Induced replication and the assessment of models

Heather Battey* and Nancy Reid†

September 5, 2026

Abstract

The paper is a first attempt at a foundation of inference for the assessment of semiparametric and other highly-parametrised models, seeking reconciliation with the low-dimensional Fisherian foundations. We highlight the possibility, in some contexts, of avoiding the usual semiparametric considerations, which typically require estimation of nuisance components through kernel smoothing or basis expansion, with the associated difficulties of tuning-parameter choice that blur the distinction between estimation and model assessment. A key aspect of the present work is the inducement of replication under the postulated model. We show via examples how this can be achieved by exploiting some non-standard inferential separations. The separations are in the vein of ancillarity/co-ancillarity and sufficiency/co-sufficiency separations, allowing the replacement of out-of-sample prediction error as a criterion for semiparametric model assessment by a type of within-sample prediction error. Framed in this light are new formulations for the proportional hazards model, for a time-dependent Poisson process with semiparametric intensity function, and for matched-pair and two-group problems. Also subsumed within the framework is a post-reduction inference approach to the construction of confidence sets of sparse regression models. Perturbative cases are probed analytically and numerically. The purpose of the paper is to highlight a new perspective on a problem of broad relevance, and to probe some of its implications. A general methodological development based on these ideas is left to others.

Some key words: co-sufficiency; exchangeability; foundations of model inference; inferential separations; model adequacy; post-reduction inference.

1 Introduction

Two widely used approaches to assessing regression models, semiparametric or otherwise, are to substitute unknown regression parameters by estimates and check residuals for any anomalous behaviour, or to assess the predictive performance of the fitted model. The visually compelling but sometimes informal approaches based on residuals can be viewed as a way of evaluating predictive accuracy within sample. The two-stage exposition for parametric models presented by Barndorff-Nielsen and Cox (1994, p. 29) clarifies this claim:

*Department of Mathematics, Imperial College London, UK, h.battey@imperial.ac.uk.
†Department of Statistical Sciences, University of Toronto, Canada.

the statistician first learns the value $S = s^o$ of the sufficient statistic for a parameter of a given statistical model, before learning the additional information that the outcome is $Y = y^o$. If y^o is extreme when calibrated against the conditional distribution of Y given $S = s^o$, a prediction within sample, then that casts doubt on the adequacy of the model. Normal-theory linear regression example illustrates this sufficiency separation in a clean and transparent form.

Example 1.1. Consider the regression model $Y \sim N(\mu, \sigma^2 I_n)$, where $\mu = X\beta \in \mathcal{X} \subset \mathbb{R}^n$, with $\mathcal{X}$ the column space of X, full rank by assumption, and $\beta \in \mathbb{R}^{d_\beta}$. The residual vector $\hat{\varepsilon}$ belongs to the orthogonal complement $\mathcal{X}^\perp \subset \mathbb{R}^n$, which has dimension $n - d_\beta$. With $y^o \in \mathbb{R}^n$ the observed outcome and $y \in \mathbb{R}^n$ an arbitrary value, the n degrees of freedom for variation of y separate, first according to $\mathcal{X} \oplus \mathcal{X}^\perp = \mathbb{R}^n$ and subsequently within $\mathcal{X}^\perp$, where one degree of freedom is used for estimation of σ^2, leaving $n - (d_\beta + 1)$ degrees of freedom for assessment of the mean model $\mu \in \mathcal{X}$.

From a sufficient statistic, $S = s(Y) = (X^{\mathrm{T}}Y, \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon})$ say, the information remaining after conditioning on the observed value s^o of S has been referred to as *co-sufficient* (e.g Lockhart et al., 2009). In general there need not exist an explicit statistic that carries all the co-sufficient information without loss or redundancy, but in the present example, a co-sufficient statistic is available.

Let $V = (v_1, \ldots, v_{d_\beta})$ be the matrix of eigenvectors corresponding to the unit eigenvalues of the projection matrix $I - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$. Based on the coordinates $V^{\mathrm{T}}Y$ in basis V of the projection $VV^{\mathrm{T}}Y \in \mathcal{X}^\perp$, a statistic independent of the squared length $\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon} = \|V^{\mathrm{T}}Y\|^2$ is

$$Q = \frac{V^{\mathrm{T}}Y}{\|V^{\mathrm{T}}Y\|} \in \mathcal{S}^{n-d_\beta}, \tag{1}$$

which is therefore unaffected by conditioning on $S = s^o$. In (1), $\mathcal{S}^{n-d_\beta}$ is the unit hypersphere embedded in $\mathbb{R}^{n-d_\beta}$, and under the postulated mean model, Q is uniformly distributed on $\mathcal{S}^{n-d_\beta}$, a conclusion that holds even if the normal distribution for Y is replaced by a student t distribution or other spherically symmetric distribution (Muirhead, 1982, Ch. 1.5).

There are $n - d_\beta - 1$ degrees of freedom for variation of the angles $a_1, \ldots, a_{n-d_\beta-1}$, specifying the position of an arbitrary q on $\mathcal{S}^{n-d_\beta}$. This shows how the n degrees of freedom for variation in y decompose, without loss or redundancy, into sufficient and co-sufficient components as $n = d_\beta + 1 + (n - d_\beta - 1)$. That the angles $A_1, \ldots, A_{n-d_\beta-1}$ are also independent in a statistical sense under the postulated mean model provides a means of assessing model adequacy as discussed in §2. □

If a model contains a genuinely infinite-dimensional component, there is no hope for either inference or model assessment from n observations. In practice, however, semiparametric models are not infinite dimensional. For example, smoothness assumptions imply an approximating basis of order smaller than n, implicitly constraining the effective parameter space. The smoothness might be imposed using a truncated basis expansion, by regularisation, or by kernel smoothing. The associated tuning-parameter selection for such estimators is typically based on cross-validation, blurring the distinction between estimation and model assessment.

The present work is a first attempt to reconcile inference for semiparametric and other highly-parametrised models with the Fisherian parametric foundations, by illustrating the feasibility and merits of an approach to model assessment that evades estimation of the high- or infinite-dimensional nuisance component. The conceptual points are conveyed via deliberately idealised examples that isolate relevant structures in an incisive form. A general theory in this vein probably entails the inducement of approximate replication, in a sense that will become clear.

The following example exposes two important points: that smoothness or other assumptions on infinite-dimensional nuisance parameters can sometimes be removed, and that model assessment can be performed without estimating nuisance parameters, and without an additional sample to assess out of sample prediction error.

Example 1.2. Let $(Y_{j1}, Y_{j0})_{j=1}^m$ be outcomes on treated and untreated units from a matched-pair design. A general semiparametric specification allows each pair to have its own parameter γ_j, specifying the distribution at baseline, and a parameter ψ quantifying a treatment effect that is stable across the pairs. This formulation is more general than a typical semiparametric one in the following sense. Had covariates $x \in \mathbb{R}^p$ been available, the same model would have been valid with $\gamma_j = h(x_j)$ for a regression function $h : \mathbb{R}^p \to \Gamma$ of arbitrary and unspecified form, conventionally belonging to a smoothness class.

Let (Y_{j1}, Y_{j0}) be independent and exponentially distributed with rates $\gamma_j\psi$, γ_j respectively. In work primarily discussing rather different points, Battey and Cox (2020, §5) proposed an approach to model assessment for this example based implicitly on the following argument, a generalisation of the co-sufficiency argument. Suppose, for an analogue of a proof by contradiction, that the true distribution belongs to the postulated model and has true parameter value ψ^*, then $Y_{j0} + Y_{j1}\psi^* =: S_j(\psi^*)$ is sufficient for γ_j and has density function

$$f_{S_j(\psi^*)}(s) = \gamma_j^2 s \exp(-\gamma_j s), \tag{2}$$

i.e., $S_j(\psi^*)$ is gamma distributed with shape parameter 2 and rate parameter γ_j. The conditional density of Y_{j1} at y_{j1}, given $S_j(\psi^*) = s_j(\psi^*)$, is $\psi^*/s_j(\psi^*)$ uniformly in y_{j1}, showing that Y_{j1} is conditionally uniformly distributed between 0 and $s_j(\psi^*)/\psi^*$. Equivalently $U_j(\psi^*) := Y_{j1}\psi^*/s_j(\psi^*)$ has a standard uniform distribution for all $j = 1, \ldots, m$. The above conclusions are invalidated if ψ^* is replaced by any other value ψ_0, or if the model is wrong, suggesting a route to assessment of model adequacy via the empirical behaviour of $U_1(\psi_0), \ldots, U_m(\psi_0)$ for a postulated value ψ_0, which can be viewed as analogues of residuals.

Some postulated values ψ_0 may produce a distribution for $U_1(\psi_0), \ldots, U_m(\psi_0)$ that is hard to distinguish from a standard uniform sample even when the model is violated. Unless m is small, it is sensible to replace ψ_0 by the realisation of a consistent estimator $\hat{\psi}$, if such exists, the logic being that only one value of ψ then needs to be refuted in order to refute the model, rather than all possible values.

In this example, the transformed random variables $Z_j = Y_{j1}/Y_{j0}$ are, under the model, identically distributed with density function

$$f_Z(z; \psi^*) = \frac{\psi^*}{(1 + \psi^* z)^2}. \tag{3}$$

Since (3) is free of all the pair-specific nuisance parameters, inference on ψ^* can be constructed from a likelihood function based on (3), the resulting maximum-likelihood estimator $\hat{\psi}$ being consistent as $m \to \infty$. An approach replacing ψ^* by $\hat{\psi}$ in $U_1(\psi^*), \ldots, U_m(\psi^*)$ can be formally justified using standard arguments recorded in §4.3. □

Example 1.2 replaces the usual semiparametric approach of assessing prediction error after regularised estimation by an in-sample assessment based on the structure of the postulated model, avoiding, without sample splitting, any post-selection model inference issues that would arise from using the same sample to choose tuning parameters and to assess the model.

2 Replication inducement

Underpinning the success of Example 1.2, and other semiparametric examples to be presented, is the possibility of inducing an internal replication, the strongest form of which is defined in Definition 2.2. The term *internal replication* refers to all types of replication of the forms described by Definitions 2.2-2.4. Some readers may prefer to see these definitions after reading the examples in §5 and §6, which motivated their formulation. With the low-dimensional foundations in mind, we are led to constructions that retain a form of co-sufficient or ancillary information in the internal replicates, as discussed in §3.2.

Definition 2.1 (**Induced exchangeability**)**.** Let $Y_1, \ldots, Y_n$, each valued in $\mathbb{F}$, be independent non-exchangeable outcomes under the postulated model m_0. Suppose there exists a function or composition of functions $h_0 : \mathbb{F}^n \to \mathbb{F}^m$, $m > 1$, depending on the postulated model but not depending on any unknown parameters, such that the joint distribution of the induced random variables satisfies $(V_1, \ldots, V_m) \stackrel{d}{=} (V_{\pi(1)}, \ldots, V_{\pi(m)})$ under the postulated model for any permutation $\pi : [m] \to [m]$, i.e. $V_1, \ldots, V_m$ are exchangeable. Then we say that h_0 is *exchangeability inducing* under the postulated model.

Definition 2.2 (**Induced replication**)**.** If, in Definition 2.1, $V_1, \ldots, V_m$ are independent and identically distributed, we say that h_0 is *replication inducing* under the postulated model. The distribution of the replicates may or may not depend on the unknown true value of a fixed low-dimensional parameter ψ.

It is not immediately clear that either an inducible exchangeability or an inducible replication is present in Example 1.1. The null marginal distribution function F_{A_j} of each angle A_j can be derived from the expression given by Muirhead (1982, p. 49). From these distribution functions, iid replicates $U_1, \ldots, U_{n-d_\beta-1}$, uniformly distributed in this case, can be constructed as $U_j = F_{A_j}(A_j)$ as in Example 1.2. The construction in Example 1.1 and here is such that, in high-dimensional problems where a preliminary reduction is needed, the post-reduction inference problem documented by Battey and Cox (2018) and Battey, Rasines & Tang (2025) is avoided without sample splitting in the assessment of a large number of possible models.

In Example 1.2 the function needed to induce the replication under the postulated model depends on a fixed-dimensional unknown parameter. This is reflected in the more general Definition 2.3.

Definition 2.3 (**Interest-dependent induced replication**). If, in Definition 2.2, h_0 depends on the true value ψ^* of an unknown parameter ψ of fixed finite dimension, then h_0 is said to be *interest-dependent replication inducing* under the postulated model, giving rise to parameter dependent replicates $V_1(\psi^*), \ldots, V_m(\psi^*)$.

Remark 2.1. If, as in Example 1.2, the unknown value ψ^* in Definition 2.3 is replaced by an estimate, the corresponding version of the induced random variables, $\hat{V}_1, \ldots, \hat{V}_m$ say, will typically not be independent, but they will be exchangeable. Provided that the estimator $\hat{\psi}$ is null-consistent for ψ^* as some quantity $n \to \infty$, $\hat{V}_1, \ldots, \hat{V}_m$ will be asymptotically independent in the same asymptotic regime.

Returning to Example 1.1, Battey, Rasines & Tang (2025) did not assess the adequacy of their models using the marginal distribution of the angles $A_1, \ldots, A_{n-d_\beta-1}$ but by injecting an additional component of randomness that produced iid copies of Q if and only if the postulated model mean model $\mu \in \mathcal{X}$ was true. Since the component of randomness was used only to assess model adequacy via the co-sufficient component Q, sufficiency was not violated. A definition that captures this external mechanism for replication inducement is as follows.

Definition 2.4 (**Randomisation-induced replication**). Let $я_1, \ldots, я_m$ be m independent external sources of randomness. These are either random variables independent of Y, or specify a component of randomness for independent replication from a suitable conditional distribution given $Y = y^o$. If there exist random variables $(\tilde{V}_j)_{j=1}^m$, constructed from Y and $я_j$ according to a rule h_0, that are iid under the postulated model, then h_0 is said to be *replication-inducing through randomisation.*

Exact replication, by whatever means, may not always be achievable, and it remains an open question whether a mechanism can always be found to approximate it. The present paper makes a more modest contribution, showing through carefully-constructed examples the diversity of ways through which an internal replication can be induced.

3 Information extraction for model assessment

3.1 Earlier literature

Bartlett (1937) was, we believe, the first to note the relevance of sufficiency in the assessment of parametric models with an explicit sufficient statistic. Barndorff-Nielsen and Cox (1994, p.29) gave an expanded explanation as outlined in §1. Engen and Lillegård (1997), Lindqvist and Taraldsen (2005), and Lockhart et al. (2007) addressed the considerable challenges in making the observation operational, proposing an approach for simulating from the conditional outcome distribution given the observed value of the sufficient statistic. The term *co-sufficient sampling* was introduced by Lockhart et al. (2009) and adopted in subsequent work, notably in work by Barber and Janson (2022) who substantially advanced the methodology and associated theory for co-sufficient sampling. There is no literature aiming to adapt information partitions for parametric models to semiparametric and other challenging contexts. We outline some relevant structures below, which in some contexts allow us to invoke the concepts of §2.

3.2 Some non-standard inferential separations

A perspective of this paper is that different model structures sometimes have inferential separations useful for model assessment, complementing Bartlett's (1937) sufficiency separation, which uses the conditional distribution of an outcome Y given the observed value of a sufficient statistic S.

A first generalisation is to a parameter dependent statistic $S(\psi)$ that is sufficient for a high-dimensional nuisance parameter under the postulated model if and only if the $\psi = \psi^*$, the value of a low-dimensional parameter.

A second generalisation allows S or $S(\psi^*)$ to be conditionally sufficient, conditional on some relevant component of the data, such as a failure having occurred at a particular time point, in the case od the proportional hazards model.

An additional classical route to model assessment is based on an ancillary statistic, part of a minimal sufficient statistic in a parametric setting. The relevant generalisation considered in the present paper is to statistics that are ancillary for the high-dimensional nuisance parameter and whose distribution typically depends on the unknown value of a low-dimensional parameter. For the proportional hazards model, the ancillary statistic for the infinite-dimensional baseline hazard function is conditionally sufficient for the low-dimensional parameter, showing that the notions can agree in some contexts.

To make the separations convenient for practical use, we suggest combining them with the notion of induced replication of §2. In some contexts there is a natural way to induce this replication because of the inherent structure of the problem, as in Example 1.2 where there was a shared parameter across pairs, even though there was no replication in the original statement of the problem. In more challenging settings, more work might be needed to uncover appropriate way to induce a replication under the postulated model, perhaps approximately. This is the context of the high-dimensional sparse regression setting of §5.3.

4 Some broad preliminary insight

4.1 Joint assessment of a model and its interest parameters

Since Definition 2.3 is more general than Definition 2.2, we discuss the situation for that case only. The discussion also covers the situation in which the replication is induced through randomisation as in Definition 2.4. For concreteness, and with no loss of generality, we take the replicates to be standard uniformly distributed when the postulated model is true with $\psi_0 = \psi^*$, and therefore use the suggestive notation $U_1(\psi_0), \ldots, U_m(\psi_0)$.

Supposing for an evidential analogue of proof by contradiction that the postulated model is true with parameter value ψ_0, the canonical way of combining the supposed independent standard-uniform replicates $U_1(\psi_0), \ldots, U_m(\psi_0)$ is via Fisher's 1932 statistic for combining p-values, whose traditional usage is in meta analysis.

The following initial discussion considers model adequacy in terms of a confidence set for the parameter ψ, before introducing in §4.3 a more powerful approach mirroring that in Example 1.2.

With G_{2m} the distribution function of a χ^2 random variable with $2m$ degrees of freedom, the two-sided p-value is $2\min\{p(\psi_0), 1 - p(\psi_0)\}$, where $p(\psi_0) := G_{2m}(-2\sum \log U_j(\psi_0))$.

The resulting α-level confidence set for the interest parameter ψ under correct specification of the model is

$$\mathcal{C}(\alpha) := \Big\{\psi_0 \in \Psi : 2\min\{p(\psi_0), 1 - p(\psi_0)\} > \alpha\Big\}. \tag{4}$$

If the confidence set is empty at a chosen level α, that casts doubt on the adequacy of the model. The use of Fisher's statistic in (4) is convenient for analytic calculations of power because the expectation and variance of $R_j = -\log U_j$ simplify, by integration by parts, to

$$\mathbb{E}(R_j) = \int_0^1 \frac{F_U(u)}{u} du, \quad \mathbb{V}(R_j) = -2\int_0^1 \frac{\log(u) F_U(u)}{u} du - \Big(\int_0^1 \frac{F_U(u)}{u} du\Big)^2, \tag{5}$$

where F_U is the distribution function of U_j.

Both moments increase as the density function f_U of U concentrates near zero, and the mean exceeds half the variance, and hence departs from the null χ^2_2 behaviour, when f_U departs from uniformity asymmetrically towards zero. When departures instead occur towards 1, the mean and variance are typically both too small to be compatible with the null distribution, although in this case sensitivity can be increased by replacing $U_j(\psi_0)$ by $1 - U_j(\psi_0)$ in $p(\psi_0)$ from (4). Thus, if the majority of departures from uniformity are in the same direction, $\mathcal{C}(\alpha)$ is empty with high probability for sufficiently large m when the postulated model is misspecified. More irregular departures from uniformity may be better detected by alternative combination rules (see e.g. Birnbaum, 1954; Heard and Rubin-Delanchy, 2018).

4.2 Two parallel analyses

In subsequent sections, we present the main conceptual development of the paper via examples. These illustrate different ways through which internal replication can be induced under the postulated model, and through which contradictions are forced when the model is violated. Before turning to those constructions, we analyse the behaviour of the confidence set (4), assuming the existence of random variables $U_j(\psi_0)$. The aim is to supply qualitative insight through a reasonably general analysis, showing how systematic departures from uniformity translate to eventual emptiness of $\mathcal{C}(\alpha)$ as $m \to \infty$. Calculations at this level of generality are inevitably idealised.

When the model is erroneous or if it contains the true distribution but $\psi_0 \neq \psi^*$, then $U_j(\psi_0)$, $j = 1, \ldots, m$, are by definition not standard uniform, and for most of the example cases we have in mind, are not identically distributed either. Since $U_j(\psi_0)$ is necessarily supported on $[0, 1]$, insight can be obtained by approximating its density function using the parametric family

$$(1 - \vartheta_j)u^{-\vartheta_j}, \quad 0 < u < 1, \quad 0 < \vartheta_j < 1, \tag{6}$$

so that the null distribution is recovered in the limit as $\vartheta_j \to 0$. The model (6) is deliberately oversimplified in that it specifies the directions of departure from uniformity to be in the same direction for all j. Specifically, the distribution (6) concentrates towards zero for all ϑ_j, the opposite behaviour achieved only by replacing $U_j(\psi_0)$ by $1 - U_j(\psi_0)$.

Under (6), the kth moment of $R_j := -\log U_j(\psi_0)$ has the convenient form

$$\mathbb{E}(R_j^k) = \frac{\Gamma(k+1)}{(1-\vartheta_j)^k}, \quad k \in \{1, 2, \ldots\}.$$

The mean and variance of $A_m := 2\sum_{j=1}^m R_j$ are

$$\mathbb{E}(A_m) = 2\sum_{j=1}^m (1-\vartheta_j)^{-1}, \quad \mathbb{V}(A_m) = 4\sum_{j=1}^m (1-\vartheta_j)^{-2}, \tag{7}$$

recovering the χ^2_{2m} mean and variance as $\vartheta_j \to 0$ for all j and showing that A_m has a larger mean and variance than a χ^2_{2m} random variable, with the discrepancies increasing as $m \to \infty$.

On letting k_α be the $1-\alpha$ quantile of the χ^2_{2m} distribution, Markov's inequality in the form presented in Appendix A shows that

$$\mathrm{pr}(A_m \geq k_\alpha) \geq 1 - \inf_{s>0}\left(e^{sk_\alpha} \prod_{j=1}^m \frac{1-\vartheta_j}{1-\vartheta_j+2s} \right). \tag{8}$$

The constituent terms are in the interval $(0,1)$ for all s in the permissible range, so the product tends to zero exponentially fast in $m \to \infty$.

For a parallel analysis that does not involve specifying a parametric family for $U_j(\psi_0)$, let $\mu_m(\psi_0) = \sum_{j=1}^m \mathbb{E}(R_j)$ and $\tau_m(\psi_0) = (\sum_{j=1}^m \mathrm{var}(R_j))^{1/2}$. Then by the central limit theorem $\tau_m(\psi_0)^{-1}(\sum_{j=1}^m R_j - \mu_m(\psi_0))$ is asymptotically standard normal for large m, thus

$$\mathrm{pr}(A_m \geq k_\alpha) \simeq 1 - \Phi\left(\frac{k_\alpha - 2\mu_m(\psi_0)}{2\tau_m(\psi_0)} \right) \simeq 1 - \Phi\left(\frac{2m + 2\sqrt{m}z_\alpha - 2\mu_m(\psi_0)}{2\tau_m(\psi_0)} \right), \tag{9}$$

as $m \to \infty$, where z_α is the $1-\alpha$ quantile of the standard normal distribution and where we have used the normal approximation to the χ^2_{2m} distribution in the form $k_\alpha \simeq 2m + 2\sqrt{m}z_\alpha$ for large m. Under the null, where the model is correct and $\psi_0 = \psi^*$, $U_j(\psi_0)$ has a standard uniform distribution for all j, R_j has a unit exponential distribution, implying that $\mu_m(\psi_0) = m$ and $\tau_m(\psi_0) = \sqrt{m}$. In this case, the previous display reduces to $\mathrm{pr}(A_m \geq k_\alpha) \simeq \alpha$, which is the exact answer for any m. When the null distribution is violated, $\mathrm{pr}(A_m \geq k_\alpha) \to 1$ if and only if $(k_\alpha - 2\mu_m(\psi_0))/2\tau_m(\psi_0) \to -\infty$, which requires that the mean grows faster than the standard deviation as $m \to \infty$. There is a parallel calculation for the left tail.

4.3 Model assessment using a null-consistent estimator of ψ^*

Example 1.2 illustrates a situation in which a consistent estimator of the interest parameter is available when the postulated model contains the true distribution, without any need to estimate the high-dimensional nuisance component. When a null-consistent estimator is available, and when the effective sample size is large enough that sampling variability of $\hat{\psi}$

around ψ^* does not materially affect the null distribution, there is no need to assess $U_j(\psi_0)$ for all plausible values ψ_0; instead $\hat{\psi}$ can be used in place of ψ_0, leading to the statistic

$$\hat{A}_m = -2\sum_{j=1}^{m} \log U_j(\hat{\psi}), \tag{10}$$

and its complementary version with $1 - U_j$ in place of U_j. The discussion following (6) illustrates that the two need not be interchangeable for the purpose of refuting a false postulated model, one version tending to have higher power than the other depending on the form of departure. If $\hat{A}_m$ is extreme when calibrated against the χ^2_{2m} distribution, that casts doubt on the adequacy of the model.

Under the postulated model with true parameter value ψ^*, let g be a function such that $U_j(\psi^*) = g(Z_j, \psi^*)$ for $j = 1, \ldots, m$ has a standard uniform distribution. The continuous random variable Z_j may itself an induced replicate according to one of the definitions in §2, in which case the only role of g is to standardise to standard uniformity. If Z_j is not an induced replicate, for instance if conditioning on a sufficient statistic for a nuisance parameter is required to eliminate them, then g is part of h_0 in Definition 2.3.

Let $\hat{\psi}$ be a null-consistent estimator of ψ^* as some quantity $n \to \infty$. For simplicity of exposition let ψ be a scalar, the only difficulties in extending to the vector case being notational. The usual linearisation argument gives

$$U_j(\hat{\psi}) = U_j(\psi^*) + o_p(1) \tag{11}$$

provided that the ψ-derivative of g is almost-surely bounded at Z_j in a neighbourhood of ψ^*. In other words, $U_j(\hat{\psi}) \rightsquigarrow U_j(\psi^*)$ where $\rightsquigarrow$ means weak convergence. If n is an increasing function of m as in Example 1.2, the above linearisation is ineffectual for showing that $\hat{A}_m$ has an approximate χ^2_{2m} distribution. In the notation of equation (9), $\mu_m(\psi^*) = m$ and $\tau_m(\psi^*) = \sqrt{m}$, thus the argument presented in equation (1.3) of Randles (1982) and the approximation $k_\alpha \simeq 2m + 2\sqrt{m}z_\alpha$ used in (9) implies

$$\mathrm{pr}(\hat{A}_m > k_\alpha) \simeq 1 - \Phi\left(\frac{k_\alpha - 2m}{2\sqrt{m}}\right) \simeq \alpha, \quad (m \to \infty) \tag{12}$$

provided that $\hat{\psi} \to_p \psi^*$ under the same asymptotic regime.

When the model is violated, $\hat{\psi}$ either converges in probability to a a fixed value $\psi_0^* \neq \psi^*$, violating standard uniformity of each $U_j(\psi_0^*)$ and therefore invalidating the mean and variance standardisation used in (12), or $U_j(\hat{\psi})$ has a non-degenerate limit distribution. This second scenario not only violates asymptotic uniformity of $U_j(\hat{\psi})$ for every j but also induces a complicated dependence between the $U_j(\hat{\psi})$, $j = 1, \ldots, m$, because of the shared dependence on $\hat{\psi}$, whose distribution depends on nuisance parameters when the postulated model is erroneous. In general, any such dependence is expected to exacerbate the distributional violations in the aggregate (10) or its null-standardised version (12), so may well be beneficial for detecting departures from the model.

5 Co-sufficient replication: non-matched examples

5.1 Introduction

The most obvious settings in which inducible replication is present under the postulated model are those in which there is a degree of replication for the nuisance parameter by design, as in Example 1.2. As illustrated in the example, there is no exchangeability across the outcomes, either within or between pair, but each nuisance parameter appears twice among the distributions of the outcomes, allowing replication to be easily induced under postulated models with an amenable structure. We start, however, with some different settings to provide the sharpest contrast with Example 1.2. Section 5.2 illustrates a more abstract form of internal replication after conditioning, §5.3 illustrates how internal replication can be achieved artificially through a carefully-designed randomisation scheme, and §5.4 illustrates the possibility of extending the notion of co-sufficiency to conditional co-sufficiency. We then return in §6 to the matched pair setting of Example 1.2, where we formalise some other constructions for achieving internal replication that do not directly correspond to co-sufficiency but follow the same broad principles.

5.2 Semiparametric time-dependent Poisson process

The following is a semi-parametric generalisation of a model proposed by Cox (1955) and best approached via Cox and Lewis (1966, pp. 45–46). Consider a time-dependent Poisson process with intensity function

$$\lambda_i(t) = e^{h(x_i)+\beta t}$$

for the ith of n individuals, where h is an unknown function of the covariate vector x_i. The model could be suitable for sequences of health events, in which $h(x_i)$ captures individual-specific effects and β captures the general effect of ageing.

To avoid assumptions on, and estimation of, h, write $\gamma_i = h(x_i)$. Suppose that in a fixed interval $(0, t_0)$, events are recorded at times $(t_{i1}, \ldots, t_{im_i})$ for individual i. The likelihood contribution for the ith individual is

$$L(\lambda_i; t_{i1}, \ldots, t_{im_i}) = \exp\Big\{-\int_0^{t_0} \lambda_i(u)du\Big\} \prod_{j=1}^{m_i} \lambda_i(t_{ij}),$$

which, for the special form $\lambda_i(t) = e^{\gamma_i+\beta t}$ becomes

$$L(\gamma_i, \beta; t_{i1}, \ldots, t_{im_i}) = \exp\Big\{m_i\gamma_i + \beta \sum_{j=1}^{m_i} t_{ij} - e^{\gamma_i}(e^{\beta t_0} - 1)/\beta\Big\}. \tag{13}$$

Thus, the jointly sufficient statistics for (γ_i, β) based on the ith individual are $(m_i, \sum_{j=1}^{m_i} t_{ij})$, and the conditional distribution of $\sum_{j=1}^{m_i} t_{ij}$ given m_i eliminates γ_i. There is, therefore, an induced replication in the conditional distributions across the n individuals. The marginal probability mass function for the number of events for the ith individual is Poisson with rate

$$\int_0^{t_0} \lambda_i(u)du = e^{\gamma_i}(e^{\beta t_0} - 1)/\beta.$$

Thus, on using the right hand side of (13), viewed as a function of $t_{i1}, \ldots, t_{im_i}$ for fixed parameter values, the conditional density is

$$f(t_{i1}, \ldots, t_{im_i} \mid m_i) = \frac{m_i! \beta^{m_i} \exp(\beta \sum_{j=1}^{m_i} t_{ij})}{(e^{\beta t_0} - 1)^{m_i}}, \tag{14}$$

which is the probability density function of an ordered sample of m_i random variables each with density function

$$f_T(t; \beta) = \frac{\beta e^{\beta t}}{(e^{\beta t_0} - 1)}, \quad 0 \leq t \leq t_0, \tag{15}$$

(Cox and Lewis, 1966, pp. 45–46). Equation (15) reduces to $1/t_0$ in the limit as $\beta \to 0$. It follows that the conditional distribution of $S_i = \sum_{j=1}^{m_i} T_{ij}$ given m_i is that of the sum of m_i independent random variables with density function (15). Cox and Lewis (1966) do not proceed further and suggest that a conditional likelihood based on (14) be used for inference on β. To assess the model via the approach of §4.3, explicit calculation of the conditional distribution of S_i given m_i is needed. Appendix B shows by way of a Laplace transform that the conditional distribution function with support $s < m_i t_0$ is

$$F_{S_i|M_i}(s \mid m_i; \beta) = \frac{\beta^{m_i}}{(e^{\beta t_0} - 1)^{m_i} \Gamma(m_i)} \sum_{v=0}^{\lfloor s/t_0 \rfloor} \binom{m_i}{v} (-1)^v e^{\beta v t_0} \int_0^{s - v t_0} e^{\beta w} w^{(m_i - 1)} dw, \tag{16}$$

where the integral is a gamma integral.

Although this alternating sum can in principle be evaluated at the observed values of S_i it is simpler, and more illustrative of general strategy when exact calculations are similarly complicated, to estimate the conditional distribution function of S_i based on Monte-Carlo replicates of sums of m_i random variables drawn from the density function (15). For large m_i, a normal approximation to the distribution can be used. Up to Monte Carlo sampling error, the resulting statistics follow a standard uniform distribution at the true value of β under correct specification of the model. The version based on the analogue of (10) uses the maximum likelihood estimate $\hat{\beta}$ in place of β, where $\hat{\beta}$ maximises the conditional log likelihood function based on the product of joint conditional density functions (14) over the n individuals.

Two adaptations of the model with intensity function $\lambda_i(t)$ as above have a quadratic trend of the form $\lambda_i(t) = \exp(\gamma_i + \beta_1 t + \beta_2 t^2)$ or a power-law trend of the form $\lambda_i(t) = e^{\gamma_i} t^{\beta}$. These are semiparametric analogues of models discussed by Cox (1955, p. 138), where the essential elements of the above argument apply on noting that, in the first case, m_i, $S_{i1} = \sum_{j=1}^{m_i} T_i$ and $S_{i2} = \sum_{j=1}^{m_i} T_i^2$ are jointly sufficient in the ith sample for γ_i, β_1 and β_2, and in the second case, m_i and $S_i = \sum_{j=1}^{m_i} \log T_i$ are jointly sufficient for γ_i and β.

5.3 Post-reduction confidence sets of models

For the construction of confidence sets of models in sparse high-dimensional regression, Battey and Cox (2018) used sample splitting to ensure appropriate coverage properties of the confidence set of models after preliminary reduction. Battey, Rasines & Tang (2025) explored a means by which, and consequences of, avoiding sample splitting using randomised

inference and the sufficiency/co-sufficiency separation presented in Example 1.1. They did not use the marginal distribution of the angles but worked instead with the random variable Q, generating iid replicates $\tilde{Q}_1, \ldots, \tilde{Q}_k$ using a generalisation of the randomisation scheme introduced by Rasines & Young (2023). Following the common logic of the present paper, the replication is induced if and only if the postulated model is correct. Battey, Rasines & Tang (2025) subsequently assessed compatibility using the Rayleigh test for uniformity on the sphere.

To reframe their approach using the unified exposition of §4, consider the $m = k(k-2)/2$ inner products between all pairs $\tilde{Q}_i$, $\tilde{Q}_j$ of synthetic replicates. Under the model hypothesis with d_0 explanatory variables, the m inner products (cosines of the angles between replicates), $Z_1, \ldots, Z_m$ say, each have density function (Fisher, 1915)

$$f_Z(z) = \frac{\Gamma((n-d_0)/2)}{\sqrt{\pi}\Gamma\{(n-d_0-1)/2\}}(1-z^2)^{(n-d_0-3)/2}, \quad -1 < z < 1. \tag{17}$$

Section 4 of the present paper provides an alternative to the Rayleigh test used by Battey, Rasines & Tang (2025). The distribution function corresponding to (17) is

$$F_Z(z) = \frac{1}{2} + \frac{2\Gamma((n-d_0)/2)\text{sign}(z)}{\sqrt{\pi}\Gamma\{(n-d_0-1)/2\}} \int_0^{z^2} u^{1/2-1}(1-u)^{(n-d_0-1)/2-1} du, \tag{18}$$

where the definite integral is the incomplete beta integral of arguments $1/2$, $(n-d_0-1)/2$ and r^2. The m inner products are converted to standard uniform random variables in the usual way $U_j = F_Z(Z_j)$, $j = 1, \ldots m$. Strictly speaking, only the angles between disjoint pairs of replicates are independent; in numerical work, however, it appears that if m is small relative to $n - d_0$, the dependence across all m of them is sufficiently weak that the χ^2_{2m} null distribution of (10) holds to close approximation; see column 1 of Table 5 in the supplementary file, where the empirical coverage probabilities of the model confidence sets and the number of false models in the set are reported.

Superficially, this example appears totally different from those introduced in previous sections, but the logic is the same: for every model postulated for inclusion in the confidence set, replication under this model is induced via the sequence of steps leading to $U_1, \ldots, U_m$. If their realised values are collectively extreme when calibrated against the distribution expected under the postulated model, this casts doubt on the model, leading to its exclusion from the confidence set.

5.4 Proportional hazards and conditional co-sufficiency

The proportional hazards model $h(y; x) = h_0(y)g(x_i^{\mathrm{T}}\beta)$ is unusual in that the infinite-dimensional nuisance parameter, the baseline hazard function h_0, can be eliminated in the partial likelihood analysis for the regression coefficient β (Cox, 1972, 1975), circumventing estimation of h_0. This section frames partial likelihood in terms of a notion of conditional sufficiency for h_0, pointing to a corresponding notion of conditional co-sufficiency for assessment of the model.

On each of a number of independent individuals there is a survival time T_i, and individuals are subject to non-informative censoring, so that the observable outcome consists

of $Y_i = \min\{T_i, c_i\}$ and an indicator $d_i = 1$ if $Y_i = T_i$, and $d_i = 0$ otherwise. We initially consider the uncensored setting $d_i = 1$ for all i. Let $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$ denote the ordered failure times and let i_j be the index of the individual who fails at time $y_{(j)}$. Thus $i = i_j$ if and only if $y_i = y_{(j)}$. Together, $y_{(1)}, \ldots, y_{(n)}$ and $i_1, \ldots, i_n$ are equivalent to the unordered failure times $y_1, \ldots, y_n$, in the sense that the latter can be recovered from the former, and vice versa. Let $\mathscr{R}_j := \{i : y_i \geq y_{(j)}\}$ be the set of individuals still at risk of failure at time $y_{(j)}$.

The following arguments implicitly underpin the development of Cox (1972). By ignoring the indices $i_1, \ldots, i_n$, the ordered survival times $y_{(1)}, \ldots, y_{(n)}$ are detached from any covariate dependence and therefore must, in the absence of the information supplied by $i_1, \ldots, i_n$, tell us nothing about how covariates influence the distribution of survival times, a version of ancillarity for β in the presence of nuisance parameters. The distribution of these ordered times depends primarily on h_0, the common instantaneous probability of failure at baseline.

Let $I_1, \ldots, I_n$ denote the random variables whose realisations are $i_1, \ldots, i_n$, and consider the conditional distribution of $I_1, \ldots, I_n$ given $y_{(1)}, \ldots, y_{(n)}$. Write

$$\mathscr{H}_j = \{y_{(1)}, \ldots, y_{(j)}, i_1, \ldots, i_{j-1}\}$$

for the history up to time $y_{(j)}$. The conditional probability that $i_j = i$, i.e. the probability that a failure occurring at time $y_{(j)}$ is contributed by individual i, conditional on $\mathscr{H}_j$, is the familiar expression appearing in the partial likelihood for β:

$$p_{j,i}(\beta^*) := \mathrm{pr}(I_j = i \mid \mathscr{H}_j) = \frac{h(y_{(j)}; x_i)}{\sum_{k \in \mathscr{R}_j} h(y_{(j)}; x_k)} = \frac{g(x_i^{\mathrm{T}} \beta^*)}{\sum_{k \in \mathscr{R}_j} g(x_k^{\mathrm{T}} \beta^*)}. \tag{19}$$

Although the conditioning set is the entire history $\mathscr{H}_j$, the expression (19) is free of $y_{(1)}, \ldots, y_{(j-1)}$, a type of conditional sufficiency, for the nuisance parameter h_0, of the increasing sets of order statistics $\{y_{(1)}, \ldots, y_{(j-1)}\}$, conditional on a failure having occurred at time $y_{(j)}$.

The above discussion suggests an assessment of the proportional hazards assumption based on the conditional distributions of the indices I_j given the relevant history $\mathscr{H}_j$. This is a version of co-sufficiency relative to the nuisance parameter, and leads to assessment of the model based on the same inferential partition that enabled inference on β^*, which is defensible provided that the sample size is large enough to reliably estimate β^*.

The discreteness of the indices and the changing risk sets present considerable difficulties in making use of the above information partitions. An approach that is applicable more generally, however, is to make use of the asymptotic normality of the score based on a version of likelihood that preserves the relevant information, in this case the conditionally co-sufficient information for the nuisance parameter.

The partial likelihood is based on

$$\begin{aligned} \mathrm{pr}(I_1 = i_1, \ldots, I_n = i_n) &= \prod_{j=1}^{n} \mathrm{pr}(I_j = i_j \mid i_1, i_2, \ldots, i_{j-1}) \\ &= \prod_{j=1}^{n} \frac{g(x_{i_j}^{\mathrm{T}} \beta^*)}{\sum_{k \in \mathscr{R}_j} g(x_k^{\mathrm{T}} \beta^*)}, \end{aligned} \tag{20}$$

leading to a consistent estimator of β^* with familiar properties (Cox, 1972; Wong, 1982). The simplest way to achieve the required replication, although almost certainly not the most efficient way to proceed and therefore not explored formally, is to split the sample at random into m blocks $j = 1, \ldots, m$ of equal size and to form the partial log-likelihood $\ell_j(\beta)$ based on (20) in each block. The partial likelihood scores $S_j(\beta) = \nabla_\beta \ell_j(\beta)$ are approximately normally distributed at $\beta = \beta^*$ as $n/m \to \infty$. Independent approximately standard uniform replicates $U_1(\beta^*), \ldots, U_m(\beta^*)$ can be obtained from the probability integral transforms for the score statistics when β^* is known. Since β^* is unknown, it is replaced by the full-sample partial likelihood estimate $\hat{\beta}$ based on (20), as discussed in §4.3.

Since the partial likelihood score on the full sample is asymptotically normally distributed, it seems possible in principle to use a form of randomised inference in the vein of Rasines & Young (2023) or Dharamshi et al. (2026), but it is unclear to what extent the theoretical guarantees are affected by violation of exact normality; this investigation requires a paper of its own and need not be restricted to the partial likelihood.

6 Matched-pair and two-group examples

6.1 Inferential structures for matched-pair examples

The most obvious settings in which inducible replication is present under the postulated model are those in which principles of experimental design have been applied, the simplest being the matched-comparison designs of Example 1.2 leading to outcomes $(Y_{j1}, Y_{j0})_{j=1}^m$ on the treated and untreated units for the m pairs. These will serve as useful cases for some perturbative power analyses in §6.2 but their primary purpose in this section is to illustrate some additional structures in an incisive form.

Example 6.1 (Continuation of Example 1.2)**.** The statistic $U_j(\psi_0)$ constructed in Example 1.2 arises also through a complementary perspective valuable in other settings. On replacing the observed value $s_j(\psi_0)$ by the corresponding random variable, we obtain

$$U_j(\psi_0) = \frac{Y_{j1}\psi_0}{S_j(\psi_0)} = \frac{\psi_0 Z_j}{1 + \psi_0 Z_j}, \tag{21}$$

where $Z_j = Y_{j1}/Y_{j0}$. The distribution function of Z_j under the proportional rates model of Example 1.2 is

$$F_Z(z; \psi^*) = \frac{\psi^* z}{1 + \psi^* z}. \tag{22}$$

Thus (21) is $F_Z(Z_j; \psi_0)$, showing through a different route that uniformity is achieved if and only if the postulated model is an adequate approximation with $\psi_0 = \psi^*$. The transformation to $Z_j = Y_{j1}/Y_{j0}$ is the most immediate way to eliminate the nuisance parameters in this scale model. It was not a priori obvious, however, that this simple route to internal replication preserves the co-sufficient information, in the generalised sense used in Example 1.2. □

The set of random variables $Z_1, \ldots, Z_m$ from Example 6.1 might be viewed from a different perspective as ancillary for the nuisance parameters, in the sense that, from observation of these statistics alone, estimation of $\gamma_1, \ldots, \gamma_m$ is impossible. There is some

ambiguity of terminology in these highly parametrised models where generalised inferential separations are needed. In particular, in a conventional parametric context, an ancillary statistic must be part of the minimal sufficient statistic, and the co-sufficient information is that left over after conditioning on the latter. When there is no exact sufficiency reduction, such as here, the separations are less clearly defined.

Example 6.2. Suppose that the true distribution of (Y_{j1}, Y_{j0}) belongs to an additive exponential model with rates $\gamma_j + \Delta$ for the treated individual in pair j and γ_j for the untreated individual, with true value Δ^* for the treatment parameter. This model would arise, for instance, for the first event time in a Poisson process, where an additional source of events is present in the treatment group. In the additive model, the statistic $S_j = Y_{j1} + Y_{j0}$ is sufficient for the nuisance parameter γ_j and the conditional distribution function of Y_{j1} given $S_j = s_j$ is

$$F_{Y_{j1}|S_j}(y \,|\, s_j \,; \Delta^*) = \frac{1 - e^{-\Delta^* y}}{1 - e^{-\Delta^* s_j}}. \tag{23}$$

Provided that the model is correctly specified, the random variables $U_j(\Delta^*) := F_{Y_{j1}|S_j}(Y_{j1} \,|\, s_j \,; \Delta^*)$ are uniformly distributed conditional on $S_j = s_j$, and therefore also unconditionally. The above conclusion is invalidated if Δ^* is replaced by another value Δ_0. □

Replication-inducing constructions for matched pairs fall into two broad classes. The first eliminates pair-specific nuisance parameters by conditioning on a co-sufficient statistic, as illustrated in Examples 1.2 and 6.2; the second eliminates them by transformation to a random variable whose distribution is free of nuisance parameters, as illustrated by Example 6.1. Each of these may or may not depend on the parameter of interest, leading to four generic inferential structures.

Structure 1. *The joint density function $f(y_1, y_0; \gamma_j, \psi)$ of (Y_{j1}, Y_{j0}) is such that there is a statistic S_j, not depending on ψ, such that $S_j := s(Y_{j1}, Y_{j0})$ is sufficient for γ_j and the conditional distribution function of $F_{Y_{j1}|S_j}(y \,|\, s_j; \psi)$ is continuous and bijective in y for any ψ, and injective in ψ for any fixed y.*

To exploit Structure 1, it is immaterial whether Y_{j1} or Y_{j0} is used in the conditional distribution.

Structure 2. *The second co-sufficiency structure parallels Structure 1, except that the sufficient statistic depends on the interest parameter ψ, as in Example 1.2.*

The two co-sufficiency structures contain all the residual information having eliminated the nuisance parameters through conditioning. Structures 1 and 2 lead therefore to natural analogues of Bartlett's (1937) argument. In view of (21), however, there is also interest in replication induced via a version of ancillarity. In an idealised parametric analysis, an ancillary statistic can be calibrated against its marginal distribution for assessment of the model. The appropriate analogue in highly parametrised settings is ancillarity for the nuisance parameters in the relaxed sense that, from observation of ancillary random variables alone, the nuisance parameters cannot be estimated. The above relaxed definition was implicit in constructions used by R. A. Fisher and D. R. Cox (e.g. Fisher, 1935; Cox, 1958). Many formalised versions of approximate ancillarity for an interest parameter

in the presence of nuisance parameters have been put forward for conditional inference on a parameter of interest (e.g. Barndorff-Nielsen and Cox, 1994, p.38). Ancillarity for a nuisance parameter is not, however, a standard concept, and its relevance for model assessment is an observation new to this work.

Structure 3. *The joint density function $f(y_1, y_0; \gamma_j, \psi)$ of (Y_{j1}, Y_{j0}) is such that there is a transformation t, not depending on ψ, such that $Z_j := t(Y_{j1}, Y_{j0})$ has a distribution function $F_Z(z; \psi)$ that is continuous and bijective in z for any fixed ψ, injective in ψ for any fixed z, and functionally independent of the nuisance parameter γ_j. While Structure 3 only applies exactly in the case of transformation models, Structure 4 extends beyond these classes, as illustrated below in Example 6.5.*

Structure 4. *Structure 4 parallels Structure 3, except that the transformation depends on the interest parameter ψ.*

The following proposition formalises the sense in which each structure induces internal replication. The propositions refer, for each of Structures 1–4, to equivalence classes of density functions within which the approach has no power to reject the postulated model. If the true density function violates the postulated model but belongs to the stated equivalence class, then the model violation will not be detected at any sample size. An analogy is to Markov equivalence classes in Gaussian graphical models, where multiple causal models give rise to the same distribution over the outcomes and therefore are statistically indistinguishable.

Proposition 6.1. Let (Y_{j1}, Y_{j0}) have a joint density function, provisionally assumed to belong to the model $\mathcal{M} = \{f(y_{j1}, y_{j0}; \gamma_j, \psi) : \gamma \in \Gamma, \psi \in \Psi\}$, where f satisfies one of Structures 1–4. When any such structure holds, define, respectively

$$
\begin{aligned}
U_j^{(1)}(\psi_0) &:= F_{Y_1|S}(Y_{j1} \,|\, s_j; \psi_0), & S_j &= s(Y_{j1}, Y_{j0}),\\
U_j^{(2)}(\psi_0) &:= F_{Y_{j1}|S_j(\psi_0)}(Y_{j1} \,|\, s_j(\psi_0)), & S_j(\psi_0) &= s(Y_{j1}, Y_{j0}; \psi_0),\\
U_j^{(3)}(\psi_0) &:= F_{Z_j}(Z_j; \psi_0), & Z_j &= t(Y_{j1}, Y_{j0}),\\
U_j^{(4)}(\psi_0) &:= F_{Z_j(\psi_0)}(Z_j(\psi_0)), & Z_j(\psi_0) &= t(Y_{j1}, Y_{j0}; \psi_0).
\end{aligned}
$$

Then $U^{(k)}(\psi_0) \sim U(0, 1)$ if and only if $\mathcal{M}$ holds with true parameter value $\psi^* = \psi_0$, or if the true distribution of (Y_0, Y_1) with density function g^* belongs to the equivalence class $\mathcal{E}^{(k)}(\psi_0)$, as specified in Appendix D.

Proof. A proof is given in the appendix. □

The four equivalence classes presented in the appendix are best explained with reference to the previous examples. Thus, in the context of Examples 1.2 and 6.1, the relevant equivalence class is $\mathcal{E}^{(3)}(\psi_0)$ and consists of all distributions over the original random variables (Y_{j1}, Y_{j0}) for which the induced distribution over the ratios $Z_j = Y_{j1}/Y_{j0}$ is identical to that postulated under $\mathcal{M}$ with the postulated value of ψ_0. It is clear from this example and others that members of the equivalence class, if any such member exists, must be such that any pair-specific heterogeneity is eliminated through the same operation as under the

postulated model, as otherwise no replication would be induced and standard uniformity for every pair could not be achieved. Elimination of pair dependence is a necessary but not a sufficient condition. Consider, for instance, a true distribution belonging to the Weibull family with multiplicative treatment effect on the rate scale and a postulated model that is exponential, a special case, then ratios $Z_j = Y_{j1}/Y_{j0}$ eliminate the pair parameters in both cases, but no members of the Weibull model with non-unit shape parameter belong to the equivalence class $\mathcal{E}^{(3)}(\psi_0)$ for any value of ψ_0, as the induced distribution over Z_j is different.

For a different perspective, consider fitting ψ by maximum likelihood in the family of densities for the induced random variable corresponding to Structure 1 or 3, i.e. the conditional and marginal density functions. If the true density function g^* belongs to $\{\mathcal{E}^{(k)}(\psi_0) : \psi_0 \in \Psi\}$ for $k = 1, 3$ respectively, then the limiting maximum likelihood solution ψ_0^* under the postulated model gives zero Kullback-Leibler divergence within the induced families. In other words, there is a point in the parameter space Ψ for which the induced model on $f_{Y_1|S}$ or f_Z respectively intersects with a "true model" on the induced random variables, this being any family of induced models that contains the true distribution over the induced random variables.

While Proposition 6.1 ensures appropriate coverage when the postulated model is correct, it does not automatically translate to a powerful method for rejecting false models, even when they do not intersect with the equivalence classes $\{\mathcal{E}^{(k)}(\psi_0) : \psi_0 \in \Psi\}$, since when the postulated model is violated at ψ_0, the notional replicates $U_j^{(k)}(\psi_0)$ typically depend on the nuisance parameters γ_j, and the summation in (4) does not necessarily produce an aggregate that is sufficiently far in the tail of a χ^2_{2m} distribution. It is partly for this reason that the version based on $\hat{\psi}$ from §4.3 is preferred. We explore the above examples in §6.2.

6.2 Some local power analyses

Here we consider examples in which the true distribution belongs to a model that departs only slightly from that postulated. This is done partly with a view to assessing the effectiveness of (4) for joint assessment of the model and its parameters relative to the approach in §4.3, where a consistent estimator of the interest parameter is used. In the first of the examples below, the same preliminary reductions eliminate the nuisance parameters under both models, so that there is no excess variability under the true model, relative to that postulated, induced through the pair-specific nuisance parameters.

Example 6.3. Let the postulated model be the proportional rates model of Example 1.2 and suppose that true distribution of the paired outcomes is Weibull with rate parameters γ_j, and $\gamma_j\psi^*$, respectively for untreated and treated individuals in the jth pair. Suppose further that the shape parameter of these Weibull distributions is $\varsigma \neq 1$, so that the postulated model is misspecified. In spite of this, the same pairwise operation eliminates γ_j under both models. Specifically, the distribution and density functions of the ratios $Z_j = Y_{j1}/Y_{j0}$ are

$$F_Z(z) = \frac{\psi^* z^\varsigma}{1 + \psi^* z^\varsigma}, \qquad f_Z(z) = \frac{\psi^* \varsigma z^{\varsigma-1}}{(1 + \psi^* z^\varsigma)^2}, \tag{24}$$

and, by monotonicity, the distribution function of $U_j(\psi_0)$ in (21) is

$$\mathrm{pr}(U_j(\psi_0) \leq u) = \mathrm{pr}\Big(Z_j \leq \frac{u}{\psi_0(1-u)}\Big) = \frac{\psi^* u^\varsigma}{\psi_0^\varsigma(1-u)^\varsigma + \psi^* u^\varsigma}, \qquad 0 \leq u \leq 1, \tag{25}$$

which is non-uniform for all $\varsigma \neq 1$ and all ψ_0.

Although no choice of ψ_0 gives uniformity, the expectation of $R_j = -\log U_j(\psi_0)$ can be shown via (5) to be monotonically decreasing in ψ_0, implying the existence of a unique value of ψ_0 at which $\mathbb{E}(R_j) = 1$. Consequently, departures from exponentiality in the direction of the Weibull model are difficult to detect using the aggregation criterion (4).

For comparison, consider maximum likelihood estimation based on the induced density based on (22). When the true model is Weibull with ς close to 1, the limiting maximiser ψ_0^* of the expected log-likelihood satisfies $\psi_0^* = \psi^* + O(\varsigma - 1)$ (see Appendix E.1). Thus, the estimator remains locally consistent despite model misspecification, and by equation (25),

$$\mathrm{pr}(U_j(\hat{\psi}) \leq u) \to \frac{\psi^{*(1-\varsigma)} u^\varsigma}{(1-u)^\varsigma + \psi^{*(1-\varsigma)} u^\varsigma} \neq u, \quad j = 1, \ldots, m. \tag{26}$$

This shows that the criterion (10) has power to detect departures from the model even though the maximum likelihood estimator of ψ^*, based on the transformed random variables Z_j, is consistent in spite of the model misspecification. In particular, from (5), $\mathbb{E}(R_j) = 1$ only when $\varsigma = 1$, the point at which the Weibull model collapses to the exponential model. Figure 1 plots the logarithm of both $\mathbb{E}(R_j)$ and $\mathbb{V}(R_j)$, obtained by numerical integration as a function of ς and ψ^*. This shows that (10) is most sensitive to departures from the postulated model in the direction of $\varsigma < 1$, with neighbourhoods of $\varsigma = 1$ and ψ^* being difficult to detect, uniformly over the values of the other variable. There is power to detect departures from the model when (ς, ψ^*) take values in the upper right quadrant of the reported range. □

Example 6.4. Another type of departure from the proportional exponential model of Example 1.2 is in the direction of the additive-rates exponential model of Example 6.2, with treatment effect Δ^*. In the additive model, the nuisance parameters are eliminated by conditioning on the pairwise sums, while for the multiplicative model, conditioning is on $Y_{j0} + Y_{j1}\psi_0$ for postulated values of ψ. Since there is no true value ψ^* when the model is erroneous, this second conditioning argument is ineffective at eliminating the nuisance parameters except at $\psi_0 = 1$.

Since Structure 3 leads to an identical inferential procedure as Structure 2 under the multiplicative model, interest lies in the behaviour of $Z_j = Y_{j1}/Y_{j0}$ under the additive model, which has distribution function

$$F_{Z_j}(z) = \mathrm{pr}(Z_j \leq z) = \frac{(\gamma_j + \Delta^*)z}{\gamma_j + (\gamma_j + \Delta^*)z}. \tag{27}$$

Since (22) and (27) agree at the null effect value $\Delta^* = 0$, which corresponds to $\psi = 1$, we expect greater difficulty in detecting inadequacy of the multiplicative model for small values of Δ^*. A more exact analysis of power echoing the general discussion of §4.2 requires

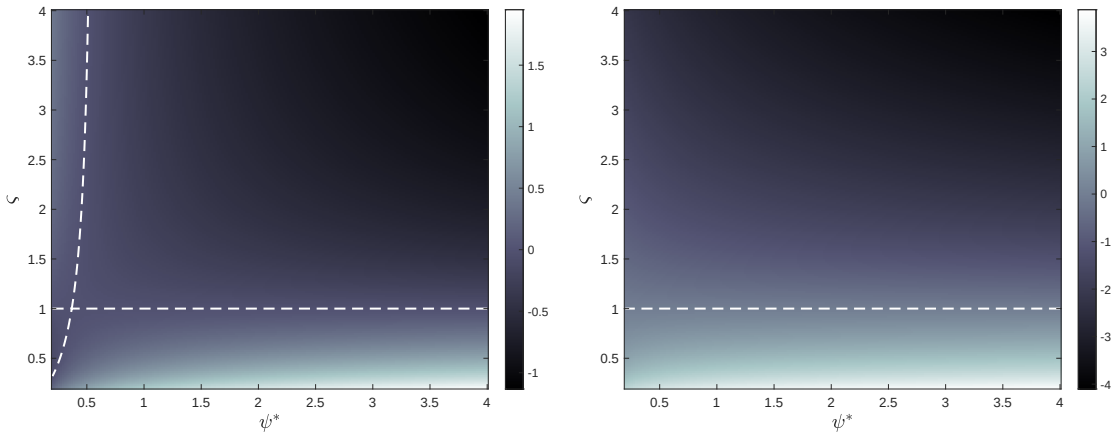

Figure 1: Logarithm of $\mathbb{E}(R_j)$ and $\mathbb{V}(R_j)$ calculated from (26) and (5), showing how the sensitivity to detect model misspecification varies with ς and ψ^*. Dashed lines indicate regions of the parameter space for the true Weibull model where $\mathbb{E}(R_j)$ and $\mathbb{V}(R_j)$ coincide with their putative values under the hypothesised exponential model.

calculation of the density function of $U_j(\psi_0) = \psi_0 Z_j/(1+\psi_0 Z_j)$ when the distribution of Z_j is specified by (27). By monotonicity,

$$\mathrm{pr}(U_j(\psi_0) \le u) = \mathrm{pr}\Big(Z_j \le \frac{u}{\psi_0(1-u)}\Big) = \frac{(\gamma_j+\Delta^*)u}{\gamma_j\psi_0(1-u)+(\gamma_j+\Delta^*)u}, \quad 0 \le u \le 1, \quad (28)$$

with corresponding density function

$$f_U(u) = \frac{\gamma_j(\gamma_j+\Delta^*)\psi_0}{(\gamma_j\psi_0(1-u)+(\gamma_j+\Delta^*)u)^2}, \quad 0 \le u \le 1.$$

This is not of the form (6), even approximately for Δ^* close to 0, but exact calculation via (5) shows that the expectation of $R_j = -\log U_j(\psi_0)$ is

$$\mathbb{E}(R_j) = \frac{\gamma_j+\Delta^*}{\gamma_j+\Delta^*-\gamma_j\psi_0}\log\Big(\frac{\gamma_j+\Delta^*}{\gamma_j\psi_0}\Big) = \frac{\eta_j\log\eta_j}{\eta_j-1}, \quad \gamma_j+\Delta^* > 0,$$

where $\eta_j = (\gamma_j+\Delta^*)/\gamma_j\psi_0$. This is monotonically increasing in $\eta_j > 0$ and crosses $\mathbb{E}(R_j) = 1$ at $\eta_j = 1$, which corresponds to $\Delta^* > \gamma_j(\psi_0 - 1)$. Thus, sensitivity in the right tail tends to increase with Δ^*, and values $\psi_0 \le 1$ are more easily detected as erroneous than $\psi_0 > 1$ when $\Delta^* > 0$, confirming intuition. The criterion $\Delta^* > \gamma_j(\psi_0 - 1)$ does, however, suggest that there are larger values of postulated ψ_0 that make the erroneous postulated model difficult to refute on the basis of the aggregation criterion (4), in spite of violation of standard uniformity; this is supported by numerical checks. It seems necessary in this example, therefore, to use a statistic based on $\hat{\psi}$, e.g. (10), echoing the conclusion from the previous example.

Appendix E.2 presents a local asymptotic expansion of the maximum likelihood solution in a neighbourhood of the point of intersection $\Delta^* = 0$ of the two models showing, as expected, that there is little sensitivity to departures from the postulated multiplicative model in the direction of the additive model at small Δ^*. □

The purpose of Examples 6.3 and 6.4 is to provide insight into the behaviour by way of some special cases where intuition can be recovered from direct calculation. These perturbative cases are ones in which the postulated model is so close to the true one that there is little harm in using it as a basis for inference. More substantial violations of modelling assumptions are probed by simulation in §F.1 of the supplement.

6.3 Extension to unbalanced strata

The experimental setting of §6 is a special case of a more general two-group problem arising in observational settings. Specifically, observations on treated and untreated individuals are stratified into groups that are as similar as possible. This typically leads to unbalanced strata, having a different number of individuals in the treated and untreated groups. Inference is based on the sufficient statistics S_{j1} and S_{j0} within treatment groups and strata, which can be treated in an analogous way to the paired observations of §6, as the strata sizes r_{j1} and r_{j0} are known.

Let $(Y_{ij1})_{i=1}^{r_{j1}}$ and $(Y_{ij0})_{i=1}^{r_{j0}}$ be observations within the jth stratum for treated and untreated individuals respectively. As a first example, if the observations $(Y_{ij1})_{i=1}^{r_{j1}}$ and $(Y_{ij0})_{i=1}^{r_{j0}}$ are normally distributed with means $\gamma_j + \psi^*$ and γ_j and variance τ, the likelihood contribution to the jth stratum depends on the data only through $S_{j1} = \sum_{i=1}^{r_{j1}} Y_{ij1}/r_{j1}$ and $S_{j0} = \sum_{i=1}^{r_{j0}} Y_{ij0}/r_{j0}$. The difference $Z_j = \sum_{i=1}^{r_{j1}} Y_{ij1}/r_{j1} - \sum_{i=1}^{r_{j0}} Y_{ij0}/r_{j0}$ is normally distributed of mean ψ^* and variance $\tau r_{j0} r_{j1}/(r_{j0} + r_{j1})$, which can be handled as discussed in §4.3 with τ also estimated. This is an example of Structure 3, but the same answer is achieved via a conditioning argument based on Structure 1.

If the individual observations are Poisson distributed counts with rates $\gamma_j\psi^*$ and γ_j, the sufficient statistics S_{j1} and S_{j0} are sums of these counts, Poisson distributed with rates $r_{j1}\gamma_j\psi^*$ and $r_{j0}\gamma_j$. The distribution of S_{j1} or S_{j0} conditional on $S_{j1} + S_{j0}$ eliminates γ_j in the usual way. This is an example of Structure 1.

Example 6.5 is more interesting as it relies on Structure 4 to eliminate the nuisance parameters.

Example 6.5. If the originating variables are exponentially distributed with rates $\gamma_j\psi^*$ and γ_j, the sufficient statistics S_{j1} and S_{j0} are gamma-distributed sums, with shape and rate parameters $(r_{j1}, \gamma_j\psi^*)$ and (r_{j0}, γ_j) respectively. Then

$$Z_j(\psi^*) := \frac{r_{j0}\psi^* S_{j1}}{r_{j1} S_{j0}}$$

has the F distribution with parameters $2r_{j1}$ and $2r_{j0}$. The strategy of §4.3 based on Proposition 6.1 then applies directly. □

7 Brief summary of numerical work

The supplementary file confirms the expected behaviour for several of the examples above, showing recovery of nominal error rates when the postulated model is true and high sensitivity to detect departures from an erroneous postulated model when a different semiparametric model generated the data. For brevity, we summarise one example that illustrates these properties.

The postulated model is the inhomogeneous Poisson process discussed in §5.2 with semiparametric intensity function $\lambda_i(t) = e^{\gamma_i + \beta t}$ for the ith individual. Simulation is not straightforward and details are available in the supplement. To assess cases where the postulated model is erroneous, we also simulated from an inhomogeneous Poisson process with semiparametric intensity function specified differently as $\lambda_i(t) = e^{\gamma_i} t^{\rho}$ with $\rho \in \{-0.5, 0, 0.5, 1\}$.

By inducing an internal replication as discussed in §5.2, the postulated model is assessed for its compatibility with the data and the results are reported in Table 1, revealing high power to detect departures from the postulated model even at values of ρ very close to the point of intersection $\rho = 0$. Left and right sensitivity refers to whether $U_j(\hat{\psi})$ or $1 - U_j(\hat{\psi})$ was used in the construction of Fisher's statistic (10).

Direction of sensitivity	ρ	Square root of number of individuals n					
		3	4	5	6	7	8
left	**0**	**0.03**	**0.04**	**0**	**0.03**	**0.02**	**0.05**
right	**0**	**0.03**	**0.03**	**0.02**	**0.03**	**0.01**	**0.03**
left	0.1	0.18	0.34	0.42	0.61	0.73	0.92
right	0.1	0.11	0.17	0.23	0.33	0.47	0.51
left	0.2	0.72	0.88	0.98	1	1	1
right	0.2	0.66	0.81	0.95	0.99	1	1

Table 1: The generating process is the inhomogeneous Poisson process with intensity function $\lambda_i(t) = e^{\gamma_i + \rho \log t}$; the postulated model has intensity function $\lambda_i(t) = e^{\gamma_i + \beta t}$. Proportion of 200 simulation runs in which the 5% test based on $F_{S_i|M_i}$ and (10) rejects the postulated model. The point of model intersection $\rho = 0$ is highlighted.

8 Discussion and open problems

The work provides new perspectives on the assessment of semiparametric and highly parametrised models, illustrating the possibility and value, in settings where the model admits this, of circumventing estimation of the infinite- or high-dimensional component. The broad principle is to use the postulated model to induce replication of known form if and only if the postulated model holds to an adequate approximation. We presented deliberately idealised examples illustrating diverse routes to such inducement, in the hope of suggesting, to others with different perspectives, new directions for research on semiparametric model assessment.

Assessment of a statistical model for its compatibility with the data is inevitably a discrete problem unless competing models are nested, or can be artificially nested, allowing model space to be continuously traversed through variation of a parameter. By contrast, the simplest approaches to inference on the parameters of a given statistical model often involve maximisation of a log-likelihood or other objective function, and therefore typically invoke a notion of continuity on the parameter space. Perhaps for this reason, valuable inferential structures, such as interest-dependent co-sufficiency, and ancillarity for a nuisance parameter, have been overlooked in the literature, yet emerge naturally for model

assessment.

There are broad principles underpinning all of the examples presented, extracted in §4, however the exact manner in which the replication is induced appears to be problem-specific. This raises the question of whether any general routes to the exact or approximate inducement of internal replication might be formulated.

Barber and Janson (2022) presented in a particular context a notion of approximate exchangeability which could in principle be used to generalise Definition 2.1. Similar formulations for approximate notions of approximation replication would facilitate a more formal development beyond the idealised examples considered in the present paper. An alternative might seek to collapse nuisance parameters into a low-dimensional summary that is both estimable and relatively insensitive to local perturbations under the postulated model.

A reader has indicated a further example (Diggle et al., 2007). In this semiparametric model for longitudinal data, possibly subject to drop-out, an internal replication is induced through transformation to a cumulative scale. In the induced model, the nonparametric component is reduced to a martingale random effect, whose distribution is arbitrary and evaded in the analysis.

The same reader has asked about the trade-off between avoiding estimation of a high-dimensional nuisance parameter, as favoured in the present paper, and having available an estimate of the nuisance parameter, even if the estimate is poor. In the context of most of our examples, this might entail a reformulation $\gamma_i = h(x_i)$ where h is now the nuisance parameter to be estimated semiparametrically. Standard approaches to model assessment would then assess prediction performance out of sample. The development of a general framework for fair comparison seems challenging.

Acknowledgements

We are grateful to Daniel Farewell for comments on a first draft. HB was supported by a UK Engineering and Physical Sciences Research Council research fellowship (EP/T01864X/1).

References

Barber, R. and Janson, L. (2022). Testing goodness of fit and conditional independence with approximate cosufficient subsampling. *Ann Statist.*, 50, 2514–2544.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. Lond. A: Math. Phys. Sci.*, 106, 268–282.

Battey, H. S. and Cox, D. R. (2018). Large numbers of explanatory variables: a probabilistic assessment. *Proc. Roy. Soc. Lond. A: Math. Phys. Sci.*, 474, 20170631.

Battey, H. S. and Cox, D. R. (2020). High-dimensional nuisance parameters: an example from parametric survival analysis. *Information Geometry*, 3, 119–148.

Battey, H. S., Rasines, D. G. and Tang, Y. (2026). Post-reduction inference for confidence sets of models. *Biometrika*, article number asag045.

Birnbaum, A. (1954). Combining independent tests of significance. *J. Amer. Statist. Assoc.*, 49, 559–574.

Cox, D. R. (1955). Some statistical methods connected with series of events (with discussion). *J. R. Statist. Soc. B*, 17, 129–164.

Cox, D. R. (1958). The regression analysis of binary sequences (with discussion). *J. R. Statist. Soc. B*, 20, 215–242.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc. B*, 34, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62, 269–276.

Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Springer, London.

Dharamshi, A., Neufeld, A., Gao, L. L., Bien, J. and Witten, D. (2026). Decomposing Gaussians with unknown covariance. *Biometrika*, 113, article number asaf057.

Diggle, P., Farewell, D. and Henderson, R. (2007). Analysis of longitudinal data with drop-out. *J. R. Statist. Soc. C*, 56, 499–550.

Engen, S. and Lillegård, M. (1997). Stochastic simulations conditioned on sufficient statistics. *Biometrika*, 84, 235–240.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507–521.

Fisher, R. A. (1932). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Fisher, R. A. (1935). The logic of inductive inference. *J. R Statist. Soc.*, 98, 39–82.

Heard, N. A. and Rubin-Delanchy, P. (2018). Choosing between methods of combining p-values. *Biometrika*, 105, 239–246.

Lindqvist, B. H. and Taraldsen, G. (2005). Monte Carlo conditioning on a sufficient statistic. *Biometrika*, 92, 451–464.

Lockhart, R. A., O'Reilly, F. J. and Stephens, M. A. (2007). Use of the Gibbs sampler to obtain conditional tests, with applications. *Biometrika*, 94, 992–998.

Lockhart, R. A., O'Reilly, F. J. and Stephens, M. A. (2009). Exact conditional tests and approximate bootstrap tests for the von Mises distribution. *J. Statist. Theory and Practice*, 3, 543–554.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York.

Randles, R. H. (2023). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.*, 10, 462–474.

Rasines, D. G. and Young, G. A. (2023). Splitting strategies for post-selection inference. *Biometrika*, 110, 597–614.

Wong, W. (1982). Theory of partial likelihood. *Ann. Statist.*, 14, 88–123.

A Derivation of equation (8)

This is a standard argument. Let k_α be the $1-\alpha$ quantile of the χ^2_{2m} distribution, Markov's inequality implies

$$\mathrm{pr}(A_m \geq k_\alpha) = 1 - \mathrm{pr}(-A_m \geq -k_\alpha) \geq 1 - e^{sk_\alpha}\mathbb{E}(e^{-sA_m}), \quad s > 0.$$

Direct calculation shows that

$$\mathbb{E}(e^{-sA_m}) = \prod_{j=1}^{m} \frac{1-\vartheta_j}{1-\vartheta_j+2s}, \quad s > 0.$$

B Derivation of equation (16)

The Laplace transform of (15) is

$$f_T^*(z) = \frac{\beta(e^{(\beta-z)t_0}-1)}{(e^{\beta t_0}-1)(\beta-z)},$$

thus the conditional density function of S_i is, for any $\tau > \beta$,

$$\begin{aligned} f_{S_i}(s \mid m_i;\beta) &= \frac{\beta^{m_i}}{(e^{\beta t_0}-1)^{m_i}} \frac{1}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} e^{zs} \frac{(e^{(\beta-z)t_0}-1)^{m_i}}{(\beta-z)^{m_i}} dz \\ &= \frac{\beta^{m_i}}{(e^{\beta t_0}-1)^{m_i}} \sum_{v=0}^{m_i} \binom{m_i}{v} (-1)^v e^{v\beta t_0} \frac{1}{2\pi i} \int_{\tau-i\infty}^{\tau+i\infty} \frac{e^{sz}e^{-vzt_0}}{(z-\beta)^{m_i}} dz \end{aligned}$$

by the binomial formula. Let

$$k^*(z) = \frac{e^{z(s-vt_0)}}{(z-\beta)^{m_i}}.$$

Its contour integral is the residue at $z = \beta$ which, for a pole of order m_i, is

$$\begin{aligned} \mathrm{Res}(k^*(z),\beta) &= \frac{1}{(m_i-1)!} \lim_{z\to\beta} \frac{d^{(m_i-1)}}{dz^{(m_i-1)}} \Big\{ (z-\beta)^{m_i} k^*(z) \Big\} \\ &= \frac{1}{(m_i-1)!} \lim_{z\to\beta} \frac{d^{(m_i-1)}}{dz^{(m_i-1)}} e^{z(s-vt_0)} \\ &= \frac{(s-vt_0)^{(m_i-1)}}{(m_i-1)!} e^{\beta(s-vt_0)}. \end{aligned}$$

Thus,

$$\begin{aligned} f_{S_i}(s \mid m_i;\beta) &= \frac{\beta^{m_i}e^{\beta s}}{(e^{\beta t_0}-1)^{m_i}\Gamma(m_i)} \sum_{v=0}^{m_i} \binom{m_i}{v} (-1)^v (s-vt_0)^{(m_i-1)} \mathbb{I}\{s > vt_0\}, \\ &= \frac{\beta^{m_i}e^{\beta s}}{(e^{\beta t_0}-1)^{m_i}\Gamma(m_i)} \sum_{v=0}^{\lfloor s/t_0 \rfloor} \binom{m_i}{v} (-1)^v (s-vt_0)^{(m_i-1)}, \quad s < m_i t_0 \end{aligned} \tag{29}$$

and zero otherwise. Integration gives

$$F_{S_i|M_i}(s \mid m_i;\beta) = \frac{\beta^{m_i}}{(e^{\beta t_0}-1)^{m_i}\Gamma(m_i)} \sum_{v=0}^{\lfloor s/t_0 \rfloor} \binom{m_i}{v} (-1)^v e^{\beta v t_0} \int_0^{s-vt_0} e^{\beta w} w^{(m_i-1)} dw.$$

This is equation (16).

C Proof of Proposition 6.1

Proof. For notational convenience, introduce $V^{(k)}$ for the, perhaps notional, random variable relevant for exploiting a postulated model with Structure k. This is Z and $Z(\psi_0)$ for $k=3$ and $k=4$ respectively and is the notional random variable $Y_\bullet(S)$ and $Y_\bullet(S(\psi_0))$ for $k=1$ and $k=2$. These are not functions of S in the conventional sense, and are notional in the sense that they typically cannot be expressed in terms of the original random variables, but once $S=s$ or $S(\psi_0)=s(\psi_0)$ is observed, $Y_\bullet(S)$ and $Y_\bullet(S(\psi_0))$ have the conditional distribution of $Y_\bullet$ given $S=s$ or $S(\psi_0)=s(\psi_0)$.

Temporarily dropping superscripts, let $F_V(v;\psi_0)$ be the distribution function of V under the assumption that the postulated model is true at ψ_0. Let $G_V^*(v)$ be the true distribution of V. Thus, if G_V^* belongs to the family $\mathcal{F}_V := \{F_V(\cdot;\psi),\ \psi\in\Psi\}$, then there exists a ψ^* such that $G_V^*(v) = F_V(v;\psi^*)$.

One direction is by direct calculation: if $\psi_0=\psi^*$, then the distribution of $F_V(V;\psi_0)$ is uniform. For the converse direction, suppose for a contradiction that $F_V(V;\psi_0)$ is uniformly distributed but that $G_V^*(v) \neq F_V(v;\psi_0)$. By uniformity,

$$u = \mathrm{pr}\big(F_V(V;\psi_0) \le u\big) \tag{30}$$

for all $u\in[0,1]$. First suppose that $G_V^*\in\mathcal{F}_V$. Since $F_V(v,\psi)$ is bijective in v for any fixed ψ, there exists a $u\in[0,1]$ such that $F_V^{-1}(u;\psi_0) \neq F_V^{-1}(u;\psi_0')$ for any $\psi_0'\neq\psi_0$. Thus, for any such value of u, $F_V\big(F_V^{-1}(u;\psi_0);\psi^*\big) \neq F_V\big(F_V^{-1}(u;\psi^*);\psi^*\big)$, where the right hand side is equal to u by definition. It follows that $F_V\big(F_V^{-1}(u;\psi_0);\psi^*\big)\neq u$, which contradicts (30).

Suppose now that uniformity is achieved and $G_V^*\notin\mathcal{F}_V$. Then $G_V^*(F_V^{-1}(u;\psi_0))$ replaces $F_V\big(F_V^{-1}(u;\psi_0);\psi^*\big)$ in the argument of the previous paragraph. Equality for all u can only be achieved at the points ψ_0 where $G_V^*(v)=F_V(v;\psi_0)$. But at such points G_V^* belongs to $\mathcal{F}_V$, a contradiction.

The remaining question is whether there are distributions over (Y_0,Y_1) that induce the same distribution $G_V^*(v)=F_V(v;\psi_0)$ over V that would hold if the postulated model for (Y_0,Y_1) were true at ψ_0. The corresponding equivalence classes are most easily expressed in terms of density functions $g^*=g^*_{Y_0,Y_1}$, inducing a density function g_V^* on V. These are the sets $\mathcal{E}^{(k)}(\psi_0)$ from Proposition 6.1, made explicit in Appendix D. □

D Equivalence classes for Proposition 6.1

Proposition 6.1 refers to equivalence classes of density functions g over pairs (Y_0,Y_1), within which the approach to achieving internal replication based on Structures 1–4 has no power

to reject the postulated model. In other words, if the true density function violates the postulated model in the directions of any model belonging to the corresponding equivalence class, then this will not be detected at any sample size. As in the statement of Proposition 6.1, we drop the pair-index subscripts. With $|J|$ the absolute Jacobian determinant corresponding to the transformation from (Y_0, Y_1) to (S, Y_1) for $k = 1$, to $(S(\psi_0), Y_1)$ for $k = 2$, to (Z, Y_1) for $k = 3$ and to $(Z(\psi_0), Y_1)$ for $k = 4$, the equivalence classes are:

$$\mathcal{E}^{(1)}(\psi_0) = \Big\{ g : \frac{g\big(y_0(s, y_1), y_1\big)\,|J|}{\int_{\mathcal{Y}_1(s)} g\big(y_0(s, y_1), y_1\big)\,|J|\,dy_1} = f_{Y_1|S}(y_1 \mid s; \psi_0) \Big\},$$

$$\mathcal{E}^{(2)}(\psi_0) = \Big\{ g : \frac{g\big(y_0(s(\psi_0), y_1), y_1\big)\,|J|}{\int_{\mathcal{Y}_1(s(\psi_0))} g\big(y_0(s(\psi_0), y_1), y_1\big)\,|J|\,dy_1} = f_{Y_1|S(\psi_0)}\big(y_1 \mid s(\psi_0)\big) \Big\},$$

$$\mathcal{E}^{(3)}(\psi_0) = \Big\{ g : \int_{\mathcal{Y}_1} g\big(y_0(z, y_1), y_1\big)\,|J|\,dy_1 = f_Z(z; \psi_0) \Big\},$$

$$\mathcal{E}^{(4)}(\psi_0) = \Big\{ g : \int_{\mathcal{Y}_1} g\big(y_0(z(\psi_0), y_1), y_1\big)\,|J|\,dy_1 = f_{Z(\psi_0)}\big(z(\psi_0)\big) \Big\}.$$

The above equivalence relations are in terms of generic expressions and usually do not specify the simplest route to calculation in particular cases. For instance, the distribution of Z_j from Example 6.1 can typically be calculated directly without explicit preliminary transformation from (Y_0, Y_1) to (Z, Y_1).

E Derivations for §6.2

E.1 Local consistency under the Weibull model

Maximisation of the log-likelihood function based on (22) when the true distribution is given by (24) produces a maximum likelihood estimator $\hat{\psi} \to_p \psi_0^*$, where ψ_0^* is the value of ψ that maximises the expected log-likelihood function

$$\psi_0^* = \underset{\psi \in \mathbb{R}^+}{\operatorname{argmax}} \left(\log \psi - 2 \int_0^\infty \log(1 + \psi z) \frac{\psi^* \varsigma z^{\varsigma - 1} dz}{(1 + \psi^* z^\varsigma)^2} \right). \tag{31}$$

This solution satisfies

$$\frac{1}{\psi_0^*} = 2\psi^* \varsigma \int_0^\infty \frac{z^\varsigma dz}{(1 + \psi^* z^\varsigma)^2 (1 + \psi_0^* z)}. \tag{32}$$

Since we are interested in the behaviour near the point of intersection of the two models $\varsigma = 1$, write $\varsigma = 1 + \varepsilon$ and expand the integral for small ε, giving

$$\frac{1}{\psi_0^*} = 2\psi^*(1 + \varepsilon) \left(\int_0^\infty \frac{z dz}{(1 + \psi^* z)^2 (1 + \psi_0^* z)} + O(\varepsilon) \right)$$

On expanding the integral in partial fractions it follows that, to first order in ε, ψ_0^* solves

$$\frac{\psi^*}{\psi_0^*} = 2\psi^*(1 + \varepsilon) \left(\frac{\psi^* \log(\psi^*/\psi_0^*) + \psi_0^* - \psi^*}{(\psi^* - \psi_0^*)^2} \right).$$

Equivalently, $x = \psi^*/\psi_0^*$ solves

$$(x-1)^2 = 2(1+\varepsilon)x(\log x + 1/x - 1). \tag{33}$$

The left and right hands sides are both convex with a unique minimum at $x = 1$ for any ε. Thus, to first order in ε, the maximum likelihood solution ψ_0^* under the postulated model is equal to the true value ψ^*.

E.2 Local asymptotic analysis for the additive exponential model

The limiting maximum likelihood solution is in this case more complicated in view of the nuisance parameters, but by the law of large numbers the average of the log-likelihood contributions converges to the average of the expected log-likelihood contributions, thus

$$\psi_0^* = \lim_{m\to\infty} \operatorname*{argmax}_{\psi\in\mathbb{R}^+} \left(\log\psi - \frac{2}{m}\sum_{j=1}^m \gamma_j(\gamma_j+\Delta^*) \int_0^\infty \frac{\log(1+\psi z)dz}{(\gamma_j + (\gamma_j+\Delta^*)z)^2} \right).$$

Differentiation shows that ψ_0^* solves

$$\frac{1}{\psi_0^*} = \lim_{m\to\infty} \frac{2}{m}\sum_{j=1}^m \gamma_j(\gamma_j+\Delta^*) \int_0^\infty \frac{zdz}{(1+\psi_0^* z)(\gamma_j + (\gamma_j+\Delta^*)z)^2}.$$

Analogously to Example 6.3 we can expand the integrand around $\Delta^* = 0$, which corresponds to the point at which the two models intersect, giving, to first order

$$\frac{1}{\psi_0^*} = \left(\frac{\psi_0^* - 1 - \log\psi_0^*}{(\psi_0^*-1)^2} + O(\Delta^*) \right) \lim_{m\to\infty} \frac{2}{m}\sum_{j=1}^m \frac{(\gamma_j+\Delta^*)}{\gamma_j}.$$

To the same order, this is of almost identical form to (33) with the term $1+\varepsilon$ replaced by the limiting average a of $(\gamma_j+\Delta^*)/\gamma_j$. The first-order solution under the notional asymptotic regime $\Delta^* \to 0$ is thus the logical one $\psi_0^* = 1$ for all a, $\psi^* = 1$ and $\Delta^* = 0$ corresponding to the null treatment effect and the point at which the two models intersect. Thus, intuitively, there is little sensitivity to departures from the postulated multiplicative model in the direction of the additive model at small Δ^*.

F Numerical checks

F.1 Matched pairs

Data were generated from a joint model for matched pairs $(Y_{j0}, Y_{j1})_{j=1}^m$ in which an additive treatment effect Δ^* operates on the rate scale for Weibull distributed outcomes with baseline rate parameters $(\gamma_j)_{j=1}^m$ and shape parameter ς, where $(\gamma_j)_{j=1}^m$ were generated from a standard uniform distribution. The assumed model is exponential with a multiplicative rate parameter as in Example 1.2. This notional parameter was estimated by maximum likelihood based on the density function of the ratios $Z_j = Y_{j1}/Y_{j0}$ which, from (22) is $\psi/(1+\psi z)^2$. Fisher's statistic $2\sum_{j=1}^m \log U_j(\hat{\psi})$ was then constructed based on (22) with

ψ^* replaced by $\hat{\psi}$ and z replaced by Z_j; Table 2 reports the results. At $(\Delta^*, \varsigma) = (0, 1)$, the true model intersects with the postulated model with null treatment effect, captured by value $\psi = 1$. Thus, the proportion of rejected tests of size α would be α by construction if ψ was perfectly estimated. Table 2 reports a beneficial under-rejection rate in that case. Elsewhere in the parameter space, power to detect departures from the model increases to 1 with increasing m except when $\varsigma = 1$, where the true Weibull distribution collapses to the exponential and therefore is closer to the point at which the true and postulated models intersect. Both the true distribution and those in the postulated model are in this example constructed from a parameter space whose dimension diverges at the same rate as the sample size m, so both models are semiparametric.

We also assessed the approach based on (4) without estimation of the notional parameter ψ, computing the proportion of confidence sets for ψ that were empty. If the parameter space of permissible values is constrained to some reasonable interval such as $[0, 4]$, then such an approach tends to be more powerful for the sample sizes reported, but a larger range, say $[0, 10]$ requires a much larger sample size in order for the confidence set to be empty with high probability; numerical results are not reported.

Direction of sensitivity	(Δ^*, ς)	Square root of number of pairs m					
		5	8	11	14	17	20
left	(0, 0.5)	0.657	0.977	1	1	1	1
right	(0, 0.5)	0.640	0.983	1	1	1	1
left	**(0, 1)**	**0.002**	**0.001**	**0**	**0**	**0**	**0**
right	**(0, 1)**	**0.001**	**0.001**	**0**	**0.001**	**0**	**0**
left	(0, 2)	0	0.221	0.994	1	1	1
right	(0, 2)	0	0.232	0.997	1	1	1
left	(1, 0.5)	0.717	0.982	1	1	1	1
right	(1, 0.5)	0.685	0.990	1	1	1	1
left	(1, 1)	0.008	0.015	0.020	0.042	0.075	0.134
right	(1, 1)	0.002	0.003	0	0.005	0.007	0.010
left	(1, 2)	0	0	0.105	0.420	0.774	0.951
right	(1, 2)	0	0.005	0.647	0.992	1	1
left	(2, 0.5)	0.731	0.983	1	1	1	1
right	(2, 0.5)	0.696	0.992	1	1	1	1
left	(2, 1)	0.010	0.025	0.030	0.073	0.134	0.236
right	(2, 1)	0.002	0.003	0.002	0.006	0.012	0.023
left	(2, 2)	0	0	0.041	0.215	0.555	0.825
right	(2, 2)	0	0.004	0.482	0.979	1	1

Table 2: Weibull generating distribution with additive treatment parameter Δ^* and shape ς. The assumed model is exponential with multiplicative treatment parameter ψ. Proportion of 1000 simulation runs in which the 5% test based on (10) rejects the postulated model. The point of model intersection $(\Delta^*, \varsigma) = (0, 1)$ is highlighted.

Table 3 reverses the roles of the two types of models, taking the baseline distribution of Y_{j0} to be the same as before but obtaining the distribution of Y_{j1} through multiplication of

Direction of sensitivity	(ψ^*, ς)	Square root of number of pairs m					
		5	8	11	14	17	20
left	(0.5, 0.5)	0.442	0.589	0.763	0.859	0.926	0.964
right	(0.5, 0.5)	0.846	0.995	1	1	1	1
left	(0.5, 1)	0.231	0.623	0.907	0.991	0.998	1
right	(0.5, 1)	0.239	0.669	0.942	0.998	1	1
left	(0.5, 2)	0.875	1	1	1	1	1
right	(0.5, 2)	0.002	0.035	0.182	0.536	0.817	0.939
left	(1, 0.5)	0.708	0.925	0.990	1	1	1
right	(1, 0.5)	0.565	0.881	0.984	1	1	1
left	**(1, 1)**	**0.079**	**0.068**	**0.082**	**0.059**	**0.070**	**0.058**
right	**(1, 1)**	**0.061**	**0.075**	**0.073**	**0.060**	**0.065**	**0.054**
left	(1, 2)	0.046	0.399	0.826	0.985	0.999	1
right	(1, 2)	0.073	0.450	0.894	0.991	1	1
left	(2, 0.5)	0.896	0.999	1	1	1	1
right	(2, 0.5)	0.329	0.466	0.606	0.740	0.873	0.936
left	(2, 1)	0.253	0.626	0.925	0.993	1	1
right	(2, 1)	0.252	0.620	0.922	0.991	1	1
left	(2, 2)	0.011	0.042	0.192	0.515	0.763	0.901
right	(2, 2)	0.889	1	1	1	1	1

Table 3: Weibull generating distribution with multiplicative treatment prameter ψ^* and shape ς. The assumed model is exponential with additive treatment parameter Δ. Proportion of 1000 simulation runs in which the 5% test based on (10) rejects the postulated model. The point of model intersection $(\psi^*, \xi) = (1, 1)$ is highlighted.

the rate parameter γ_j by a value ψ^*. The postulated model, by contrast, is the exponential additive rates model of Example 6.2. There is high power to detect the erroneous model, and at the point $(\psi^*, \varsigma) = (1, 1)$ at which the true and postulated models coincide, the nominal level is recovered asymptotically.

F.2 Time-dependent Poisson process

Simulation of an inhomogeneous Poisson process can be performed most simply using the argument of Cox and Lewis (1966, pp. 27–28) that on the transformed time scale $\tau_i(t) = \int_0^t \lambda_i(u)du$ the process is Poisson of unit constant rate. It follows that the sequence $T_{i1}, \ldots, T_{im_i}$ of ordered event times for individual i satisfy $\tau_i(T_{ik_i}) = \sum_{j=1}^{k_i} E_{ij}$, where E_{ij} are independent unit exponential random variables. For $\lambda_i(t) = e^{\gamma_i + \beta t}$, we have $\tau_i(t) = e^{\gamma_i}(e^{\beta t} - 1)/\beta$, and the event times are distributed as

$$T_{ik_i} \stackrel{d}{=} \tau_i^{-1}\Big(\sum_{j=1}^{k_i} E_{ij}\Big) = \log\Big(1 + (\beta/e^{\gamma_i})\sum_{j=1}^{k_i} E_{ij}\Big)/\beta, \tag{34}$$

the limit as $\beta \to 0$ being $T_{ik_i} = e^{-\gamma_i}\sum E_{ij}$ as expected. To check that the nominal error rates are recovered under the postulated model, we used this generating process with γ_i

drawn from a standard normal distribution, generating unit exponential random variables to use in (34) until their sum exceeded the endpoint of the transformed timescale $\tau_i(t_0)$ with $t_0 = 5$. Since we wish to assess cases where the model is misspecified, we also simulated using the power-law intensity function $\lambda_i(t) = e^{\gamma_i} t^\rho$ with $\rho \in \{-0.5, 0, 0.5, 1\}$. We generated

$$T_{ik_i} \stackrel{d}{=} \tau_i^{-1}\Big(\sum_{j=1}^{k_i} E_{ij}\Big) = \left(\frac{(\rho+1)\sum_{j=1}^{k_i} E_{ij}}{e^{\gamma_i}}\right)^{1/(\rho+1)} \tag{35}$$

until the corresponding sum of unit exponentials exceeded $\tau_i(t_0) = e^{\gamma_i} t_0^{\rho+1}/(\rho+1)$. At $\rho = \beta = 0$ the true and postulated models intersect, but since the likelihood function (14) is indeterminate there due to division by zero, we do not necessarily expect to recover the nominal level.

From n individuals we estimated β in the postulated model by maximising the conditional log-likelihood function

$$\ell(\beta) = \sum_{i=1}^{n} m_i \log\Big(\frac{\beta}{e^{\beta t_0} - 1}\Big) + \beta \sum_{i=1}^{n}\sum_{j=1}^{m_i} t_{ij} \tag{36}$$

based on (14). The model was subsequently assessed using the statistic $-2\sum_{i=1}^{n} \log U_i(\hat\beta)$ where $U_i(\hat\beta) = F_{S_i|M_i}(S_i \,|\, m_i; \hat\beta)$ and the conditional distribution was approximated either through Monte Carlo sampling using 1000 replicates under the postulated model if m_i was below 40, or by a normal approximation to the distribution of the sum under the postulated model if m_i exceeded 40. An analogous statistic was computed with $\hat\beta$ replaced by its true value. The results are reported in Table 4 for data generated under the postulated model and in Table 1 for data generated under the power-law model.

Table 4 shows that the procedure rarely rejects the true model when the conditional maximum likelihood estimate is used in the construction of each $U_i(\hat\beta)$. When the true value of β is used, it attains a rejection rate close to the nominal level, the discrepancy being primarily due to Monte Carlo sampling error in the approximation to $F_{S_i|M_i}(S_i \,|\, m_i; \hat\beta)$. Table 1 of the main paper reports the power to detect departures from the postulated model.

F.3 Confidence sets of regression models

This section presents numerical performance of the approach discussed in §5.3. The original procedure of Battey, Rasines & Tang (2025) was intended for the situation in which the full set of variables contemplated is inordinately large, necessitating some preliminary reduction by variable screening prior to assessment of low-dimensional subsets of variables. Since, in the present paper, we are illustrating a particular point, we simplify the problem by assuming that the number of starting variables is 15, so that no preliminary reduction is needed. If that were actually the case, there would be no advantage to using anything other than a likelihood-ratio test of every low-dimensional model against the comprehensive model, but it is reassuring to see coverage probabilities for the confidence sets based on $U_j = F_Z(Z_j)$ from (18). These are reported as a function of the number of synthetic replicates k and the sample size n. We also report the simulation-average size of the

Direction of sensitivity	β	estimated/true	Square root of number of individuals n					
			3	4	5	6	7	8
left	0	true	0.05	0.08	0.02	0.07	0.05	0.06
right	0	true	0.05	0.07	0.04	0.05	0.02	0.05
left	0	estimated	0.03	0.04	0	0.03	0.02	0.05
right	0	estimated	0.03	0.03	0.02	0.03	0.01	0.03
left	1	true	0.02	0.07	0.06	0.04	0.07	0.09
right	1	true	0.03	0.04	0.03	0.03	0.05	0.05
left	1	estimated	0.01	0.01	0	0	0	0
right	1	estimated	0	0	0.01	0	0	0
left	2	true	0.06	0.03	0.06	0.05	0.06	0.05
right	2	true	0.06	0.04	0.04	0.07	0.03	0.04
left	2	estimated	0	0	0	0	0	0
right	2	estimated	0.01	0	0	0.01	0	0

Table 4: The generating process is the inhomogeneous Poisson process with intensity function $\lambda_i(t) = e^{\gamma_i + \beta t}$ as postulated. Proportion of 200 simulation runs in which the 5% test based on $F_{S_i|M_i}$ and (10) rejects the true model.

resulting confidence set of models, i.e. the number of false models that are included in the confidence set on average over simulation runs.

The experiment was conducted as follows. In each of 500 simulations, the n rows of the $n \times d$ covariate matrix X were drawn from a mean-zero normal distribution with correlation 0.9 between $s + a$ of the d variables, and correlation zero elsewhere, where $d = 15$, $s = 5$ is the number of signal variables and $a = 3$ is the number of noise variables that are correlated with signal variables. Associated with X is a vector θ of regression coefficients with entries 1 in the positions corresponding to the signal variables, and zeros elsewhere. The outcome vector was constructed as $Y = X\theta + \varepsilon$, where the entries of ε were taken as standard normally distributed.

For every possible model of size $d_0 \leq 5$, the corresponding columns X_0 of X were extracted. For this postulated model we constructed a basis for the orthogonal projection onto the null space of X_0 by taking the $n - d_0$ eigenvectors of $I - X_0(X_0^{\mathrm{T}}X_0)^{-1}X_0$ corresponding to the unit eigenvalues. Let V_0 denote this $n \times (n - d_0)$ matrix of eigenvectors. From the vector of outcomes Y, $k \in \{4, 8, 12, 16\}$ synthetic replicates $\tilde{Y}_1, \ldots, \tilde{Y}_k$ were generated according to equation (4.3) of Battey, Rasines & Tang (2025) and the corresponding co-sufficient replicates were constructed as $\tilde{Q}_j = V_0^{\mathrm{T}}\tilde{Y}_j / \|V_0^{\mathrm{T}}\tilde{Y}_j\|_2$. The statistics $Z_1, \ldots, Z_m$ were then computed as the $m = k(k-1)/2$ inner products $\langle \tilde{Q}_j, \tilde{Q}_i \rangle$, $j \neq i$. When the postulated model is true, these follow the distribution (18) and $U_j = F_Z(Z_j)$ has a standard uniform distribution. The rejection regions for the model were thus defined in the usual way via (10). Table 5 reports the resulting coverage probabilities of the 0.95 nominal-coverage confidence sets and the average number of false models in the set from 500 simulations.

Direction of sensitivity	$(n,\ k)$	Coverage	$\hat{\mathbb{E}}\lvert\mathcal{M}\backslash\mathcal{S}\rvert$	$\frac{\hat{\mathbb{E}}(\#\text{ false models})}{\#\text{ models tested}}$
left	(100, 4)	0.944	613	0.124
right	(100, 4)	0.954	603	0.122
left	(100, 8)	0.954	508	0.103
right	(100, 8)	0.962	504	0.102
left	(100, 12)	0.964	477	0.097
right	(100, 12)	0.956	476	0.096
left	(100, 16)	0.956	462	0.093
right	(100, 16)	0.968	460	0.093
left	(200, 4)	0.950	270	0.055
right	(200, 4)	0.954	250	0.051
left	(200, 8)	0.956	213	0.043
right	(200, 8)	0.958	206	0.041
left	(200, 12)	0.958	197	0.040
right	(200, 12)	0.958	192	0.039
left	(200, 16)	0.956	190	0.038
right	(200, 16)	0.950	186	0.038

Table 5: Simulated coverage probability and expected size of the confidence sets of models from 500 simulations. These were constructed from (10) based on the distribution (18) of the cosine angles between projected synthetic replicates under each postulated sparse model.

References

Battey, H. S., Rasines, D. G. and Tang, Y. (2025). Post-reduction inference for confidence sets of models. *arXiv:2507.10373.*

Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events.* Springer, London.