Constantine E. Frangakis*

The method of envelopes to concisely calculate semiparametric efficient scores under parametric restrictions

<https://doi.org/10.1515/ijb-2019-0043>

Received April 17, 2019; accepted December 1, 2020; published online December 24, 2020

Abstract: When addressing semiparametric problems with parametric restrictions (assumptions on the distribution), the efficient score (ES) of a parameter is often important for generating useful estimates. However, usual derivation of ES, although conceptually simple, is often lengthy and with many steps that do not help in understanding why its final form arises. This drawback often casts onto semiparametric estimation a mantle that can turn away otherwise able doctoral students or researchers. Here we show that many ESs can be obtained as a one-step derivation after we characterize those features (envelopes) of the unrestricted problem that are constrained in the restricted problem. We demonstrate our arguments in three problems with known ES but whose usual derivations are lengthy. We show that the envelope-based derivation is dramatically explanatory and compact, needing essentially two lines where the standard approach needs 10 or more pages. This suggests that the envelope method can add useful intuition and exegesis to both teaching and research of semiparametric estimation.

Keywords: efficiency; envelope; score; semiparametric.

1 The problem

Semiparametric problems are useful when their “restrictions” (i.e., assumptions about the distribution) reflect existing knowledge or a simplified expression of a complex reality. When addressing semiparametric problems, the efficient score (ES) of a parameter is often important for generating useful estimators (e.g., [1–4]).

One important example is the “restricted conditional moment” problem, where we measure a random sample of predictors and a response, (X_i, Y_i) , $i = 1, \dots, n$. In such a problem, we model the regression $E(Y_i | X_i) = g(X_i, \beta_0)$ and we wish to estimate β_0 ; and the rest of the distribution is unspecified and indexed by λ_0 (e.g., [3]). Another important example is the “treatment effect regression” problem, where a random sample of patients with covariates X_i is assigned one of two treatments, $T_i = 0$ or 1, based on X_i , and outcome Y_i is measured. In such a problem, we often model the difference $E(Y_i | T_i = 1, X_i) - E(Y_i | T_i = 0, X_i) = g(X_i, \beta_0)$ and wish to estimate β_0 ; and the rest of the distribution is, again, unspecified and indexed by λ_0 (e.g., [5, 6]).

The above examples are cases of a more general problem in which we measure a random sample of data D_i , $i = 1, \dots, n$ from an unknown true distribution $\text{pr}_0(D_i)$; we assume $\text{pr}_0(D_i)$ is characterized by a finite dimensional vector β_0 that we wish to estimate (say $\beta_0 \in \text{Set}_\beta$); and leave the rest of the distribution unspecified and indexed by λ_0 (say $\lambda_0 \in \text{Set}_\lambda$). Such a semiparametric model is summarized (e.g., [1]) as:

$$(\beta_0, \lambda_0) \in M^{\text{sem}} = \{(\beta, \lambda) \in \text{Set}_\beta \times \text{Set}_\lambda\}. \quad (1)$$

Derivation of the ES, say $S_\beta^{\text{eff}}(D_i, \beta, \lambda)$, is useful because it can lead to efficient estimation of β . Specifically, under some regularity conditions, the estimator for β_0 that zeros out the empirical sum $\sum_i S_\beta^{\text{eff}}(D_i, \beta, \lambda)$ over

*Corresponding author: Constantine E. Frangakis, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA, E-mail: cfrangak@jhsph.edu

individuals is consistent and efficient if λ , when needed, is replaced by a consistent estimator with convergence rates larger than $n^{1/4}$ [4].

In what follows, by notation such as $S_{\beta,i}^{\text{eff}}$ we mean the random variable $S_{\beta}^{\text{eff}}(D_i, \beta, \lambda)$, with the understanding that it arises from a deterministic function taking a data value and parameters β, λ as arguments. The usual derivation of ES is summarized in the following steps (e.g., [2]):

- (i) consider a parametric submodel, say, M_s^{sem} indexed by β and a finite dimensional parameter λ_s ;
- (ii) find the scores S_{λ_s} ;
- (iii) find, across all submodels s , the set of all linear combinations of scores S_{λ_s} , and the closure of this set, known as the nuisance tangent space and labeled as Λ ;
- (iv) find the efficient score S_{β}^{eff} for β as the residual from the score S_{β} after projecting it on the space Λ .

$$S_{\beta}^{\text{eff}} := S_{\beta} - \text{proj}(S_{\beta} | \Lambda). \quad (2)$$

2 The method of envelopes

The key idea here is that, instead of constructing the score for β with the “bottom up” argument (using smaller parametric submodels M_s^{sem}), it can be easier to construct it with a “top down” argument, using a model *larger* than M^{sem} . To proceed, we assume the support of data D_i is discrete, even though the dimension can be large. This is often reasonable on its own, e.g., for arguments that take sample size to a limit but keep measurement instruments with fixed, finite number of values. Here, discreteness is used mostly as a device to uncover the form of the efficient score when this is understood to exist.

Consider the larger model, say M^{env} , that is as M^{sem} except that it poses no restrictions using β . Consider then *all* those features of the distribution $\text{pr}_0(D_i)$ in M^{env} that *do* get constrained by β in the semiparametric model M^{sem} . Those features, which we denote by “env”, can be seen as envelopes, in the larger model M^{env} , of the consequences of placing the restrictions β in the semiparametric model M^{sem} . Denote the larger model as:

$$(\text{env}_0, \lambda_0^{\text{env}}) \in M^{\text{env}} := \{(\text{env}, \lambda^{\text{env}}) \in \text{Set}_{\text{env}} \times \text{Set}_{\lambda^{\text{env}}}\}. \quad (3)$$

We then have the following useful observations:

- (i) the meaning of the parameters λ and the set Set_{λ} are the same in model M^{sem} as the parameters λ^{env} and the set $\text{Set}_{\lambda^{\text{env}}}$ in model M^{env} ; therefore, the efficient score for the envelope in M^{env} is

$$S_{\text{env}}^{\text{eff}} = S_{\text{env}} - \text{proj}(S_{\text{env}} | \Lambda), \quad \text{for the same } \Lambda \text{ as in (2);} \quad (4)$$

- (ii) from the semiparametric model we obtain that $S_{\beta} = \frac{\partial \text{env}}{\partial \beta} \times S_{\text{env}}$; here, $\frac{\partial \text{env}}{\partial \beta}$ is a deterministic matrix of partial derivatives of the $\dim(\text{env})$ envelopes with respect to the $\dim(\beta)$ parameters, evaluated at the true values.
- (iii) substituting, using the last expression, $\frac{\partial \text{env}}{\partial \beta} \times S_{\text{env}}$ for S_{β} in (2), we obtain the following connection between the two efficient scores:

Result. The efficient score S_{β}^{eff} in the smaller model can be obtained from the efficient score $S_{\text{env}}^{\text{eff}}$ in the larger model as

$$\underbrace{S_{\beta}^{\text{eff}}}_{\dim(\beta) \times 1} = \underbrace{\frac{\partial \text{env}}{\partial \beta}}_{\dim(\beta) \times \dim(\text{env})} \times \underbrace{S_{\text{env}}^{\text{eff}}}_{\dim(\text{env}) \times 1}$$

$$\text{Proof: } S_{\beta}^{\text{eff}} = \frac{\partial \text{env}}{\partial \beta} \cdot S_{\text{env}} - \text{proj}\left(\frac{\partial \text{env}}{\partial \beta} \cdot S_{\text{env}} | \Lambda\right) \left\{ \begin{array}{l} \text{here } \Lambda \text{ acts as the nuis. tang.} \\ \text{space for the smaller model} \end{array} \right\}$$

$$\begin{aligned}
&= \frac{\partial \text{env}}{\partial \beta} \{S_{\text{env}} - \text{proj}(S_{\text{env}}|\Lambda)\} \left\{ \begin{array}{l} \text{now } \Lambda \text{ can act as the nuis.} \\ \text{tang. space for the larger model} \end{array} \right\} \\
&\stackrel{\text{from (4)}}{=} \frac{\partial \text{env}}{\partial \beta} \times S_{\text{env}}^{\text{eff}}.
\end{aligned} \tag{5}$$

This last expression is important because it gives the efficient score S_{β}^{eff} as a simple function of the score for the envelopes $S_{\text{env}}^{\text{eff}}$ in the larger model, and bypasses long calculations within model M^{sem} . The scores for $S_{\text{env}}^{\text{eff}}$ are often known either directly or through their relation to the efficient influence functions (EIF) $\phi_{\text{env}}^{\text{eff}}$ of the envelopes:

$$S_{\text{env}}^{\text{eff}} = E\left(\phi_{\text{env}}^{\text{eff}} \phi_{\text{env}}^{\text{eff}'}\right)^{-1} \cdot \phi_{\text{env}}^{\text{eff}} \tag{6}$$

where the latter are easily derived as Gateaux derivatives [7]. Once the efficient scores are derived, the efficient influence function for β , say $\phi_{\beta}^{\text{eff}}$ can be derived in the usual way (e.g., [3]; p. 66) as

$$\phi_{\beta}^{\text{eff}} = E\left(S_{\beta}^{\text{eff}} S_{\beta}^{\text{eff}'}\right)^{-1} \cdot S_{\beta}^{\text{eff}}.$$

3 Examples

3.1 Restricted conditional moment problem

The first problem of Section 1 is the efficient score for β modeling $E(Y_i | X_i) = \mu(X_i, \beta_0)$, with no other assumptions. The standard derivation of this efficient score uses the steps (i)–(iv) of Section 1 (e.g., [3]). These, typically, take long arguments of 10 or more pages of text, but also involve conjectures (and then verifications), which leave the question of how one first arrives to such conjectures more generally.

To see the practicality of the ideas in Section 2, first observe that, since β only restricts the conditional means $E(Y_i | X_i = x) = \mu(x)$, the vector of all these means $\text{env} := [\mu(x) : \text{all } x]$ is an envelope parameter. Therefore, in the right hand side of (6), the EIF $\phi_{\text{env}, i}^{\text{eff}}$ is the vector of the nonparametric EIFs for each conditional mean, $\phi_{\mu(x), i}^{\text{eff}}$, which we know is the residual $\frac{\{Y_i - \mu(x)\}1(X_i=x)}{p(x)}$, where $p(x)$ is the mass function at x . Also, the variance matrix in (6) is diagonal with variances $\text{var}(Y_i | X_i = x)/p(x)$, for each x . Substituting these in (6) and then in (5), and replacing $\frac{\partial \text{env}(\beta)}{\partial \beta}$ by $\left[\frac{\partial \mu(x, \beta)}{\partial \beta}, \text{ all } x \right]$, we get

$$S_{\beta, i}^{\text{eff}} = \frac{\partial \mu(X_i, \beta)}{\partial \beta} \frac{\{Y_i - \mu(X_i, \beta)\}}{\text{var}(Y_i | X_i)}$$

which is the well known efficient score, derived in essentially one argument. This short derivation is paralleled by an analogously short intuition: the EIF elements $[\phi_{\mu(x), i}^{\text{eff}} : \text{all } x]$ of the envelope in the larger model contain the sufficient information also for the smaller model's parameter β ; then these elements are combined through the derivative of the envelope and inversely proportional to their variance.

3.2 Treatment effect regression problem

The second problem of Section 1 is the efficient score for β modeling $E(Y_i | T_i = 1, X_i) - E(Y_i | T_i = 0, X_i) = \delta(X_i, \beta_0)$, with no other assumptions. The standard derivation of the efficient score, like in problem 1, also uses the steps (i)–(iv) of Section 1, needs conjectures and is also analogously long.

The use of envelopes is again straightforward. Since β only restricts $\delta(X_i) = E(Y_i | T_i = 1, X_i) - E(Y_i | T_i = 0, X_i)$, the set of differences $\text{env} = [\delta(x) : \text{all } x]$ forms an envelope for β . We know that for each conditional mean $\mu(t, x) = E(Y_i | T_i = t, X_i = x)$, the nonparametric (in M^{env}) EIF $\phi_{\mu(t, x), i}^{\text{eff}}$ is the residual $\frac{Y_i - \mu(t, x)1(T_i = t, X_i = x)}{p(t, x)}$ where $p(t, x)$ is the mass function at t, x . Therefore, in the right hand side of (6), the EIF $\phi_{\text{env}, i}^{\text{eff}}$ is the vector of the nonparametric EIFs for each difference,

$$[\phi_{\delta(x), i}^{\text{eff}} : \text{all } x], \quad \text{where } \phi_{\delta(x), i}^{\text{eff}} = \left\{ \frac{\{Y_i - \mu(1, x)\}1(T_i = 1)}{e(x)} - \frac{\{Y_i - \mu(0, x)\}1(T_i = 0)}{1 - e(x)} \right\} \frac{1(X_i = x)}{p(x)},$$

and $e(x)$ is the propensity score $pr(T_i = 1 | X_i = x)$. Also, the variance matrix in (6) is again diagonal with variances $v(x)/p(x)$, where

$$v(x) = \frac{\text{var}(Y_i | T_i = 1, X_i = x)}{e(x)} + \frac{\text{var}(Y_i | T_i = 0, X_i = x)}{1 - e(x)}.$$

Substituting these in (6) and then in (5), and replacing $\frac{\partial \text{env}(\beta)}{\partial \beta}$ by $\left[\frac{\partial \delta(x, \beta)}{\partial \beta}, \text{all } x \right]$, we get

$$S_{\beta, i}^{\text{eff}} = \frac{\partial \delta(X_i, \beta)}{\partial \beta} \frac{1}{v(x)} \left\{ \frac{\{Y_i - \mu(0, X_i) - \delta(X_i, \beta)\}1(T_i = 1)}{e(X_i)} - \frac{\{Y_i - \mu(0, X_i)\}1(T_i = 0)}{1 - e(X_i)} \right\}.$$

As with the first problem, the intuition that parallels the derivation is that the EIF elements of the envelope in the larger model contain the sufficient information also for the smaller model's parameter β .

3.3 Restricted conditional median problem

To show the ease of applying this idea also to other functionals, consider a variation of problem 1 where now we are interested in β that models the conditional medians of Y_i given $X_x = x$ as, say, $m(X_i, \beta_0)$, with no other assumptions.

Since β only restricts the conditional medians, $m(x)$, of Y_i given $X_i = x$, the vector of medians $\text{env} = [m(x) : \text{all } x]$ is an envelope parameter. Therefore, in the right hand side of (6), the EIF $\phi_{\text{env}, i}^{\text{eff}}$ is the vector of the nonparametric EIFs for each conditional median, $\phi_{m(x), i}^{\text{eff}}$, which we know is $[1\{Y_i > m(x)\} - 0.5] \cdot 1(X_i = x) / \{f(0; x)p(x)\}$, where $f(0; x)$ is the conditional mass function of $Y_i - m(x)$ at 0 given $X_i = x$, and $p(x)$ is the mass function at x . The variance matrix in (6) is diagonal with variances $1/\{4f^2(0; x)p(x)\}$, for each x . Substituting these in (6) and then in (5), and replacing $\frac{\partial \text{env}(\beta)}{\partial \beta}$ by $\left[\frac{\partial m(x, \beta)}{\partial \beta}, \text{all } x \right]$, we get

$$S_{\beta, i}^{\text{eff}} = \frac{\partial m(X_i, \beta)}{\partial \beta} 4f(0; X_i) [1\{Y_i > m(X_i)\} - 0.5]$$

which is the known efficient score ([8]; p. 106).

4 Comments

A conceptual comparison of the usual method to the envelope method to derive a semiparametric efficient score is the following. The usual method relates the semiparametric restricted model to smaller, submodels with more restrictions. The envelope method relates the restricted model to a larger model that relaxes those restrictions. Here, we showed how this latter relation can provide a much shorter derivation and understanding of the efficient scores for teaching and research.

Restrictions in semiparametric models can take forms other than through a finite dimensional parameter, for example through conditional independence or symmetry. If the functional of interest in the smaller model is only a function of easily estimable features of the larger, non-restricted model, then the envelope method can still be useful.

An alternative approach to calculating efficient scores is through Target Maximum Likelihood Estimation (TMLE, e.g., [9]). Ways to use TMLE while avoiding knowledge of the efficient influence curve have been explored in some cases, although their generality is still under investigation.

Acknowledgments: We thank the editor, and Dan Scharfstein, Betsy Ogburn, Mark van der Laan, and Stijn Vansteelandt for useful comments.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This article was funded by National Institutes of Health.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Newey WK. Semiparametric efficiency bounds. *J Appl Econom* 1990;5:99–135.
2. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc* 1994;89:846–66.
3. Tsiatis AA. Semiparametric inference and missing data. New York: Springer; 2007.
4. van der Vaart AW. Asymptotic statistics. New York: Cambridge University Press; 2000.
5. Yu Z, van der Laan MJ. Measuring treatment effects using semiparametric models. U.C. Berkeley Division of Biostatistics Working Paper Series, paper 136; 2003.
6. Vansteelandt S, Joffe M. Structural nested models and g-estimation: the partially realized promise. *Stat Sci* 2014;20:707–31.
7. Hampel FR. The influence curve and its role in robust estimation. *J Am Stat Assoc* 1974;69:383–93.
8. van der Laan M, Robins J. Unified methods for censored longitudinal data and causality. New York: Springer; 2003.
9. van der Laan M, Rose S. Targeted learning: causal inference for observational and experimental data. New York: Springer; 2011.