Mind the Efficiency Gap[1]

Tobias Fissler

Vienna University of Economics and Business

13 March 2020

[1] Based on joint ongoing work with Timo Dimitriadis and Johanna Ziegel.

Plan

- Problem: Parameter estimation in a regression framework
- Review: M-estimation and Z-estimation
- Structural results on M- and Z-estimators.
- Mind the Gap!
- Implications I: Equivariance.
- Implications II: Efficiency.
- Simulation results.
- Discussion, conclusion and outlook

Parametric Regression Framework

Regression framework

$(Y_t, X_t)_{t\in\mathbb{N}}$ time series such that ...

- Y_t – real-valued response variable,
- X_t – $\mathbb{R}^p$-valued covariates / regressors
- Γ – functional of interest for the conditional distribution $F_t = F_{Y_t|X_t}$ with values in $\mathbb{R}^k$
- Classes of distributions: $F_{Y_t} \in \mathcal{F}_{\mathcal{Y}}$, $F_{X_t} \in \mathcal{F}_{\mathcal{X}}$, $F_{Y_t|X_t} \in \mathcal{F}_{\mathcal{Y}|\mathcal{X}}$, $F_{Y_t,X_t} \in \mathcal{F}_{\mathcal{Y},\mathcal{X}}$.
- $m : \mathbb{R}^p \times \Theta \to \mathbb{R}^k$ – parametric model
- $\Theta \subseteq \mathbb{R}^q$ – parameter space

Assumption (1): Unique model specification

Assume that there is some unique $\theta_0 \in \Theta$ such that

$$\Gamma(F_{Y_t|X_t}) = m(X_t, \theta_0), \quad \mathbb{P}\text{-a.s.} \quad \text{for all } t \in \mathbb{N}$$

- θ_0 is a functional of the joint distribution F_{Y_t,X_t}.
- We do not need strong stationarity, but only semiparametric stationarity.

Key task in statistics and econometrics:

Given $(Y_t, X_t)_{t=1,\dots T}$, find a 'good' estimator $\widehat{\theta}_T$ for θ_0.

Desirable properties of $\widehat{\theta}_T$

- Consistency: $\widehat{\theta}_T \xrightarrow{\mathbb{P}} \theta_0$
- Unbiasedness: $\mathbb{E}[\widehat{\theta}_T] = \theta_0$
- Asymptotic normality: $\sqrt{T}(\widehat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$
- (Asymptotic) efficiency: $\widehat{\theta}_T$ is more efficient if its (asymptotic) covariance matrix is smaller in the Loewner order
- Equivariance properties
- Robustness
- Computational aspects
- ...

We will mainly consider efficiency and equivariance properties.

M- and Z-estimation

M- and Z-estimation

M-estimator: $$\widehat{\theta}_{M,T} = \arg\min_{\theta\in\Theta} \frac{1}{T}\sum_{t=1}^{T} \rho_t\big(Y_t, m(X_t,\theta)\big)$$

Z-estimator: $$\widehat{\theta}_{Z,T} = \arg\min_{\theta\in\Theta} \Big\| \frac{1}{T}\sum_{t=1}^{T} \psi_t(Y_t, X_t, \theta)\Big\|^2$$

Assumption on serial dependence:
(Y_t, X_t) is iid, or stationary and ergodic, or mixing.

$\rightsquigarrow$ Consistency and asymptotic normality of $\widehat{\theta}_{M,T}$ and $\widehat{\theta}_{Z,T}$ under
strict unconditional model-consistency:

$$\mathbb{E}\big[\rho_t\big(Y_t, m(X_t,\theta_0)\big)\big] < \mathbb{E}\big[\rho_t\big(Y_t, m(X_t,\theta)\big)\big] \quad \text{for all } \theta \neq \theta_0$$

strict unconditional model-identification:

$$\mathbb{E}\big[\psi_t(Y_t, X_t, \theta)\big] = 0 \quad \Longleftrightarrow \quad \theta = \theta_0 \quad \text{for all } \theta \in \Theta$$

M- and Z-estimation

- If Γ is one-dimensional, there is a (roughly speaking) a one-to-one correspondence between M- and Z-estimators:
  - differentiate ρ_t wrt θ to obtain ψ_t
  - integrate ψ_t to obtain ρ_t.
- This one-to-one correspondence means that that there is no difference in terms equivariance and efficiency properties.
- If Γ is vector-valued, there are (roughly speaking) more Z-estimators than M-estimators.
- **Reason:** Not every identification function ψ_t has an antiderivative due to integrability conditions.

Integrability conditions

Integrability conditions

Let $U \subset \mathbb{R}^n$ be open and $f_1, \ldots, f_n \colon U \to \mathbb{R}$ continuously differentiable.

If there is a potential $f \colon U \to \mathbb{R}$ such that

$$\partial_i f = f_i \qquad \text{for all } i = 1, \ldots, n,$$

then f is twice continuously differentiable and it holds that

$$\partial_i f_j = \partial_j f_i \qquad \text{for all } i, j = 1, \ldots, n.$$

In order to establish this gap between the classes of Z- and M-estimators we need to establish some structural results.

Construction of loss ρ

Definition 1 (Consistency)

(i) The loss ρ is strictly $\mathcal{F}$-consistent for Γ if

$$\mathbb{E}\big[\rho\big(Y, \Gamma(F_Y)\big)\big] < \mathbb{E}\big[\rho\big(Y, z\big)\big]$$

for all Y such that $F_Y \in \mathcal{F}$ and for all $z \neq \Gamma(F_Y)$.

(ii) The loss $\rho\colon \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ is strictly unconditionally $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-model-consistent for the model m if

$$\mathbb{E}\big[\rho\big(Y, m(X, \theta_0)\big)\big] < \mathbb{E}\big[\rho\big(Y, m(X, \theta)\big)\big]$$

for all (Y, X) such that $F_{Y,X} \in \mathcal{F}_{\mathcal{Y},\mathcal{X}}$ and for all $\theta \in \Theta$, $\theta \neq \theta_0$.

(iii) The loss ρ is strictly conditionally $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-model-consistent for the model m if

$$\mathbb{E}\big[\rho\big(Y, m(X, \theta_0)\big)\big|X\big] < \mathbb{E}\big[\rho\big(Y, m(X, \theta)\big)\big|X\big] \qquad \mathbb{P}\text{-a.s.}$$

for all (Y, X) such that $F_{Y,X} \in \mathcal{F}_{\mathcal{Y},\mathcal{X}}$ and for all $\theta \in \Theta$, $\theta \neq \theta_0$.

(i) $\Longrightarrow$ (iii) $\Longrightarrow$ (ii)

Construction of loss ρ

Theorem 2 (Dimitriadis, F and Ziegel (2020))

Sufficiency: *Under assumption (1) any strictly $F_{\mathcal{Y}|\mathcal{X}}$-consistent loss for Γ is strictly unconditionally $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-model-consistent for m.*
Necessity: *Under assumption (1) and richness assumptions on $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$,[a] any strictly unconditionally $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-model-consistent loss for m is strictly $F_{\mathcal{Y}|\mathcal{X}}$-consistent for Γ.*

[a] For any given conditional distribution $F_{Y|X}$, the marginal of the regressors F_X can vary sufficiently.

$\rightsquigarrow$ We can characterise M-estimators in terms of strictly consistent losses ρ for Γ.

Γ	$\rho(y,z)$	
mean	$\phi(y)-\phi(z)+\phi'(z)(z-y)$	ϕ strictly convex
α-quantile	$\lvert\mathbb{1}\{y \leqslant z\}-\alpha\rvert\,\lvert g(z)-g(y)\rvert$	g strictly increasing

Z-estimators are a bit trickier ...

Identification function

Definition 3 (Identification functions)

(i) The function $\varphi\colon \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}^k$ is a strict $\mathcal{F}$-identification function for Γ if

$$\mathbb{E}[\varphi(Y,z)] = 0 \quad \iff \quad z = \Gamma(F_Y)$$

for all Y such that $F_Y \in \mathcal{F}$ and for all z.

(ii) The function $\psi\colon \mathbb{R} \times \mathbb{R}^p \times \Theta \to \mathbb{R}^q$ is a strict unconditional $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-identification function for $\theta_0\colon \mathcal{F}_{\mathcal{Y},\mathcal{X}} \to \Theta$ if

$$\mathbb{E}[\psi(Y,X,\theta)] = 0 \quad \iff \quad \theta = \theta_0(F_{Y,X})$$

for all Y, X such that $F_{Y,X} \in \mathcal{F}_{\mathcal{Y},\mathcal{X}}$ and for all $\theta \in \Theta$.

(iii) The function ψ is a strict conditional $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-identification function for $\theta_0\colon \mathcal{F}_{\mathcal{Y},\mathcal{X}} \to \Theta$ if

$$\mathbb{E}\big[\psi(Y,X,\theta)\big|X\big] = 0 \quad \mathbb{P}\text{-a.s.} \quad \iff \quad \theta = \theta_0(F_{Y,X})$$

for all Y, X such that $F_{Y,X} \in \mathcal{F}_{\mathcal{Y},\mathcal{X}}$ and for all $\theta \in \Theta$.

(ii) $\not\Longrightarrow$ (iii) $\not\Longrightarrow$ (ii)

Identification functions

Idea: Use $\varphi\big(Y, m(X,\theta)\big)$ as identification functions for θ_0.

Γ	$\varphi(y,z)$
mean	$z - y$
α-quantile	$\mathbb{1}\{y \leqslant z\} - \alpha$

If φ is a strict identification function for Γ, then $\varphi\big(Y, m(X,\theta)\big)$ is a strict conditional identification function for θ_0.

However: It is in general not possible to establish an equivalence between strict conditional and strict unconditional identification functions due to possible cancellation effects.

Construction of identification functions

- Starting with a strict $\mathcal{F}_{\mathcal{Y}|\mathcal{X}}$-identification function φ for Γ, the function $\varphi(y, m(x,\theta))$ is only a strict conditional $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-identification function for θ_0:

$$\mathbb{E}\big[\varphi\big(Y, m(X,\theta)\big)\big|X\big] = 0 \quad \mathbb{P}\text{-a.s.} \qquad \Longleftrightarrow \qquad \theta = \theta_0(F_{Y,X})\,.$$

- Recall that $\mathbb{E}\big[\varphi\big(Y, m(X,\theta)\big)\big|X\big] = 0$ $\mathbb{P}$-a.s. is equivalent to

$$\mathbb{E}\big[a(X)^\intercal\varphi\big(Y, m(X,\theta)\big)\big] = 0 \qquad \text{for all measurable } a\colon \mathbb{R}^p \to \mathbb{R}^k$$

- Unless σ-algebra $\sigma(X)$ is very simple (e.g. if X assumes only finitely many values), we need infinitely many test functions...
- Construction: Stack test functions into an instrument matrix $A(x,\theta) \in \mathbb{R}^{q\times k}$ and consider

$$\psi_A(y,x,\theta) = A(x,\theta)\varphi\big(y, m(x,\theta)\big).$$

- One needs to check strict unconditional identification on a case by case basis (sometimes there are primitive conditions).

Conditional vs. unconditional identifiability

Conditional identifiability $\implies$ unconditional identifiability:

Example 4 (Mean regression with a linear model)

Let $k = 1$, $q = p \geqslant 1$ and consider

$$Y = X^\intercal \theta_0 + \varepsilon, \quad \theta_0 \in \Theta = \mathbb{R}^q,\ \mathbb{E}[\varepsilon|X] = 0\,.$$

Recall Econometrics I course: no perfect multicollinearity, i.e. $\mathbb{E}[XX^\intercal]$ has full rank.
Indeed, this full rank condition implies a uniquely identified model parameter:

$$0 < (\theta - \theta')^\intercal \mathbb{E}\big[XX^\intercal\big](\theta - \theta') = \mathbb{E}\big[\|m(X,\theta) - m(X,\theta')\|^2\big] \quad \forall \theta' \neq \theta.$$

Setting $A(X,\theta) = X$, and using $\varphi(y,z) = z - y$, we obtain a strict unconditional identification function:

$$\mathbb{E}\big[A(X,\theta)\varphi\big(Y, m(X,\theta)\big)\big] = \mathbb{E}\big[A(X,\theta)X^\intercal\big](\theta - \theta_0) = \mathbb{E}\big[XX^\intercal\big](\theta - \theta_0).$$

$\rightsquigarrow$ We can also use other instrument matrices $A(X,\theta)$. Crucial condition is

$$\mathbb{E}\big[A(X,\theta)X^\intercal\big] \quad \text{has full rank for all } \theta \in \Theta\,.$$

Conditional vs. unconditional identifiability

Conditional identifiability $\implies$ unconditional identifiability:

Proposition 5 (Dimitriadis, F, Ziegel (2020))

Under Assumption (1), let $\varphi\colon \mathbb{R}\times\mathbb{R}^k \to \mathbb{R}^k$ *be a strict* $\mathcal{F}_{\mathcal{Y}|\mathcal{X}}$*-identification function. Let* $A\colon \mathbb{R}^p \times \Theta \to \mathbb{R}^{q\times k}$ *be an instrument matrix such that*

$$\mathbb{E}\big[A(X,\theta)D(X,\theta')\big] \quad \text{has full rank, where}$$

$$D(X,\theta') = \nabla_\theta \mathbb{E}\big[\varphi\big(Y, m(X,\theta)\big)\,\big|\,X\big]\,\Big|_{\theta=\theta'}$$

for all (Y,X) *such that* $F_{Y,X} \in \mathcal{F}_{\mathcal{Y},\mathcal{X}}$ *and for all* $\theta,\theta' \in \Theta$ *such that there is a* $\lambda \in [0,1]$ *with* $\theta' = (1-\lambda)\theta_0 + \lambda\theta$.
Then $A(x,\theta)\varphi\big(y, m(x,\theta)\big)$ *is a strict unconditional identification function for* θ_0.

Proof: Clearly, $\mathbb{E}\big[A(X,\theta_0)\varphi\big(Y,m(X,\theta_0)\big)\big] = 0$. For $\theta \neq \theta_0$ use the mean value theorem:

$$\begin{aligned}\mathbb{E}\big[\varphi\big(Y,m(X,\theta)\big)\,\big|\,X\big] &= \mathbb{E}\big[\varphi\big(Y,m(X,\theta)\big)\,\big|\,X\big] - \mathbb{E}\big[\varphi\big(Y,m(X,\theta_0)\big)\,\big|\,X\big] \\ &= \nabla_\theta \mathbb{E}\big[\varphi\big(Y,m(X,\theta)\big)\,\big|\,X\big]\,\big|_{\theta=\theta'}(\theta-\theta_0) = D(X,\theta')(\theta-\theta_0)\end{aligned}$$

Therefore

$$\mathbb{E}\big[A(X,\theta)\varphi\big(Y,m(X,\theta)\big)\big] = \mathbb{E}\big[A(X,\theta)D(X,\theta')\big](\theta-\theta_0)\,.$$

□

Summary

- Given a strictly $\mathcal{F}_{\mathcal{Y}|\mathcal{X}}$-consistent loss function $\rho\colon \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ for Γ,
$$\rho\big(Y, m(X,\theta)\big)$$
is an unconditional strictly $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-model-consistent loss.

- Given a strict $\mathcal{F}_{\mathcal{Y}|\mathcal{X}}$-identification function $\varphi\colon \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}^k$ for Γ,
$$\varphi\big(Y, m(X,\theta)\big)$$
is a strict conditional $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-identification function for θ_0.

- If we use an instrument matrix $A\colon \mathbb{R}^p \times \Theta \to \mathbb{R}^{q\times k}$ such that
$$\mathbb{E}\big[A(X,\theta)D(X,\theta')\big] \quad \text{has full rank}$$
for all $\theta, \theta' \in \Theta$ such that there is a $\lambda \in [0,1]$ with $\theta' = (1-\lambda)\theta_0 + \lambda\theta$. Then
$$A(X,\theta)\varphi\big(Y, m(X,\theta)\big)$$
is a strict unconditional $\mathcal{F}_{\mathcal{Y},\mathcal{X}}$-identification function for θ_0.

Mind the Gap!

Gap between M- and Z-estimators

What is the relation between the building blocks ρ and φ?

Theorem 6 (Osband (1985), Gneiting (2011), F and Ziegel (2016), Dimitriadis, F and Ziegel (2020))

Let $\varphi\colon \mathbb{R}\times\mathbb{R}^k \to \mathbb{R}^k$ be some strict $\mathcal{F}$-identification function for Γ and $\mathcal{F}$ be sufficiently rich.

(i) *$\rho\colon \mathbb{R}\times\mathbb{R}^k \to \mathbb{R}$ is a strictly $\mathcal{F}$-consistent loss for $\Gamma\colon \mathcal{F}\to\mathbb{R}^k$ only if there is some matrix-valued function $h(z)\in\mathbb{R}^{k\times k}$ such that*

$$\nabla_z \rho(y,z) = h(z)\varphi(y,z)\,. \tag{1}$$

(ii) *$\tilde\varphi\colon \mathbb{R}\times\mathbb{R}^k \to \mathbb{R}^k$ is a strict $\mathcal{F}$-identification function for $\Gamma\colon \mathcal{F}\to\mathbb{R}^k$ if and only if there is some matrix-valued function $h(z)\in\mathbb{R}^{k\times k}$ with full rank such that*

$$\tilde\varphi(y,z) = h(z)\varphi(y,z)\,. \tag{2}$$

$$\rightsquigarrow \nabla_\theta \rho\big(Y, m(X,\theta)\big) = \underbrace{\big(\nabla_\theta m(X,\theta)\big)^\intercal h\big(m(X,\theta)\big)}_{=A(X,\theta)} \varphi\big(Y, m(X,\theta)\big).$$

Gap between M- and Z-estimators

For one-dimensional functionals $k = 1$, the class of strictly consistent losses and strict identification functions is very similar.

Mean:

$$\nabla_z \rho(y, z) = \underbrace{h(z)}_{=\phi''(z) \geqslant 0} (y - z),$$

$$\tilde{\varphi}(y, z) = h(z)(y - z), \quad h(z) \neq 0.$$

α-quantile:

$$\nabla_z \rho(y, z) = \underbrace{h(z)}_{=g'(z) \geqslant 0} \big(\mathbb{1}\{y \leqslant z\} - \alpha\big),$$

$$\tilde{\varphi}(y, z) = h(z)\big(\mathbb{1}\{y \leqslant z\} - \alpha\big), \quad h(z) \neq 0.$$

If the expected losses don't have saddle points, then $h > 0$. If the expected identification functions are continuous, then either $h > 0$ or $h < 0$.
Therefore, ignoring the sign, there is a one-to-one relation between strictly consistent losses and strict identification functions for Γ.

Gap between M- and Z-estimators

There is a substantial gap between the classes of M-estimators and Z-estimators for higher dimensional functionals! The reason are integrability conditions.[2]

Examples: Double quantile:

$$\nabla_{z_1,z_2}\rho(y,z_1,z_2) = \begin{pmatrix} g_1'(z_1) & 0 \\ 0 & g_2'(z_2) \end{pmatrix} \begin{pmatrix} \mathbb{1}\{y \leqslant z_1\} - \alpha \\ \mathbb{1}\{y \leqslant z_2\} - \beta \end{pmatrix}, \; g_1'(z_1), g_2'(z_2) \geqslant 0$$

$$\tilde{\varphi}(y,z_1,z_2) = h(z_1,z_2) \begin{pmatrix} \mathbb{1}\{y \leqslant z_1\} - \alpha \\ \mathbb{1}\{y \leqslant z_2\} - \beta \end{pmatrix}, \qquad \det h(z_1,z_2) \neq 0$$

(mean, variance):

$$h(z_1,z_2)\begin{pmatrix} z_1 - y \\ z_2 - (z_1 - y)^2 \end{pmatrix} = \begin{cases} \nabla_{z_1,z_2}\rho(y,z_1,z_2), & h \text{ symmetric and positive semi-definite} \\ & \partial h_{12} = \partial_1 h_{22}, \; \partial_2 h_{11} = \partial_1 h_{21} - 2h_{22} \\ \tilde{\varphi}(y,z_1,z_2), & \det h(z_1,z_2) \neq 0. \end{cases}$$

[2] The Hessian of the expected score must be symmetric.

Gap between M- and Z-estimators

$(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$:

$$\varphi(y, z_1, z_2) = \begin{pmatrix} \mathbb{1}\{y \leqslant z_1\} - \alpha \\ z_2 + (\mathbb{1}\{y \leqslant z_1\} - \alpha)z_1/\alpha - \mathbb{1}\{y \leqslant z_1\}y/\alpha \end{pmatrix}.$$

$$\nabla_{z_1,z_2}\rho(y, z_1, z_2) = \begin{pmatrix} g'(z_1) + \phi'(z_2)/\alpha & 0 \\ 0 & \phi''(z_2) \end{pmatrix} \varphi(y, z_1, z_2),$$

where $g'(z_1) + \phi'(z_2)/\alpha, \phi''(z_2) \geqslant 0$.

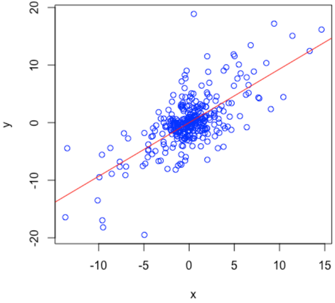
$$\tilde{\varphi}(y, z_1, z_2) = h(z_1, z_2)\varphi(y, z_1, z_2), \qquad \det h(z_1, z_2) \neq 0$$

Considering vector-valued functionals Γ might be of direct applied interest (e.g. when fitting prediction intervals). On the other hand, the non-elicitability of functionals such as variance or ES requires to use their co-elicitability with other functionals (mean or VaR).

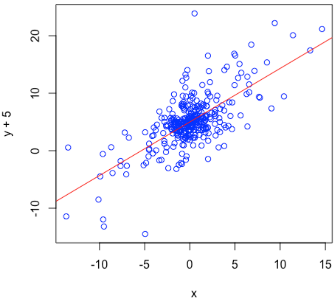
Implications I: Equivariance

Equivariance – Example: mean regression with linear model

Response Y



Response $Y + 5$




```
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
  -0.003084     0.931221
```

```
lm(formula = y + 5 ~ x)

Coefficients:
(Intercept)            x
  4.9969        0.931221
```

Equivariance

- Previous example reflects: $\mathbb{E}[Y+c] = \mathbb{E}[Y] + c$ for all $c \in \mathbb{R}$.
- Recall that for all $c \in \mathbb{R}$

$$\big(\mathrm{VaR}_\alpha(Y+c), \mathrm{ES}_\alpha(Y+c)\big) = \big(\mathrm{VaR}_\alpha(Y) + c, \mathrm{ES}_\alpha(Y) + c\big),$$
$$\big(\mathbb{E}[Y+c], \mathrm{Var}(Y+c)\big) = \big(\mathbb{E}[Y] + c, \mathrm{Var}(Y)\big).$$

- Assume we model $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ with a common or two distinct intercepts. Alternatively, assume we model $(\mathbb{E}, \mathrm{Var})$ with an intercept for the mean component.
- For fixed regressors, if the responses Y_t are shifted by some constant $c \in \mathbb{R}$, an equivariant estimator $\widehat{\theta}$ should have the same shift in the intercept components.
- If the model is correctly specified and the estimator is consistent, its limit will exhibit this equivariance property.
- We would like the estimator to be equivariant also on finite samples or under model misspecification.

Equivariance

- There are no translation invariant losses for $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ and for $(\mathbb{E}, \mathrm{Var})$ (F and Ziegel, 2019).
- But there are translation invariant identification functions:

$$(\mathbb{E}, \mathrm{Var}): \qquad \varphi(y, z_1, z_2) = \begin{pmatrix} z_1 - y \\ z_2 - (z_1 - y)^2 \end{pmatrix},$$

$$(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha): \qquad \varphi(y, z_1, z_2) = \begin{pmatrix} \mathbb{1}\{y \leqslant z_1\} - \alpha \\ z_2 + \frac{1}{\alpha}\mathbb{1}\{y \leqslant z_1\}(z_1 - y) - z_1 \end{pmatrix}.$$

- Given that $A(X, \theta)$ is constant in the intercept component[3], we obtain a translation equivariant Z-estimator.

[3] This can usually be used for linear models.

Implications II: Efficiency

Efficiency

$$\widehat{\theta}_{M,T} = \arg\min_{\theta\in\Theta} \frac{1}{T}\sum_{t=1}^{T} \rho_t\big(Y_t, m(X_t,\theta)\big),$$

$$\widehat{\theta}_{Z,T} = \arg\min_{\theta\in\Theta} \Big\|\frac{1}{T}\sum_{t=1}^{T} A_t(X_t,\theta)\varphi\big(Y_t, m(X_t,\theta)\big)\Big\|^2,$$

Question: What is a good / optimal choice of for ρ_t and ψ_t / A_t?

- $\rightsquigarrow$ Consider asymptotic efficiency of estimators:

$$\sqrt{T}\Lambda_{M,T}^{1/2}\big(\widehat{\theta}_{M,T} - \theta_0\big) \xrightarrow{d} \mathcal{N}_q(0, I_q), \quad \sqrt{T}\Lambda_{Z,T}^{1/2}\big(\widehat{\theta}_{Z,T} - \theta_0\big) \xrightarrow{d} \mathcal{N}_q(0, I_q)$$

- Estimator with lower asymptotic covariance matrix wrt the Loewner order is more efficient:

For two positive semi-definite matrices B and C we say that

$$B \succcurlyeq C \iff B - C \text{ is positive semi-definite,}$$
$$B \succ C \iff B - C \text{ is positive definite.}$$

Asymptotic efficiency

Asymptotic covariance of Z-estimator is $\Lambda_{Z,T}^{-1}$ where:

$$\Lambda_{Z,T} = \Delta_T^{-1}\Sigma_T\left(\Delta_T^{-1}\right)^\intercal, \qquad \text{(Sandwich form!)}$$

$$\Sigma_T = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[A_t(X_t,\theta_0)S_t(X_t,\theta_0)A_t(X_t,\theta_0)^\intercal\right] \in \mathbb{R}^{q\times q} \qquad \text{and}$$

$$\Delta_T = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[A_t(X_t,\theta_0)D_t(X_t,\theta_0)\right] \in \mathbb{R}^{q\times q}$$

where, for any $\theta \in \Theta$,

$$S_t(X_t,\theta) = \mathbb{E}\left[\varphi\big(Y_t, m(X_t,\theta)\big)\varphi\big(Y_t, m(X_t,\theta)\big)^\intercal \Big| X_t\right] \in \mathbb{R}^{k\times k} \qquad \text{and}$$

$$D_t(X_t,\theta) = \nabla_\theta\mathbb{E}\Big[\varphi\big(Y_t, m(X_t,\theta)\big) \mid X_t\Big] \in \mathbb{R}^{k\times q}.$$

Asymptotic covariance of the M-estimator has the same structure, considering

$$\psi_t(Y_t,X_t,\theta) = \nabla_\theta\rho_t\big(Y_t, m(X_t,\theta)\big) = \underbrace{\big(\nabla_\theta m(X_t,\theta)\big)^\intercal h_t\big(m(X_t,\theta)\big)}_{=A_t(X_t,\theta)}\varphi\big(Y_t, m(X_t,\theta)\big).$$

Efficiency bound

Theorem 7 (Dimitriadis, F, Ziegel (2020))

$$A^*_{t,C}(X_t,\theta_0) = CD_t(X_t,\theta_0)^\intercal S_t(X_t,\theta_0)^{-1} \quad \text{for all } t = 1,\ldots,T, \qquad (3)$$

for some invertible matrix $C \in \mathbb{R}^{q\times q}$. Then

(i) *The Z-estimator based on $A^*_{t,C}$ has asymptotic covariance matrix*

$$(\Lambda^*_T)^{-1} := \left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[D_t(X_t,\theta_0)^\intercal S_t(X_t,\theta_0)^{-1}D_t(X_t,\theta_0)\right]\right)^{-1}.$$

(ii) *Any other choice of instrumental matrices is at most as efficient:* $\Delta_T^{-1}\Sigma_T\left(\Delta_T^{-1}\right)^\intercal \succcurlyeq (\Lambda^*_T)^{-1}$.

(iii) *The form at (3) is necessary: If for some $t \in \{1,\ldots,T\}$ and for any non-singular and deterministic matrix C* $\mathbb{P}\big(A_t(X_t,\theta_0) \neq A^*_{t,C}(X_t,\theta_0)\big) > 0$, *then* $\Delta_T^{-1}\Sigma_T\left(\Delta_T^{-1}\right)^\intercal \succ (\Lambda^*_T)^{-1}$.

Efficiency bound

Idea of the proof:

- (i) is direct calculation.
- (ii) define the vectors

$$\chi_{t,T} = \Big(\Delta_{T,A}^{-1} A_t(X_t, \theta_0) - \Lambda_T^{-1} A_t^*(X_t, \theta_0)\Big)\, \varphi_0\big(Y_t, m(X_t, \theta_0)\big)$$

  Then one can show that

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\chi_{t,T}\chi_{t,T}^{\intercal}\right] = \Delta_T^{-1}\Sigma_T\left(\Delta_T^{-1}\right)^{\intercal} - \left(\Lambda_T^*\right)^{-1}$$

- (iii) More involved, but same starting point as (ii).

Comments:

- Time series generalisation of the result in Newey (1993).
- Necessary assertion (iii) is novel.

Example: Mean regression

- Let $\Gamma(F_{Y_t|X_t}) = \mathbb{E}[Y_t|X_t]$.
- $\rho_t\big(Y_t, m(X_t,\theta)\big) = \phi_t(Y_t) - \phi_t(m(X_t,\theta)) + \phi_t'(m(X_t,\theta))(m(X_t,\theta) - Y_t)$ where ϕ is strictly convex.
- $\varphi\big(Y_t, m(X_t,\theta)\big) = Y_t - m(X_t,\theta)$.
- $S_t(X_t,\theta_0) = \mathbb{E}\left[\varphi\big(Y_t, m(X_t,\theta)\big)\varphi\big(Y_t, m(X_t,\theta)\big)^\intercal \big| X_t\right] = \mathrm{Var}(Y_t|X_t)$
- $D_t(X_t,\theta_0) = \nabla_\theta \mathbb{E}\big[\varphi\big(Y_t, m(X_t,\theta)\big) \,|\, X_t\big] = \nabla_\theta m(X_t,\theta_0)$
- $A^*_{t,I}(X_t,\theta_0) = D_t(X_t,\theta_0)^\intercal S_t(X_t,\theta_0)^{-1} = \frac{(\nabla_\theta m(X_t,\theta_0))^\intercal}{\mathrm{Var}(Y_t|X_t)}$
- This efficiency bound can be achieved with an M-estimator, using the loss functions

$$\rho_t\big(Y_t, m(X_t,\theta)\big) = \frac{1}{2}\frac{\big(Y_t - m(X_t,\theta)\big)^2}{\mathrm{Var}(Y_t|X_t)} \quad \rightsquigarrow \phi_t(z) = \frac{1}{2}\frac{z^2}{\mathrm{Var}(Y_t|X_t)}\,.$$

- Classical result. But: Estimating the conditional mean efficiently requires to solve the more involved problem of estimating the conditional variance!

Example: Quantile regression

- Let $\Gamma(F_{Y_t|X_t}) = q_\alpha(F_{Y_t|X_t})$
  $\rho_t\big(Y_t, m(X_t,\theta)\big) = \big(\mathbb{1}\{Y_t \leqslant m(X_t,\theta)\} - \alpha\big)\big(g_t(m(X_t,\theta)) - g_t(Y_t)\big)$
  where g_t is strictly increasing.
- $\varphi\big(Y_t, m(X_t,\theta)\big) = \mathbb{1}\{Y_t \leqslant m(X_t,\theta)\} - \alpha.$
- $S_t(X_t,\theta_0) = \mathbb{E}\left[\varphi\big(Y_t, m(X_t,\theta)\big)\varphi\big(Y_t, m(X_t,\theta)\big)^\intercal \big| X_t\right] = \alpha(1-\alpha)$
- $D_t(X_t,\theta_0) = \nabla_\theta \mathbb{E}\big[\varphi\big(Y_t, m(X_t,\theta)\big) \mid X_t\big] = f_{Y_t|X_t}\big(m(X_t,\theta_0)\big)\nabla_\theta m(X_t,\theta_0)$
- $A^*_{t,I}(X_t,\theta_0) = D_t(X_t,\theta_0)^\intercal S_t(X_t,\theta_0)^{-1} = f_{Y_t|X_t}\big(m(X_t,\theta_0)\big)\frac{(\nabla_\theta m(X_t,\theta_0))^\intercal}{\alpha(1-\alpha)}$
- This efficiency bound can be achieved with an M-estimator, using $g_t = F_{Y_t|X_t}$, resulting in the loss function (Komunjer and Vuong, 2010)

$$\rho_t\big(Y_t, m(X_t,\theta)\big) = \big(\mathbb{1}\{Y_t \leqslant m(X_t,\theta)\} - \alpha\big)\big(F_{Y_t|X_t}(m(X_t,\theta)) - F_{Y_t|X_t}(Y_t)\big)\,.$$

- Drawback: Estimating the conditional quantile efficiently requires to solve the more involved problem of estimating the whole conditional distribution! (Actually, estimating the conditional density at the correct quantile would be sufficient. But still very involved!)

Example: Mean–Variance regression

- Let $\Gamma(F_{Y_t|X_t}) = \big(\mathbb{E}[Y_t|X_t], \mathrm{Var}(Y_t|X_t)\big)^\intercal$
- $\rho_t\big(Y_t, m_1(X_t,\theta), v(X_t,\theta)\big) =$
  $-\phi_t\begin{pmatrix} m_1(X_t,\theta) \\ v(X_t,\theta) + m_1^2(X_t,\theta)\end{pmatrix} + (\nabla\phi_t)\begin{pmatrix} m_1(X_t,\theta) \\ v(X_t,\theta) + m_1^2(X_t,\theta)\end{pmatrix}\varphi\big(Y_t, m(X_t,\theta)\big)$
- $\varphi\big(Y_t, m(X_t,\theta)\big) = \begin{pmatrix} m_1(X_t,\theta) - Y_t \\ v(X_t,\theta) + m_1^2(X_t,\theta) - Y_t^2\end{pmatrix}.$
- $S_t(X_t,\theta_0) = \mathrm{Var}\left(Y_t \quad Y_t^2 \,\middle|\, X_t\right)$
- $D_t(X_t,\theta_0) = \begin{pmatrix} \nabla_\theta m_1(X_t,\theta_0) \\ \nabla_\theta v(X_t,\theta_0) + 2m_1(X_t,\theta_0)\nabla_\theta m_1(X_t,\theta_0)\end{pmatrix}$
- No Efficiency Gap: Efficiency bound can be achieved using the strictly convex function

$$\phi_t(z) = \frac{1}{2} z^\intercal \left(\mathrm{Var}\left(Y_t \quad Y_t^2 \,\middle|\, X_t\right)\right)^{-1} z$$

Example: Double Quantile regression

- Let $\Gamma(F_{Y_t|X_t}) = \left(q_\alpha(F_{Y_t|X_t}), q_\beta(F_{Y_t|X_t})\right)^\intercal$, $\alpha < \beta$
  (For example when interested in prediction intervals).
- $$\begin{aligned}\varphi_t(y, z_1, z_2) = &\left(\mathbb{1}\{y \leqslant z_1\} - \alpha\right)\left(g_{1,t}(z_1) - g_{1,t}(y)\right)\\ &+ \left(\mathbb{1}\{y \leqslant z_2\} - \beta\right)\left(g_{2,t}(z_2) - g_{2,t}(y)\right) + \kappa_t(y),\end{aligned}$$
  where $g_{1,t}, g_{2,t}$ are strictly increasing (F and Ziegel, 2016).
- $$\varphi(y, z_1, z_2) = \begin{pmatrix}\mathbb{1}\{y \leqslant z_1\} - \alpha\\ \mathbb{1}\{y \leqslant z_2\} - \beta\end{pmatrix}$$
- Efficiency Gap: There are DGPs where this efficiency bound cannot be achieved by an M-estimator!

Theorem 8 (Dimitriadis, F, Ziegel (2020))

Assume that

(dqr1) *the parameters of the individual models are separated*

$$m(X_t, \theta) = \left(m_\alpha(X_t, \theta^\alpha), m_\beta(X_t, \theta^\beta)\right)^\intercal,$$

where $\theta = \left(\theta^\alpha, \theta^\beta\right) \in \Theta^\alpha \times \Theta^\beta = \Theta \subseteq \mathbb{R}^q$*, where* $\theta^\alpha \in \mathbb{R}^{q_1}$ *and* $\theta^\beta \in \mathbb{R}^{q_2}$*;*

(dqr2) *the support of* $\nabla m_\alpha(X_t, \theta_0^\alpha)$ *contains at least* $q_1 + 1$ *different values* $v_1, \dots, v_{q_1+1}$*, such that any subset of cardinality* q_1 *of* $\{v_1, \dots, v_{q_1+1}\}$ *is linearly independent. Similarly, the support of* $\nabla m_\beta(X_t, \theta_0^\beta)$ *contains at least* $q_2 + 1$ *such values.*

Then, the following statements hold:

(i) *If not*

$$\exists c_1, c_2, c_3 > 0 \ \forall t = 1, \dots, T : \begin{cases} f_t(m_\alpha(X_t, \theta_0^\alpha)) = c_1 f_t(m_\beta(X_t, \theta_0^\beta)) \ \mathbb{P}\text{-a.s.} & \text{and} \\ g'_{1,t}(m_\alpha(X_t, \theta_0^\alpha)) = c_2 f_t(m_\alpha(X_t, \theta_0^\alpha)) \ \mathbb{P}\text{-a.s.} & \text{and} \\ g'_{2,t}(m_\beta(X_t, \theta_0^\beta)) = c_3 f_t(m_\beta(X_t, \theta_0^\beta)) \ \mathbb{P}\text{-a.s.} \end{cases}$$

then the M-estimator cannot attain the Z-estimation efficiency bound.

(ii) *If the condition above holds and* $\nabla m_\alpha(X_t, \theta_0^\alpha) = \nabla m_\beta(X_t, \theta_0^\beta)$ *almost surely for all* $t = 1, \dots, T$*, then the M-estimator achieves the Z-estimation efficiency bound.*

Example: Double Quantile regression

Proposition 9 (Dimitriadis, F, Ziegel (2020) – Working version)

For a parametric and linear model with separated parameters such that

$$\frac{f_t\big(m_\alpha(X_t,\theta_0)\big)}{f_t\big(m_\beta(X_t,\theta_0)\big)}$$

is deterministic for all $t = 1, \dots, T$, the most efficient Z-estimator is based on a strict global identification function.

Example: VaR–ES regression

- Let $\Gamma(F_{Y_t|X_t}) = \big(Q_\alpha(F_{Y_t|X_t}), \mathrm{ES}_\alpha(F_{Y_t|X_t})\big)^\intercal$
-

$$\rho_t(y, z_1, z_2) = (\mathbb{1}\{y \leqslant z_1\} - \alpha)\, g_t(z_1) - \mathbb{1}_{\{y \leqslant z_1\}} g_t(y) + \kappa_t(y) \\ + \phi_t'(z_2)\left(z_2 - z_1 + \frac{(z_1 - y)\mathbb{1}\{y \leqslant z_1\}}{\alpha}\right) - \phi_t(z_2),$$

where g_t is increasing and $\phi', \phi'' > 0$ (F and Ziegel, 2016).

-

$$\varphi(y, z_1, z_2) = \begin{pmatrix} \mathbb{1}\{y \leqslant z_1\} - \alpha \\ z_2 - z_1 + \frac{1}{\alpha}\big(z_1 - y\big)\mathbb{1}\{y \leqslant z_1\} \end{pmatrix},$$

- Efficiency Gap: There are DGPs where this efficiency bound cannot be achieved by an M-estimator!

Theorem 10 (Dimitriadis, F, Ziegel (2020))

Assume that

(qesr1) *the parameters of the individual models are separated*

$$m(X_t, \theta) = \big(q_\alpha(X_t, \theta^q),\ e_\alpha(X_t, \theta^e)\big)^\intercal, \quad \theta = \big(\theta^q, \theta^e\big) \in \Theta^q \times \Theta^e = \Theta \subseteq \mathbb{R}^q$$

(qesr2) *the support of* $\nabla q_\alpha(X_t, \theta_0^q)$ *contains at least* $q_1 + 1$ *different values* $v_1, \ldots, v_{q_1+1}$, *such that any subset of cardinality* q_1 *of* $\{v_1, \ldots, v_{q_1+1}\}$ *is linearly independent. Similarly, the support of* $\nabla e_\alpha(X_t, \theta_0^e)$ *contains at least* $q_2 + 1$ *such values.*

Then, the following statements hold:

(i) *If not* $\exists c_1, c_2, c_3 > 0, c_4 \in \mathbb{R}\ \forall t = 1, \ldots, T:$

$$\begin{cases} \mathbb{E}_t\Big[\big(q_\alpha(X_t, \theta_0^q) - Y_t\big)^2 \mathbb{1}_{\{Y_t \leqslant q_\alpha(X_t, \theta_0^q)\}}\Big] = c_1\big(q_\alpha(X_t, \theta_0^q) - e_\alpha(X_t, \theta_0^e)\big)^2, \ \mathbb{P}\text{-a.s.} \quad \text{and} \\ \phi_t''(e_\alpha(X_t, \theta_0^e)) = \frac{\alpha c_2}{\big(q_\alpha(X_t, \theta_0^q) - e_\alpha(X_t, \theta_0^e)\big)^2} \ \mathbb{P}\text{-a.s.} \quad \text{and} \\ g_t'(q_\alpha(X_t, \theta_0^q)) = c_3 f_t(q_\alpha(X_t, \theta_0^q)) - \frac{c_2}{\big(q_\alpha(X_t, \theta_0^q) - e_\alpha(X_t, \theta_0^e)\big)} - \frac{c_4}{\alpha} \ \mathbb{P}\text{-a.s.} \end{cases}$$

then the M-estimator cannot attain the Z-estimation efficiency bound.

(ii) *If the condition above holds and* $\nabla q_\alpha(X_t, \theta_0^q) = \nabla e_\alpha(X_t, \theta_0^e)$ *almost surely for all* $t = 1, \ldots, T$, *then the M-estimator achieves the Z-estimation efficiency bound.*

Simulation: Double Quantile Regression

$X_t = (1, X_{t,2})^\intercal$, $\gamma_0 = (10, 0.5)^\intercal$ and $\eta_0 = (0.5, 0.5)^\intercal$, $T = 2000$

$$X_{t,2} \overset{iid}{\sim} 3 \cdot \mathrm{Beta}(3, 1.5), \quad \text{and} \quad Y_t = X_t^\intercal \gamma_0 + (X_t^\intercal \eta_0) u_t,$$

for a conditional location model $X_t^\intercal \gamma_0$, conditional scale model $X_t^\intercal \eta_0$ and residuals u_t independent of regressors X_t such that,

homoscedastic $u_t \overset{iid}{\sim} \mathcal{N}(0, 1)$, or

heteroscedastic $u_t \overset{iid}{\sim} t_{\nu_t}(\mu_t, \sigma_t)$ with time-varying location μ_t, scale σ_t and degrees of freedom ν_t.

For the heteroscedastic case, we consider a break model:

$\nu_t = 3\mathbb{1}_{\{t \leqslant T/2\}} + 100\mathbb{1}_{\{t > T/2\}}$

$$\mu_t = Q_\beta(t_{\nu_1}) - \sigma_t Q_\beta(t_{\nu_t}) \quad \text{and} \quad \sigma_t = \frac{Q_\alpha(t_{\nu_1}) - Q_\beta(t_{\nu_1})}{Q_\alpha(t_{\nu_t}) - Q_\beta(t_{\nu_t})},$$

$$\rightsquigarrow Q_\alpha(Y_t | X_t) = X_t^\intercal \big(\gamma_0 + \eta_0 z_\alpha\big) \quad \text{and} \quad Q_\beta(Y_t | X_t) = X_t^\intercal \big(\gamma_0 + \eta_0 z_\beta\big),$$

where $z_\alpha = F_u^{-1}(\alpha)$, $z_\beta = F_u^{-1}(\beta)$.

Simulation: Double Quantile Regression

For homoscedastic scenario:

$$\exists \text{ deterministic constant } c > 0 : \quad c = \frac{f_t\big(q_\alpha(X_t, \theta_0^\alpha)\big)}{f_t\big(q_\beta(X_t, \theta_0^\beta)\big)} = \frac{f_{u_t}(z_\alpha)}{f_{u_t}(z_\beta)}.$$

For heteroscedastic scenario:

$$\frac{f_t\big(q_\alpha(X_t, \theta_0^\alpha)\big)}{f_t\big(q_\beta(X_t, \theta_0^\beta)\big)} = \frac{f_{u_t}(z_\alpha)}{f_{u_t}(z_\beta)}$$

is deterministic, but time varying!

Simulation: Double Quantile Regression – True Asymptotic Standard Deviations

	Hom				Het			
	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
	Panel B: $(\alpha, \beta) = (1\%, 2.5\%)$							
id	14.220	7.865	10.175	5.627	51.446	24.383	31.131	15.319
eff	13.619	7.546	9.745	5.399	48.632	22.643	29.936	14.636
eff.bound	13.619	7.546	9.745	5.399	47.871	22.289	29.466	14.409
	Panel C: $(\alpha, \beta) = (5\%, 10\%)$							
id	8.238	4.517	6.664	3.654	17.012	8.929	11.996	6.389
eff	7.895	4.337	6.387	3.508	16.605	8.672	11.643	6.208
eff.bound	7.895	4.337	6.387	3.508	16.453	8.585	11.536	6.145
	Panel D: $(\alpha, \beta) = (25\%, 50\%)$							
id	5.306	2.910	4.880	2.677	6.306	3.450	5.436	2.975
eff	5.091	2.797	4.683	2.573	6.037	3.309	5.211	2.857
eff.bound	5.091	2.797	4.683	2.573	6.031	3.306	5.206	2.854

Simulation: VaR–ES regression

$X_t = (1, X_{t,2})^\intercal$, $\gamma_0 = (-1, -0.5)^\intercal$ and $\eta_0 = (0.5, 0.5)^\intercal$, $T = 2000$

$$X_{t,2} \overset{iid}{\sim} 3 \cdot \mathrm{Beta}(3, 1.5), \quad \text{and} \quad Y_t = X_t^\intercal \gamma_0 + (X_t^\intercal \eta_0) u_t,$$

for a conditional location model $X_t^\intercal \gamma_0$, conditional scale model $X_t^\intercal \eta_0$ and residuals u_t independent of regressors X_t such that,

homoscedastic $u_t \overset{iid}{\sim} \mathcal{N}(0, 1)$, or

heteroscedastic $u_t \overset{iid}{\sim} t_{\nu_t}(\mu_t, \sigma_t)$ with time-varying location μ_t, scale σ_t and degrees of freedom ν_t.

For the heteroscedastic case, we consider a break model:

$\nu_t = 3\mathbb{1}_{\{t \leqslant T/2\}} + 100\mathbb{1}_{\{t > T/2\}}$

$$\mu_t = Q_\alpha(t_{\nu_1}) - \sigma_t Q_\alpha(t_{\nu_t}) \quad \text{and} \quad \sigma_t = \frac{Q_\alpha(t_{\nu_1}) - ES_\alpha(t_{\nu_1})}{Q_\alpha(t_{\nu_t}) - ES_\alpha(t_{\nu_t})}$$

$$\rightsquigarrow Q_\alpha(Y_t|X_t) = X_t^\intercal\big(\gamma_0 + \eta_0 z_\alpha\big), \quad \text{and} \quad ES_\alpha(Y_t|X_t) = X_t^\intercal\big(\gamma_0 + \eta_0 z_\alpha\big),$$

where $z_\alpha = \mathrm{VaR}_\alpha(u_t)$ and $z_\alpha = \mathrm{ES}_\alpha(u_t)$ are true quantile and ES of u_t.

Simulation: VaR–ES regression

For homoscedastic scenario:

$$\mathrm{Var}_t\left(Y_t \big| Y_t \leqslant q_\alpha(X_t, \theta_0^q)\right) = (X_t^\intercal \eta_0)^2 \left[1 - z_\alpha \frac{\phi(z_\alpha)}{\Phi(z_\alpha)} - \left(\frac{\phi(z_\alpha)}{\Phi(z_\alpha)}\right)^2\right],$$

$$\left(q_\alpha(X_t, \theta_0^q) - e_\alpha(X_t, \theta_0^e)\right)^2 = (z_\alpha - z_\alpha)^2 \left(X_t^\intercal \eta_0\right)^2, \quad \text{and}$$

$$f_t(q_\alpha(X_t, \theta_0^q)) = \frac{1}{X_t^\intercal \eta_0} f_{u_t}(z_\alpha),$$

For heteroscedastic scenario:

$$\mathrm{Var}_t\left(Y_t | Y_t \leqslant q_\alpha(X_t, \theta_0)\right) = (X_t^\intercal \eta_0)^2 \,\mathrm{Var}\left(u_t | u_t \leqslant Q_\alpha(u_t)\right),$$

which is time varying.

Simulation: VaR–ES regression – True Asymptotic Standard Deviations

		Hom				Het			
		θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
		Panel A: $\alpha = 1\%$							
zero	exp	17.760	11.411	19.282	11.525	238.471	216.677	323.986	304.832
zero	log	13.780	7.611	16.966	9.370	70.491	32.788	153.331	84.508
eff	eff	13.772	7.607	16.926	9.349	61.588	28.192	153.252	84.468
eff.bound	eff.bound	13.772	7.607	16.926	9.349	55.103	24.923	123.186	70.744
		Panel B: $\alpha = 2.5\%$							
zero	exp	12.274	7.610	13.107	7.632	83.024	63.349	113.068	85.285
zero	log	9.950	5.479	11.942	6.574	36.121	17.699	71.808	39.550
eff	eff	9.943	5.475	11.908	6.557	33.381	16.354	71.747	39.517
eff.bound	eff.bound	9.943	5.475	11.908	6.557	31.153	15.235	59.236	33.869
		Panel C: $\alpha = 10\%$							
zero	exp	7.268	4.299	7.474	4.227	18.627	11.866	27.838	16.970
zero	log	6.356	3.505	7.182	3.959	13.186	6.970	23.749	13.108
eff	eff	6.350	3.501	7.154	3.945	12.760	6.795	23.706	13.085
eff.bound	eff.bound	6.350	3.501	7.154	3.945	12.473	6.669	19.888	11.173

Summary and Outlook

Summary

- Structural results of conditional and unconditional consistency / identifiability.
- For vector-valued functionals, class of Z-estimators is substantially larger than the class of M-estimators, due to integrability conditions.
- Implication I: M-estimators may fail to be translation equivariant, where Z-estimators are.
- Implication II: Efficiency Gap:
  - Generalisation of the classical Z-estimation efficiency bound into a time series framework.
  - For double quantile regression and VaR–ES regression, there are DGPs where the M-estimator does not reach the Z-estimation efficiency bound.
  - Simulated data support the result and hint and a more pronounced gap for 'extreme' levels of α for VaR–ES regression

Discussion and Outlook

- Generalise separated parameter condition.
- Semiparametric Z-estimation efficiency bound does not necessarily coincide with semiparametric efficiency bound due to Stein (1956).
- Investigate more situation, such as double expectile regression. Suspicion that there is an efficiency gap.
- Examine efficiency considerations for Diebold-Mariano tests as well, which aim at forecast comparison and forecast selection.

Further Reading

- Main reference:

  T. Dimitriadis, T. Fissler, and J. F. Ziegel. The Efficiency Gap.

  Work in progess, 2020

- Z-estimation Efficiency Bound for mean:

  W. K. Newey. Efficient semiparametric estimation via moment restrictions.

  Econometrica, 72(6):1877–1897, 1993

- Z-estimation Efficiency Bound for quantile:

  I. Komunjer and Q. Vuong. Semiparametric efficiency bound in time-series models for conditional quantiles.

  Econometric Theory, 26(02):383–405, 2010

- Elicitability of vector-valued functionals and elicitability of (VaR, ES):

  T. Fissler and J. F. Ziegel. Higher order elicitability and Osband's principle.

  Ann. Statist., 44(4):1680–1707, 2016

Thank you for your attention!