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

July 9, 2026

Introduction

Inference in the presence of nuisance parameters is known to be 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 many and/or poorly estimated, (first-order) Neyman orthogonality does not suffice to salvage inference.

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

(6. Limitations and future work)

Motivation

Interested in inference on parameter θ_0, defined through moment condition

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

where η_0 is a nuisance parameter and $z_1, \ldots, z_n$ are i.i.d. random variables.

Throughout we take everything to be scalar-valued for notational simplicity.

In a likelihood setting with density $f(z; \theta, \eta)$ a natural choice for u would be

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

the score.

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

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

for θ.

Here, $\hat{\eta}$ is an auxiliary estimators of η_0.

Supposed to be consistent, but potentially converging very slowly.

Then

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

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

Hence, in general,

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

(ignoring smaller-order terms).

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

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

in general.

This puts very strong requirements on the auxiliary estimator.

Corresponds to the usual standard-error adjustment for two-step estimators.

With (first-order) Neyman orthogonality,

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

and so

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u(z_i;\theta_0,\hat{\eta})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u(z_i;\theta_0,\eta_0)+\mathbb{E}\left(\frac{d^2u(z_i;\theta_0,\eta_0)}{d\eta^2}\right)\frac{\sqrt{n}\,(\hat{\eta}-\eta_0)^2}{2}$$

(again omitting higher-order terms).

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

For score for θ,

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

so first-order orthogonality corresponds to classic information orthogonality.

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 now well understood.

Not for higher-order versions.

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

Today: likelihood framework.

In work in progress (w/ Whitney Newey): method-of-moment framework.

The Neyman-Scott example

$N \times T$ panel data on

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

The maximum-likelihood estimator uses

$$\hat{\eta}_i = \bar{z}_i \sim \mathbf{N}(\eta_{i0}, {}^{\theta_0}/_{T})$$

and equals

$$\hat{\theta} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{it} - \hat{\eta}_i)^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{it} - \eta_{i0})^2 - \frac{1}{N}\sum_{i=1}^{N}(\hat{\eta}_i - \eta_{i0})^2.$$

Note that

$$\mathbb{E}(\hat{\theta} - \theta_0) = -\frac{1}{N}\sum_{i=1}^{N}\text{var}(\hat{\eta}_i) = -\frac{\theta_0}{T}$$

while $\text{std}(\hat{\theta}) \asymp (NT)^{-1/2}$. For bias to be negligible we need that ${}^{N}/_{T} \to 0$.

This example is the panel data version of the degrees-of-freedom correction.

The problem factors across strata.

The score contribution of stratum i is

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

Observe that

$$\mathbb{E}\left(\frac{d\, u(z_i; \theta_0, \eta_{i0})}{d\eta}\right) = -\mathbb{E}\left(\frac{\sum_{t=1}^{T}(z_{it} - \eta_{i0})}{\theta_0^2}\right) = 0,$$

so the score is already orthogonal to first order here.

To connect to the literature, note that $n = N \times T$ and $\hat{\eta}_i - \eta_{i0} = O_P(T^{-1/2})$.

For $\hat{\eta}_i - \eta_{i0} = o_P(n^{-1/4})$ to hold we need that $N = o(T)$.

This is the usual rate to get correct inference based on maximum likelihood.

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

In this example, $u(z_i;\theta_0,\eta_i) - u(z_i;\theta_0,\eta_{i0})$ is equal to

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

Both terms have expectations that are $O(1)$, in general, so both need to be handled to reduce bias. (At $\eta_i = \bar{z}_i$, these expectations are $1/2\theta_0$ and $-2/2\theta_0$.)

The second term can be handled using sample splitting.

Dealing with the first term requires second-order Neyman orthogonality.

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 p let

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

For example,

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

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

Orthogonality to order q

Collect the leading q functions $v_1, v_2, \ldots, v_q$ in the vector function $w_q(z; \theta, \eta)$.

We look for coefficient vector a such that

$$u_q^*(z; \theta, \eta) = u(z; \theta, \eta) - a(\theta, \eta)^\top w_q(z; \theta, \eta)$$

is orthogonal to order q.

The solution is

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

and is known up to (θ, η).

This is a projection coefficient:

u_q^* is the residual from a (population) regression of u on w_q.

Intuition: First-order orthogonality

Consider function u so that $\mathbb{E}_{\theta,\eta}(u(z;\theta,\eta)) = 0$.

Look for linear combination of u and v_1

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

for which

$$\mathbb{E}_{\theta,\eta}\left(\frac{du_1^*(z;\theta,\eta)}{d\eta}\right) = \mathbb{E}_{\theta,\eta}\left(\frac{du(z;\theta,\eta)}{d\eta}\right) - a(\theta,\eta)\,\mathbb{E}_{\theta,\eta}\left(\frac{dv_1(z;\theta,\eta)}{d\eta}\right) = 0.$$

Note that

$$-\frac{da(\theta,\eta)}{d\eta}\,\mathbb{E}_{\theta,\eta}\left(v_1(z;\theta,\eta)\right) = 0$$

and so this term does not appear.

The solution is

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

which is $\mathbb{E}_{\theta,\eta}\left(u(z;\theta,\eta)\, v_1(z;\theta,\eta)\right)\mathbb{E}_{\theta,\eta}\left(v_1(z;\theta,\eta)^2\right)^{-1}$ by Bartlett identities.

Intuition: Second-order orthogonality

Now look for linear combination of u and (v_1, v_2)

$$u_2^*(z;\theta,\eta) = u(z;\theta,\eta) - \begin{pmatrix} a_1(\theta,\eta) & a_2(\theta,\eta) \end{pmatrix} \begin{pmatrix} v_1(z;\theta,\eta) \\ v_2(z;\theta,\eta) \end{pmatrix}$$

so that the first two derivatives have mean zero.

Can do this sequentially.

Same way as before, first-derivative condition is

$$\mathbb{E}_{\theta,\eta}\left(\frac{du(z;\theta,\eta)}{d\eta}\right) - \begin{pmatrix} a_1(\theta,\eta) & a_2(\theta,\eta) \end{pmatrix} \begin{pmatrix} \mathbb{E}_{\theta,\eta}\left(\frac{dv_1(z;\theta,\eta)}{d\eta}\right) \\ \mathbb{E}_{\theta,\eta}\left(\frac{dv_2(z;\theta,\eta)}{d\eta}\right) \end{pmatrix} = 0,$$

and so

$$a_1(\theta,\eta) = \frac{\mathbb{E}_{\theta,\eta}\left(\frac{du(z;\theta,\eta)}{d\eta}\right)}{\mathbb{E}_{\theta,\eta}\left(\frac{dv_1(z;\theta,\eta)}{d\eta}\right)} - a_2(\theta,\eta)\frac{\mathbb{E}_{\theta,\eta}\left(\frac{dv_2(z;\theta,\eta)}{d\eta}\right)}{\mathbb{E}_{\theta,\eta}\left(\frac{dv_1(z;\theta,\eta)}{d\eta}\right)}$$

must hold for any a_2.

Substituting back into the original problem yields

$$u_2^*(z;\theta,\eta) = u_1^*(z;\theta,\eta) - a_2(\theta,\eta)\left(v_2(z;\theta,\eta) - \frac{\mathbb{E}_{\theta,\eta}\left(\frac{dv_2(z;\theta,\eta)}{d\eta}\right)}{\mathbb{E}_{\theta,\eta}\left(\frac{dv_1(z;\theta,\eta)}{d\eta}\right)} v_1(z;\theta,\eta)\right)$$

Here, u_1^* is Neyman orthogonal to first order.

The term in brackets is a first-order orthogonalized version of v_2.

Hence, u_2^* is orthogonal to first order for any choice of a_2.

This implies that the second-derivative condition on u_2^* becomes linear in a_2:

The terms involving the derivatives ${da_2(\theta,\eta)}/{d\eta}$ and ${d^2a_2(\theta,\eta)}/{d\eta^2}$ both drop out.

The vector $a = (a_1, a_2)$ is thus the solution to a linear system of equations.

Bartlett identities can be used to write the solution as a projection residual.

Generalization

Interest in parameter μ_0 defined through

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

Here, the adjustment is

$$u_q^*(z; \theta, \eta; \mu) = u(z; \theta, \eta; \mu) - a(\theta, \eta; \mu)^\top\, w_q(z; \theta, \eta)$$

where, now, a satisfies

$$\mathbb{E}_{\theta,\eta}\left(w_q(z; \theta, \eta)\, u_q^*(z; \theta, \eta; \mu)\right) = \beta_q(\theta, \eta; \mu)$$

for $\beta_q(\theta, \eta; \mu)$ vector containing leading q derivatives (with respect to η) of

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

The adjustment is needed because the moment condition is zero only at μ_0

Reveals the u_q^* to be higher-order influence functions (van der Vaart 2014).

Example: Neyman-Scott problem

Recall that

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

Here, can look at contributions of individual strata, so

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

and

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

Here $a = (0, {}^{1}\!/\!_{2T})^\top$ and a calculation gives

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

which does not depend on η_i.

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

We could also be interested in

$$\mu_0 = \frac{1}{N}\sum_{i=1}^{N} h(\eta_{i0})$$

for a smooth function h. A useful example would be higher-order moments

$$\mu_0 = \frac{1}{N}\sum_{i=1}^{N} \eta_{i0}^{s}$$

for some $s \geq 1$.

The plug-in estimator is $1/N \sum_{i=1}^{N} h(\hat{\eta}_i)$.

We have

$$\mathbb{E}(h(\hat{\eta}_i) - h(\eta_{i0})) = \sum_{p=1}^{q} \frac{d^p h(\eta_{i0})}{d\eta_i^p} \frac{\mathbb{E}((\hat{\eta}_i - \eta_{i0})^p)}{p!} + o_P(|\hat{\eta}_i - \eta_{i0}|^q)$$

so here the bias depends on the leading moments of the plug-in estimator $\hat{\eta}_i$.

In the normal case we can write the elements of the Bhattacharyya basis as

$$v_p(z_i;\theta,\eta_i) = \left(\frac{T}{\theta}\right)^{p/2} \mathrm{H}_p\left(\frac{\bar{z}_i - \eta_i}{\sqrt{\theta/T}}\right),$$

where $\mathrm{H}_1(x) = x, \mathrm{H}_2(x) = x^2 - 1, \ldots$ are the standard Hermite polynomials.

The cross-sectional average of

$$h(\hat{\eta}_i) + \sum_{p=1}^{q} \frac{d^p h(\hat{\eta}_i)}{d\eta_i^p} \frac{v_p(z_i;\hat{\theta},\hat{\eta}_i)}{p!}$$

is correct to order q.

For $h(\eta) = \eta^2$ the bias of the plug-in estimator is θ_0/T and, for $q = 2$, we use

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

To see how the adjustment works, note that

$$h(\eta_i) + \sum_{p=1}^{q} \frac{d^p h(\eta_i)}{d\eta_i^p} \frac{v_p(z_i;\theta,\eta_i)}{p!}$$

has derivative

$$\frac{dh(\eta_i)}{d\eta_i} + \sum_{p=1}^{q} \frac{d^{p+1} h(\eta_i)}{d\eta_i^{p+1}} \frac{v_p(z_i;\theta,\eta_i)}{p!} + \sum_{p=1}^{q} \frac{d^p h(\eta_i)}{d\eta_i^p} \frac{1}{p!} \frac{dv_p(z_i;\theta,\eta_i)}{d\eta_i}$$

Now, the key is that

$$\frac{dv_p(z_i;\theta,\eta_i)}{d\eta_i} = -p\, v_{p-1}(z_i;\theta,\eta_i),$$

as this yields (by a change of variable in the second term)

$$\sum_{p=1}^{q} \frac{d^{p+1} h(\eta_i)}{d\eta_i^{p+1}} \frac{v_p(z_i;\theta,\eta_i)}{p!} - \sum_{p=1}^{q-1} \frac{d^{p+1} h(\eta_i)}{d\eta_i^{p+1}} \frac{v_p(z_i;\theta,\eta_i)}{p!} = \frac{d^{q+1} h(\eta_i)}{d\eta_i^{q+1}} \frac{v_q(z_i;\theta,\eta_i)}{q!},$$

which has mean zero for any $q > 0$.

We can take further derivatives.

The second derivative, for example, is (by using the same argument) equal to

$$\frac{d^{q+2}h(\eta_i)}{d\eta_i^{q+2}}\,\frac{v_q(z_i;\theta,\eta_i)}{q!} - \frac{d^{q+1}h(\eta_i)}{d\eta_i^{q+1}}\,\frac{v_{q-1}(z_i;\theta,\eta_i)}{(q-1)!}.$$

Observing that $\mathrm{H}_0(x) = 1$, this is mean zero only if $q > 1$,

Or if the function satisfies

$$\frac{d^{q+1}h(\eta_i)}{d\eta_i^{q+1}} = 0.$$

These calculations generalize to arbitrary order, as one works back to H_0.

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.

Maximum-likelihood estimator suffers from well-known Nickell (1981) bias.

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.

See Dhaene and Jochmans (2016).

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

Team production

Analysis motivated by Ahmadpoor and Jones (2019).

CES production function

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

for a team of m workers $i_1, \ldots i_m$.

Inputs are worker 'quality'.

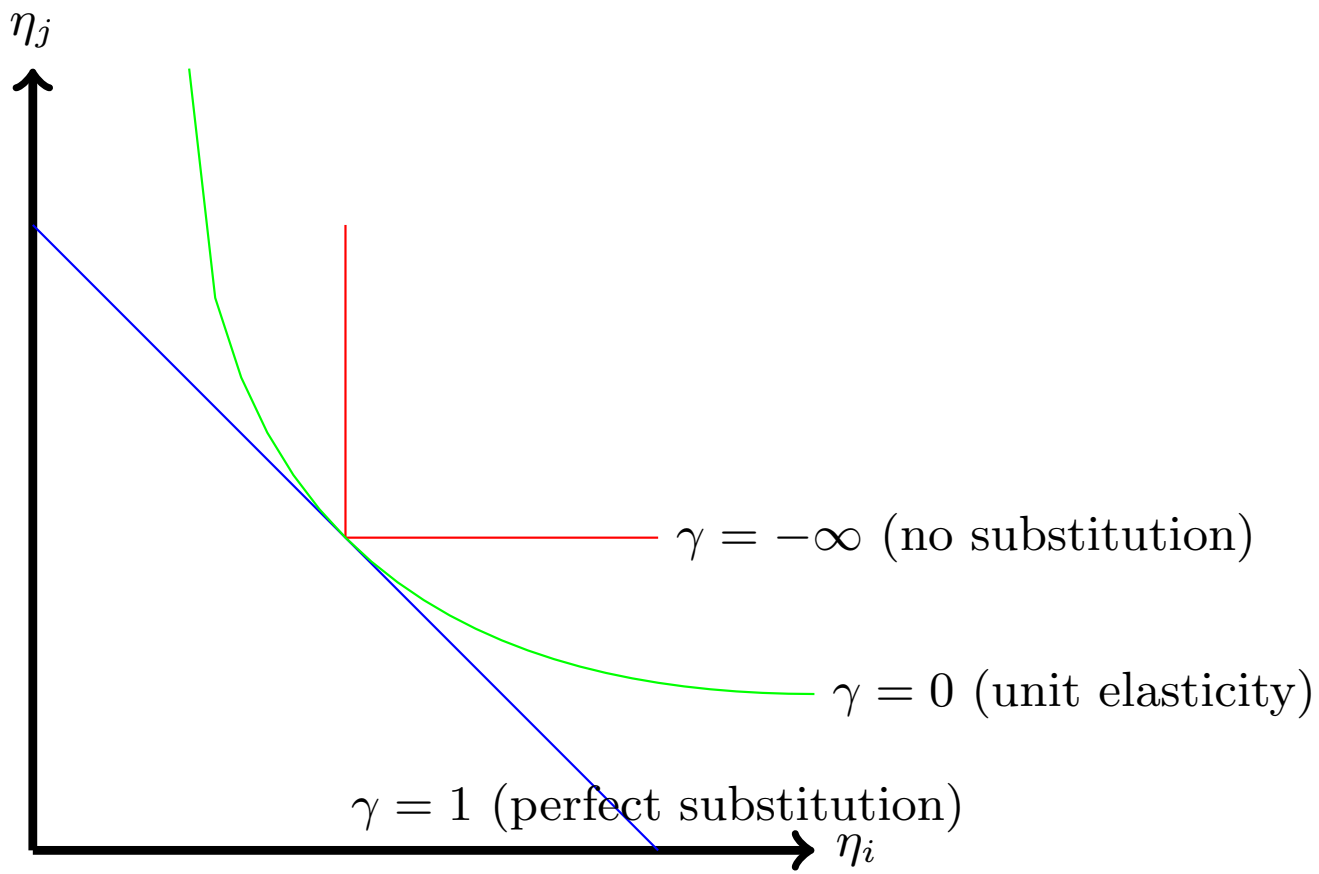
Here, α_m is total factor productivity and γ_m measures complementarity.

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

This gives a nonlinear regression model in logs.

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

Dyadic production



The network setting does not yield a clean factorization of the likelihood.

We consider (many) small subnetworks:

Units that produce on their own and together in a team of size two.

We normalize $\alpha_1 = 1$:

-Single production follows the Neyman-Scott problem.

-Use single-unit output as hold-out sample to obtain auxiliary estimator $\hat{\eta}_i$.

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.

	γ_2 (Substitution)	α_2 (Team size)	σ_2^2 (Variance)	σ_1^2 (Variance)
Plug-in	0.1233 (0.0466)	1.2890 (0.0217)	1.6230 (0.0249)	1.6404 (0.0204)
$q = 1$	−1.9979 (0.2123)	1.3344 (0.0279)	1.6797 (0.0264)	1.7379 (0.0265)
$q = 2$	0.7268 (0.2367)	1.3046 (0.0359)	1.4341 (0.0260)	1.4679 (0.0260)
$q = 3$	0.4467 (0.2034)	1.2936 (0.0369)	1.4399 (0.0254)	1.4423 (0.0236)
$q = 4$	0.3976 (0.1763)	1.2931 (0.0362)	1.4346 (0.0254)	1.4238 (0.0232)
$q = 5$	0.3947 (0.1730)	1.2930 (0.0361)	1.4328 (0.0254)	1.4209 (0.0231)
$q = 6$	0.3944 (0.1770)	1.2930 (0.0365)	1.4316 (0.0254)	1.4194 (0.0230)

Limitations and future work

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.

Ongoing work is about the construction of orthogonalized moments in the semiparametric setting.

As an example, consider $n \times m$ panel data on

$$z_{it} = \eta_{i0} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \text{independent over } i, t; \text{ identically distributed over } t,$$

and again interested in moments of $^1/_n \sum_{i=1}^n h(\eta_{i0})$.

A correction of order q to $^1/_n \sum_{i=1}^n h(\hat{\eta}_i)$ uses

$$h(\hat{\eta}_i) + \sum_{p=1}^{q} \frac{d^p h(\hat{\eta}_i)}{d\eta_i^p} \, \frac{v_p(z_i; \hat{\eta}_i)}{p!}$$

where, now,

$$v_p(z_i; \eta_i) = \frac{p}{m(m-1)\cdots(m-p)} \sum_{t_1 < \cdots < t_p} (z_{it_1} - \eta_i) \cdots (z_{it_p} - \eta_i)$$

are degenerate U-statistics; their derivatives here again satisfy the recursion

$$\frac{dv_p(z_i; \eta_i)}{d\eta_i} = -p \, v_{p-1}(z_i; \eta_i).$$

Full correction for $h(\eta_{i0}) = \eta_{i0}^s$ is equal to $^s/_{m\cdots(m-s)} \sum_{t_1 < \cdots < t_s} z_{it_1} \cdots z_{it_s}$.