Post ANOVA *Practical Efficiency* for a Quadratic Model with Numerical Factors, Few Significant Factors + Normal Dist.

Pedro H R Bruel

February 11, 2019

Optimizing problems with a large number of configurable parameters, or *factors*, is one of the main objectives of our transparent Design of Experiments approach to autotuning. Despite the high dimensionality of these problems, it is usually the case that only a small number of factors impacts the measured performance metrics significantly. Our approach aims to identify the most significant factors using an initial performance model, D-Optimal designs, and Analysis of Variance.

add reference

Our experiments with SPAPT kernels obtained results similar to a uniform sampling strategy, while using significantly less evaluations. Since it was unfeasible to evaluate all possible configurations of each SPAPT kernel due to their large search spaces, the experimental designs used in each iteration of our iterative approach were selected from a limited sample of configurations. The size of these samples were always small relative to the size of each search space, due to the high dimension and number of values for each factor.

1 Sampling Strategies

A practical concern derived from the limitation on the size of samples to choose a design from is the large amount of time it takes to evaluate a relatively large number of candidate configurations. Each search space also had several constraints on factor levels, reducing the valid search space and slowing the process of finding candidates to test. Another concern is the intuition that the quality of the designs constructed from uniformly sampled candidates should always be poorer than in the case where designs are

improve, add reference

selected from complete search spaces. This property should be more pronounced when certain types of performance models are selected, such as quadratic models, due to the fact that practically none of the points sampled uniformly from a high-dimensional search space would be in the central region of the space.

Table 1: Design construction techniques and target performance models

Sampling & Construction Strategy	Performance Models
Uniform	Linear & quadratic
Biased	Linear
Biased + Normal	Quadratic
Uniform + *Federov's*	Linear & quadratic
Biased + *Federov's*	Linear
Biased + Normal + *Federov's*	Quadratic

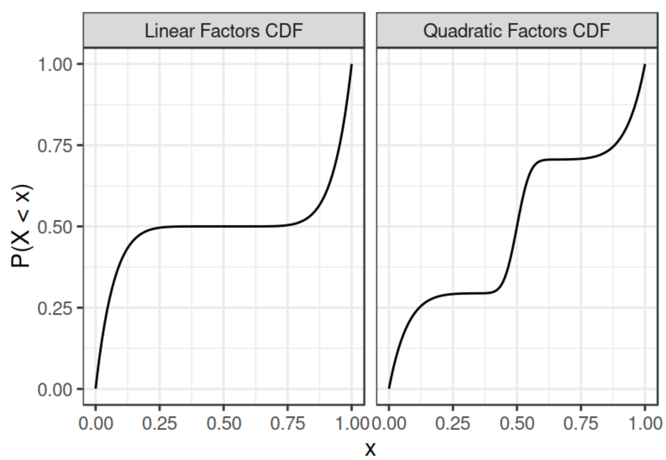


Figure 1: Cumulative Distribution Functions used for Biased Samples

These concerns motivated the development of a methodology to evaluate the quality of the designs produced by *Fedorov's algorithm*, the chosen D-Optimal construction technique, in the cases where designs were selected from samples with different properties. We studied design efficiency, measured by the *D-criterion*, for designs constructed using the performance models and construction techniques listed in Table 1. Sampling strategies consisted of obtaining samples using the *cumulative distribution functions* in Figure 1, and adding samples from a normal distribution to the quadratic

add reference

model biased samples.

2 Comparing Sampling Strategies

In addition to the D-Criterion value, we also measured the quality of the designs generated by each technique using the *euclidean distance* between the vector of responses generated by different models fitted using the results of a design's experiments. We compared the results of a model using all factors and of a re-fitted model using only significant factors identified with ANOVA. We performed this comparison for all sampling strategies in Table 1.

We decided to sample new functions for each repetition of our experiments with D-Criteria and model fits. This was initially done by sampling values for the coefficients for each model term, generating a new function that was used to compute the results corresponding to the factor values of each of a design's experiments. Coefficient sampling is described in more detail in the following Section.

2.1 Sampling Function Coefficients

Consider the functions $f_{\boldsymbol{\alpha},\boldsymbol{\beta}} \colon \mathbf{X} \in \mathbb{R}^n \to \mathbb{R}$ of the form

$$f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{X}) = \alpha_0 + \sum_{i=1}^{n} \left(\alpha_i x_i + \beta_i x_i^2\right) + \varepsilon,$$

where $x \in [-1, 1]\ \forall x \in \mathbf{X}$, with *coefficients of linear terms* $\boldsymbol{\alpha} \in \mathbb{R}^{n+1}$, *coefficients of quadratic terms* $\boldsymbol{\beta} \in \mathbb{R}^n$, and normally distributed error ε. The procedure we used to sample new coeficients at each experiment starts by choosing which of the $\alpha_i \in \boldsymbol{\alpha}$, $\beta_i \in \boldsymbol{\beta}$ will be significant. With our target scenario in mind, each coefficient had a 15% chance of being significant. A uniformly sampled value in the interval $[-x_{\text{sup}}, -x_{\text{inf}}] \cup [x_{\text{inf}}, x_{\text{sup}}]$ is chosen for each significant coefficient, where $x_{\text{sup}}, x_{\text{inf}} \in \mathbb{R}$, $x_{\text{sup}} > x_{\text{inf}} > 0$. Small normally distributed values with mean zero and standard deviation x_{sd} are assigned for the other factors. Figure 2 shows one coefficient sample obtained by this process, where $|\boldsymbol{\alpha}| + |\boldsymbol{\beta}| = 2|\boldsymbol{X}| + 1 = 121$, $x_{\text{inf}} = 1$, $x_{\text{sup}} = 7$, and $x_{\text{sd}} = 0.04$.

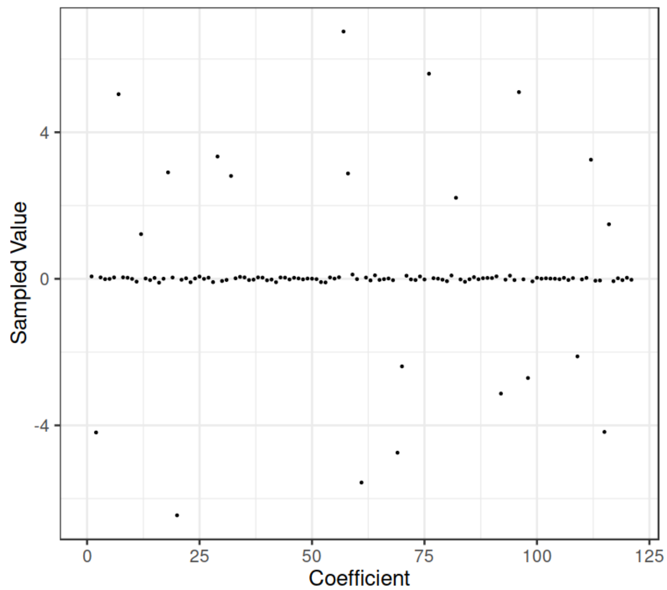


Figure 2: Sampled coefficients

2.2 Comparing Response Vectors

Consider a design $\xi_{n,k}$ with factors $x_{k1}, \ldots, x_{kn}$, experiments $\boldsymbol{X_1}, \ldots, \boldsymbol{X_k}$, and design matrix given by

$$\xi_{n,k} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \ldots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \ldots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \ldots & x_{kn} \end{bmatrix},$$

where the k experiments were chosen by Fedorov's algorithm from a relatively large sample of size K.

Using a function $f_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ sampled as described in the previous section, we can compute the value or *response vector* $\boldsymbol{y} = (y_1, \ldots, y_k)$ for each experiment vector $\boldsymbol{X_i} = (x_{i1}, \ldots, x_{in})$, obtaining the design $\xi'_{n,k}$ with design matrix given by

$$\xi'_{n,k} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \ldots & x_{1n} & f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{X}_1) \\ x_{21} & x_{22} & x_{23} & \ldots & x_{2n} & f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{X}_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \ldots & x_{kn} & f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{X}_k) \end{bmatrix}.$$

We then use $\xi'_{n,k}$ to perform a linear regression with model

$$f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{X}_i) = y_i = \alpha'_0 + \boldsymbol{X}_i^{\top}(\alpha'_2, \ldots, \alpha'_{n+1}) + \boldsymbol{X}_i^{2\top}(\beta'_1, \ldots, \beta'_n) + \varepsilon,$$

obtaining coefficients that define the function $f_{\boldsymbol{\alpha}',\boldsymbol{\beta}'}$ which approximates $f_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. Evaluating the fitted model in the original large sample of size K gives the response vector

$$\boldsymbol{Y}' = (f_{\boldsymbol{\alpha}',\boldsymbol{\beta}'}(\boldsymbol{X}_1), \ldots, f_{\boldsymbol{\alpha}',\boldsymbol{\beta}'}(\boldsymbol{X}_K)),$$

and evaluating the true sampled function gives the response vector

$$\boldsymbol{Y} = (f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{X}_1), \ldots, f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{X}_K)).$$

The *response vector distance* $d(\mathbf{Y}, \mathbf{Y}') = \|\mathbf{Y} - \mathbf{Y}'\|^2$ provides a measure of how well the fitted model approximates the real function in the initial sample of size K.

2.3 Comparing D-Criteria

3 Results

The following results show 100 repetitions of the experiments described below. The first step of each iteration is generating model coefficients. Since it better mirrors our targeted use case, we decided to generate coefficient sets where only a small amount of coefficients is allowed to be significant, 20% in this case. Significant coefficients are picked from an uniform distribution in the interval $[-5, -3] \cup [3, 5]$, and unsignificant coefficients are picked from a normal distribution in the interval $[-0.2, 0.2]$.

After generating the coefficients, we generate a set of experiments for comparing model fits. Then, each iteration generates designs of size 70 for 30 numerical factors in the interval $[-1.0, 1.0]$, using the strategies listed below:

Strategy	Description
Uniform Sample	Pick 70 experiments from a uniform distribution
Biased Sample	Pick 70 experiments from a biased distribution
Federov w/ Uniform Sample	Generate uniformly a set of size 1000 exp.; pick 70 w.r.t. D
Federov w/ Biased Sample	Generate biasedly a set of size 1000 exp.; pick 70 w.r.t. D

Responses are computed using the generated coefficient set, with added noise from a normal distribution with standard deviation $sd = 2$ and mean zero. The results are used to fit linear models, which are compared with the actual model generated for that iteration using 2 model metrics and 1 design metric. The figures below compare the `coefficient_distance`, the `linear_fit_distance` and the `D` values of the 100 repetitions. These values are computed as follows:

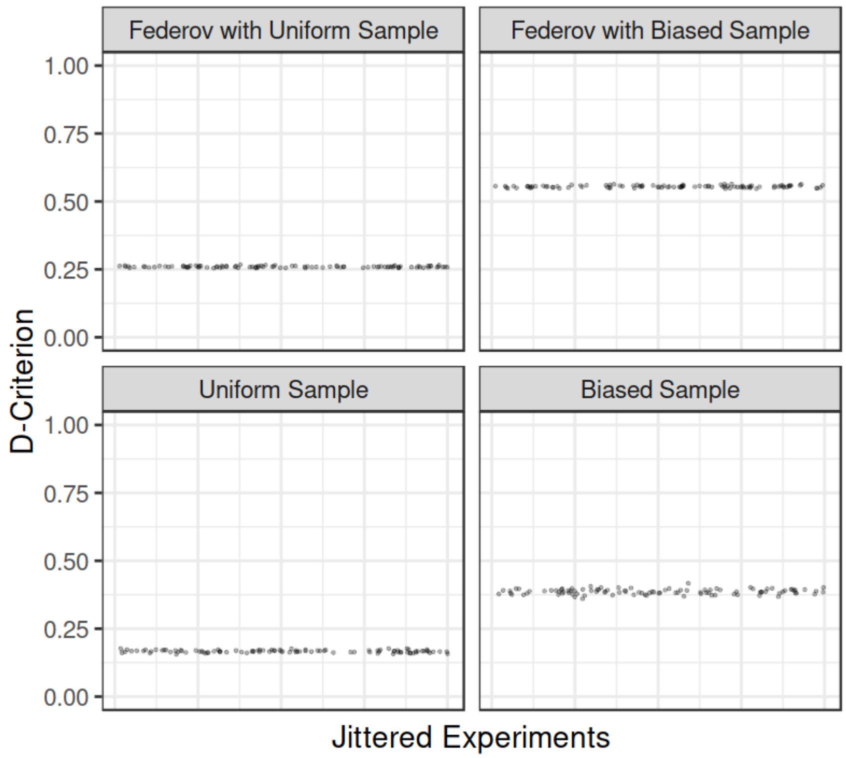
Value	Computation
`coefficient_distance`	Euclidean distance between fitted & real model coefficients
`linear_fit_distance`	Euclidean distance between predicted & real response
`D`	Computed as usual by `AlgDesign::eval.design()`

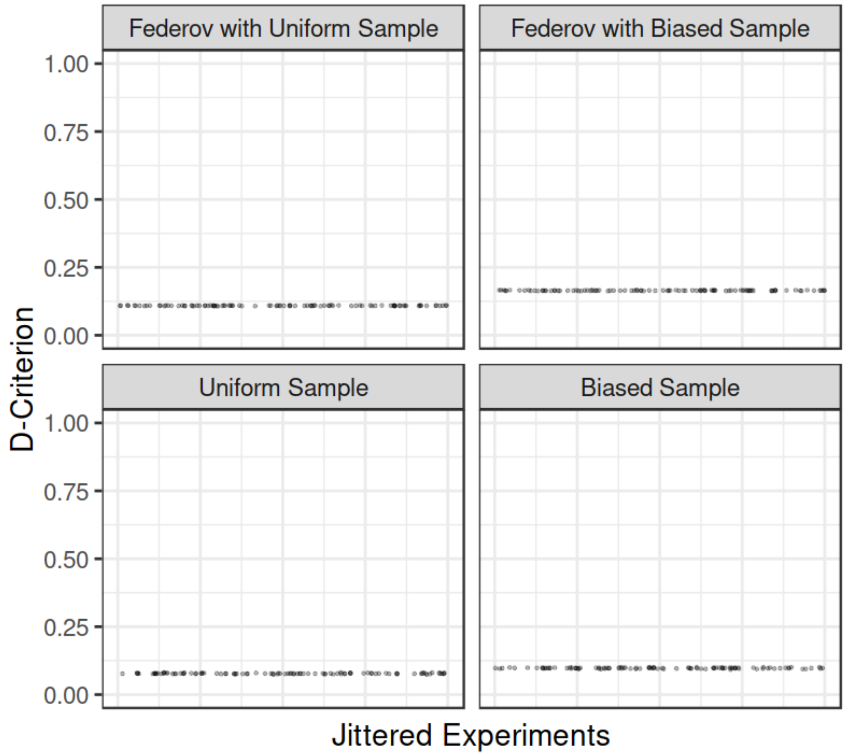
I've also computed the distance between the following two metrics, in an attempt to detect the model's hability to identify significant coefficients in a set were most coefficients do not impact the response.

Value	Computation
`real_coefficients_prf_sum`	Sum of the fitted model's Pr(>F) values for actual coeffs.
`noise_coefficients_prf_sum`	Sum of the fitted model's Pr(>F) values for noise coeffs.

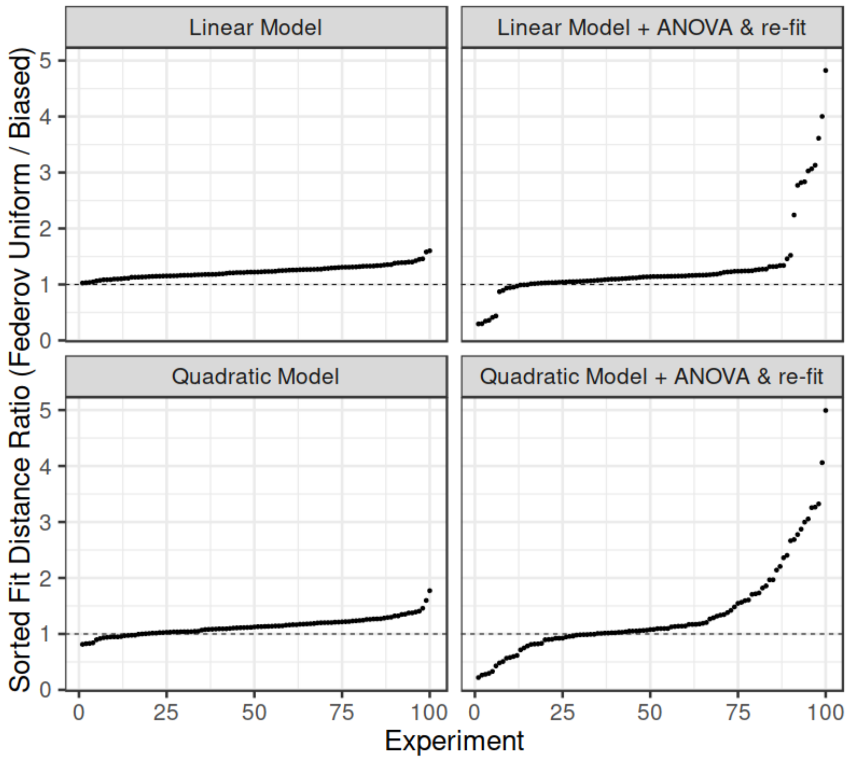
3.1 Plots

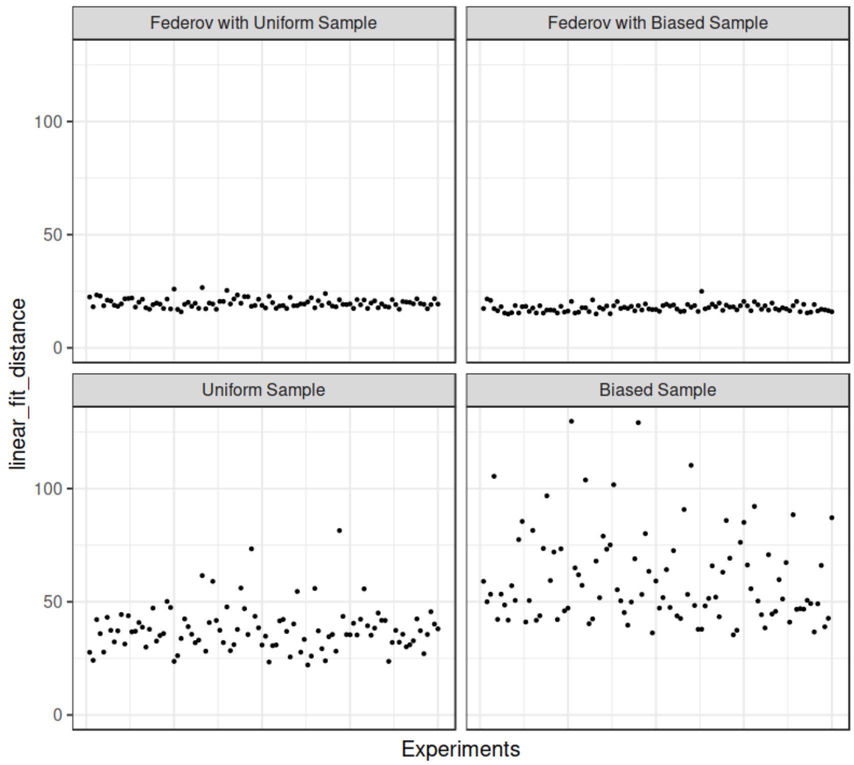
3.1.1 D-Criterion



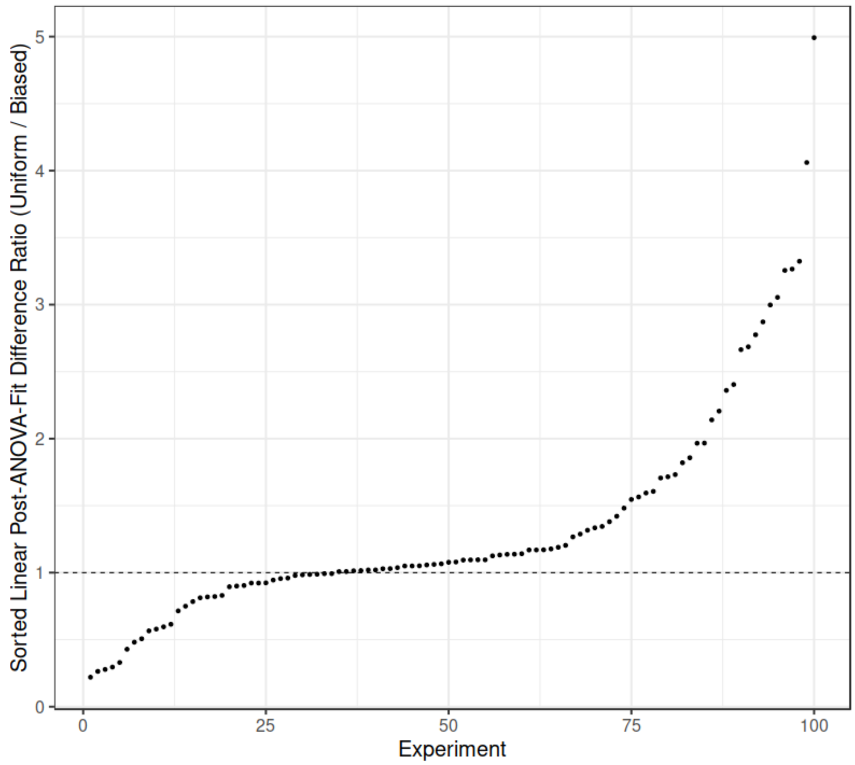

Federov with Uniform Sample
Federov with Biased Sample
Uniform Sample
Biased Sample
D-Criterion
1.00
0.75
0.50
0.25
0.00
Jittered Experiments

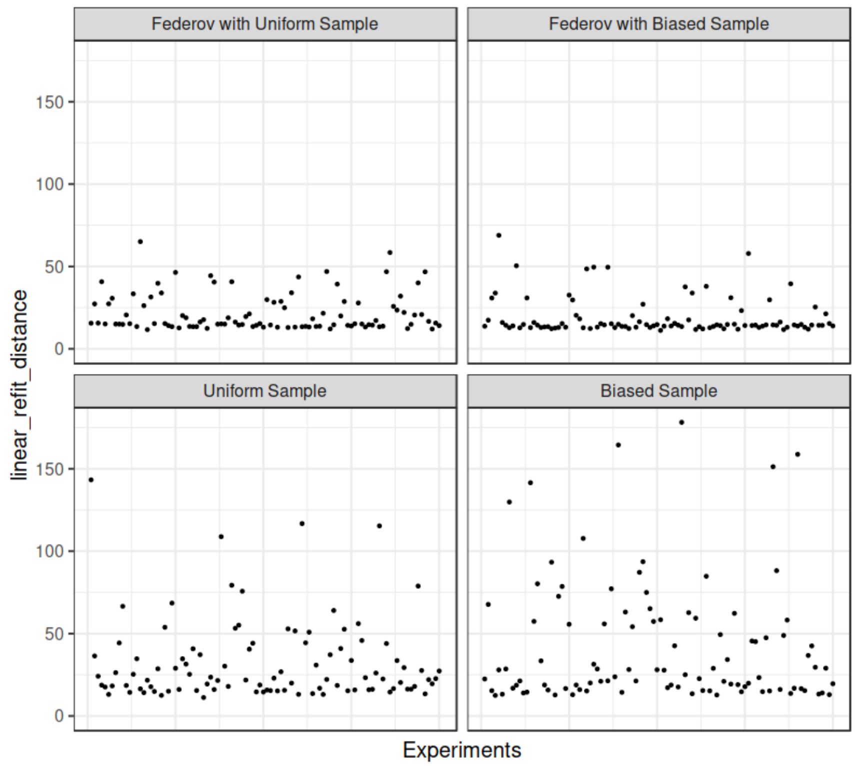
3.1.2 Response Vector Distance




Federov with Uniform Sample
Federov with Biased Sample
Uniform Sample
Biased Sample
linear_fit_distance
100
50
0
Experiments

3.1.3 Post-ANOVA+Refit Response Vector Distance





Federov with Uniform Sample
Federov with Biased Sample
Uniform Sample
Biased Sample
linear_refit_distance
0
50
100
150
Experiments