Statistical Science
2019, Vol. 34, No. 4, 591–598
https://doi.org/10.1214/19-STS747
Main article: https://doi.org/10.1214/18-STS693, https://doi.org/10.1214/18-STS694

Comment: Models as (Deliberate) Approximations

David Whitney, Ali Shojaie and Marco Carone

1. OVERVIEW

We applaud Buja and coauthors for drawing further attention to the important problem of model misspecification in regression and to the study of its ramifications. In their interesting piece, they advocate for viewing model-based regression coefficients as nonparametric functionals of the data-generating mechanism. This viewpoint has the advantage of clarifying the definition of the estimand and formalizing how to perform model-robust inference based upon influence functions. In our note, we would like to continue the conversation along these lines. We wish to highlight additional considerations that arise in the context of model misspecification in a broader range of scenarios. Our main points are as follows:

(i) the model-robust interpretation of model-based estimands may not always be appealing, particularly when there is significant model misspecification or the sampling scheme includes some form of coarsening;
(ii) when the model fitting procedure involves data-adaptive estimation of nuisances, valid model-robust inference may be much more difficult to achieve;
(iii) these difficulties can be preempted by defining deliberate projection parameters and using suitable non or semiparametric techniques for inference.

2. MODEL-ROBUST INTERPRETATION

Framing regression coefficients as indices for the 'projection' of the true regression function onto the specified model is intuitively appealing. In our experience, most practitioners are aware that this is implicitly what they are doing when fitting regression models. However, it must be stressed that not all projections are useful projections. Below, we highlight that model-based regression coefficients may have a poor interpretation when (a) the model used is overly parsimonious, or (b) when the data are subject to some form of coarsening.

2.1 Targeted Versus Indiscriminate Parsimony

A primary reason for the popularity of regression models is their ability to summarize parsimoniously key relationships. However, parsimony can have several impacts on the interpretation of regression coefficients. For example, it can mask effect modification—this occurs if the portion of the model pertaining to the exposure of interest is parsimonious. This may be desirable if the goal is to succinctly summarize population-averaged relationships. This targeted form of parsimony is what renders regression models attractive. However, parsimony could also result in poor confounding control—this occurs when the portion of the model that involves potential confounders is too inflexible to allow sufficient deconfounding. This is an example of indiscriminate parsimony, which is both unnecessary—it can often be mitigated by the use of regression models with parsimonious exposure involvement but flexible confounding adjustment—and possibly harmful.

As an illustration, we expand upon a simple example stemming from the discussion of Section 10 in Part I. There, the authors note that when the underlying associations exhibit symmetry, there may be little to no linear trend. To be concrete, suppose that the data unit consists of the triple (W, X, Y), including a continuous outcome Y, exposure of interest X, and confounder W, generated from data-generating distribution P. Ordinary least-squares (OLS) regression may often be used in this context, with exposure and confounder both included as main terms, and reported upon with the appropriate caveat that the model coefficients represent

David Whitney is Teaching Fellow in Statistics, Department of Mathematics, Imperial College London, Huxley Building, South Kensington Campus, Imperial College, London SW7 2AZ, United Kingdom (e-mail: d.whitney@imperial.ac.uk). Ali Shojaie is Associate Professor, Department of Biostatistics, University of Washington, 1705 NE Pacific Street, Seattle, Washington 98195, USA (e-mail: ashojaie@uw.edu). Marco Carone is Assistant Professor and Norman Breslow Endowed Faculty Fellow, Department of Biostatistics, University of Washington, 1705 NE Pacific Street, Seattle, Washington 98195, USA (e-mail: mcarone@uw.edu).

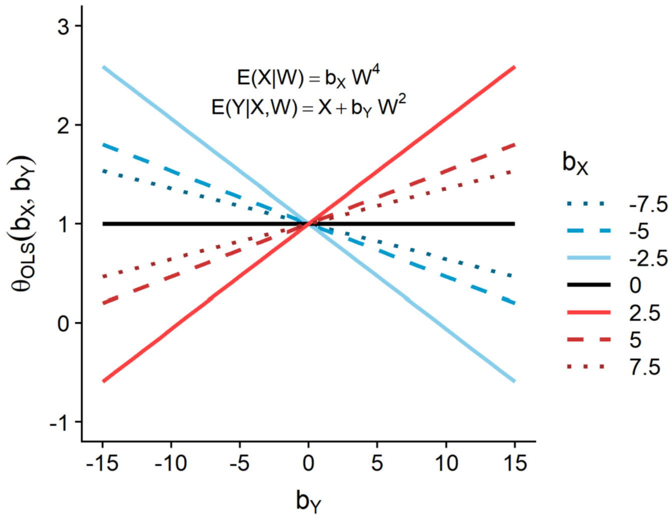

FIG. 1. *Exposure-specific OLS regression functional value when the true regression model follows a partially linear additive model. The line* $\theta_{\mathrm{OLS}} = 1$ *(black) denotes the exposure-specific regression coefficient for the true data-generating mechanism. The other lines represent the value of the OLS regression functional for different strengths of the exposure-confounder relationship.*

indices of the least-squares projection. We show with a numerical example that the resulting estimand may not be particularly useful. Specifically, we consider P to be specified by $W \sim U(-2, 2)$,

$$X \mid W \sim \mathcal{N}(b_X W^4, 1)$$

and

$$Y \mid (X, W) \sim \mathcal{N}(X + b_Y W^2, 1).$$

Coefficients b_X and b_Y control the strength of the exposure-confounder and (nonlinear) outcome-confounder relationships, respectively. In this example, the deconfounded linear relationship between Y and X is unambiguous: the regression slope equals one. However, the OLS estimand has explicit form

$$\theta_{\mathrm{OLS}}(b_X, b_Y) = 1 + \frac{1}{7}\left(\frac{7680 b_X b_Y}{225 + 4096 b_X^2}\right);$$

a range of numerical values are displayed in Figure 1 for various b_X and b_Y values. Depending on the strength of the underlying associations, the resulting estimand can be stronger or weaker, and of possibly the opposite sign as the true slope. This emphasizes that not all projections are useful—in fact, when the postulated regression model is strongly misspecified, there is a risk of inadequate deconfounding, and the regression functional may not be reflective of the underlying association of interest.

In the particular example considered, inclusion of polynomial confounder terms of sufficient degree in the linear model would have resolved the issue. However, this would likely not have been known a priori. The issue may have been discovered in a post-fit diagnostic analysis, but model revisions based on diagnostics are known to render calibrated inference difficult to perform. As an alternative, it would have been possible to consider a model with more flexible confounding control. In other words, a model without unnecessary (and possibly harmful) parsimony could have been used instead. For instance, as a flexible alternative to the linear model-based regression functionals, the partially linear additive model (PLAM) specifying that $E_P(Y \mid X = x, W = w) = \theta x + g(w)$ for some scalar θ and univariate real-valued function g could be considered. The estimand would then be the index θ_* of the least-squares projection of the true regression function onto the PLAM.

This simple example underscores that more flexible semiparametric models can lead to estimands with a more useful interpretation than provided by restrictive parametric models, without sacrificing parsimony relative to the association of interest. Nevertheless, this improved model-robust interpretation can come at a cost, as more involved procedures may be required to achieve valid inference. Indeed, performing calibrated model-robust inference requires additional considerations when data-adaptive techniques are used in the construction of the regression coefficient estimator. We discuss these challenges in Section 3.

2.2 The Impact of Coarsening in the Data Collection Mechanism

In many applications, the observed data consist of a coarsening of the full data, for instance, due to missingness or censoring. Regression models are typically imposed on the full data distribution, since it is a feature of this distribution that is generally of scientific interest. Although in this context estimands resulting from misspecified models can still be interpreted as projections, the latter generally involve the coarsening mechanism.

As an illustration, it is instructive to consider the use of maximum likelihood (ML) with coarsened data. When the full data are available, the ML approach is known to yield consistent estimators of the index of the model element closest to the true data-generating distribution in a Kullback–Leibler sense. When instead the data are subject to coarsening, we must distinguish between the space of distributions for the observed versus full data. The ML approach in this case will identify the member of the model for the observed data (as induced by the model for the full data and the coarsening mechanism) closest to the true distribution of the observed data (as induced by the true distribution for the full data and the coarsening mechanism). As such, in these settings, model-based estimators often correspond to regression functionals that depend not only on the full data distribution but also the coarsening mechanism.

This phenomenon generalizes the notion of mis/well-specification introduced by the authors to include the distribution of coarsening variables in addition to that of regressors. However, the coarsening mechanism is usually a study-specific nuisance rather than an inherent feature of the population of interest. As such, dependence of the regression functional on the coarsening mechanism is particularly troublesome. Indeed, two investigators studying the same population and fitting the same regression models may be estimating very different quantities simply because of differences in the coarsening affecting their study samples.

As a concrete illustration of this phenomenon, it is informative to consider the case of proportional hazards (PH) regression under right-censoring. In the simplest of scenarios, where the full data consist of observations on the time-to-event variable T and a single binary covariate X, the PH model stipulates that the conditional hazard function h_x of the distribution of T given $X = x$ satisfies

$$h_x(t) = h_0(t)\exp(\theta x) \quad \text{for all } t > 0,$$

where h_0 is an unspecified baseline hazard and θ is the scalar regression coefficient of interest. Instead of complete observations, it is common to observe possibly right-censored event times. When the censoring variable is conditionally independent of T given X, and the PH model indeed holds, the maximizer of the partial likelihood is known to be a consistent estimator of the true regression coefficient.

When instead the PH model is misspecified, one may hope that the resulting regression functional perhaps represents an average of the time-varying hazard ratio (on a logarithmic scale). It has been shown that this is indeed approximately true, though the limit in probability θ_* of the maximum partial likelihood estimator (MPLE) depends not only on the conditional time-to-event and marginal covariate distributions but also on the conditional censoring distribution (Struthers and Kalbfleisch, 1986) in a complicated manner. The fact that the censoring distribution defines the estimand is particularly alarming. In commenting on this finding, O'Quigley (2008) states that the partial likelihood-based regression functional is not itself particularly useful nor interpretable—we agree with this viewpoint.

To emphasize this point numerically, we may consider the hazard functions $h_1(t) := \alpha t^{\alpha-1}$ for arbitrary $\alpha \geq 1$ and $h_0(t) := 1$. In such case, the PH model holds if and only if $\alpha = 1$. The further α is from this value, the more time-varying the hazard ratio $h_1(t)/h_0(t)$ becomes, thereby increasingly violating the PH model assumption. In Figure 2, we display the value θ_* of the partial likelihood regression functional as a function of α for various censoring distributions. For simplicity, we have considered exponential distributions for the conditional censoring distribution, with exposure-specific rate parameters $\gamma_0, \gamma_1 \in \{0.2, 1.1, 2.0\}$. As is readily apparent, the dependence of θ_* on the censoring distribution increases with α.

Noting this dependence, under differing independence assumptions, Xu and O'Quigley (2000), Schemper, Wakounig and Heinze (2009) and Hattori and Henmi (2012) have studied weighted partial likelihood-based estimators whose corresponding estimands do not depend on the censoring distribution. Nevertheless, their estimands represent interpretable weighted averages of log-hazard ratios only in an *approximate* sense. In Section 4, we propose a novel estimator *exactly* targeting a weighted average log-hazard ratio.

3. VALID INFERENCE IN THE PRESENCE OF IRREGULAR NUISANCES

In Part II, the authors define the regression functional broadly as the solution of (a set of) population-level

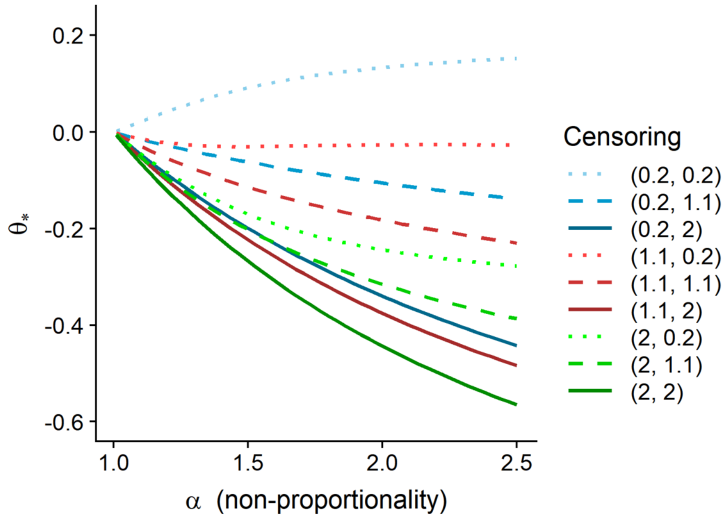

FIG. 2. *The model-robust interpretation of the partial likelihood regression functional depends on the censoring distribution. For a binary covariate, the population hazard ratio at time t is assumed to be $\alpha t^{\alpha-1}$, with larger α values corresponding to greater departures from the PH model. The exposure group-specific censoring distributions are exponential with rates γ_0 and γ_1.*

model-derived estimating equation(s), possibly arising from the minimization of a risk function. For the examples explicitly considered, the estimating function is entirely parametric, being indexed by a vector including the parameter of interest and possibly nuisance parameters. In such cases, under regularity conditions, the resulting estimator can be shown to be asymptotically linear using a standard Taylor expansion; thus, asymptotic normality at the parametric rate holds, even when the model is misspecified. In contrast, in the context of certain semi-parametric models, some of the indexing nuisances may be infinite-dimensional and irregular, in the sense that they are not estimable at the parametric rate without strong (e.g., parametric) assumptions. The asymptotic linearity of model-based estimators may in such cases rely on correct specification of the model. This happens because the nuisance estimator may not contribute in first order to the behavior of the regression functional estimator *when the model is correctly specified*, but indeed does so *when the model is misspecified*. In the latter case, the regression functional estimator may inherit the slow convergence rate of the nuisance estimator, and fail to be asymptotically normal at the parametric rate, thereby rendering inference difficult.

For concreteness, suppose that an estimating function ψ is available for the parametric index θ of a semiparametric regression model. Suppose further that this estimating function is indexed by an infinite-dimensional nuisance η. To simplify notation, we consider θ to be scalar. Suppose that η_N is a consistent estimator of the true nuisance value η_0, defined unambiguously when the semi-parametric model holds, and that η_N tends to some $\eta_*(P)$ in general, where P denotes the data-generating distribution. If P is in the model, then $\eta_*(P) = \eta_0$. In this case, the model-robust regression functional is the solution $\theta_*(P)$ of the population equation $E_P\{\psi(\theta, \eta_*(P); Z)\} = 0$. In practice, any solution θ_N of the empirical equation

$$\frac{1}{N}\sum_{i=1}^{N}\psi(\theta, \eta_N; Z_i) = 0$$

may be taken as estimator of $\theta_*(P)$. In what follows, we will simply write θ_* and η_*, dropping the explicit dependence on P for convenience. As before, under regularity conditions, a Taylor expansion results in the first-order approximation

$$\theta_N - \theta_* \approx -\left[\frac{\partial}{\partial\theta}E_P\{\psi(\theta, \eta_*; Z)\}\bigg|_{\theta=\theta_*}\right]^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N}\psi(\theta_*, \eta_*; Z_i) + \Phi(\eta_N)\right],$$

where Φ is the functional $\eta \mapsto \int \psi(\theta_*, \eta; z)\, dP(z)$. If $\Phi(\eta_N)$ is asymptotically negligible in the sense that $\Phi(\eta_N) = o_P(N^{-1/2})$, then θ_N is an asymptotically linear estimator with influence function proportional to the estimating function. Otherwise, $N^{1/2}(\theta_N - \theta_*)$ may fail to converge in law to a nondegenerate limit.

If Φ is sufficiently smooth, we may use the first-order approximation

$$\Phi(\eta_N) = \Phi(\eta_N) - \Phi(\eta_*) \approx \dot{\Phi}(\eta_*; \eta_N - \eta_*),$$

where the expression on the right-hand side denotes the Gâteaux derivative of Φ at η_* in the direction of $\eta_N - \eta_*$. If $h \mapsto \dot{\Phi}(\eta_*; h)$ is identically zero for h ranging in a set in which $\eta_N - \eta_*$ concentrates, then $\Phi(\eta_N)$ can be expected to be $o_P(N^{-1/2})$ provided $\eta_N - \eta_*$ vanishes quickly enough, as may be needed to guarantee the asymptotic linearity of θ_N. In such case, the estimating function is said to be orthogonalized with respect to the nuisance η. In contrast, if the estimating function is not orthogonalized, then $\dot{\Phi}(\eta_*; \eta_N - \eta_*)$ will generally contribute in first-order to the behavior of $\theta_N - \theta_*$. Since estimators of irregular parameters have a rate of convergence that is slower than the parametric rate, convergence of $N^{1/2}(\theta_N - \theta_*)$ to a nondegenerate limit distribution cannot be expected, at least with standard tuning of the involved data-adaptive nuisance estimators.

The above discussion relates to key ideas in efficiency theory. In that literature, influence functions often serve as estimating functions, in part because they are typically pre-orthogonalized relative to the nuisances involved; see, for example, Lemma 1.3 of van der Laan and Robins (2003) for results in the finite-dimensional case. However, this orthogonalization generally only holds in the model under which the influence function is derived. As such, it may well be that $\Phi(\eta_N)$ is a higher-order term when the semiparametric regression model holds but is a first-order term otherwise.

We highlight the above phenomenon in the context of an example. Suppose that we wish to evaluate the association between outcome Y and binary exposure X adjusting for W, and we denote the data unit by $Z := (W, X, Y)$. We focus on the coefficient θ_0 in the PLAM $E_P(Y \mid X = x, W = w) = \theta_0 x + g_0(w)$ mentioned in Section 2.1. Defining $\pi_0(w) := E_P(X \mid W = w)$, we consider model-based estimation approaches built upon two candidate estimating functions

$$\psi_0(\theta, \pi; z) := \{x - \pi(w)\}(y - \theta x)$$

and

$$\psi(\theta, g, \pi; z) := \{x - \pi(w)\}\{y - \theta x - g(w)\}.$$

It can be shown that ψ is proportional to the efficient influence function for θ_0 in the PLAM under homoscedasticity, but that ψ_0 is not even an influence function (Yu and van der Laan, 2003). These estimating functions require estimation of π_0 and g_0. The nuisance π_0 is defined irrespective of whether the PLAM holds and can be estimated using a non-parametric estimator π_N. This allows us to define the model-based estimator $\theta_{0,N}$ based on ψ_0 and π_N, namely

$$\theta_{0,N} := \frac{\sum_{i=1}^N Y_i\{X_i - \pi_N(W_i)\}}{\sum_{i=1}^N X_i\{X_i - \pi_N(W_i)\}},$$

with corresponding model-robust estimand $\theta_* := E_P[Y\{X - \pi_0(W)\}]/E_P[X\{X - \pi_0(W)\}]$. The nuisance g_0 is only well-defined under the PLAM. Two different estimators of g_0 consistent under the PLAM, say $g_{1,N}$ and $g_{2,N}$, may each converge to distinct limits $g_{1,*}$ and $g_{2,*}$ outside the PLAM. Noting that $g_{1,*}(w) := E_P(Y \mid X = 0, W = w) = g_0(w)$ under the PLAM, any nonparametric estimator $g_{1,N}(w)$ of $g_{1,*}(w)$ could be used as estimator of $g_0(w)$. Alternatively, we may consider the more elaborate backfitting approach of Buja, Hastie and Tibshirani (1989), setting $g_{2,N}(w)$ to be a nonparametric estimator of the regression of $Y - \theta_{0,N}X$ onto W evaluated at w. We note that $g_{2,N}(w)$ is then a consistent estimator of $g_{2,*}(w) := E_P(Y \mid W = w) - \theta_*\pi_0(w)$, which also coincides with $g_0(w)$ under the PLAM. The resulting model-based estimators of θ_0 based on ψ and $(\pi_N, g_{j,N})$ are then

$$\theta_{j,N} := \frac{\sum_{i=1}^N \{Y_i - g_{j,N}(W_i)\}\{X_i - \pi_N(W_i)\}}{\sum_{i=1}^N X_i\{X_i - \pi_N(W_i)\}}$$

for $j = 1, 2$. Both $\theta_{1,N}$ and $\theta_{2,N}$ also tend to θ_* outside the PLAM.

The respective (first-order) nuisance contributions of $(\pi_N, g_{j,N})$ when using ψ_0 and ψ are $\Phi_0(\pi_N) \approx -\int \{\pi_N(w) - \pi_0(w)\}E_P(Y - \theta_* X \mid W = w)\, dP(z)$ and

$$\Phi(\pi_N, g_{j,N}) \approx \int \{\pi_N(w) - \pi_0(w)\} \times \{g_{j,*}(w) - E_P(Y - \theta_* X \mid W = w)\}\, dP(z).$$

It is clear that, whether the PLAM holds or not, $\Phi_0(\pi_N)$ makes a first-order contribution to the behavior of $\theta_{0,N} - \theta_*$. This fact is not surprising since ψ_0 is not orthogonalized relative to π. If the PLAM holds, then $E_P(Y - \theta_* X \mid W = w) = E_P(Y - \theta_0 X \mid W = w) = g_0(w)$ and the first-order approximation of $\Phi(\pi_N, g_N)$ is zero with g_N taken to be either $g_{1,N}$ or $g_{2,N}$. If instead the PLAM does not hold, the situation is more complex. In general, $g_{1,*}(w) - E_P(Y - \theta_* X \mid W = w) \neq 0$, whereas $g_{2,*}(w) - E_P(Y - \theta_* X \mid W = w) = 0$ for each w. Thus, when the PLAM does not hold, the nuisance estimator will make a first-order

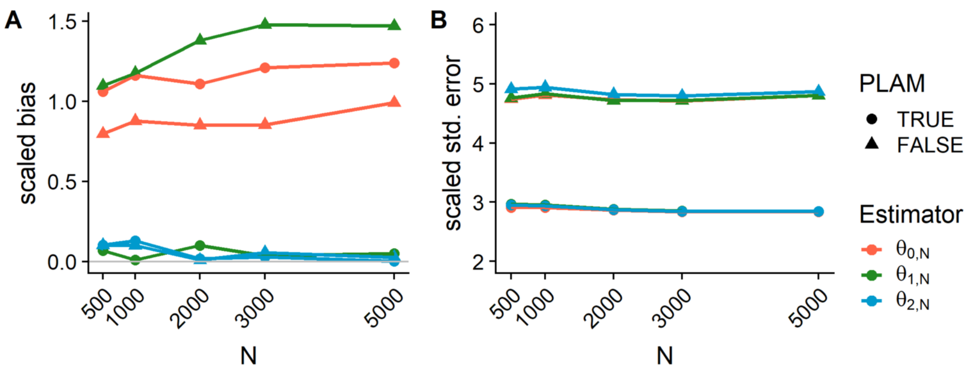

FIG. 3. *Empirical bias and standard error of model-based estimators* $\theta_{0,N}$, $\theta_{1,N}$ *and* $\theta_{2,N}$ *scaled by* $N^{1/2}$ *for sample sizes* $N \in \{500, 1000, 2000, 3000, 5000\}$ *computed using* 5000 *simulated datasets for each sample size under correct and incorrect model specifications*.

contribution to the behavior of $\theta_{1,N} - \theta_*$ but not to that of $\theta_{2,N} - \theta_*$. In other words, valid model-robust inference can be easily carried out using $\theta_{2,N}$ but not $\theta_{1,N}$.

We illustrate this phenomenon in a simulation study. We set $W \sim U(-2, 2)$, $X \mid W \sim \text{Bernoulli}(\pi_0(W))$ with $\pi_0(w) = \text{expit}(0.5 + 2w - w^2)$, and $Y|(W, X) \sim \mathcal{N}(\mu(X, W), 1)$, where $\mu(x, w) = 0.2x + w^2$ or $\mu(x, w) = (2x - 1)w$ to simulate a valid versus misspecified PLAM, respectively. We generated 5000 datasets for each sample size $N \in \{500, 1000, 2000, 3000, 5000\}$ and computed estimators $\theta_{0,N}$, $\theta_{1,N}$ and $\theta_{2,N}$ with Nadaraya–Watson kernel estimator (with cross-validated bandwidth selection) used for nuisance estimation whenever nonparametric regression was required. Results are depicted in Figure 3.

This simulation study confirms the expected behavior of the estimators considered. The bias of $\theta_{0,N}$ does not tend to zero sufficiently fast to allow convergence of $N^{1/2}(\theta_{0,N} - \theta_*)$ to a nondegenerate distribution regardless of whether the PLAM holds—this occurs because ψ_0 is not orthogonalized, and so, the behavior of the kernel regression estimator dominates. When the PLAM holds, the bias of both $\theta_{1,N}$ and $\theta_{2,N}$ tends to zero faster than $N^{-1/2}$. However, only the bias of $\theta_{2,N}$ remains small when the PLAM is misspecified. These results demonstrate the importance of estimating function orthogonality when evaluating the model-agnostic sampling behavior of regression model-based estimators. Additionally, they highlight that if a model-based procedure is not suitably orthogonalized, it does not suffice to devise an improved variance estimator—bias that does not vanish quickly enough results in the lack of a nondegenerate distribution at the parametric rate.

4. DELIBERATE MODEL-AGNOSTIC PARAMETER EXTENSIONS

In their article, the authors focus on regression functionals that arise as the limit of model-based procedures, and propose strategies for model-robust inference. As an alternative, it may be fruitful to first define a model-agnostic parameter extension (or projection) based on the considered regression model, and then develop robust inferential procedures for this estimand. By deliberately defining the projection of interest rather than letting it be dictated by some model-based estimator, issues pertaining to parameter interpretation, as highlighted in Section 2, can largely be circumvented. Furthermore, by using non or semiparametric tools, valid inference can be performed for this projection parameter while avoiding the potentially poor behavior of model-based procedures in the presence of irregular nuisances and model misspecification, as discussed in Section 3.

To illustrate what we mean by model-agnostic parameter extension, we return to the proportional hazards model—we refer interested readers to Chambaz, Neuvial and van der Laan (2012), Graham and Pinto (2018) and references therein for a treatment of projections onto the PLAM. Under proportional hazards, the regression coefficient θ_0 is the (constant) hazard ratio value. A natural summary of a time-varying hazard ratio is

$$\theta_{**} := \int \log\left\{\frac{h_1(t)}{h_0(t)}\right\}\nu(dt),$$

where h_x is the true hazard function corresponding to $X = x$, and ν represents a weight function, possibly dependent on components of the data-generating distribution. If the PH model holds, θ_{**} coincides with the usual PH regression coefficient θ_0. If the PH model does not hold, θ_{**} remains a transparent and interpretable estimand. While the usual Cox estimand is often claimed to represent a quantity such as θ_{**} when the PH model fails to hold, we see from Figure 2 that the weight function depends to a large extent on the interplay between the censoring distribution and degree of model misspecification.

Consider N independent triples $Z_i := (Y_i, \Delta_i, X_i)$, where Y_i is the follow-up time, Δ_i the event indicator, and X_i the exposure group indicator for study participant i. Suppose that the right-censoring mechanism is uninformative within exposure groups. In that case, under identification conditions, θ_{**} represents a pathwise differentiable parameter of the data-generating distribution. A regular and asymptotically linear estimator of θ_{**} can thus be constructed. For example, for the important case in which ν is the marginal time-to-event distribution, we may consider the one-step bias-corrected estimator

$$\theta_{\mathrm{OS},N} := \int \theta_N(t) F_N(dt) + \frac{1}{N}\sum_{i=1}^{N} \phi_N(Z_i),$$

where $\theta_N(t) := \log h_{1,N}(t) - \log h_{0,N}(t)$ with $h_{x,N}$ a nonparametric estimator of h_x, $F_N := (1-\pi_N)F_{0,N} + \pi_N F_{1,N}$ is a nonparametric estimator of the marginal time-to-event distribution function, $F_{x,N}$ is the Kaplan–Meier estimator of the distribution function corresponding to $X = x$, and π_N is the proportion of study participants with $X = 1$. Here, ϕ_N is a plug-in estimator of the (nonparametric) efficient influence function of θ_{**} when the weight function is considered fixed:

$$\phi_N(z) := \left(\frac{x}{\pi_N}\right)\gamma_{1,N}(y,\delta) - \left(\frac{1-x}{1-\pi_N}\right)\gamma_{0,N}(y,\delta),$$

$$\gamma_{x,N}(y,\delta) := \frac{\delta \exp\{-x\theta_N(y)\} Q_N(y)}{R_{x,N}(y)} - \int_{u\le y} \frac{1}{R_{x,N}(u)} F_N(du)$$

with

$$Q_N(y) := (1-\pi_N)\{1-F_{0,N}(y)\} + \pi_N\{1-F_{1,N}(y)\}\exp\{\theta_N(y)\}$$

and

$$R_{x,N}(y) := N^{-1}\sum_{i=1}^{N} I(Y_i \ge y).$$

It can be shown that $N^{1/2}(\theta_{\mathrm{OS},N} - \theta_{**})$ tends to a mean-zero Gaussian variable under regularity conditions and rate conditions on $h_{x,N}$.

A simulation study was conducted to evaluate the finite-sample behavior of this estimator. We generated exposure $X \sim \mathrm{Bernoulli}(2/3)$ and time-to-event $T \mid X \sim \mathrm{Weibull}(1, 1+\alpha X/2)$, with $\alpha \in \{0, 1\}$ yielding correct and incorrect PH specifications. Censoring time $C \sim \mathrm{exponential}(0.2)$ was generated independently of (T, X). For each scenario, we generated 5000 datasets of size $N \in \{500, 1000, 2000, 3000, 5000\}$, and evaluated the empirical bias (relative to the projection estimand) and standard error of $\theta_{\mathrm{OS},N}$ and of the maximum partial likelihood estimator $\theta_{\mathrm{MPLE},N}$. Hazard functions were estimated with kernel regression. Results are displayed in Figure 4. Under correct PH specification,

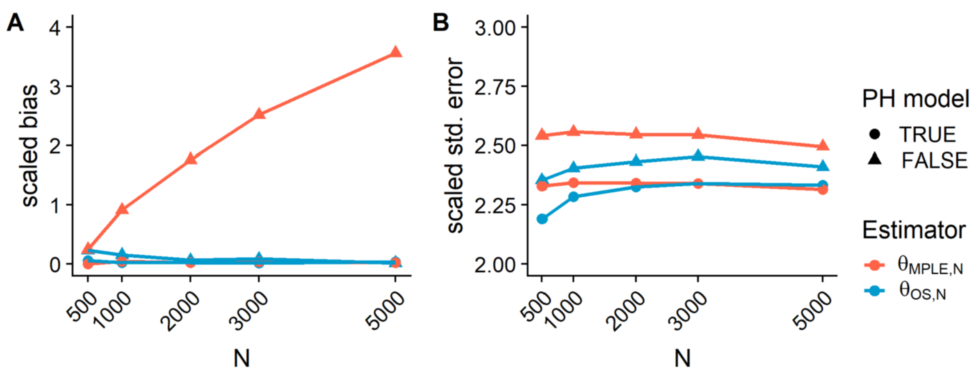

FIG. 4. *Empirical bias and standard error of estimators $\theta_{\mathrm{MPLE},N}$ and $\theta_{\mathrm{OS},N}$ scaled by $N^{1/2}$ for sample sizes $N \in \{500, 1000, 2000, 3000, 5000\}$ computed using 5000 simulated datasets for each sample size under correct and incorrect model specifications.*

both estimators have negligible bias; interestingly, they also have similar standard errors. When the PH assumption fails, only $\theta_{\mathrm{OS},N}$ tends to the projection parameter. It also has bias tending to zero faster than $N^{-1/2}$ and variance stabilizing at rate N^{-1}. In contrast to the Cox estimand, the projection parameter is an interpretable summary of the hazard ratio invariant to the censoring distribution.

This simple example highlights that it is possible to define deliberate model-agnostic extensions of regression coefficients, and to construct (nonparametric efficient) estimators that minimize the need for unrealistic assumptions about the data-generating mechanism. We emphasize that if knowledge on the data-generating distribution is available (e.g., known moment conditions or conditional independences), it should be incorporated into the inferential process. In such case, the efficient influence function used to construct the estimator of the regression functional should be relative to this greater state of knowledge. While we used the simple one-step construction in our illustration, more recent strategies with possibly improved performance also exist; see, for example, van der Laan and Rose (2011). These strategies naturally allow the use of flexible, data-adaptive tools (e.g., machine learning) for nuisance estimation yet allow valid inference for the deliberate target of scientific interest. A potential challenge is the reliance of these strategies on analytic objects whose derivation requires specialized skills, though there have been recent efforts to overcome this difficulty using computational tools (Carone, Luedtke and van der Laan, 2019). In addition to better understanding the ramifications of regression model misspecification, and devising model-robust inferential procedures for available estimators, as Buja and coauthors have done, we hope to see further efforts to develop and vet estimators of natural model-agnostic parameter extensions based upon common regression models.

ACKNOWLEDGMENTS

The authors thank the Editor for the opportunity to contribute to this discussion. This work was partly funded by NIH Grants R01HL137808 and R01GM114029, and by NSF Grant DMS-1915855.

REFERENCES

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249

CARONE, M., LUEDTKE, A. R. and VAN DER LAAN, M. J. (2019). Toward computerized efficient estimation in infinite-dimensional models. *J. Amer. Statist. Assoc.* **114** 1174–1190. MR4011771

CHAMBAZ, A., NEUVIAL, P. and VAN DER LAAN, M. J. (2012). Estimation of a non-parametric variable importance measure of a continuous exposure. *Electron. J. Stat.* **6** 1059–1099. MR2988439

GRAHAM, B. S. and PINTO, C. C. D. X. (2018). Semiparametrically efficient estimation of the average linear regression function. Technical report, National Bureau of Economic Research.

HATTORI, S. and HENMI, M. (2012). Estimation of treatment effects based on possibly misspecified Cox regression. *Lifetime Data Anal.* **18** 408–433. MR2973773

O'QUIGLEY, J. (2008). *Proportional Hazards Regression. Statistics for Biology and Health*. Springer, New York. MR2400249

SCHEMPER, M., WAKOUNIG, S. and HEINZE, G. (2009). The estimation of average hazard ratios by weighted Cox regression. *Stat. Med.* **28** 2473–2489. MR2751535

STRUTHERS, C. A. and KALBFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73** 363–369. MR0855896

VAN DER LAAN, M. J. and ROBINS, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York.

VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data. Springer Series in Statistics*. Springer, New York. MR2867111

XU, R. and O'QUIGLEY, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics* **1** 423–439.

YU, Z. and VAN DER LAAN, M. J. (2003). Measuring treatment effects using semiparametric models. Technical report, Univ. California, Berkeley, Dept. Biostatistics Working Paper Series.