Conditional Influence Functions

Whitney K. Newey

HKU Econometric Society Winter School in DSE 2025
December 10-11

with Victor Chernozhukov and Vasilis Syrgkanis;

https://arxiv.org/pdf/2412.18080.

INTRODUCTION

There are many nonparametric, conditional objects of interest.

An example is the conditional average treatment effect, conditional on one or more key covariates.

An economic example is average welfare effects of increasing price of a good, conditional on income.

Objects that condition on a few variables can provide useful, low dimensional summaries of high dimensional unknown functions and have causal or structural interpretations.

We give influence functions for conditional objects of interest.

Similarly to the classic influence function of a functional of a cumulative distribution function (CDF), conditional influence functions can be used to:

–Construct Neyman orthogonal estimating equations for conditional, nonparametric objects, so that machine learning can be used to estimate functions on which the objects depend.

–Quantify local effects of misspecification on conditional objects.

–Define conditional, local policy effects.

Here we mostly focus on constructing Neyman orthogonal estimating equations but also touch on other uses for conditional influence functions.

Estimators of conditional objects of interest based on machine learners have previously been developed.

Convergence rates for conditional objects given in Foster and Syrgkanis (2019).

Lee (2019) gave estimator of the average potential outcome for continuous treatment.

Linear projection estimators in Semenova and Chernozhukov (2021).

Kernel regression estimators with automatic debiasing in Chernozhukov, Newey, and Singh (2022).

Estimators of the Conditional Average Treatment Effect in Kennedy et al. (2024).

Some efficiency calculations are given in Luedtke and Chung (2023).

Some estimators based on conditionally Neyman orthogonal estimating equations given in Yang, Kuchibhotla, and Tchetgen-Tchetgen (2024) and Dalal et al. (2024).

We give the conditional influence function and show how it can be used to construct Neyman orthogonal conditional moment functions for estimating conditional objects of interest.

This provides a general way of constructing Neyman orthogonal moment functions rather than having to guess their form in particular cases.

We give conditional influence functions.

Show how they can be used to construct Neyman orthogonal estimating equations.

Briefly discuss other possible uses.

Describe estimators based on Neyman orthogonality.

CONDITIONAL INFLUENCE FUNCTIONS

Recall the classical influence function.

$$\begin{array}{rcl} W & : & \text{One data observation in i.i.d. sequence.} \\ F & : & \text{CDF of one observation having true value } F_0, \\ \theta(F) & : & \text{parameter of interest having true value } \theta_0 = \theta(F_0), \\ F_\varepsilon & = & (1-\varepsilon)F_0 + \varepsilon H \text{ for CDF } H \text{ represents local variation.} \end{array}$$

The influence function (when it exists) is $\psi_0(W)$ satisfying.

$$\frac{\partial \theta(F_\varepsilon)}{\partial \varepsilon} = \int \psi_0(W) dH, \;\; E[\psi_0(W)] = 0.$$

$\psi_0(W)$ is Gateaux derivative of $\theta(F)$ at F_0.

To describe a conditional influence function, let $V = T(W)$ be some random variable, and define

$$\begin{aligned} F^v &: \text{ CDF of one observation conditional on } V = v, \\ &\quad\ \text{having true value } F_0^v; \\ \theta(F^v) &: \text{ parameter of interest, having true value } \theta_0(v) = \theta(F_0^v); \\ F_\varepsilon^v &= (1-\varepsilon)F_0^v + \varepsilon H^v \text{ for alternate conditional CDF } H^v. \end{aligned}$$

The conditional influence function (when it exists) is $\psi_0^v(W)$ satisfying

$$\frac{\partial \theta(F_\varepsilon^v)}{\partial \varepsilon} = \int \psi_0^v(W) dH^v, \ E[\psi_0^v(W)|V] = 0.$$

Running Example: Conditional average potential outcome.

There is an outcome variable $Y = DY(1) + (1 - D)Y(0)$ where $D \in \{0, 1\}$ is treatment, $Y(1)$, $Y(0)$ are potential outcomes with and without treatment, are mean independent of treatment conditional on covariates Z, and $\pi_0(Z) = \Pr(D = 1|Z) > 0$ with probability one.

Let $V = t(Z)$ some function of the covariates (e.g. a single covariate), $X = (D, Z)$, $W = (Y, X)$, and $\gamma_0(X) = E[Y|X]$.

An object of interest is

$$\theta_0(v) = E[Y(1)|V = v] = E[\gamma_0(1, Z)|V = v].$$

The conditional influence function of $\theta_0(V)$ is shown here to be

$$\gamma_0(1, Z) - \theta_0(V) + \alpha_0(X)\{Y - \gamma_0(X)\}, \ \alpha_0(X) = \frac{D}{\pi_0(Z)}.$$

Conditional influence functions can be used to:

–Construct Neyman orthogonal estimating equations for a conditional, nonparametric object of interest, where unknown functions have zero first-order effect on conditional objects of interest.

–Quantify local effects of misspecification on conditional object.

–Define conditional, local policy effects.

It seems they cannot be used to characterize asymptotic variances or make efficiency comparisons for conditional parameters, when these are nonparametric, meaning conditional on continuous variables.

For example series and kernel estimators of conditional means have different asymptotic variances because they employ different localizations.

Neyman Orthogonal Estimating Equations

These are conditional estimating equations for objects of interest where the first-order effect of nuisance functions on the conditional estimating equation is zero.

We construct these by starting with an identifying conditional moment equation.

Let γ denote a first step function that helps identify $\theta_0(V)$, e.g. $\gamma(X)$ is a possible $E[Y|X]$ as in running example.

Let $g(W, \gamma, \theta)$ be a function of data observation W, γ, and θ, a possible $\theta_0(V)$ function.

Assume that $\theta_0(V)$ solves the condtional moment restriction

$$0 = E[g(W, \gamma_0, \theta)|V].$$

In the running example

$$g(W, \gamma, \theta) = \gamma(1, Z) - \theta(V),$$

so $\theta_0(V) = E[\gamma_0(1, Z)|V]$.

We construct Neyman orthogonal moment functions by adding to $g(W,\gamma,\theta)$ the conditional influence function $\phi(W,\gamma,\alpha,\theta)$ of $E[g(W,\gamma(F^v),\theta)|V]$, where $\gamma(F^v)$ is the plim of some first step, nonparametric object that helps identify $\theta_0(v)$, F^v is unrestricted, and α is a first step function additional to γ.

We specify that $\phi(W,\gamma,\alpha,\theta)$ satisfies, for $F^v_\varepsilon = (1-\varepsilon)F^v_0 + \varepsilon H^v$,

$$\begin{aligned}\frac{\partial}{\partial\varepsilon}E[g(W,\gamma(F^v_\varepsilon),\theta)|V &= v] = \int \phi(w,\gamma_0,\alpha_0,\theta)dH^v, \\ E[\phi(W,\gamma_0,\alpha_0,\theta)|V] &= 0.\end{aligned}$$

A conditionally Neyman orthogonal moment function is

$$\psi(W,\gamma,\alpha,\theta) = g(W,\gamma,\theta) + \phi(W,\gamma,\alpha,\theta).$$

In the running example we have $\phi(W,\gamma,\alpha,\theta) = \alpha(X)[Y-\gamma(X)]$ so that

$$\psi(W,\gamma,\alpha,\theta) = \gamma(1,Z) - \theta(V) + \alpha(X)[Y-\gamma(X)],$$

where $\alpha_0(X) = D/\pi_0(Z)$.

The function $\psi(W, \gamma, \alpha, \theta)$ has two orthogonality properties:

$$(1)\ \frac{\partial}{\partial \varepsilon} E[\psi(W, \gamma(F_\varepsilon^V), \alpha_0, \theta)|V] = 0,\ (2)\ E[\phi(W, \gamma_0, \alpha, \theta)|V] = 0,$$

for any α satisfying $\alpha = \alpha(F^v)$ for some F^v and $\gamma_0 = \gamma((1-\varepsilon)F^v + \varepsilon F_0^v)$ for ε small enough.

Property 1 follows from differentiating with respect to ε the identity

$$E_{F_\varepsilon^v}[\phi(W, \gamma(F_\varepsilon^v), \alpha_0, \theta)|V = v] \equiv 0,$$

for any F_ε^v with $\alpha_0 = \alpha(F_\varepsilon^v)$.

Property 2 follows from

$$\begin{aligned} E[\phi(W, \gamma_0, \alpha, \theta)|V] &= \int \phi(W, \gamma_0, \alpha(F^v), \theta) dF_0^v \\ &= \frac{\partial}{\partial \varepsilon} E[g(W, \gamma(\bar{F}_\varepsilon^v), \alpha_0(F^v), \theta)|V] \\ &= \frac{\partial}{\partial \varepsilon} E[g(W, \gamma_0, \alpha_0(F^v), \theta)|V] = 0, \end{aligned}$$

where $\bar{F}_\varepsilon^v = (1-\varepsilon)F^v + \varepsilon F_0^v$.

The two orthogonality properties are:

$$(1)\ \frac{\partial}{\partial \varepsilon} E[\psi(W, \gamma(F_\varepsilon^V), \alpha_0, \theta)|V] = 0,\ \ (2)\ E[\phi(W, \gamma_0, \alpha, \theta)|V] = 0,$$

for any α.

Under regularity conditions (1) implies

$$\frac{\partial}{\partial \delta} E[\psi(W, \gamma_0 + \delta(\gamma - \gamma_0), \alpha_0, \theta)|V] = 0,$$

i.e. orthogonality (zero Gateaux derivative) with respect to γ.

In this sense γ has zero first order effect on $E[\psi(W, \gamma, \alpha_0, \theta)|V]$.

Also, (2) implies that α has no effect on $E[\psi(W, \gamma_0, \alpha, \theta)|V]$, so that $\psi(W, \gamma_0, \alpha, \theta)$ is globally conditionally robust with respect to α.

Global conditional robustness with respect to γ and α is conditional double robustness.

$\psi(W, \gamma, \alpha, \theta)$ is conditionally doubly robust if and only if $E[\psi(W, \gamma, \alpha_0, \theta)|V]$ is linear affine in γ.

In the running example the conditionally Neyman orthogonal moment function is

$$\psi(W, \gamma, \alpha, \theta) = \gamma(1, Z) - \theta(V) + \alpha(X)[Y - \gamma(X)].$$

By iterated expectations for $\alpha_0(X) = D/\pi_0(Z)$,

$$\begin{aligned}
E[\psi(W, \gamma, \alpha, \theta_0)|V] &= E[\gamma(1, Z)|V] - E[\gamma_0(1, Z)|V] \\
&\quad + E[\alpha(X)\{Y - \gamma(X)\}|V] \\
&= E[E[\frac{D}{\pi_0(Z)}\{\gamma(1, Z) - \gamma_0(1, Z)\}|Z]|V] \\
&\quad + E[\alpha(X)\{\gamma_0(X) - \gamma(X)\}|V] \\
&= -E[\{\alpha(X) - \alpha_0(X)\}\{\gamma(X) - \gamma_0(X)\}|V].
\end{aligned}$$

Other Uses of Conditional Influence Functions

Can be used to quantify the effect of misspecification on conditional objects of interest.

Gateaux derivative

$$\frac{\partial\theta(F_\varepsilon^v)}{\partial\varepsilon} = \int \psi_0^v(W)dH^v$$

gives the local effect of changing the conditional distribution of a data observations away from F_0^v in the direction $(1-\varepsilon)F_0^v + \varepsilon H^v$.

Could be used to evaluate conditional versions of local misspecification effects like those in Andrews, Gentzkow, and Shapiro (2017).

Alternatively if changes in F^v represent policy shifts then $\partial\theta(F_\varepsilon^v)/\partial\varepsilon$ can be interpreted as a local policy effect.

Conditional Influence Function of Regression Functionals

Generalize running example to an object of interest

$$\theta_0(V) = E[m(W,\gamma_0)|V],\ \gamma_0(X) = E[Y|X],\ V = T(X).$$

In running example, $m(W,\gamma) = \gamma(1,Z)$, for $X = (D,Z)$.

The key condition needed for conditional influence function to exist is there is $\alpha_0(X)$ such that

$E[m(W,\gamma)|V] = E[\alpha_0(X)\gamma(X)|V]$ for all $\gamma(X)$ with $E[\gamma(X)^2|V] < \infty$ w.p.1.

This is a conditional Riesz representation.

In the running example, where $m(W,\gamma) = \gamma(1,Z)$, we have $\alpha_0(X) = D/\pi_0(Z)$.

Another example is

$$\theta_0(V) = E[\frac{\partial}{\partial d}\gamma_0(D,Z)|V],\ V = T(Z).$$

Here, if $V = T(Z)$, for $m(W,\gamma) = \partial\gamma(D,Z)/\partial d$, iterated expectations and integration by parts conditional on Z give

$$\begin{aligned} E[m(W,\gamma)|V] &= E[E[\partial\gamma(D,Z)/\partial d|Z]|V] = E[\alpha_0(D,Z)\gamma(D,Z)|V], \\ \alpha_0(D,Z) &= -\partial \ln f_0(D|Z)/\partial d, \end{aligned}$$

where $f_0(D|Z)$ is conditional pdf of D conditional on Z.

Thus, the key condition is satisfied for the average derivative conditional on a function V of the covariates Z.

Key condition requires that V cannot be a function of variables in X that $m(W,\gamma)$ changes or operates on.

The conditional influence function of a conditionally mean-square continuous functional $\theta_0(V) = E[m(W, \gamma_0)|V]$ of $\gamma_0(X) = E[Y|X]$ is

$$m(W, \gamma_0) - \theta_0(V) + \alpha_0(X)[Y - \gamma_0(X)].$$

There is also a conditional influence functions for objects that are conditional expectations of a linear functional of

$$\gamma_0 = \arg\min_{\gamma(X)} E[Q(W, \gamma)|V],$$

such as $\gamma_0(X)$ a conditional quantile.

There is also a conditional influence function for average potential outcome for continuous treatment, as considered for example by Lee (2019).

ESTIMATION FOR CONDITIONAL MEAN γ

To estimate $\theta_0(V)$ from observations $W_1, ..., W_n$ we can used the debiased pseudo outcomes

$$\hat{S}_i = m(W_i, \hat{\gamma}) + \hat{\alpha}(X_i)[Y_i - \hat{\gamma}(X_i)].$$

We can run a nonparametric regression of $\hat{S}_i$ on V_i at v to estimate $\theta_0(v) = E[m(W, \gamma_0)|V]$.

For theoretical reasons use cross fitting where partition observation indices $i = 1, ..., n$ into distinct sets $I_\ell, (\ell = 1, ..., L)$, and let $\hat{\gamma}_\ell$ and $\hat{\alpha}_\ell$ be computed from observations not in I_ℓ.

Let $\hat{S}_{i\ell} = m(W_i, \hat{\gamma}_\ell) + \hat{\alpha}_\ell(X_i)[Y_i - \hat{\gamma}_\ell(X_i)]$, $i \in I_\ell$.

We run a nonparametric regression of $\hat{S}_{i\ell}$ on V_i at v to estimate $\theta_0(v)$.

For expositional example consider a kernel regression estimator.

Let $K(u)$ be a pdf, h a bandwidth, and r the dimension of v.

A cross-fit kernel regression estimator is

$$\hat{\theta}(v) = \frac{\sum_{\ell=1}^{L} \sum_{i \in I_\ell} K((v - V_i)/h^r) \hat{S}_{i\ell}}{\sum_i K((v - V_i)/h^r)}.$$

In practice a locally linear regression estimator is preferable for all the usual reasons locally linear are preferred to kernel regression.

Theory for locally linear regression given in paper.

Could also use random forests.

Straightforward to give conditions for this estimator to have same limiting distribution as estimator where $\hat{S}_{i\ell}$ is replaced by $S_i = m(W_i, \gamma_0) + \alpha_0(X_i)[Y_i - \gamma_0(X_i)]$.

This is possible because of Neyman orthogonality of the conditional influence function.

For bounded kernel (including locally linear regression) a key condition is that, for $\|a\| = \{E[a(W)^2]\}^{1/2}$.

$\|\hat{\gamma} - \gamma_0\| = o_p(h^{r/2})$, $\|\hat{\alpha} - \alpha_0\| = o_p(h^{r/2})$, $\sqrt{n}\,\|\hat{\gamma} - \gamma_0\|\,\|\hat{\alpha} - \alpha_0\| = o_p(h^{r/2})$;see paper for details; the conditions are stronger the larger is the dimension r of V, requiring faster mean square convergence rates for $\hat{\gamma}$ and $\hat{\alpha}$ when r is larger.

If strengthen L_2 rates for $\hat{\gamma}$ and $\hat{\alpha}$ to L_4 rates then can weaken h conditions; see Foster and Syrgkanis (2019) and Zhia (2025).

For an asymptotic approximation based on treating $\hat{\alpha}$ and $\hat{\gamma}$ as known to result in a good distribution approximation the dimension of X_i should exceed the dimension of V_i; the larger the dimension difference the better this approximation should be.

Estimation is based on

$$\hat{S}_{i\ell} = m(W_i, \hat{\gamma}_\ell) + \hat{\alpha}_\ell(X_i)[Y_i - \hat{\gamma}_\ell(X_i)].$$

Get $\hat{\gamma}_\ell(X_i)$ from nonparametric regression of Y_i on X_i.

The needed $\hat{\alpha}_\ell(X_i)$ can be obtained as minimum of Riesz regression loss

$$\hat{\alpha}_\ell = \arg\min_{\alpha \in \mathcal{A}} \sum_{i \notin I_\ell} [-2m(W_i, \alpha) + \alpha(X_i)^2],$$

for a set $\mathcal{A}$ of approximating functions, with penalty possibly added.

Can minimize over neural net and other specifications for α and generalize from functionals of conditional means to other conditional location functions, like quantiles; see Chernozhukov, Newey, Quintas-Martinez, and Syrgkanis (2021).

These are automatic methods for constructing $\hat{\alpha}$, requiring only $m(W, \alpha)$ and not using a specific formula for α_0.

SUMMARY

We have given the conditional influence function.

Showed how to construct conditionally Neyman orthogonal moment functions.

Touched on other uses of conditional influence functions.

Described debiased machine learning of conditional nonparametric parameters.

Given some rate conditions.