = \mathcal{P}_{E_{\text{obs}}} \begin{bmatrix} 1 & 0 \\ \frac{1}{\sigma_j^2}\Sigma_{-j,j} & I_{d-1,d-1} \end{bmatrix}, \qquad Q = \begin{bmatrix} I_{p,p} \\ 0_{d-p,p} \end{bmatrix}, \qquad r = A_{\text{obs}}.$$

PROPOSITION 8. *Let* $\text{Pivot}^{(j)}(V_n^{(j)}, U_n^{(j)})$ *assume the value*

$$\left(\mathrm{D}(U_n^{(j)}; \sqrt{n}\beta_n^{(j)})\right)^{-1} \cdot \int_{V_n^{(j)}}^{\infty} \phi\left( \frac{1}{\sigma_j}(v - \sqrt{n}\beta_n^{(j)}) \right) \cdot \mathrm{F}\left( R^{(j)} \begin{pmatrix} v & (U_n^{(j)})' \end{pmatrix}' \right) dv.$$

*Then it holds that* $\text{Pivot}^{(j)}(V_n^{(j)}, U_n^{(j)})$ *is distributed as a* $\text{Unif}(0,1)$ *conditional on* $\{E_n = E_{\text{obs}}, A_n = A_{\text{obs}}\}$.

Clearly, this pivot has the same representation as our running example.

Using the generating scheme from the preceding example, we selectively infer for the effect that corresponds to the larger sample mean. A similar comparison between carving and data splitting unfolds in Table 2 for different models.

EXAMPLE 6.3. We turn to inference for the selected regression coefficients after solving the LASSO. Let $y_n$ and $X_n$ denote our response vector and our design matrix with $d$ predictors, respectively.

To begin, we derive a pivot using a randomized rule with Gaussian variables. Consider solving

$$\underset{\beta \in \mathbb{R}^d}{\text{minimize}} \frac{1}{2\sqrt{n}} \|y_n - X_n\beta\|_2^2 + \lambda\|\beta\|_1 - W_n'\beta, \tag{6.2}$$

where $W_n$ is a Gaussian randomization variable. This problem has been termed as the randomized LASSO in [23].

After observing the selected set of variables $E_n = E_{\text{obs}}$, a common model for inference is the selected model

$$y_n \sim \mathcal{N}(X_{n,E_{\text{obs}}}\beta_n, \sigma^2 I).$$

Define

$$\widehat{\beta}_n^{(E_{\text{obs}})} = \big((X_n^{(E_{\text{obs}})})' X_n^{(E_{\text{obs}})}\big)^{-1}(X_n^{(E_{\text{obs}})})' y_n,$$

the refitted least squares estimator, which is obtained by regressing our response against the selected variables. Based on the least squares estimator and the selected set of variables, let

$$\begin{pmatrix} V_n^{(E_{\text{obs}})} \\ V_n^{(E_{\text{obs}}^c)} \end{pmatrix} = \begin{pmatrix} \sqrt{n}\widehat{\beta}_n^{(E_{\text{obs}})} \\ \frac{1}{\sqrt{n}}(X_n^{(E_{\text{obs}}^c)})'(y_n - X_n^{(E_{\text{obs}})}\widehat{\beta}_n^{(E_{\text{obs}})}) \end{pmatrix}, \tag{6.3}$$

and let

$$V_n^{(j)} = e_j' \sqrt{n}\widehat{\beta}_n^{(E_{\text{obs}})},$$

which is the $j$th regression coefficient in the selected set.

Fixing some more notation, let

$$\begin{pmatrix} \widehat{\beta}_{n,\lambda} \\ 0_{d-p} \end{pmatrix}$$

denote the coefficients of the LASSO solution, where $\widehat{\beta}_{n,\lambda}$ collects its nonzero coefficients. Let $S_n^{(E_n)}$ collect the signs of the nonzero LASSO coefficients. Let $\mathcal{G}_n^{(E_n^c)}$ collect the components of the subgradient from the LASSO penalty present in the inactive set $E_n^c$ at the solution. Define

$$A_n = \begin{pmatrix} A_{1,n}' & A_{2,n}' \end{pmatrix}' = \begin{pmatrix} \lambda \cdot (S_n^{(E_n)})' & (\mathcal{G}_n^{(E_n^c)})' \end{pmatrix}',$$

which we note is equal to subgradient of the LASSO penalty at the solution. Finally, let $T_n = \operatorname{diag}(S_n^{(E_n)})\widehat{\beta}_{n,\lambda}$ collect the magnitudes of the nonzero LASSO coefficients.

Based on these notation, fix the following matrices:

$$P_n = \begin{bmatrix} \frac{1}{n}(X_n^{(E_{\text{obs}})})' X_n^{(E_{\text{obs}})} & 0_{p,d-p} \\ \frac{1}{n}(X_n^{(E_{\text{obs}}^c)})' X_n^{(E_{\text{obs}})} & I_{d-p,d-p} \end{bmatrix}, \qquad Q_n = \begin{bmatrix} \frac{1}{n}(X_n^{(E_{\text{obs}})})' X_n^{(E_{\text{obs}})} \\ \frac{1}{n}(X_n^{(E_{\text{obs}}^c)})' X_n^{(E_{\text{obs}})} \end{bmatrix} \operatorname{diag}(S_n^{(E_{\text{obs}})}).$$

Let $P = \mathbb{E}_{\mathbb{P}_n}[P_n]$ and $Q = \mathbb{E}_{\mathbb{P}_n}[Q_n]$, and also let $\sigma_j^2 = \sigma^2 \cdot \Sigma_{j,j}^{(E_{\text{obs}})}$ where

$$\Sigma^{(E_{\text{obs}})} = \left(\mathbb{E}_{\mathbb{P}_n}\left[\frac{1}{n}(X_n^{(E_{\text{obs}})})' X_n^{(E_{\text{obs}})}\right]\right)^{-1}.$$

Suppose that the randomization variable $W_n$ in (6.2) is drawn from the Gaussian distribution $\mathcal{N}(0_d, \rho^2\Sigma)$, independently of data, where

$$\Sigma = \sigma^2 \cdot \mathbb{E}_{\mathbb{P}_n}\left[\frac{1}{n}X_n' X_n\right].$$

For now, we assume that:

(i) the variables in (6.3) are distributed as Gaussian variables, where $V_n^{(E_{\text{obs}})}$ has mean $\sqrt{n}\beta_n$ and covariance

$$\sigma^2 \cdot \left( \mathbb{E}_{\mathbb{P}_n} \left[ \frac{1}{n} (X_n^{(E_{\text{obs}})})' X_n^{(E_{\text{obs}})} \right] \right)^{-1} = \sigma^2 \cdot \Sigma^{(E_{\text{obs}})},$$

and $V_n^{(E_{\text{obs}})}$ is independent of $V_n^{(E_{\text{obs}}^c)}$.

(ii) the magnitudes of the nonzero LASSO coefficients satisfy

$$\left( W_n^{(E_{\text{obs}})'} \quad W_n^{(E_{\text{obs}}^c)'} \right)' = QT_n + \left( A_{1,n}' \quad A_{2,n}' \right)' - P \left( V_n^{(E_{\text{obs}})'} \quad V_n^{(E_{\text{obs}}^c)'} \right)'.$$

Later we show that the variables $V_n$ have an asymptotic Gaussian distribution with the properties listed in (i), and the equality in (ii) holds only up to an $o_p(1)$ remainder term.

Proposition 9 gives a pivot that yields exactly-valid selective inference under the above-stated randomized rule and assumptions.

PROPOSITION 9. *Let* $\text{Pivot}^{(j)}(V_n^{(j)}, U_n^{(j)})$ *assume the value*

$$\left(\mathrm{D}(U_n^{(j)}; \sqrt{n}\beta_n^{(j)})\right)^{-1} \cdot \int_{V_n^{(j)}}^{\infty} \phi\left( \frac{1}{\sigma_j} (v - \sqrt{n}\beta_n^{(j)}) \right) \cdot \mathrm{F}\left( \mathrm{PR}^{(j)} \left( v \quad (U_n^{(j)})' \right)' \right) dv,$$

*where*

$$\mathrm{D}(U; \sqrt{n}\beta_n^{(j)}) = \int_{-\infty}^{\infty} \phi\left( \frac{1}{\sigma_j} (v - \sqrt{n}\beta_n^{(j)}) \right) \cdot \mathrm{F}\left( \mathrm{PR}^{(j)} \left( v \quad U' \right)' \right) dv.$$

*Conditional on* $\{E_n = E_{\text{obs}}, A_n = A_{\text{obs}}\}$, $\text{Pivot}^{(j)}(V_n^{(j)}, U_n^{(j)})$ *is distributed as a* $\text{Unif}(0, 1)$ *variable.*

Suppose that our data contains $n$ independent and identically distributed observations. We solve the LASSO problem on a randomly drawn subsample of size $n_1$:

$$\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \frac{(1+\rho^2)}{2\sqrt{n}} \|y_{n_1} - X_{n_1}\beta\|_2^2 + \lambda \|\beta\|_1. \tag{6.4}$$

We define

$$W_n = \frac{\partial}{\partial \beta} \left\{ \frac{1}{2\sqrt{n}} \|y_n - X_n\beta\|_2^2 - \frac{(1+\rho^2)}{2\sqrt{n}} \|y_{n_1} - X_{n_1}\beta\|_2^2 \right\} \Big|_{\widehat{\beta}^{\lambda}}.$$

According to previous work by [11, 18], the LASSO optimization problem can be rewritten as (6.2). The randomization variable $W_n$ is distributed asymptotically as $\mathcal{N}(0_d, \rho^2\Sigma)$, where $\rho^2 = \frac{n_2}{n_1}$. Furthermore, it is asymptotically independent of $V_n$. It is also worth noting that the variables $V_n$ follow an asymptotic Gaussian distribution with the properties listed in (i). See, for example, Proposition 4.1 in [18], which gives the joint distribution of $W_n$ and $V_n$.

Furthermore, we can verify that

$$\left( W_n^{(E_{\text{obs}})'} \quad W_n^{(E_{\text{obs}}^c)'} \right)' + \mathrm{O}_n = QT_n + \left( A_{1,n}' \quad A_{2,n}' \right)' - P \left( V_n^{(E_{\text{obs}})'} \quad V_n^{(E_{\text{obs}}^c)'} \right)',$$

where $\mathrm{O}_n = o_p(1)$. In what follows, we ignore the $o_p(1)$ remainder term without a loss of generality. We can always work with the variable

$$\widetilde{W}_n = W_n + \mathrm{O}_n,$$

which has the same asymptotic distribution as $W_n$.

The theory in our paper confirms that the pivot in Proposition 9 enables us to draw asymptotically-valid inference for the selected regression coefficients. Below, we provide

TABLE 3
*Comparison of inference between carving and data splitting*

| | Gaussian | | Model-1 | | Model-2 | | Model-3 | | Model-4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho^2 = 1$ | Cov | Len | Cov | Len | Cov | Len | Cov | Len | Cov | Len |
| snr = 0.10 | | | | | | | | | | |
| Carve | 88.15% | 0.43 | 88.05% | 0.42 | 89.59% | 0.42 | 90.39% | 0.42 | 90.73% | 0.43 |
| Split | 88.56% | 0.52 | 88.15% | 0.50 | 90.24% | 0.50 | 88.75% | 0.50 | 88.75% | 0.50 |
| snr = 0.15 | | | | | | | | | | |
| Carve | 88.61% | 0.43 | 88.28% | 0.42 | 88.56% | 0.42 | 89.06% | 0.44 | 87.69% | 0.42 |
| Split | 89.27% | 0.50 | 90.26% | 0.51 | 89.78% | 0.50 | 89.00% | 0.53 | 85.51% | 0.50 |
| snr = 0.20 | | | | | | | | | | |
| Carve | 91.22% | 0.43 | 88.53% | 0.42 | 89.18% | 0.43 | 92.88% | 0.43 | 89.94% | 0.42 |
| Split | 88.65% | 0.51 | 88.15% | 0.50 | 90.54% | 0.52 | 88.72% | 0.51 | 89.97% | 0.51 |

empirical evidence to support our theory by demonstrating the performance of our pivot in both synthetic and real data experiments.

**Synthetic data**. Fix $n = 100$ and $d = 50$. In each round of our simulations, we draw an $n \times d$ design matrix $X$ such that the rows $x_i \sim \mathcal{N}(0_d, \Sigma)$ and $\Sigma_{j,k} = 0.40^{|j-k|}$. We then draw our response according to the model

$$y_i = x_i'\beta + \sigma \cdot \mathrm{e}_{i,n},$$

by generating the model errors $\mathrm{e}_{i,n}$ in an i.i.d. fashion from Models (1)–(4) and the baseline Gaussian model. We let $\beta \in \mathbb{R}^d$ be a sparse vector with $s = 5$ signals, all of the same strength and positioned randomly in the $d$-length vector. Each signal is assigned a positive sign with probability 0.5. We fix $\sigma^2 = 1$, $\rho^2 = 1$ and vary $\beta$ such that the signal-to-noise ratio $\mathrm{snr} = \frac{1}{\sigma^2}\beta'\Sigma\beta$ takes values in the set

$$\{0.10, 0.15, 0.20\}.$$

In this example, the function F and our pivot no longer have a closed-form expression. To alleviate this computational barrier, we use a Laplace-type probabilistic approximation proposed by [14] to compute F. Inverting the approximate pivot yields asymptotic confidence intervals based on our carved pivot. The cells in Table 3 compare the 90%-confidence intervals based on carving and data splitting. We note that our asymptotic intervals not only cover the selected regression parameters at the desired level, but also provide tighter bounds than data splitting. Furthermore, selective inference is valid even at lower values of signal-to-noise ratio, where rare outcomes are more likely.

**Real data**. We apply our carved pivot on real data. Our data comes from 441 patients in the publicly available The Cancer Genome Atlas (TCGA) database [27]. Carving is applied to infer for the selected associations between gene expression values and log-transformed survival times for Gliomas, a common type of brain tumor. We include 2500 predictors with the highest variability in the observed samples and solve the LASSO on a randomly drawn subsample of the full data. The $\ell_1$ penalty tuning parameter is fixed at a theoretical value that was sugggested by [12].

We obtain confidence intervals for the selected regression coefficients by inverting the carved pivot. Figure 2 shows the distribution of lengths of the confidence intervals based on carving and data splitting. On the x-axis, we vary the ratio $1/(1 + \rho^2)$. The plot demonstrates the advantages of conducting selective inference with the carved pivot, which reuses data from selection steps. Interval estimates for both procedures grow wider when fewer hold-out samples are available for inference. However, the benefits of carving only become more pronounced as more data is used at the selection step.

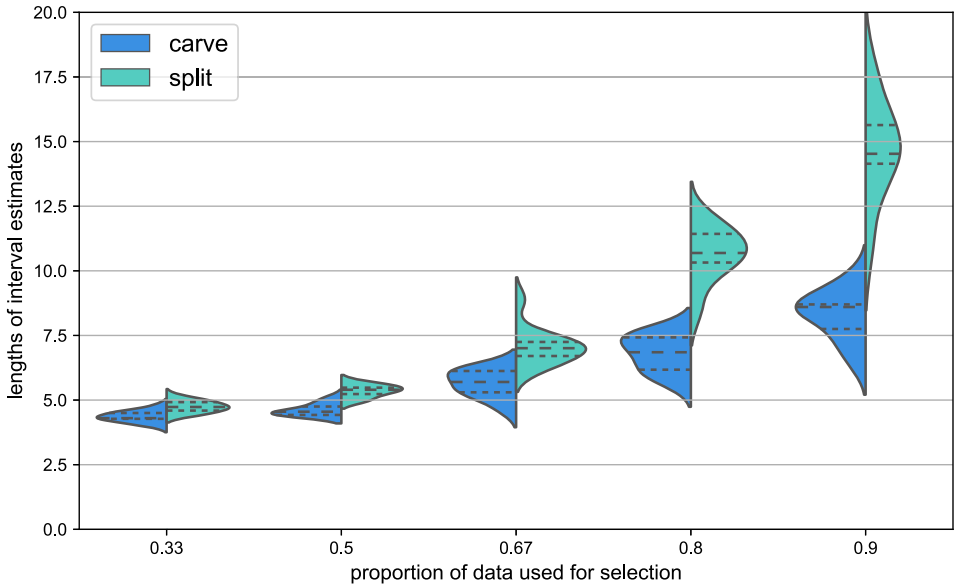


FIG. 2. *Distribution of lengths of interval estimates for the selected regression coefficients.*

**7. Conclusion.** Our paper presents an asymptotic framework for carving as we move away from Gaussian data. We consider two data sets: one for selection and the other reserved for inference. This situation often occurs when researchers select promising findings from pilot data. Later on, when new data is available, the goal is to make inferences based on the selected findings.

Carving helps to adjust for overoptimism resulting from selection, and also allows for efficient reuse of pilot data for inference. We show that pivots formed by conditioning on the selection outcome in the pilot data provide valid asymptotic inference. Our theory also supports the use of pivots based on Gaussian randomized selection rules. Recent studies, such as [15, 19, 21, 22], have explored the potential of randomized selection rules for improved inference, in theory and various applications.

Although we have mainly focused on conditional pivots based on the standard recipe in this paper, there is still room for further research in this area. In the future, we plan to investigate other types of pivots that have been developed for conditional inference. Two examples are the approximate Gaussian pivot by [17], which uses the maximum likelihood estimator, and the pivots proposed by [10] in the full model with less conditioning than the earlier work by [9]. However, for such pivots, new theoretical results are needed to study the rate of weak convergence and to examine whether asymptotically-valid selective inference still holds when self-normalized statistics are used to form the pivots.

**Acknowledgments.** S.P. would like to thank Jonathan Taylor, Liza Levina, Xuming He and two anonymous referees for providing several insightful comments on an initial draft of the paper.

**Funding.** S.P. acknowledges support in part by NSF Grants DMS-1951980 and DMS-2113342.

## SUPPLEMENTARY MATERIAL

**Supplement to "Carving model-free inference"** (DOI: 10.1214/23-AOS2318SUPP; .pdf). The Supplementary Material includes proofs for all results stated in the manuscript, along with additional results supporting the main theory.

## REFERENCES


[1] BENJAMINI, Y. (2020). Selective inference: The silent killer of replicability. *Harv. Data Sci. Rev.* **2**.

[2] CHATTERJEE, S. and MECKES, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 257–283. MR2453473

[3] CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications* (*New York*). Springer, Heidelberg. MR2732624 https://doi.org/10.1007/978-3-642-15007-4

[4] DE ACOSTA, A. (1992). Moderate deviations and associated Laplace approximations for sums of independent random vectors. *Trans. Amer. Math. Soc.* **329** 357–375. MR1046015 https://doi.org/10.2307/2154092

[5] FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. arXiv preprint. Available at arXiv:1410.2597.

[6] GUO, X. and HE, X. (2021). Inference on selected subgroups in clinical trials. *J. Amer. Statist. Assoc.* **116** 1498–1506. MR4309288 https://doi.org/10.1080/01621459.2020.1740096

[7] HASHORVA, E. and HÜSLER, J. (2003). On multivariate Gaussian tails. *Ann. Inst. Statist. Math.* **55** 507–522. MR2007795 https://doi.org/10.1007/BF02517804

[8] KIVARANOVIC, D. and LEEB, H. (2020). A (tight) upper bound for the length of confidence intervals with conditional coverage. arXiv preprint. Available at arXiv:2007.12448.

[9] LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

[10] LIU, K., MARKOVIC, J. and TIBSHIRANI, R. (2018). More powerful post-selection inference. with application to the lasso. arXiv preprint. Available at arXiv:1801.09037.

[11] MARKOVIC, J. and TAYLOR, J. (2016). Bootstrap inference after using multiple queries for model selection. arXiv preprint. Available at arXiv:1612.07811.

[12] NEGAHBAN, S., YU, B., WAINWRIGHT, M. J. and RAVIKUMAR, P. K. (2009). A unified framework for high-dimensional analysis of $m$-estimators with decomposable regularizers. In *Advances in Neural Information Processing Systems* 1348–1356.

[13] PANIGRAHI, S. (2023). Supplement to "Carving model-free inference." https://doi.org/10.1214/23-AOS2318SUPP

[14] PANIGRAHI, S., MARKOVIC, J. and TAYLOR, J. (2017). An MCMC free approach to post-selective inference. arXiv preprint. Available at arXiv:1703.06154.

[15] PANIGRAHI, S., MOHAMMED, S., RAO, A. and BALADANDAYUTHAPANI, V. (2023). Integrative Bayesian models using post-selective inference: A case study in radiogenomics. *Biometrics* **79** 1801–1813. MR4643957 https://doi.org/10.1111/biom.13740

[16] PANIGRAHI, S. and TAYLOR, J. (2018). Scalable methods for Bayesian selective inference. *Electron. J. Stat.* **12** 2355–2400. MR3832095 https://doi.org/10.1214/18-EJS1452

[17] PANIGRAHI, S. and TAYLOR, J. (2022). Approximate selective inference via maximum likelihood. *J. Amer. Statist. Assoc.* 1–11.

[18] PANIGRAHI, S., TAYLOR, J. and WEINSTEIN, A. (2021). Integrative methods for post-selection inference under convex constraints. *Ann. Statist.* **49** 2803–2824. MR4338384 https://doi.org/10.1214/21-aos2057

[19] PANIGRAHI, S., WANG, J. and HE, X. (2022). Treatment effect estimation with efficient data aggregation. arXiv preprint. Available at arXiv:2203.12726.

[20] PANIGRAHI, S., ZHU, J. and SABATTI, C. (2021). Selection-adjusted inference: An application to confidence intervals for *cis*-eQTL effect sizes. *Biostatistics* **22** 181–197. MR4207150 https://doi.org/10.1093/biostatistics/kxz024

[21] RASINES, D. G. and YOUNG, G. A. (2023). Splitting strategies for post-selection inference. *Biometrika* **110** 597–614. MR4627773 https://doi.org/10.1093/biomet/asac070

[22] SCHULTHEISS, C., RENAUX, C. and BÜHLMANN, P. (2021). Multicarving for high-dimensional post-selection inference. *Electron. J. Stat.* **15** 1695–1742. MR4255311 https://doi.org/10.1214/21-ejs1825

[23] TIAN, X., PANIGRAHI, S., MARKOVIC, J., BI, N. and TAYLOR, J. (2016). Selective sampling after solving a convex problem. arXiv preprint. Available at arXiv:1609.05609.

[24] TIAN, X. and TAYLOR, J. (2018). Selective inference with a randomized response. *Ann. Statist.* **46** 679–710. MR3782381 https://doi.org/10.1214/17-AOS1564

[25] TIBSHIRANI, R. J., RINALDO, A., TIBSHIRANI, R. and WASSERMAN, L. (2018). Uniform asymptotic inference and the bootstrap after model selection. *Ann. Statist.* **46** 1255–1287. MR3798003 https://doi.org/10.1214/17-AOS1584

[26] TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. MR3538689 https://doi.org/10.1080/01621459.2015.1108848

[27] TOMCZAK, K., CZERWIŃSKA, P. and WIZNEROWICZ, M. (2015). Review the cancer genome atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol./Współczesna Onkologia* **2015** 68–77.

[28] ZRNIC, T. and JORDAN, M. I. (2023). Post-selection inference via algorithmic stability. *Ann. Statist.* **51** 1666–1691. MR4658572 https://doi.org/10.1214/23-aos2303