A NEYMAN-ORTHOGONALIZATION APPROACH TO THE INCIDENTAL-PARAMETER PROBLEM

Stéphane Bonhomme Koen Jochmans Martin Weidner

University of Chicago

Toulouse School of Economics

University of Oxford

June 12, 2024

Introduction

Inference in the presence of nuisance parameters is complicated.

Long history of 'modifying' estimating equations to make them less sensitive to estimation noise in nuisance parameters.

A key concept is orthogonality—due to work of Neyman (1959, 1979)—which has found renewed applicability in recent work on high-dimensional inference.

When nuisance parameters are very poorly estimated, (first-order) Neyman orthogonality is insufficient.

We consider higher-order generalizations of Neyman orthogonality that yield increased robustness.

Approach is (conditional) likelihood based and general.

Useful in settings with many fixed effects, such as panel data or network data.

Outline

1. Motivation
2. Panel data example
3. Constructing Neyman-orthogonal estimating equations
4. Examples
5. Application on team production

Motivation

Consider a likelihood-based setting where we have i.i.d. data $z_1, \ldots, z_n$ from density

$$f(z_i; \theta_0, \eta_0).$$

Here, η_0 is the nuisance parameter.

Interested in inference on parameter μ_0 defined through

$$\mathbb{E}(u(z_i; \theta_0, \eta_0; \mu_0)) = 0.$$

We take everything to be scalar valued for simplicity of presentation.

For now we simply set $\mu_0 = \theta_0$, in which case a natural choice for u is

$$u(z_i; \theta, \eta) := u(z_i; \theta, \eta; \theta) = \frac{d \log f(z_i; \theta, \eta)}{d\theta},$$

the score.

Let $\hat{\eta}$ be an auxiliary estimators of η_0.

Consistent but potentially converging very slowly.

The plug-in estimator $\hat{\theta}$ solves

$$\frac{1}{n}\sum_{i=1}^{n} u(z_i; \theta, \hat{\eta}) = 0$$

for θ.

Then

$$\left(-\mathbb{E}\left(\frac{du(z; \theta_0, \eta_0)}{d\theta}\right) + o_P(1)\right)(\hat{\theta} - \theta_0) = \frac{1}{n}\sum_{i=1}^{n} u(z_i; \theta_0, \hat{\eta})$$

and properties of $\hat{\theta} - \theta_0$ are dictated by behavior of sample average on the right.

An expansion and the role of sample splitting

Under regularity conditions the difference

$$\frac{1}{n}\sum_{i=1}^{n} u(z_i;\theta_0,\hat{\eta}) - \frac{1}{n}\sum_{i=1}^{n} u(z_i;\theta_0,\eta_0)$$

has the expansion

$$\sum_{p=1}^{q}\left\{\frac{1}{p!}\mathbb{E}\left(\frac{d^p u(z_i;\theta_0,\eta_0)}{d\eta^p}\right)(\hat{\eta}-\eta_0)^p\right\} + O_P(|\hat{\eta}-\eta_0|^{q+1}),$$

up to the term

$$\sum_{p=1}^{q}\left\{\frac{1}{p!}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{d^p u(z_i;\theta_0,\eta_0)}{d\eta^p} - \mathbb{E}\left(\frac{d^p u(z_i;\theta_0,\eta_0)}{d\eta^p}\right)\right)(\hat{\eta}-\eta_0)^p\right\}.$$

This final term can generally be ensured to be $o_P(n^{-1/2})$ by using sample splitting and cross fitting.

The role of orthogonalization

If we want $\sqrt{n}(\hat{\theta} - \theta_0) = O_P(1)$ then, in general,

$$\hat{\eta} - \eta_0 = O_P(n^{-1/2})$$

is required.

This puts strong requirements on the auxiliary estimator.

With first-order (Neyman) orthogonality,

$$\mathbb{E}\left(\frac{du(z;\theta_0,\eta_0)}{d\eta}\right) = 0,$$

and $\hat{\eta} - \eta_0 = o_P(n^{-1/4})$ suffices to remove the first-order effect of estimation noise in the nuisance parameter.

For score for θ, first-order orthogonality is simply information orthogonality:

$$\mathbb{E}\left(\frac{du(z_i;\theta_0,\eta_0)}{d\eta}\right) = \mathbb{E}\left(\frac{d^2 \log f(z_i;\theta_0,\eta_0)}{d\theta\, d\eta}\right) = 0.$$

We say that a function is (Neyman) orthogonal to order q when its first q derivatives with respect to the nuisance parameter have zero mean at true values:

$$\mathbb{E}\left(\frac{d^p u(z_i;\theta_0,\eta_0)}{d\eta^p}\right)=0 \qquad \text{for all } 1\leq p\leq q.$$

In this case, we need only that

$$\hat{\eta}-\eta_0=o_P(n^{-1/2(q+1)})$$

for estimation noise in $\hat{\eta}$ to not affect the limit distribution of $\hat{\theta}$

Constructing functions that are first-order orthogonal is well understood (even outside the likelihood setting).

Look for modifications to u that deliver u_q^* which are orthogonal to order q.

The Neyman-Scott example

We have

$$z_{it} \sim \mathbf{N}(\eta_{i0}, \theta_0).$$

The score contribution of stratum i is

$$u(z_i; \theta, \eta) = -\frac{1}{2\theta}\left(T - \frac{\sum_{t=1}^{T}(z_{it} - \eta_i)^2}{\theta}\right),$$

Note that

$$\mathbb{E}_{\theta,\eta_i}\left(\frac{d\, u(z_i; \theta, \eta)}{d\eta}\right) = -\mathbb{E}_{\theta,\eta_i}\left(\frac{\sum_{t=1}^{T}(z_{it} - \eta_i)}{\theta^2}\right) = 0.$$

The score in this problem already is orthogonal to order 1.

First-order orthogonality is insufficient to deal with incidental-parameter bias.

By an expansion $u(z_i;\theta_0,\eta_i) - u(z_i;\theta_0,\eta_{i0})$ is equal to

$$
\begin{aligned}
& \mathbb{E}\left(\frac{d^2 u(z_i;\theta_0,\eta_{i0})}{d\eta_i^2}\right)\frac{(\eta_i-\eta_{i0})^2}{2} + \frac{du(z_i;\theta_0,\eta_{i0})}{d\eta_i}(\eta_i-\eta_{i0}) \\
= & \frac{T}{\theta_0^2}\frac{(\eta_i-\eta_{i0})^2}{2} - \frac{\sum_{t=1}^T (z_{it}-\eta_{i0})}{\theta_0^2}(\eta_i-\eta_{i0}).
\end{aligned}
$$

Both terms have expectations that are $O(1)$, in general, so both need to be handled to reduce bias.

If we evaluate this in maximum-likelihood estimator $\hat{\eta}_i = \bar{z}_i \sim \mathbf{N}(\eta_{i0}, {}^{\theta_0}/_T)$ and take expectations we get

$$
\mathbb{E}(u(z_i;\theta_0,\hat{\eta}_i)) = \frac{T}{\theta_0^2}\frac{\operatorname{var}(\bar{z}_i)}{2} - \sum_{t=1}^T \frac{\operatorname{cov}(z_{it},\bar{z}_i)}{\theta_0^2} = \frac{1}{2\theta_0} - \frac{1}{\theta_0} = -\frac{1}{2\theta_0} = O(1).
$$

The terms represent estimation noise in $\hat{\eta}_i$ and dependence between $\hat{\eta}_i$ and z_{it}, respectively.

Comment: Bias correction in panel data

In a general panel data problem $n = N \times T$ and $\dim \eta \propto N$, and we can at best construct

$$\hat{\eta}_i - \eta_{i0} = O_P(T^{-1/2}).$$

For $T^{-1/2} = o(n^{-1/4})$ we need that

$$N = o(T).$$

This is the same rate requirement to ensure asymptotic unbiasedness of the (uncorrected) maximum-likelihood estimator.

With orthogonality to order q (combined with sample splitting in the time series dimension) the bias is reduced from $O(T^{-1})$ to $O(T^{-q})$ and we require only that

$$N = o(T^{2q-1})$$

for a correctly-centered limit distribution.

This connects to the literature on (higher-order) bias correction.

Bhattacharyya basis

Collect data in $z = (z_1, \ldots, z_n)$ and write $\ell(z; \theta_0, \eta_0)$ for the likelihood.

For any integer o let

$$v_o(z; \theta, \eta) = \frac{1}{\ell(z; \theta, \eta)} \frac{d^o \ell(z; \theta, \eta)}{d\eta^o}.$$

For example,

$$\begin{aligned} v_1(z; \theta, \eta) &= \frac{d \log \ell(z; \theta, \eta)}{d\eta}, \\ v_2(z; \theta, \eta) &= \frac{d^2 \log \ell(z; \theta, \eta)}{d\eta^2} + \left(\frac{d \log \ell(z; \theta, \eta)}{d\eta} \right)^2. \end{aligned}$$

Note that $\mathbb{E}_{\theta,\eta}(v_o(z; \theta, \eta)) = 0$ for any o.

Orthogonality to order q

Collect the leading o functions $v_1, v_2, \ldots, v_o$ in the vector function $w_o(z; \theta, \eta)$.

We look for coefficient vector c such that

$$u_q^*(z; \theta, \eta) = u(z; \theta, \eta) - \boldsymbol{a}_q(\theta, \eta)' w_q(z; \theta, \eta)$$

is orthogonal to order q.

Using Bartlett identities the solution is

$$\boldsymbol{a}_q(\theta, \eta; \mu) = \mathbb{E}_{\theta,\eta}(w_q(z; \theta, \eta)\, w_q(z; \theta, \eta)')^{-1}\, \mathbb{E}_{\theta,\eta}(w_q(z; \theta, \eta)\, u(z; \theta, \eta))$$

This is a projection coefficient.

We recover the projected score of Small and McLeish (1989) and Waterman and Lindsay (1996).

Intuition: Orthogonality to order 1

For any (scalar) coefficient a_1 in

$$u_1^*(z;\theta,\eta) = u(z;\theta,\eta) - a_1(\theta,\eta)\, v_1(z;\theta,\eta)$$

we immediately have that

$$\mathbb{E}_{\theta,\eta}(u_1^*(z;\theta,\eta)) = \mathbb{E}_{\theta,\eta}(u(z;\theta,\eta)) \qquad (= 0 \text{ here}).$$

We next solve

$$\mathbb{E}_{\theta,\eta}\left(\frac{du(z;\theta,\eta)}{d\eta}\right) - a_1(\theta,\eta)\, \mathbb{E}_{\theta,\eta}\left(\frac{dv_1(z;\theta,\eta)}{d\eta}\right) = 0$$

to find

$$a_1(\theta,\eta) = \left(\mathbb{E}_{\theta,\eta}\left(\frac{du(z;\theta,\eta)}{d\eta}\right)\right)\left(\mathbb{E}_{\theta,\eta}\left(\frac{dv_1(z;\theta,\eta)}{d\eta}\right)\right)^{-1}.$$

Intuition: Orthogonality to order 2

Look for coefficient $\boldsymbol{a}_2 = (a_{21}, a_{22})'$ in

$$u_2^*(z;\theta,\eta) = u(z;\theta,\eta) - a_{21}(\theta,\eta)\, v_1(z;\theta,\eta) - a_{22}(\theta,\eta)\, v_2(z;\theta,\eta)$$

so that the resulting function is orthogonal to order 2.

From the constraint on the first derivative we find that

$$a_{21}(\theta,\eta) = a_1(\theta,\eta) - a_{22}(\theta,\eta)\, b_1(\theta,\eta),$$

where a_1 is as before and

$$b_1(\theta,\eta) = \left(\mathbb{E}_{\theta,\eta}\left(\frac{dv_2(z;\theta,\eta)}{d\eta}\right)\right)\left(\mathbb{E}_{\theta,\eta}\left(\frac{dv_1(z;\theta,\eta)}{d\eta}\right)\right)^{-1}.$$

Plugging this back in gives

$$u_2^*(z;\theta,\eta) = u_1^*(z;\theta,\eta) - a_{22}(\theta,\eta)\, v_2^*(z;\theta,\eta),$$

where

$$v_2^*(z;\theta,\eta) = v_2(z;\theta,\eta) - b_1(\theta,\eta)\, v_1(z;\theta,\eta).$$

Recall that u_1^* is orthogonal to order 1.

In the same way, v_2^* is orthogonal to order 1.

It follows that u_2^* is orthogonal to order 1 for any a_{22}.

Taking second derivatives and expectations shows that

$$a_{22}(\theta,\eta) = \left(\mathbb{E}_{\theta,\eta}\left(\frac{d^2 u_1^*(z;\theta,\eta)}{d\eta^2}\right)\right)\left(\mathbb{E}_{\theta,\eta}\left(\frac{d^2 v_2^*(z;\theta,\eta)}{d\eta^2}\right)\right)^{-1}.$$

The terms involving ${}^{da_{22}(\theta,\eta)}\!/_{d\eta}$ and ${}^{d^2 a_{22}(\theta,\eta)}\!/_{d\eta^2}$ both drop out.

The vector $\boldsymbol{a}_2$ is the solution to linear system.

Comment: Ancillarity to order q

An implication of the above is that (for the case of the score for θ) looking for Neyman orthogonality to order q is equivalent to choosing u_q^* such that:

For all $1 \leq o \leq q$,

$$\mathbb{E}(u_q^*(z;\theta_0,\eta_0)\, v_o(z;\theta_0,\eta_0)) = 0,$$

which is a least-squares problem.

For all $1 \leq o \leq q$,

$$\frac{d^o}{d\eta^o}\mathbb{E}_{\theta_0,\eta}(u_q^*(z;\theta_0,\eta_0))\bigg|_{\eta=\eta_0} = 0,$$

which is E-ancillarity to order q.

The equivalence follows from the fact that

$$\frac{d^o}{d\eta^o}\mathbb{E}_{\theta_0,\eta}(u_q^*(z;\theta_0,\eta_0)) = \int u_q^*(z;\theta_0,\eta_0)\,\frac{d^o \ell(z;\theta_0,\eta)}{d\eta^o}\,dz.$$

Other parameters of interest

When interest lies in μ_0 defined through

$$\mathbb{E}(u(z;\theta_0,\eta_0;\mu_0)) = 0$$

we proceed as before and obtain the coefficient

$$\begin{aligned}\boldsymbol{a}_q(\theta,\eta;\mu) = {} & \mathbb{E}_{\theta,\eta}(w_q(z;\theta,\eta)\, w_q(z;\theta,\eta)')^{-1}\, \mathbb{E}_{\theta,\eta}(w_q(z;\theta,\eta)\, u(z;\theta,\eta;\mu)) \\ & - \mathbb{E}_{\theta,\eta}(w_q(z;\theta,\eta)\, w_q(z;\theta,\eta)')^{-1}\, \beta_q(\theta,\eta;\mu)\end{aligned}$$

for $\beta_q(\theta,\eta;\mu)$ the leading q derivatives (wrt η) of

$$\beta_0(\theta,\eta;\mu) := \mathbb{E}_{\theta,\eta}(u(z;\theta,\eta;\mu)).$$

No longer projection coefficient, as $\beta_0(\theta_0,\eta_0;\mu) = 0$ only at $\mu = \mu_0$.

Now,

$$\mathbb{E}(u_q^*(z;\theta_0,\eta_0;\mu_0)\, w_q(z;\theta_0,\eta_0)) = \beta_q(\theta_0,\eta_0;\mu_0);$$

the adjusted ‘score’ has an interpretation of a higher-order influence function.

Example: Neyman-Scott problem

Recall that

$$z_{it} \sim \mathbf{N}(\eta_{i0}, \theta_0).$$

Here, can look at contributions of individual strata, so

$$u(z_i; \theta, \eta) = -\frac{1}{2\theta}\left(T - \frac{\sum_{t=1}^{T}(z_{it} - \eta_i)^2}{\theta}\right),$$

and

$$v_1(z_i; \theta, \eta) = \frac{\sum_{t=1}^{T}(z_{it} - \eta_i)}{\theta}, \qquad v_2(z_i; \theta, \eta) = -\frac{T}{\theta} + \left(\frac{\sum_{t=1}^{T}(z_{it} - \eta_i)}{\theta}\right)^2.$$

We find $a_{21} = 0$ and $a_{22} = {}^1/_{2T}$ so that

$$u_2^*(z_i; \theta, \eta_i) = \frac{1}{2\theta}\left(\frac{\sum_{t=1}^{T}(z_{it} - \bar{z}_i)^2}{\theta} - (T-1)\right)$$

which does not depend on η_i.

The implied estimator performs the usual degrees-of-freedom correction.

We may also be interested in orthogonalizing functions other than the score.

An example is $\mu_0 = {}^1\!/_N \sum_{i=1}^N \eta_{i0}^2$. This fits our framework, with

$$u(z_1, \ldots, z_N; \theta, \eta_1, \ldots, \eta_N; \mu) = \frac{1}{N} \sum_{i=1}^N \eta_i^2 - \mu.$$

We find that, for given θ,

$$\frac{1}{N} \sum_{i=1}^N \bar{z}_i^2 - \frac{\theta}{T}$$

is an estimator that is second-order orthogonal.

The bias of the maximum-likelihood estimator is ${}^{\theta_0}\!/_T$.

The plug-in version of our estimator based on maximum-likelihood has bias ${}^{\theta_0}\!/_{T^2}$, and so is bias reducing.

The plug-in version of our estimator based on the corrected estimator of θ_0 is exactly unbiased.

Example: Linear autoregression

Now suppose that

$$z_{it} = \eta_{i0} + \rho_0 z_{it-1} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathbf{N}(0, \sigma_0^2).$$

Can recenter the data by working with $z_{it} - z_{i0}$, so initial condition is set to zero.

Here, $\theta = (\rho, \sigma^2)$. The adjustment for σ^2 is the same as before, so we focus on the score for ρ.

Now,

$$u(z_i; \theta, \eta_i; \rho) = \frac{\sum_{t=1}^{T} z_{it-1}(z_{it} - \eta_i - \rho z_{it-1})}{\sigma^2},$$

and the score is not orthogonal to any order.

The second-order orthogonal score takes the form

$$u_2^*(z_i;\theta,\eta_i;\rho) = u(z_i;\theta,\eta_i;\rho) + c(\rho) + c(\rho)\,T\,\hat{\eta}(\rho)\,(\eta_i - \hat{\eta}_i(\rho)).$$

for

$$c(\rho) := \frac{1}{1-\rho}\left(1 - \frac{1}{T}\frac{1-\rho^T}{1-\rho}\right)$$

and $\hat{\eta}_i(\rho) = \bar{z}_i - \rho\bar{z}_{i-}$.

At the maximum-likelihood estimator (for given θ) this yields

$$u(z_i;\theta,\hat{\eta}_i(\rho);\rho) + c(\rho) = \frac{\sum_{t=1}^T z_{it-1}((z_{it}-\bar{z}_i) - \rho(z_{it-1}-\bar{z}_{i-}))}{\sigma^2} + c(\rho)$$

which is known to be unbiased for fixed T.

Connects to Cox and Reid (1987), Lancaster (2000), Woutersen (2002), and Arellano (2003) where information orthogonality is used (but not in isolation)

First-order orthogonal score is

$$u_1^*(z_i;\theta,\eta_i) = u(z_i;\theta,\eta_i) + c(\rho)\,\frac{T}{\sigma^2}\,\eta_i\,(\eta_i - \hat{\eta}_i(\rho)).$$

Comment: profiled estimation

For first-order bias correction of $\hat{\theta}$ sample splitting is not needed.

This follows from the fact that, for $q \geq 2$,

$$\mathbb{E}\left(\frac{du_q^*(z_i;\theta_0,\eta_{i0})}{d\eta_i}\, v_1(z_i;\theta_0,\eta_{i0}) \right) = 0$$

so the influence function of $\hat{\eta}_i(\theta)$ in

$$\hat{\eta}_i(\theta_0) - \eta_{i0} \approx -\mathbb{E}\left(\frac{dv_1(z_i;\theta_0,\eta_{i0})}{d\eta_i} \right)^{-1} v_1(z_i;\theta_0,\eta_{i0})$$

is uncorrelated with ${du_q^*(z_i;\theta_0,\eta_{i0})}/{d\eta_i}$ and their dependence on the same data is irrelevant.

Differentiating with respect to η twice the zero-mean property $\mathbb{E}_{\theta,\eta}(u_q^*(z;\theta,\eta)) = 0$ an re-arranging yields

$$\mathbb{E}_{\theta,\eta}\left(\frac{du_q^*(z;\theta,\eta)}{d\eta}\, v_1(z;\theta,\eta) \right) = -\frac{1}{2}\mathbb{E}_{\theta,\eta}\left(\frac{d^2 u_q^*(z;\theta,\eta)}{d\eta^2} + u_q^*(z;\theta,\eta)\, v_2(z;\theta,\eta) \right)$$

from which the result follows.

Example: Network regression

Consider n-dimensional outcome vector y generated through

$$y = X\eta_0 + \varepsilon, \qquad \varepsilon \sim \mathbf{N}(0, \theta_0 I_n).$$

Approach for θ_0 boils down to the usual degrees-of-freedom correction.

Interest lies in

$$\mu_0 = \eta_0' Q\, \eta_0$$

for chosen matrix Q.

Here,

$$u(z; \theta, \eta; \mu) = \mu - \eta' Q\, \eta.$$

The plug-in estimator uses $\hat{\eta} = (X'X)^{-1}X'y = \eta_0 + (X'X)^{-1}X'\varepsilon$ and is biased:

$$\mathbb{E}(\hat{\eta}' Q\, \hat{\eta}) = \eta_0' Q\, \eta_0 + \theta_0 \operatorname{tr}(Q(X'X)^{-1}).$$

A first-order adjustment is

$$u_1^*(z;\theta,\eta,\mu) = \mu - \eta' Q\,\eta - 2\eta' Q(\hat{\eta} - x\eta).$$

A second-order adjustment is

$$u_2^*(z;\theta,\eta,\mu) = \mu - \hat{\eta}' Q\,\hat{\eta} + \theta\,\mathrm{tr}(Q(x'x)^{-1})$$

which no longer depends on η.

The implied estimator (using degrees-of-freedom corrected estimator of θ_0) is the Andrews et al. (2008) estimator.

Gives exactly unbiased estimator.

Team production

Let

$$y_{i_1,\ldots,i_m} = \alpha_m \left({}^{\eta_{i_1}^{\gamma_m}}/m + {}^{\eta_{i_m}^{\gamma_m}}/m \right)^{1/\gamma_m} \varepsilon_{i_1,\ldots,i_m}$$

be the production of the team of m workers $i_1, \ldots i_m$.

We take log-normal errors with variance that can depend on m.

This is a CES production function that depends on team size.

Inputs are worker 'quality'.

Here, α_m is total factor productivity and γ_m measures complementarity.

Here, we do not get a clean factorization of the likelihood.

We look at units that produce on their own and together in a team of size two.

The normality assumption allows for tractable computation (using Faà di Bruno).

We normalize $\alpha_1 = 1$:

-Single production follows the Neyman-Scott problem.

-Use a random subset of such team output as hold-out sample.

Data and results

Data on scientific output of academic researchers (Ductor et al. 2014).

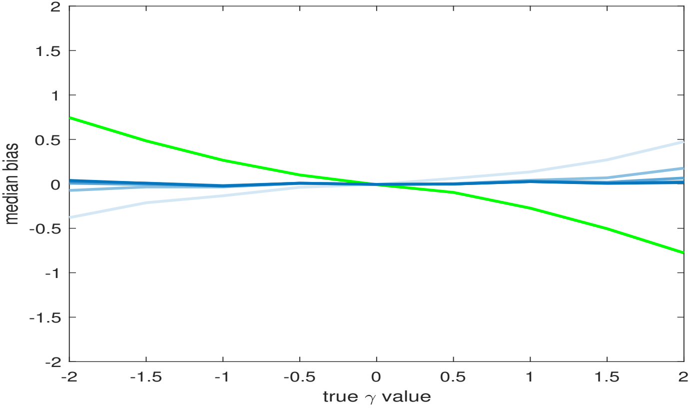
Co-authorship network, based on EconLit.

55k papers for 6.5k authors.

Results for teams of size two (all with sample splitting):

q	SUBSTITUTION γ_2	ELASTICITY $1/(1-\gamma_2)$	TFP α_2	VARIANCE σ_2^2
0	0.116	1.131	1.370	1.617
1	0.116	1.131	1.386	1.617
	[-0.536, 0.605]	–	[1.295, 1.488]	[1.552, 1.725]
2	0.449	1.815	1.360	1.454
	[-0.074, 0.920]	–	[1.257, 1.480]	[1.380, 1.536]
3	0.371	1.590	1.360	1.454
	[-0.142, 0.817]	–	[1.258, 1.481]	[1.381, 1.540]
4	0.374	1.598	1.360	1.450
	[-0.139, 0.817]	–	[1.257, 1.480]	[1.377, 1.536]
5	0.377	1.605	1.360	1.450
	[-0.136, 0.820]	–	[1.257, 1.480]	[1.376, 1.534]
6	0.377	1.605	1.360	1.450
	[-0.136, 0.821]	–	[1.257, 1.480]	[1.376, 1.534]

Simulation



2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
median bias
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
true γ value

Limitations

The parametric setting is important in our derivations.

First-order orthogonality can be achieved outside the likelihood setting for any moment equation $u(z_i; \theta, \eta; \mu)$ using any estimating equation $v(z_i; \theta, \eta)$ for the nuisance parameter.

Can treat a in

$$u(z_i; \theta, \eta; \mu) - a\, v(z_i; \theta, \eta)$$

as an additional nuisance parameter. The modified score is orthogonal to it!

This does not extend to higher-order setting: The implied system of equations becomes inconsistent, in general.

In certain settings other modifications can be done, but no discussion on this today.