Statistica Sinica Preprint No: SS-2025-0089	
Title	Robust Design and Misspecification Detection for Count Models
Manuscript ID	SS-2025-0089
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0089
Complete List of Authors	Linglong Kong, Rui Hu and Jordan Slessor
Corresponding Authors	Linglong Kong
E-mails	lkong@ualberta.ca
Notice: Accepted author version.	

Statistica Sinica: Newly accepted Paper

Statistica Sinica

Robust Design and Misspecification Detection for Count Models

Jordan Slessor[1], Rui Hu[12], and Linglong Kong[1]

[1]*Department of Mathematical and Statistical Sciences, University of Alberta*

[2]*Department of Mathematics and Statistics, MacEwan University*

Abstract: Classical optimal designs for count data are highly efficient but degrade severely when the assumed generalized linear model is misspecified. To address this, we propose a two-stage robust design framework for Poisson and negative binomial regressions. First, we adapt the nonparametric XICOR correlation measure to formally detect functional dependence between residuals and design points. Second, we introduce two robust design criteria—one utilizing pilot data to target specific deviations, and a minimax-style averaging over a functional class of misspecifications when prior data are absent. We optimize these non-convex objective functions using a simulated annealing algorithm. Extensive simulations and a real-world toxicological application demonstrate that our robust designs yield substantial efficiency gains over classical approaches under misspecification, while maintaining low premium costs under correct specification.

Key words and phrases: Generalized linear models, Model misspecification, Optimal design, Robust statistics.

Statistica Sinica: Newly accepted Paper

1. Introduction

Generalized linear models (GLMs) are common in the analysis of count data, with Poisson and negative binomial regressions serving as the standard analytical workhorses across fields such as ecology, epidemiology, and clinical trials. A critical step in the deployment of these models is experimental design — specifically, determining the optimal allocation of a finite number of experimental trials to maximize information yield. Classical optimal design strategies, such as D-optimal and A-optimal designs, are highly efficient when the underlying model is perfectly specified (Russell et al., 2009; Yang, 2008).

However, in practice, the assumed mean structure of a GLM is rarely perfect. A common practical problem arises when the linear predictor omits a structured, non-linear trend or when the chosen link function misrepresents the true data-generating process. Under such model misspecification, the asymptotic covariance matrix of the maximum likelihood estimator deviates from the inverse of the Fisher information matrix (Fahrmeir, 1990). Consequently, classical designs generated under the assumption of perfect specification can suffer severe degradation in efficiency, leading to biased inference and suboptimal resource allocation.

To address the limitations of existing optimal design methods, we pro-

Statistica Sinica: Newly accepted Paper

pose a robust optimal design framework for count models that explicitly guards against model misspecification. Our contribution is threefold:

1. **Misspecification Detection:** We adapt the recently developed XI-COR correlation measure (Chatterjee, 2021), applying it for the first time as a formal, nonparametric diagnostic test to detect potential model misspecification through functional dependence between model residuals and design points.

2. **Robust Design Criteria:** We propose two robust loss functions based on the asymptotic mean square error (AMSE). The first ($\mathcal{L}_{\text{full}}$) utilizes pilot data to estimate the misspecification directly, while the second ($\mathcal{L}_{\text{ave}}$) provides a minimax-style robust penalty by averaging over a broad functional class of possible misspecifications when no pilot data is available (Adewale and Wiens, 2009; Kong and Wiens, 2015).

3. **Algorithmic Implementation:** We provide a simulated annealing approach to optimize these non-convex, discrete design problems, demonstrating substantial efficiency gains over classical designs in both correctly specified and misspecified scenarios.

The remainder of this paper is structured as follows. Section 2 es-

tablishes the framework for count models and optimal design. Section 3 introduces the XICOR-based misspecification test. Section 4 derives the robust optimal design criteria. Section 5 presents simulation studies and a real-data application. Technical proofs and algorithmic details are deferred to the Supplementary Material.

2. Framework and Methodology

2.1 Count models

The Poisson and negative binomial (NB) distributions are standard generalized linear models (GLMs) for count data. For a sequence of independent responses Y_i with means μ_i, the Poisson probability mass function in exponential family form is $f(y_i|\mu_i) = \exp(y_i \log \mu_i - \mu_i - \log y_i!)$.

We relate the mean to the linear predictor η_i via a link function $g(\mu_i) = \eta_i$. For both the Poisson model and the standard NB model—which introduces a dispersion parameter κ to handle the overdispersion commonly found in count data—we adopt the canonical log-link, $g(\mu_i) = \log \mu_i$, such that the mean function becomes $\mu(\eta_i) = \exp(\eta_i)$.

Suppose we are presented with a set of N candidate design points $\{\boldsymbol{x}_i\}_{i=1}^N$ of dimension $q \times 1$, for which we define a $p \times 1$ vector of regressors $\boldsymbol{z}(\boldsymbol{x}_i)$, and a $p \times 1$ vector of model parameters $\boldsymbol{\beta}$. By defining the linear

Statistica Sinica: Newly accepted Paper

predictor as $\eta_i = \boldsymbol{z}^\top(\boldsymbol{x}_i)\boldsymbol{\beta}$, both models can be fitted using traditional GLM techniques such as iteratively reweighted least squares (IRWLS).

2.2 Optimal designs

Because observing response variables $\{y_i\}_{i=1}^N$ is often expensive or practically constrained, experiments typically restrict observations to a budget of size n (where usually, but not strictly, $n < N$). Optimal design seeks the most informative allocation of these n observations.

Formally, given a candidate set $\mathcal{S} = \{\boldsymbol{x}_i\}_{i=1}^N$, an experimenter allocates $n_i \geq 0$ observations to each $\boldsymbol{x}_i$ such that $\sum_{i=1}^N n_i = n$. Following Adewale and Wiens (2009), we formulate this allocation as a probability distribution $\{p_i\}_{i=1}^N$ on $\mathcal{S}$, where $p_i = n_i/n$ and $\sum_{i=1}^N p_i = 1$.

Optimal designs are typically constructed by optimizing a scalar function of the Fisher information matrix. The information matrix for $\boldsymbol{\beta}$ from a design consisting of $\{\boldsymbol{x}_i\}_{i=1}^n$ is

$$\sum_{i=1}^{n} w(\boldsymbol{x}_i, \boldsymbol{\beta})\boldsymbol{z}(\boldsymbol{x}_i)\boldsymbol{z}^\top(\boldsymbol{x}_i) = \boldsymbol{Z}^\top \boldsymbol{W}\boldsymbol{Z},$$

where $\boldsymbol{Z} = (\boldsymbol{z}(\boldsymbol{x}_1), \ldots, \boldsymbol{z}(\boldsymbol{x}_n))^\top$,

$$w(\boldsymbol{x}_i, \boldsymbol{\beta}) = \frac{(d\mu/d\eta_i)^2}{Var(Y|\boldsymbol{x}_i)},$$

and $\boldsymbol{W} = diag\{w(\boldsymbol{x}_i, \boldsymbol{\beta})\}$. The conditional variance of the response given

Statistica Sinica: Newly accepted Paper

a design point for the Poisson and negative binomial models are μ_i and $\mu_i + \mu_i^2/\kappa$, respectively, and $d\mu/d\eta_i$ is the derivative of the mean function with respect to the linear term.

Two common forms of optimal designs are **D**-optimal designs which minimize the determinant of the information matrix (Russell et al., 2009; King and Wong, 2000; Zhang and Ye, 2014), and **A**-optimal designs which minimize the trace of the information matrix (Yang, 2008), although there exist alternatives. The optimality of the designs, however, may not hold true if the model, specifically the linear term, η_i, is misspecified.

2.3 Model misspecification

Suppose a count model GLM assumes the mean structure

$$\mu_i = \mu(\eta_i), \ i = 1, \ldots, n, \tag{2.1}$$

when the true data-generating process is actually

$$\mu_{T,i} = \mu(\eta_i + f(\boldsymbol{x}_i)), \ i = 1, \ldots, n, \tag{2.2}$$

where $\eta_i = \boldsymbol{z}^\top(\boldsymbol{x}_i)\boldsymbol{\beta}_0$ and $f(\boldsymbol{x}_i)$ is an unmodeled misspecification function. The target parameter $\boldsymbol{\beta}_0$ is defined to minimize the squared discrepancy between the true and assumed conditional means:

$$\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}} \frac{1}{N} \sum_{i=1}^{N} \{\mathbb{E}(Y_i|\boldsymbol{x}_i) - \mu(\boldsymbol{z}^\top(\boldsymbol{x}_i)\boldsymbol{\beta})\}^2.$$

Statistica Sinica: Newly accepted Paper

To bound the asymptotic bias introduced by f, we define $\mathcal{F}$ as the class of misspecification functions satisfying a bounded variance condition

$$\frac{1}{N}\sum_{i=1}^{N} f^2(\boldsymbol{x}_i) \leq \tau^2, \tag{2.3}$$

where $\tau^2 = O(n^{-1})$, alongside an identifiability condition ensuring orthogonality to the model regressors:

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{z}(\boldsymbol{x}_i) f(\boldsymbol{x}_i) = \mathbf{0}_{p\times 1}. \tag{2.4}$$

In practice, the exact functional form of $f(\boldsymbol{x}_i)$ is rarely known (Adewale and Wiens, 2009). To handle this uncertainty, one approach—adopted in this paper—is to average the target loss function over the functional class $\mathcal{F}$ satisfying (2.3) and (2.4), a method Adewale and Wiens (2009) successfully applied to misspecified logistic models. An alternative is the minimax approach (King and Wong, 2000; Fang and Wiens, 2000; Kong and Wiens, 2015; Selvaratnam et al., 2021; Hu et al., 2023; Hu and Wiens, 2017), which constructs optimal designs by minimizing the maximum expected loss over a neighborhood of possible model deviations.

If the misspecification function $f(\boldsymbol{x}_i)$ is strictly linear, the Pearson correlation coefficient between the design points and the model residuals—defined as

$$d(\boldsymbol{x}_i) = y(\boldsymbol{x}_i) - \mu(\hat{\eta}_i),\ \ i = 1, \ldots, N, \tag{2.5}$$

Statistica Sinica: Newly accepted Paper

where $\hat{\eta}_i = \boldsymbol{z}^\top(\boldsymbol{x}_i)\hat{\boldsymbol{\beta}}$—sufficiently flags the deviation. However, to detect arbitrary functional dependencies, we utilize the XICOR coefficient (Chatterjee, 2021). This nonparametric measure possesses established asymptotic properties that make it ideal for formally testing and detecting generalized model misspecification.

Building on these foundations, we propose a comprehensive diagnostic and design framework. We leverage the XICOR coefficient to detect linear predictor misspecification, extend the robust averaging principles of Adewale and Wiens (2009) to Poisson and negative binomial models, and deploy a simulated annealing algorithm to navigate the resulting non-convex optimization landscape.

While our primary focus remains on robust, non-sequential optimal design for count models, this methodology naturally complements adaptive trial designs. In response-adaptive randomization, power and ethical trade-offs heavily depend on target allocations and procedural variability (Hu and Rosenberger, 2003; Rosenberger and Sverdlov, 2008; Sun et al., 2025). Model-based optimal designs generalize Neyman allocations to accommodate covariates (Atkinson, 2015; Hu et al., 2014), while covariate-adjusted response-adaptive (CARA) designs seek to balance baseline factors and control multi-level imbalances (Rosenberger and Sverdlov, 2008; Hu and

Hu, 2012; Ma et al., 2024). Our proposed robust optimal designs offer a principled foundation within this ecosystem: they provide highly efficient, misspecification-resistant initial allocations that can serve as reliable building blocks for subsequent adaptive or CARA assignment updates.

3. Misspecification Detection Using XICOR

3.1 Motivation and Problem Setup

Consider observations $\{(x_i, y_i)\}_{i=1}^N$ generated from a generalized linear model (GLM) with the conditional mean given by

$$\mu_i = \exp(\eta_i), \quad \eta_i = z(x_i)^\top \boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is the parameter vector to be estimated. Misspecification occurs if the true model is

$$\mu_{T,i} = \exp(\eta_i + f(x_i)),$$

where $f(x_i)$ is a nonzero misspecification function. If the assumed model is correct, the residuals $d_i = y_i - \mu_i$ should not exhibit any systematic dependence on the design point x_i. Our goal is to formally test whether the residuals are functionally dependent on the design points, indicating misspecification.

Statistica Sinica: Newly accepted Paper

3.2 Hypothesis Test

We propose a formal hypothesis test for misspecification defined in terms of the population conditional expectations. Let the true conditional mean of the response given the design point be $\mu_T(x) = \mathbb{E}[Y \mid X = x]$. The assumed generalized linear model specifies the mean as $\mu(x; \boldsymbol{\beta}) = \exp(z(x)^\top \boldsymbol{\beta})$.

We define the true population discrepancy function as $\Delta(x) = \mu_T(x) - \mu(x; \boldsymbol{\beta}_0)$, where $\boldsymbol{\beta}_0$ is the target parameter minimizing the squared discrepancy. The hypothesis test for misspecification is formalized as:

$$H_0 : \Delta(x) = 0 \text{ almost everywhere (the model is correctly specified),}$$

$$H_1 : \Delta(x) = \psi(x) \text{ for some non-zero measurable function } \psi.$$

In practice, the unobservable population discrepancy $\Delta(x_i)$ is estimated by the observed residuals $d_i = y_i - \hat{\mu}_i$. Under H_0, the residuals should consist only of random noise and exhibit no functional dependence on x_i. The presence of functional dependence is assessed using the XICOR coefficient (Chatterjee, 2021), which detects both linear and nonlinear associations between the residuals and design points.

Statistica Sinica: Newly accepted Paper

3.3 XICOR Correlation Coefficient

Let $\{(x_i, d_i)\}_{i=1}^N$ be the pairs of design points and residuals. First, order the design points as $x_{(1)} \leq \cdots \leq x_{(N)}$, and denote the corresponding reordered residuals as $d_{(i)}$. Define ranks

$$r_i = \#\{j : d_{(j)} \leq d_{(i)}\}, \quad l_i = \#\{j : d_{(j)} \geq d_{(i)}\}.$$

The XICOR coefficient Chatterjee (2021) is defined as

$$\xi_N(x, d) = 1 - \frac{N \sum_{i=1}^{N-1} |r_{i+1} - r_i|}{2 \sum_{i=1}^{N} l_i(N - l_i)}.$$

The coefficient $\xi_N(x, d)$ takes values in $[0, 1]$, where 0 indicates independence and values approaching 1 indicate strong functional dependence.

3.4 Asymptotic Distribution and Test Statistic

Chatterjee (2021, Theorem 2.2) established that under the null hypothesis of independence, the XICOR coefficient satisfies:

$$\sqrt{N}\xi_N(x, d) \xrightarrow{d} \mathcal{N}(0, \tau^2), \quad \text{as } N \to \infty,$$

with τ^2 being a distribution-specific variance given by

$$\tau^2 = \frac{\mathbb{E}\phi(d_1, d_2)^2 - 2\mathbb{E}[\phi(d_1, d_2)\phi(d_1, d_3)] + (\mathbb{E}\phi(d_1, d_2))^2}{(\mathbb{E}[G(d)(1 - G(d))])^2},$$

where for $t \in \mathbb{R}$, $G(t) = P(d \geq t)$, $\phi(d, d') = \min\{F(d), F(d')\}$ for the cumulative distribution function F of residuals d, and d_1, d_2 and d_3 are

Statistica Sinica: Newly accepted Paper

independent copies of d. Chatterjee (2021, Theorem 2.3) further provides a consistent estimator τ^2, denoted $\hat{\tau}_N^2$, computable in $O(N \log N)$ time.

The standardized test statistic is thus

$$Z_N = \frac{\sqrt{N}\xi_N(x,d)}{\hat{\tau}_N} \xrightarrow{d} \mathcal{N}(0,1), \quad \text{under } H_0,$$

and the corresponding p-value for this test is $p = 2[1 - \Phi(|Z_N|)]$, where Φ is the standard normal cumulative distribution function.

3.5 Algorithm for Misspecification Detection

The practical steps for implementing the XICOR misspecification detection test are:

1. Fit the assumed GLM to obtain estimates $\hat{\boldsymbol{\beta}}$ and calculate fitted values $\hat{\mu}_i$.
2. Compute residuals $d_i = y_i - \hat{\mu}_i$.
3. Calculate the XICOR coefficient $\xi_N(x,d)$.
4. Estimate $\hat{\tau}_N^2$ using the procedure given by (Chatterjee, 2021).
5. Compute the test statistic Z_N and the corresponding p-value.
6. If the p-value is smaller than the chosen significance level α, reject H_0 and conclude that model misspecification is detected.

Statistica Sinica: Newly accepted Paper

If the test indicates misspecification (i.e., $p < \alpha$), we recommend using the robust design approach developed in subsequent sections to account for the identified misspecification. If the test fails to reject independence, the traditional optimal design methods remain applicable. Although we focus on count models in this work, the XICOR-based misspecification detection procedure is applicable to any GLM where residuals can be computed.

4. Robust Optimal Designs

As previously mentioned, an optimal design is usually taken as a minimizer or maximizer of some function of the information matrix, but this optimality may be lost when the model is misspecified. For example, Russell et al. (2009) compute a D-optimal design by maximizing the determinant of the information matrix. However, as Fahrmeir (1990) shows, when the model is misspecified, the asymptotic covariance of the maximum likelihood estimator, $\hat{\boldsymbol{\beta}}$, is no longer equal to the inverse of the information matrix, and thus the D-optimal design is no longer optimal.

In this section we propose methodologies for generating optimal designs robust against possible model misspecification for the estimation of regression parameters for the Poisson and negative binomial generalized linear models. The procedure is heavily inspired by the work of Adewale and

Statistica Sinica: Newly accepted Paper

Wiens (2009).

4.1 Asymptotic Bias and Mean Square Error

In this subsection we derive two loss functions based on the asymptotic mean square error of $\hat{\boldsymbol{\beta}}$, which includes the asymptotic bias due to the model misspecification if it exists, and the asymptotic variance of $\hat{\boldsymbol{\beta}}$.

Suppose that $\{\boldsymbol{x}_i, y_i\}$ are given, with n_i observations at each $\boldsymbol{x}_i$ and y_i is the count of events at $\boldsymbol{x}_i$. Then, the asymptotic bias and covariance of $\hat{\boldsymbol{\beta}}$ are given in Theorem 1. This Theorem is a restatement of Theorem 1 from Adewale and Wiens (2009) under the different response distributions.

Theorem 1. *Define* $w_i = d\mu_i/d\eta_i$, *where for the Poisson and negative binomial response,* $w_i = e^{\eta_i}$. *Also let* $\boldsymbol{Z}_{N\times p}$ *be the matrix with rows* $\boldsymbol{z}^\top(\boldsymbol{x}_i)$. *Recall* (2.1) *and* (2.2); *let* $\boldsymbol{\gamma}$ *and* $\boldsymbol{\gamma}_T$ *be* $N \times 1$ *vectors with elements* μ_i *and* $\mu_{T,i}$, *respectively. Let* $\boldsymbol{P}, \boldsymbol{W}$, *and* $\boldsymbol{W}_T$ *be* $N \times N$ *diagonal matrices with diagonal elements* $n_i/n, w_i$, *and* $W_{T,i} = \mu_{T,i}$, *respectively. Finally, define* $\boldsymbol{b} = \boldsymbol{Z}^\top \boldsymbol{P}(\boldsymbol{\gamma}_T - \boldsymbol{\gamma})$, $\boldsymbol{H}_n = \boldsymbol{Z}^\top \boldsymbol{P}\boldsymbol{W}\boldsymbol{Z}$, *and* $\tilde{\boldsymbol{H}}_n = \boldsymbol{Z}^\top \boldsymbol{P}\boldsymbol{W}_T\boldsymbol{Z}$. *The asymptotic bias and covariance matrix of the maximum likelihood estimator* $\hat{\boldsymbol{\beta}}$ *of the model parameter vector* $\boldsymbol{\beta}$ *from the misspecified model are*

$$bias(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{H}_n^{-1}\boldsymbol{b} + o(n^{-1/2}),$$

Statistica Sinica: Newly accepted Paper

and

$$cov(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)) = \boldsymbol{H}_n^{-1}\tilde{\boldsymbol{H}}_n\boldsymbol{H}_n^{-1} + o(1),$$

respectively.

The standard goal for generalized linear models is the estimation of the mean, and thus as in Adewale and Wiens (2009), we take the normalized average mean squared error, AMSE, as the loss function. For the predictions $\mu(\hat{\eta}_i)$ where $\hat{\eta}_i = \boldsymbol{z}^\top(\boldsymbol{x}_i)\hat{\boldsymbol{\beta}}$, this loss is

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} E[\{\mu(\hat{\eta}_i) - \mu(\eta_i + f(\boldsymbol{x}_i))\}^2].$$

Next, using the asymptotic results from Theorem 1, the AMSE can be asymptotically approximated as shown in the following theorem.

4.2 The Full Loss Criterion $\mathcal{L}_{\text{full}}$

Theorem 2. *The AMSE can be approximated as* $\mathcal{L} = \mathcal{L}_{full}(\boldsymbol{P}, \boldsymbol{f}) + o(n^{-1})$, *where*

$$\mathcal{L}_{full}(\boldsymbol{P}, \boldsymbol{f}) = \frac{1}{nN} tr[\boldsymbol{W}\boldsymbol{Z}\boldsymbol{H}_n^{-1}\tilde{\boldsymbol{H}}_n\boldsymbol{H}_n^{-1}\boldsymbol{Z}^\top\boldsymbol{W}] + \frac{1}{N}||\boldsymbol{W}(\boldsymbol{Z}\boldsymbol{H}_n^{-1}\boldsymbol{b} - \boldsymbol{f})||^2, \tag{4.6}$$

where $\boldsymbol{f} = (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N))^\top$.

Equation (4.6) isolates the leading-order contribution of model misspecification to the AMSE. Expanding the expected squared error in terms of

Statistica Sinica: Newly accepted Paper

f yields a decomposition involving the deterministic matrices $\boldsymbol{W}$, $\boldsymbol{H}_n$, and $\tilde{\boldsymbol{H}}_n$. While these matrices depend on $\mu_i = \exp(\eta_i)$ and f, they can be approximated if pilot data is available (Adewale and Wiens, 2009). Let the discrepancy between the true and assumed response at location $\boldsymbol{x}$ be

$$d(\boldsymbol{x}) = \mu(\boldsymbol{z}^\top(\boldsymbol{x})\boldsymbol{\beta}_0 + f(\boldsymbol{x})) - \mu(\boldsymbol{z}^\top(\boldsymbol{x})\boldsymbol{\beta}_0).$$

Approximating $d(\boldsymbol{x})$ to first order yields $d(\boldsymbol{x}) \approx (d\mu/d\eta)f(\boldsymbol{x})$. By substituting the empirical residual $\hat{d}(\boldsymbol{x}) = y(\boldsymbol{x}) - \mu(\boldsymbol{z}^\top(\boldsymbol{x})\hat{\boldsymbol{\beta}})$, we obtain the estimate $\hat{f}(\boldsymbol{x}) = \hat{d}(\boldsymbol{x})/(d\mu/d\eta_{|\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}})$. Plugging this estimate into (4.6) renders the objective function fully computable for subsequent optimization.

4.3 The Average Loss Criterion $\mathcal{L}_{\text{ave}}$

When pilot data is unavailable to estimate $f(\boldsymbol{x})$, we provide a robust minimax-style alternative by averaging the approximate loss (4.6) over $\mathcal{F}$, the class of functions satisfying (2.3) and (2.4). This procedure and subsequent theorem are both proposed by Adewale and Wiens (2009). The strength of this theorem is the removal of the dependence on the misspecification term from the loss.

First begin with the singular value decomposition of the regressor matrix

$$\boldsymbol{Z} = \boldsymbol{U}_{N\times p}\boldsymbol{\Lambda}_{p\times p}\boldsymbol{V}^\top_{p\times p}, \tag{4.7}$$

Statistica Sinica: Newly accepted Paper

where $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I}_p$ and $\boldsymbol{\Lambda}$ is diagonal and invertible, followed by the augmentation of $\boldsymbol{U}$, $\tilde{\boldsymbol{U}}_{N\times(N-p)}$, such that $[\boldsymbol{U}\vdots\tilde{\boldsymbol{U}}]_{N\times N}$ is orthogonal. Then by assumptions (2.3) and (2.4), there exists an $(N-p)\times 1$ vector $\boldsymbol{t}$, with $||\boldsymbol{t}|| \leq 1$ satisfying

$$\boldsymbol{f}(=\boldsymbol{f}_t) = \tau\sqrt{N}\tilde{\boldsymbol{U}}\boldsymbol{t}. \tag{4.8}$$

Here, $\boldsymbol{f}_t$ denotes $\boldsymbol{f}$ expressed as a function of the parameter vector $\boldsymbol{t}$. The average loss is the expected value of (4.6) with respect to the uniform measure on the sphere and its interior in $\mathbb{R}^{N-p}$. The measure being used has the density $p(\boldsymbol{t}) = (1/\kappa_{N,p})I(||\boldsymbol{t}|| \leq 1)$ where $\kappa_{N,p} = \pi^{(N-p)/2}/\Gamma((N-p)/2+1)$ is the volume of the unit sphere.

Theorem 3. *The average loss over the misspecification class $\mathcal{F}$ is, apart from $o(n^{-1})$ terms,*

$$\begin{aligned}\mathcal{L}_{ave}(\boldsymbol{P},\tau^2) &= \int \mathcal{L}_{full}(\boldsymbol{P},\boldsymbol{f}_{\boldsymbol{t}})p(\boldsymbol{t})d\boldsymbol{t}\\ &= \frac{1}{nN}tr[(\boldsymbol{U}^\top\boldsymbol{P}\boldsymbol{W}\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{W}^2\boldsymbol{U}]\\ &+ \frac{\tau^2}{N-p+2}tr[\boldsymbol{W}(\boldsymbol{R}-\boldsymbol{I}_N)(\boldsymbol{R}^\top-\boldsymbol{I}_N)\boldsymbol{W}],\end{aligned} \tag{4.9}$$

where $\boldsymbol{R}_{N\times N} = \boldsymbol{U}(\boldsymbol{U}^\top\boldsymbol{P}\boldsymbol{W}\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{P}\boldsymbol{W}$. Note that when computing (4.9), it is more efficient to compute the second trace as

$$tr[\boldsymbol{W}(\boldsymbol{R}-\boldsymbol{I}_N)(\boldsymbol{R}^\top-\boldsymbol{I}_N)\boldsymbol{W}] = \sum_{i=1}^{N} w_i^2||\tilde{\boldsymbol{r}}_i||^2,$$

where $\tilde{\boldsymbol{r}}_i^\top$ is the ith row of $\boldsymbol{R}-\boldsymbol{I}_N$.

Statistica Sinica: Newly accepted Paper

This result provides a fully operational loss criterion that no longer depends on a specific form of the misspecification function f. By averaging over the misspecification class $\mathcal{F}$, the expression in Theorem 3 yields a tractable approximation to the average mean squared error under model deviation. The first term corresponds to nominal variance, while the second introduces a smooth robustness penalty scaled by τ^2. Importantly, this penalty does not require knowing or estimating f, and instead integrates over its possible directions orthogonal to the design space. This makes the criterion directly usable in practice when no pilot data are available, and it forms the basis of the minimax robust design procedure developed in the next section.

In the definition of the average loss function $\mathcal{L}_{\text{ave}}$, the expectation over $\boldsymbol{t}$ is taken with respect to the uniform distribution on the unit ball in $\mathbb{R}^{N-p}$, i.e., $p(\boldsymbol{t}) = (1/\kappa_{N,p})I(||\boldsymbol{t}|| \leq 1)$, where $\kappa_{N,p}$ is the volume of the ball. This choice reflects a non-informative prior over the direction and magnitude of the misspecification vector $\boldsymbol{t}$, treating all admissible perturbations $||\boldsymbol{t}|| \leq 1$ as equally likely. Other choices are possible. For example, a Gaussian prior $\boldsymbol{t} \sim N(0, \sigma^2 I)$ would down-weight extreme perturbations, while a prior supported on a lower-dimensional manifold would focus on structured forms of misspecification. If the prior is misspecified, $\mathcal{L}_{\text{ave}}$ would still yield designs

Statistica Sinica: Newly accepted Paper

optimal for the assumed prior, but efficiency could degrade for directions of $\boldsymbol{t}$ given higher (or lower) prior weight under the truth. In practice, the uniform prior offers a neutral baseline that avoids favoring any particular misspecification direction or magnitude without substantive prior knowledge.

The loss functions $\mathcal{L}_{\text{full}}$ and $\mathcal{L}_{\text{ave}}$ are intended for different experimental settings and reflect the amount of information available about potential model misspecification. When pilot data are available, $\mathcal{L}_{\text{full}}$ is preferable, as it incorporates an estimate of the misspecification function f and directly targets the contribution of model deviation to the AMSE, yielding designs that are more efficient when $\hat{f}$ is reasonably accurate. When no such data are available, or when the form of the misspecification cannot be reliably estimated, $\mathcal{L}_{\text{ave}}$ provides a robust alternative by averaging over a broad class of possible misspecification functions. In practice, this suggests a two-stage strategy in which a small initial portion of the experimental budget is allocated using a classical or space-filling design to obtain pilot data, followed by the misspecification diagnostic as described in Section 3.5. If misspecification is detected, the remaining design points may be selected by minimizing $\mathcal{L}_{\text{full}}$, while $\mathcal{L}_{\text{ave}}$ remains appropriate when pilot data are unavailable or insufficient.

Statistica Sinica: Newly accepted Paper

Because the loss functions derived above possess discrete parameter spaces and are non-convex, traditional gradient-based optimization techniques are inapplicable. We utilize a simulated annealing algorithm, incorporating the perturbation method proposed by Fang and Wiens (2000), to identify the optimal design points. A detailed description of the initialization, iteration, cooling schedule, and convergence guarantees of this algorithm is provided in Section S2 of the Supplementary Materials.

5. Simulation and Real Data Studies

This section evaluates the performance of the proposed robust designs against classical designs across correctly specified and misspecified scenarios. To conserve space, the main text focuses on three representative settings: (i) no misspecification, (ii) quadratic misspecification, and (iii) a real-world earthworm toxicity study. Additional scenarios including cubic misspecification, link function misspecification, higher-dimensional settings, and extended algorithmic stability checks are provided in the supplementary material.

Statistica Sinica: Newly accepted Paper

5.1 Simulation Specifications

To isolate the performance of the objective functions, all simulations adhere to the following baseline configuration unless otherwise stated:

- **Design Space:** $N = 80$ candidate points evenly spaced over $[-1, 1]$.
- **Sample Size:** $n = 40$ total observations to be allocated.
- **Regressors:** $\boldsymbol{z}(x_i) = (1, x_i)^\top$.
- **Optimization:** All optimal designs are computed using the identical simulated annealing algorithm (detailed in Section S2). While non-heuristic algorithms exist for classical designs (Harman et al., 2020; Yang et al., 2013), deploying a common stochastic optimization procedure with fixed random seeds ensures direct comparability by isolating differences to the loss functions rather than the optimizers.

We compare the classical design, which assumes the model is perfectly specified ($f = 0$) and minimizes the standard asymptotic loss, against robust designs optimizing the $\mathcal{L}_{\text{full}}$ (Eq. (4.6), pilot data available) and $\mathcal{L}_{\text{ave}}$ (Eq. (4.9), no pilot data) criteria. Throughout, the working parameter $\boldsymbol{\beta}$ (and, for the negative binomial models, the working dispersion θ) used by the loss machinery is held fixed at its assumed value during both de-

Statistica Sinica: Newly accepted Paper

sign optimization and evaluation; pilot data are used only to construct the residual-based contamination estimate $\hat{f}(x)$ that feeds into $\mathcal{L}_{\text{full}}$.

Because $\mathcal{L}_{\text{full}}$'s design depends on a random pilot draw – a new pilot sample yields a new $\hat{f}(x)$ and hence potentially a new optimal design – its performance is evaluated via a two-stage Monte Carlo procedure: each of M_{outer} replicates draws fresh pilot data, fits $\hat{\boldsymbol{\beta}}$ and $\hat{f}(x)$, re-optimizes the robust design, and then simulates fresh production data at that replicate's design to compute a simulated AMSE. The classical and $\mathcal{L}_{\text{ave}}$ designs do not depend on pilot data, so their simulated AMSE requires only a single source of randomness (production data simulated at the fixed design); these are evaluated with a larger number of replicates since no re-optimization is needed per draw. All three designs are reported on the same n-scaled simulated AMSE, making the loss column directly comparable across methods.

5.2 Performance Evaluation Metrics

It is critical to distinguish between the actual realized loss evaluated post-design and the theoretical mathematical criteria upon which the designs were optimized. We report the following metrics to capture both practical performance and theoretical efficiency.

1. **Loss ($\mathcal{L}$):** The primary measure of experimental performance, re-

Statistica Sinica: Newly accepted Paper

ported as the n-scaled simulated AMSE described above (mean over Monte Carlo replicates, with simulation standard errors in parentheses). This evaluates each design under the true data-generating process, explicitly incorporating any underlying model misspecification.

2. **Gain:** The relative theoretical reduction in the design's optimization criterion when comparing the robust design (P_{robust}) to the classical design (P_{class}):

$$\text{Gain} = \frac{\mathcal{L}_{\text{full}}(P_{\text{class}} \mid f) - \mathcal{L}_{\text{full}}(P_{\text{robust}} \mid f)}{\mathcal{L}_{\text{full}}(P_{\text{class}} \mid f)}$$

We report two versions of this quantity. *Gain (estimated)* plugs in the information actually available to the experimenter: $f = \hat{f}$ for $\mathcal{L}_{\text{full}}$-based designs, and the admissible-class radius ρ for $\mathcal{L}_{\text{ave}}$-based designs. *Gain (oracle)* substitutes the true generating contamination in place of the estimate, and is reported as a diagnostic to show how each design would have fared had the truth been known exactly. A positive gain indicates that the robust design minimized its target criterion more effectively than the classical allocation would have.

3. **Premium:** The theoretical "insurance cost" of deploying a robust design if the true data-generating process turns out to be perfectly specified. It is the relative inflation in the unpenalized loss criterion

Statistica Sinica: Newly accepted Paper

($f = 0$) incurred by using the robust design instead of the classical design:

$$\text{Premium} = \frac{\mathcal{L}_{\text{full}}(P_{\text{robust}} \mid f = 0) - \mathcal{L}_{\text{full}}(P_{\text{class}} \mid f = 0)}{\mathcal{L}_{\text{full}}(P_{\text{class}} \mid f = 0)}$$

Because Premium is evaluated at $f = 0$ in both the numerator and denominator, it is insensitive to whether $\hat{f}$ or the true f is used, and so a single value is reported regardless of the Gain column. A low premium indicates that the robust design provides cheap insurance: it protects against misspecification without severely degrading experimental efficiency if the assumed model was actually correct.

5.3 Simulation 1: No misspecification

This simulation study investigates the behavior of robust designs in the absence of model misspecification. This serves as a control setting to assess whether the robust methods incur an unnecessary premium when there is nothing to be robust against. Maintaining the baseline simulation parameters ($N = 80$, $n = 40$), we consider both Poisson and NB ($\kappa = 1$) responses. The true model is perfectly specified as $\mu_{T,i} = \exp(\boldsymbol{z}(x_i)^\top \boldsymbol{\beta})$ with $\boldsymbol{\beta} = (1/2, 1/2)^\top$, i.e., $f_{\text{true}}(x) \equiv 0$.

As expected, the XICOR test fails to reject the null hypothesis of independence between residuals and design points in both models. For Poisson:

Statistica Sinica: Newly accepted Paper

$\xi = -0.029$, $Z = -0.404$, $p = 0.686$; for NB: $\xi = -0.110$, $Z = -1.558$, $p = 0.119$. This confirms that the assumed models are correctly specified, and that classical optimal design methods remain applicable in this setting.

Optimal designs were computed under three criteria. The classical design assumes $f(x) = 0$. The $\mathcal{L}_{\text{full}}$ design uses pilot data to estimate $f(x)$ and minimizes the corresponding full loss. The $\mathcal{L}_{\text{ave}}$ design ($\rho = 10$) assumes no access to $f(x)$ and instead minimizes the average loss over a class of admissible misspecification functions. Figure 1 displays the resulting design weight allocations for the Poisson model.

Table 1 summarizes the simulated AMSE under the true model, along with Gain and Premium. In both models, the realized loss of the robust designs is only modestly higher than that of the classical design – e.g., for Poisson, 2.7220 for $\mathcal{L}_{\text{ave}}$ versus 2.5354 for the classical design – confirming that the robust designs do not incur unnecessary loss when no misspecification is present. As expected when $f_{\text{true}} = 0$, the oracle Gain is negative for every robust design (e.g., -9.8% for Poisson $\mathcal{L}_{\text{full}}$, -16.8% for Poisson $\mathcal{L}_{\text{ave}}$): once the true (null) contamination is substituted for the estimate, the robust designs no longer outperform the classical design, exactly mirroring the Premium paid for that protection. This is the expected behavior of a control setting: the estimated Gain is driven entirely by sampling noise in

Statistica Sinica: Newly accepted Paper

$\hat{f}$ and ρ rather than by genuine misspecification, and the oracle comparison confirms there is nothing to be gained.

Table 1: Simulated AMSE, Gain, and Premium for Simulation 1 (No Misspecification). Simulation standard errors (SE) are in parentheses. Gain (Est.) is evaluated at $\hat{f}$ (for $\mathcal{L}_{\text{full}}$) or $\rho = 10$ (for $\mathcal{L}_{\text{ave}}$); Gain (Oracle) is evaluated at the true contamination, $f_{\text{true}} \equiv 0$.

Design	Loss ($\mathcal{L}$)	Gain Est. (%)	Gain Oracle (%)	Premium (%)
Poisson				
Classical	2.5354 (0.1278)	–	–	–
Robust ($\mathcal{L}_{\text{full}}$)	3.2297 (0.4932)	24.1	-9.8	9.8
Robust ($\mathcal{L}_{\text{ave}}$)	2.7220 (0.1314)	37.1	-16.8	16.8
Negative Binomial				
Classical	7.2839 (0.3980)	–	–	–
Robust ($\mathcal{L}_{\text{full}}$)	8.8021 (1.6434)	36.0	-15.8	15.8
Robust ($\mathcal{L}_{\text{ave}}$)	8.1962 (0.4173)	32.6	-11.4	11.4

5.4 Simulation 2: Quadratic misspecification

This simulation study evaluates the performance of robust designs under quadratic model misspecification, demonstrating how well the methods

Statistica Sinica: Newly accepted Paper

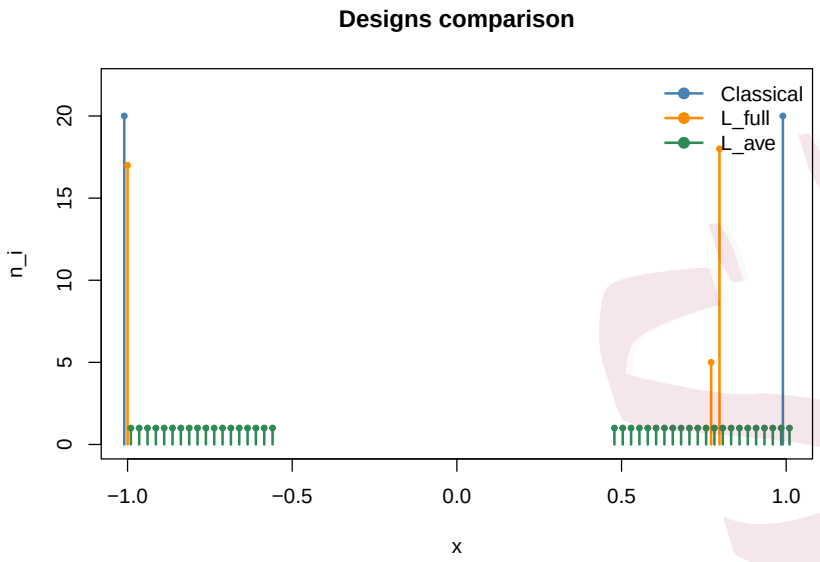

Figure 1: Design weight allocations for the Poisson model under $\mathcal{L}_{\text{full}}$, $\mathcal{L}_{\text{ave}}$, and classical criteria.

adapt when the assumed linear predictor omits an unmodeled nonlinear trend. Maintaining the baseline experimental parameters, we simulate responses from both Poisson and negative binomial ($\kappa = 1$) models. The true response introduces a quadratic misspecification component $f(x) = \beta_2\, g(x)$, where $g(x) = (x^2 - \mu_2)/\sqrt{\mu_4 - \mu_2^2}$ and $\mu_k = N^{-1}\sum_i x_i^k$. The function $g(x)$ is normalized so that its mean squared value over the design space equals one, and on this symmetric, equally spaced grid it is exactly orthogonal to $\{1, x\}$, satisfying the bounded bias and identifiability conditions of the misspecification neighborhood. Here $\beta_2 = 0.5$, with $\boldsymbol{\beta} = (1/2, 1/2)^\top$ as

Statistica Sinica: Newly accepted Paper

before.

The XICOR test detects significant nonlinear dependence in both models: for Poisson, $\xi = 0.701$, $Z = 9.915$, $p < 0.001$; for negative binomial, $\xi = 0.534$, $Z = 7.542$, $p < 0.001$. Unlike in Simulation 1, the oracle-truth Gain figures here are genuinely informative, since real contamination is present to detect and protect against.

Figure 2 shows the estimated contamination and loess smoothed estimate that was used in the construction of the robust designs.

Table 2 reports the simulated AMSE, Gain, and Premium. The classical design's realized loss deteriorates sharply under quadratic misspecification, rising to 439.73 for Poisson and 442.91 for negative binomial – roughly two orders of magnitude above its Simulation 1 values – since it has no mechanism to accommodate the omitted curvature. Both robust designs substantially mitigate this: for Poisson, $\mathcal{L}_{\text{full}}$ reduces the realized loss to 48.49 and $\mathcal{L}_{\text{ave}}$ to 80.07, an order-of-magnitude improvement over the classical design. The oracle Gain values confirm this is genuine protection against misspecification rather than sampling noise: e.g., the Poisson $\mathcal{L}_{\text{full}}$ design achieves a 93.6% oracle gain (versus 96.9% estimated), and even $\mathcal{L}_{\text{ave}}$, which has no access to pilot data, achieves an 84.4% oracle gain. This comes at a real cost if the model happens to be correctly specified: the Premium for

Statistica Sinica: Newly accepted Paper

$\mathcal{L}_{\text{full}}$ is 38.6% (Poisson) and 49.0% (NB), markedly higher than the modest premiums seen in Simulation 1, while $\mathcal{L}_{\text{ave}}$'s premium is unchanged from Simulation 1 (16.8% Poisson, 11.4% NB) since it does not depend on pilot data or the realized contamination.

Table 2: Simulated AMSE, Gain, and Premium for Simulation 2 (Quadratic Misspecification). Simulation standard errors (SE) are in parentheses. Gain (Est.) is evaluated at $\hat{f}$ (for $\mathcal{L}_{\text{full}}$) or $\rho = 10$ (for $\mathcal{L}_{\text{ave}}$); Gain (Oracle) is evaluated at the true contamination, $\beta_2 = 0.5$.

Design	Loss ($\mathcal{L}$)	Gain Est. (%)	Gain Oracle (%)	Premium (%)
Poisson				
Classical	439.7253 (4.1531)	–	–	–
Robust ($\mathcal{L}_{\text{full}}$)	48.4903 (0.9097)	96.9	93.6	38.6
Robust ($\mathcal{L}_{\text{ave}}$)	80.0728 (0.7836)	37.1	84.4	16.8
Negative Binomial				
Classical	442.9089 (11.0270)	–	–	–
Robust ($\mathcal{L}_{\text{full}}$)	75.3624 (3.4167)	99.4	99.0	49.0
Robust ($\mathcal{L}_{\text{ave}}$)	123.1184 (2.9998)	32.6	80.8	11.4

As seen in this simulation, quadratic misspecification leads to substan-

Statistica Sinica: Newly accepted Paper

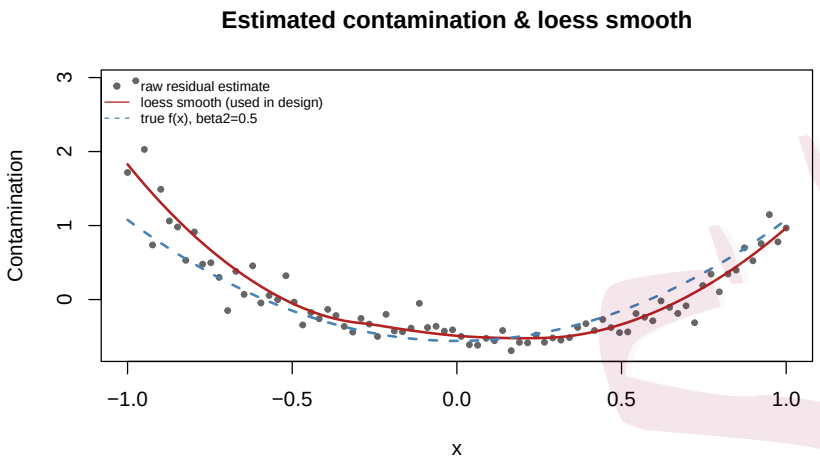

Figure 2: Estimated and loess smoothed contamination functions (Poisson model).

tial deterioration in the classical design's realized performance. Robust designs, particularly $\mathcal{L}_{\text{full}}$, substantially mitigate this risk by explicitly accounting for the impact of model deviation at the design stage. The $\mathcal{L}_{\text{ave}}$ design performs reasonably well without requiring pilot data, though it trades off some of $\mathcal{L}_{\text{full}}$'s realized-loss reduction in exchange for a lower premium.

5.5 Simulation 3: Earthworm data

We now demonstrate the proposed robust design methodology on a real dataset obtained from a toxicity experiment involving earthworms, available

Statistica Sinica: Newly accepted Paper

in the R package `drc` (Ritz et al., 2016). The study recorded the number of earthworms remaining in a container contaminated with varying doses of a toxic substance. At dose zero, roughly half the earthworms are expected to remain, as the rest are free to migrate to a neighboring uncontaminated container. Due to this structure, traditional binomial or logistic regression models are inappropriate at dose zero, motivating the use of a Poisson model with a log offset for the total number of earthworms initially placed in the container.

The observed dataset consists of 35 observations across 7 unique dose levels. We model the count of remaining earthworms as a Poisson response with a log offset for the total number initially placed in the container. To standardize the design space, the dose regressor is rescaled to lie in $[0, 1]$, and a simple linear model with intercept and dose is used: $\boldsymbol{z}(x) = (1, x)^\top$. Because the candidate design space here is the discrete set of $N = 7$ unique dose levels actually used in the data (rather than a fine grid), the design problem allocates $n = 20$ observations across these 7 points. The regression coefficients $\hat{\boldsymbol{\beta}}$ are estimated once by fitting this Poisson model (with offset) to the full 35-row dataset, and are then treated as fixed working parameters for every design loss computed below. As there is no repeatable data-generating process to resimulate from in a real-data setting, we report

Statistica Sinica: Newly accepted Paper

theoretical losses only; no simulated AMSE or oracle-truth diagnostics are computed for this example.

The XICOR test applied to residuals from the fitted model suggests misspecification, with $\xi = 0.280$, $Z = 2.602$, $p = 0.009$, leading to rejection of the null hypothesis of independence. This highlights the practical value of a design strategy robust to deviations from the assumed model.

Figure 3 shows the resulting designs and estimated contamination function. The classical design concentrates its $n = 20$ replicates at just two dose levels (12 at dose 0, 8 at dose 0.25), while the $\mathcal{L}_{\text{full}}$ design spreads its mass across four points, retaining substantial weight at dose 0 (7 replicates) but shifting allocation toward two additional interior doses (0.031 and 0.062). The $\mathcal{L}_{\text{ave}}$ design spreads its support most broadly, placing at least 3 replicates at each of five dose levels.

The classical design's theoretical loss (0.2926) is the smallest of the three, since it is evaluated at the assumption $f = 0$ under which it was optimized. Because there is no ground truth available for this real dataset, this figure does not admit the same realized-loss comparison reported for Simulations 1 and 2, and Gain and Premium offer the more informative comparison here instead: the $\mathcal{L}_{\text{full}}$ design achieves a 15.5% gain over the classical design when evaluated under its own criterion, at a Premium of

16.4% if the simple linear model had in fact been correctly specified. The $\mathcal{L}_{\text{ave}}$ design achieves a larger 21.3% gain, at a comparable 16.1% premium. Because $\mathcal{L}_{\text{ave}}$'s theoretical loss scales linearly with the admissible-class radius ρ (here $\rho = 10$), its raw loss value (0.7016) is not directly comparable in magnitude to the classical or $\mathcal{L}_{\text{full}}$ losses, and should be interpreted only relative to other $\mathcal{L}_{\text{ave}}$-criterion values.

Table 3: Theoretical loss, Gain, and Premium for the Earthworm data example.

Design	Loss ($\mathcal{L}$)	Gain (%)	Premium (%)
Classical	0.2926	–	–
Robust ($\mathcal{L}_{\text{full}}$)	0.3656	15.5	16.4
Robust ($\mathcal{L}_{\text{ave}}$)	0.7016	21.3	16.1

6. Final Remarks

This paper developed a robust optimal design framework for count models, motivated by the practical reality that assumed mean structures are rarely exactly correct. By pairing a XICOR-based misspecification diagnostic with two complementary robust design criteria, $\mathcal{L}_{\text{full}}$ and $\mathcal{L}_{\text{ave}}$, we provide a principled way to guard experimental efficiency against model uncertainty at

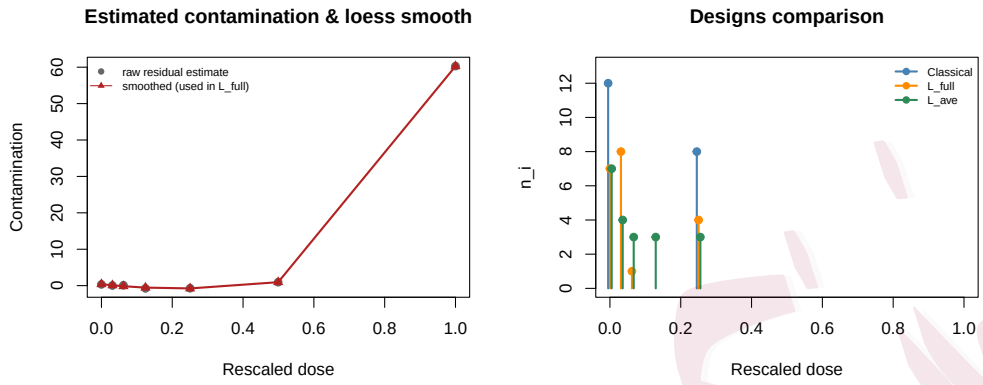

Figure 3: Estimated contamination function and design weight allocations for the Earthworm data example using $\mathcal{L}_{\text{full}}$, $\mathcal{L}_{\text{ave}}$, and classical criteria.

the design stage itself, rather than only correcting for it after the data are in. Our simulations and the earthworm toxicity example show that the proposed designs remain competitive when the assumed model is correct and substantially outperform classical designs once it is not.

The framework extends naturally beyond the settings explored here. It accommodates designs involving both treatment indicators and baseline covariates, making it directly relevant to clinical trials with count outcomes such as adverse events or tumor counts, where $\mathcal{L}_{\text{full}}$ can exploit available pilot data and $\mathcal{L}_{\text{ave}}$ offers protection when little is known in advance. It also complements covariate-adjusted response-adaptive (CARA) designs

Statistica Sinica: Newly accepted Paper

(Rosenberger and Sverdlov, 2008; Hu et al., 2014; Ma et al., 2024): our robust allocations offer a sound initial design for adaptive updating to build on, and the XICOR diagnostic provides a natural mechanism for guiding such updates in real time, a direction we leave for future work.

Open questions remain, including extensions to richer distributional families, higher-dimensional design spaces, and non-heuristic optimization routines. We hope this framework offers both a useful methodological contribution to robust design and a practical tool for experimenters.

Supplementary Materials

The online supplementary materials contain proofs for Theorems 1, 2, and 3. Moreover, they contain the detailed algorithmic formulation of the simulated annealing procedure, as well as the results and discussions for the negative binomial simulations and additional misspecification scenarios.

Acknowledgment

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number 02589]. Linglong Kong was partially supported by grants from the Canada CIFAR AI Chairs program, the Alberta Machine Intelligence Institute (AMII), the

Statistica Sinica: Newly accepted Paper

Canadian AI Safety Institute Research Program at CIFAR, the Natural Sciences and Engineering Research Council of Canada (NSERC), NSERC Alliance - Alberta Innovates Advance Program, and the Canada Research Chair program from NSERC.

References

Adewale, A. J. and D. P. Wiens (2009). Robust designs for misspecified logistic models. *Journal of Statistical Planning and Inference 139*(1), 3–15. Special Issue on Metaheuristics, Combinatorial Optimization and Design of Experiments.

Atkinson, A. C. (2015). Optimum designs for two treatments with unequal variances in the presence of covariates. *Biometrika 102*(2), 494–499.

Box, G. E. P. and N. R. Draper (1959). A Basis for the Selection of a Response Surface Design. *Journal of the American Statistical Association 54*(287), 622–654.

Chaloner, K. and K. Larntz (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference 21*(2), 191–208.

Chatterjee, S. (2021). A New Coefficient of Correlation. *Journal of the American Statistical Association 116*(536), 2009–2022.

Chatterjee, S. and S. Holmes (2023). *XICOR: Robust and generalized correlation coefficients.* CRAN. https://github.com/spholmes/XICOR.

Chernoff, H. (1953). Locally Optimal Designs for Estimating Parameters. *The Annals of*

Statistica Sinica: Newly accepted Paper

Mathematical Statistics 24(4), 586–602.

Fahrmeir, L. (1990). Maximum likelihood estimation in misspecified generalized linear models. *Statistics 21*(4), 487–502.

Fang, Z. and D. P. Wiens (2000). Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models. *Journal of the American Statistical Association 95*(451), 807–818.

Harman, R., L. Filová, and P. Richtárik (2020). A Randomized Exchange Algorithm for Computing Optimal Approximate Designs of Experiments.

Hu, F., Y. Hu, Z. Ma, and W. F. Rosenberger (2014). Adaptive randomization for balancing over covariates. *WIREs Computational Statistics 6*(4), 288–303.

Hu, F. and W. F. Rosenberger (2003). Optimality, Variability, Power. *Journal of the American Statistical Association 98*(463), 671–678.

Hu, R., I. Bica, and Z. Zhai (2023, September). Robust Optimal Design When Missing Data Happen at Random. *Journal of Statistical Theory and Practice 17*, 43.

Hu, R. and D. P. Wiens (2017). Robust discrimination designs over Hellinger neighbourhoods. *The Annals of Statistics 45*(4), 1638 – 1663.

Hu, Y. and F. Hu (2012). Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics 40*(3), 1794 – 1815.

King, J. and W.-K. Wong (2000). Minimax D-Optimal Designs for the Logistic Model. *Bio-*

Statistica Sinica: Newly accepted Paper

metrics 56(4), 1263–1267.

Kong, L. and D. P. Wiens (2015). Model-Robust Designs for Quantile Regression. *Journal of the American Statistical Association 110*(509), 233–245.

Ma, W., P. Li, L.-X. Zhang, and F. Hu (2024). A New and Unified Family of Covariate Adaptive Randomization Procedures and Their Properties. *Journal of the American Statistical Association 119*(545), 151–162.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ritz, C., F. Baty, J. C. Streibig, and D. Gerhard (2016). Dose-Response Analysis Using R. *PLOS ONE 10*(12), 1–13.

Rodríguez-Torreblanca, C. and J. Rodríguez-Díaz (2007). Locally D- and c-optimal designs for Poisson and negative binomial regression models. *Metrika: International Journal for Theoretical and Applied Statistics 66*(2), 161–172.

Rosenberger, W. F. and O. Sverdlov (2008). Handling Covariates in the Design of Clinical Trials. *Statistical Science 23*(3), 404 – 419.

Russell, K. G., D. C. Woods, S. M. Lewis, and J. A. Eccleston (2009). D-OPTIMAL DESIGNS FOR POISSON REGRESSION MODELS. *Statistica Sinica 19*(2), 721–730.

Selvaratnam, S., L. Kong, and D. P. Wiens (2021). Model-robust designs for nonlinear quantile regression. *Statistical Methods in Medical Research 30*(1), 221–232.

Statistica Sinica: Newly accepted Paper

REFERENCES

Sinha, S. and D. P. Wiens (2002). Robust sequential designs for nonlinear regression. *Canadian Journal of Statistics 30*(4), 601–618.

Sun, K., L. Kong, H. Zhu, and C. Shi (2025). ARMA-Design: Optimal Treatment Allocation Strategies for A/B Testing in Partially Observable Time Series Experiments.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer. ISBN 0-387-95457-0.

Yang, M. (2008). A-optimal designs for generalized linear models with two parameters. *Journal of Statistical Planning and Inference 138*(3), 624–641.

Yang, M., S. Biedermann, and E. Tang (2013). On Optimal Designs for Nonlinear Models: A General and Efficient Algorithm. *Journal of the American Statistical Association 108*(504), 1411–1420.

Zhang, Y. and K. Ye (2014). Bayesian D-Optimal Designs for Poisson Regression Models. *Communications in Statistics - Theory and Methods 43*(6), 1234–1247.

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, T6G 2N8

E-mail: jaslesso@ualberta.ca

Department of Mathematics and Statistics, MacEwan University, Edmonton, AB, T5H 0K9

E-mail: hur3@macewan.ca

Statistica Sinica: Newly accepted Paper

REFERENCES

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, T6G 2N8

E-mail: lkong@ualberta.ca

Statistica Sinica