The Efficiency Gap

Timo Dimitriadis [1] Tobias Fissler [2] Johanna F. Ziegel [3]

[1]Graduate School of Decision Sciences, University of Konstanz

[2]Imperial College London, Department of Mathematics

[3]University of Bern, Department of Mathematics and Statistics

4th Konstanz - Lancaster Workshop on Finance and Econometrics 2018

July 31, 2018

Notation Regression Models

- One common probability space: $\big(\Omega, \mathcal{A}, \mathbb{P}\big)$.
- $X_t : \Omega \to \mathbb{R}^p$ with distributions $G_t \in \mathcal{G}$ for all $t = 1, \ldots, T$.
- $Y_t : \Omega \to \mathbb{R}$ such that (X_t, Y_t) has distribution $H_t \in \mathcal{H}$.
- $F_t = F_{Y_t|X_t} \in \mathcal{F}$: conditional distribution of Y_t given X_t in some class of distributions $\mathcal{F}$.
- Assume that the process $\big\{(X_t, Y_t), t = 1, \ldots, T\big\}$ is α-mixing.
- $\Gamma : \mathcal{F} \to \mathbb{R}^k$ is some k-dimensional functional.
- Parametric model $m(X_t, \theta)$ where $m(X_t, \theta_0) = \Gamma(F_t)$ a.s. for all t for some $\theta_0 \in \Theta \subseteq \mathbb{R}^q$.

Motivation: M- and Z-estimation

- Z-/MM-/GMM-estimation based on some identification function $\psi\big(Y_t, X_t, \theta)\big)$:

$$\mathbb{E}\left[\psi\big(Y_t, X_t, \theta\big)\right] = 0 \qquad \Leftrightarrow \qquad \theta = \theta_0. \tag{1}$$

- M-estimation based on some loss function $\rho\big(Y_t, m(X_t, \theta)\big)$:

$$\mathbb{E}\left[\rho\big(Y_t, m(X_t, \theta_0)\big)\right] < \mathbb{E}\left[\rho\big(Y_t, m(X_t, \theta)\big)\right] \tag{2}$$

  for all $\theta \neq \theta_0$.
- Given some regression model $m(X_t, \theta)$, the possible loss and identification functions ρ and ψ are not unique.
- Open Question 1: What is the full class of loss functions for M-estimation of a regression model?
- Open Question 2: What is the full class of identification functions for Z-estimation of a regression model?

Motivation: Newey (1993) Efficiency Bound I

- Some k-dim conditional identification function for the k-dim model $m(X_t, \theta)$:

$$\mathbb{E}\left[\varphi_0\Big(Y_t, m(X_t,\theta)\Big)\,\Big|\, X_t\right] = 0 \text{ a.s.} \qquad \Leftrightarrow \qquad \theta = \theta_0. \tag{3}$$

- We consider the class of estimators, given by some q-dim unconditional identification function induced by *some* $q \times k$ matrix $A(X_t)$:

$$\psi_A\Big(Y_t, X_t, \theta\Big) = A(X_t) \cdot \varphi_0\Big(Y_t, m(X_t,\theta)\Big), \qquad \text{such that} \tag{4}$$

$$\mathbb{E}\left[\psi_A\Big(Y_t, X_t, \theta\Big)\right] = 0 \qquad \Leftrightarrow \qquad \theta = \theta_0, \tag{5}$$

- Given that some regularity conditions (Newey and McFadden, 1994) are satisfied,

$$\Delta_{T,A}\Sigma_{T,A}^{-1/2}\sqrt{T}\Big(\hat{\theta}_{T,A} - \theta_0\Big) \stackrel{d}{\to} \mathcal{N}(0, I_q), \tag{6}$$

where

$$\Delta_{T,A} = 1/T \sum \mathbb{E}\left[A(X_t)D(X_t)\right] \tag{7}$$

$$\Sigma_{T,A} = 1/T \sum \mathbb{E}\left[A(X_t)S(X_t)A(X_t)^\top\right] \tag{8}$$

$$S(X_t) = \mathbb{E}\left[\varphi_0\Big(Y_t, m(X_t,\theta_0)\Big)\cdot \varphi_0\Big(Y_t, m(X_t,\theta_0)\Big)^\top \,\Big|\, X_t\right] \tag{9}$$

$$D(X_t) = \nabla_\theta \mathbb{E}\left[\varphi_0\Big(Y_t, m(X_t,\theta_0)\Big)\,\Big|\, X_t\right] \tag{10}$$

Motivation: Newey (1993) Efficiency Bound II

- Efficiency bound is reached by $A_C^*(X_t) = C \cdot D(X_t)^\top \cdot S(X_t)^{-1}$, where C is non-singular and deterministic and

$$S(X_t) = \mathbb{E}\left[\varphi_0\Big(Y_t, m(X_t, \theta_0)\Big) \cdot \varphi_0\Big(Y_t, m(X_t, \theta_0)\Big)^\top \Big| X_t\right] \quad (11)$$

$$D(X_t) = \nabla_\theta \mathbb{E}\left[\varphi_0\Big(Y_t, m(X_t, \theta_0)\Big) \Big| X_t\right] \quad (12)$$

- Then $\Delta_{T,A^*} = \Sigma_{T,A^*} =: \Lambda_T$ and

$$\Lambda_T^{1/2}\sqrt{T}\Big(\hat\theta_{T,A_C^*} - \theta_0\Big) \xrightarrow{d} \mathcal{N}(0, I_q), \quad (13)$$

- Open Question 3: Can this efficiency bound also be attained by some other $A(X_t) \neq A_C^*(X_t)$?
- Open Question 4: Can this efficiency bound also be attained by some M-estimator?

Loss and Identification Functions

In the classical elicitability framework, Y is stochastic and the functional Γ is deterministic:

- A function $\rho(Y,x)$ is called a strictly consistent loss function for the functional Γ with respect to the distributions $F \in \mathcal{F}$ (of Y), when

$$\mathbb{E}\left[\rho(Y,\Gamma(F))\right] < \mathbb{E}\left[\rho(Y,x)\right] \qquad \forall x \in \mathbb{R}. \tag{14}$$

- A function $\psi(Y,x)$ is called a strict identification function for the functional Γ with respect to the distributions $F \in \mathcal{F}$ (of Y), when

$$\mathbb{E}\left[\rho(Y,x)\right] = 0 \qquad \Leftrightarrow \qquad x \in \Gamma(F) \tag{15}$$

Strictly (Un)conditionally Consistent Loss Functions I

For regressions with stochastic regressors, the functional $\Gamma(F_{Y_t|X_t}) = m(X_t, \theta_0)$ is stochastic. Now, we make the more general definitions:

Definition 1

- A loss function $\rho\big(Y_t, m(X_t, \theta)\big)$ is called **strictly conditionally consistent** for the functional Γ with respect to the distributions $H_t \in \mathcal{H}$, if

$$\mathbb{E}\left[\rho\big(Y_t, m(X_t, \theta_0)\big) \,\middle|\, X_t\right] < \mathbb{E}\left[\rho\big(Y_t, m(X_t, \theta)\big) \,\middle|\, X_t\right] \; a.s. \qquad \forall \theta \neq \theta_0. \tag{16}$$

- A loss function $\rho\big(Y_t, m(X_t, \theta)\big)$ is called **strictly unconditionally consistent** for the functional Γ with respect to the distributions $H_t \in \mathcal{H}$, if

$$\mathbb{E}\left[\rho\big(Y_t, m(X_t, \theta_0)\big)\right] < \mathbb{E}\left[\rho\big(Y_t, m(X_t, \theta)\big)\right] \qquad \forall \theta \neq \theta_0. \tag{17}$$

Strictly (Un)conditionally Consistent Loss Functions II

Theorem 2

- *It holds that (**strict conditional consistency**) $\Rightarrow$ (**strict unconditional consistency**)*
- *Given Assumption 1, it holds that (**strict unconditional consistency**) $\Rightarrow$ (**strict conditional consistency**)*

- Assumption 1 ensures that the class $\mathcal{G}$ is large enough.

Technical Condition

Assumption 1

For all $(X_t, Y_t) \sim H_t \in \mathcal{H}$*,* $X_t \sim G_t \in \mathcal{G}$*, and for all measurable functions* $a : \mathbb{R}^p \to \mathbb{R}$ *with* $\operatorname{supp}(G) \cap a^{-1}\big((-\infty, 0]\big) \neq \emptyset$*, there exists a pair* $(\tilde{X}_t, \tilde{Y}_t) \sim \tilde{H}_t \in \mathcal{H}$*,* $\tilde{X}_t \sim \tilde{G}_t$ *such that*

- $F_{\tilde{Y}_t|\tilde{X}_t=x} = F_{Y_t|X_t=x}$ *for all* $x \in \operatorname{supp}(\tilde{G}_t)$*,*
- *and* $\operatorname{supp}(\tilde{G}_t) \subseteq \operatorname{supp}(G_t) \cap a^{-1}\big((-\infty, 0]\big)$.

- We can change $G_t \in \mathcal{G}$ such that there is zero mass on the region where a is positive.
- This basically means that the class $\mathcal{G}$ is large enough!
- In line with the logic of strict consistency: $\mathcal{F}$ (class of distributions of Y_t given X_t) must be large enough.
- This is reasonable: we want to have a consistent estimator for a large class of processes.

Strictly (Un)-conditional Identification Functions I

$$\varphi_0 : \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}^k \qquad (Y_t, m(X_t, \theta)) \mapsto \varphi_0(Y_t, m(X_t, \theta)) \tag{18}$$

$$\psi_A : \mathbb{R} \times \mathbb{R}^p \times \Theta \to \mathbb{R}^q \qquad (Y_t, X_t, \theta) \mapsto \psi_A(Y_t, X_t, \theta) \tag{19}$$

Definition 3

- A function φ_0 is called a **strictly conditional identification function** for the functional Γ with respect to the distributions $H_t \in \mathcal{H}$, if

$$\mathbb{E}[\varphi_0(Y_t, m(X_t, \theta_0))|X_t] = 0 \quad a.s. \quad \text{and} \tag{20}$$

$$\Big\{\forall \theta \in \Theta : \ \big[(\mathbb{E}[\varphi_0(Y_t, m(X_t, \theta))|X_t] = 0 \quad a.s.) \implies \theta = \theta_0\big]\Big\} \tag{21}$$

- A function ψ_A is called a **strictly unconditional identification function** for the functional Γ with respect to the distributions $H_t \in \mathcal{H}$, if

$$\mathbb{E}[\psi_A(Y_t, X_t, \theta_0)] = 0 \qquad \text{and} \tag{22}$$

$$\{\forall \theta \in \Theta : \ \mathbb{E}[\psi_A(Y_t, X_t, \theta)] = 0 \implies \theta = \theta_0\}\,. \tag{23}$$

Strictly (Un)conditional Identification Functions II

Theorem 4

- *Given Assumption 1 and given that*

$$\mathbb{P}(A(X_t) \text{ has full rank}) = 1 \tag{24}$$

*for all $t = 1, \ldots, T$, it holds that (**strict unconditional identification**) $\Rightarrow$ (**strict conditional identification**)*

- *Given that*

$$\mathbb{E}\left[A(X_t)\nabla\mathbb{E}_t\left[\varphi_0(Y_t, X_t, \tilde{\theta})\right]\right] \text{ has full rank } \forall\theta, \tilde{\theta} \in \Theta, \tag{25}$$

*and for all $t = 1, \ldots, T$, it holds that (**strict conditional identification**) $\Rightarrow$ (**strict unconditional identification**)*

Loss and Identification Functions

Theorem 5

Assume that $\rho(Y_t, m(X_t, \theta))$ is smooth, and strictly conditionally consistent for some k-dim functional Γ and has no saddle points. Then, $\varphi_0(Y_t, m(X_t, \theta)) = \nabla_m \rho(Y_t, m(X_t, \theta))$ is a strict conditional identification function for Γ.

1-dim functionals: $\Gamma(F_t) \in \mathbb{R}$

- $\varphi_0(Y_t, m(X_t, \theta))$ strict cond id function. Then, $\rho(Y_t, m(X_t, \theta)) = \int_{m_0}^{m(X_t,\theta)} \varphi_0(Y_t, t)\mathrm{d}t + c(Y_t)$ is a strictly consistent loss function.
- Reason: Fundamental theorem of calculus.
- $\{\text{loss fct}\} = \{\text{identification fct}\}$

Multiv. functionals: $\Gamma(F_t) \in \mathbb{R}^k$

- The inverse direction does not hold in general!
- For multivariate functionals, we need some *integrability conditions* to hold in order to get this result, Königsberger (2004), p.184.
- $\{\text{loss fct}\} \subseteq \{\text{identification fct}\}$

Conclusions of this Section

We have learned that (informal notation):

$$
\begin{aligned}
&\textbf{(Z-estimator)} \\
&= (\textbf{strict unconditional identification}) \\
&\supseteq (\textbf{strict conditional identification}) \\
&\supseteq (\textbf{strict conditional consistency}) \\
&\supseteq (\textbf{strict unconditional consistency}) \\
&= \textbf{(M-estimator)}
\end{aligned}
$$

- $\supseteq$: given Assumption 1
- $\supseteq$: given that the loss ρ does not have any saddle points
- $\supseteq$: given that $\mathbb{E}\left[A(X_t)\nabla\mathbb{E}_t\left[\varphi_0(Y_t, X_t, \tilde{\theta})\right]\right]$ has full rank $\forall\theta, \tilde{\theta} \in \Theta$

Strict Efficiency Bounds

- Recall the Z-estimator based on $A(X_t)$:

$$\psi_A\Big(Y_t, X_t, \theta\Big) = A(X_t) \cdot \varphi_0\Big(Y_t, m(X_t, \theta)\Big), \qquad \text{such that} \tag{26}$$

$$\mathbb{E}\Big[\psi_A\Big(Y_t, X_t, \theta\Big)\Big] = 0 \qquad \Leftrightarrow \qquad \theta = \theta_0, \tag{27}$$

- *Any* Z-estimator: $\Delta_{T,A}\Sigma_{T,A}^{-1/2}\sqrt{T}\Big(\hat{\theta}_{T,A} - \theta_0\Big) \xrightarrow{d} \mathcal{N}(0, I_q)$
- *Efficient* Z-estimator: $\Lambda_T^{1/2}\sqrt{T}\Big(\hat{\theta}_{T,A_C^*} - \theta_0\Big) \xrightarrow{d} \mathcal{N}(0, I_q)$

Theorem 6

- *(Newey (1993)) For all $q \times k$ matrices $A(X_t)$, it holds that $\Delta_{T,A}^{-1}\Sigma_{T,A}\Delta_{T,A}^{-1} - \Lambda_T^{-1}$ is positive semi-definite.*
- *If $A(X_t) \neq A_C^*(X_t)$ with positive probability for some t and for all deterministic and non-singular matrices C, then the matrix*

$$\Delta_{T,A}^{-1}\Sigma_{T,A}\Delta_{T,A}^{-1} - \Lambda_T^{-1} \tag{28}$$

is positive semi-definite with at least one strictly positive Eigenvalue.

A Regression for the first two moments I

- We jointly model the first two moments by the joint model

$$m(X_t,\theta) = \big(m_1(X_t,\theta^{(1)}), \quad m_2(X_t,\theta^{(2)})\big)\,. \tag{29}$$

- Multivariate Bregman type class of loss functions:

$$\rho\big(Y_t, m_1(X_t,\theta^{(1)}), m_2(X_t,\theta^{(2)})\big) \tag{30}$$

$$= -\,\phi_t\big(m(X_t,\theta)\big) + \nabla\phi_t\big(m(X_t,\theta)\big)\cdot\begin{pmatrix} m_1(X_t,\theta^{(1)}) - Y_t \\ m_2(X_t,\theta^{(2)}) - Y_t^2\end{pmatrix}, \tag{31}$$

where $\phi_t : \mathbb{R}^2 \to \mathbb{R}$ are strictly convex functions.

- The *associated* identification functions are given by

$$\psi_\phi\big(Y_t, m_1(X_t,\theta^{(1)}), m_2(X_t,\theta^{(2)})\big) \tag{32}$$

$$= \begin{pmatrix} \nabla_\theta m_1(X_t,\theta^{(1)}) \\ \nabla_\theta m_2(X_t,\theta^{(2)})\end{pmatrix}^\top \cdot H_\phi(m(X_t,\theta)) \cdot \begin{pmatrix} m_1(X_t,\theta^{(1)}) - Y_t \\ m_2(X_t,\theta^{(2)}) - Y_t^2\end{pmatrix} \tag{33}$$

$$= A_\phi(X_t)\begin{pmatrix} m_1(X_t,\theta^{(1)}) - Y_t \\ m_2(X_t,\theta^{(2)}) - Y_t^2\end{pmatrix}. \tag{34}$$

A Regression for the first two moments II

- Thus

$$A_\phi(X_t) = \begin{pmatrix} \nabla_\theta m_1(X_t, \theta^{(1)}) & 0 \\ 0 & \nabla_\theta m_2(X_t, \theta^{(2)}) \end{pmatrix}^\top \cdot H_\phi(m(X_t, \theta)) \tag{35}$$

- Efficient choice:

$$A^*_C(X_t) = C \cdot \begin{pmatrix} \nabla_\theta m_1(X_t, \theta_0^{(1)}) & 0 \\ 0 & \nabla_\theta m_2(X_t, \theta_0^{(2)}) \end{pmatrix}^\top \cdot \mathrm{Var}\left(\begin{pmatrix} Y_t \\ Y_t^2 \end{pmatrix} \middle| X_t \right)^{-1}. \tag{36}$$

- Thus $C = I_q$ and

$$H_\phi(z) = \mathrm{Var}\left(\begin{pmatrix} Y_t \\ Y_t^2 \end{pmatrix} \middle| m(X_t, \theta) = z \right)^{-1}. \tag{37}$$

- This can e.g. be realized by using the quadratic form

$$\phi_t(z) = z^\top \mathrm{Var}\left(\begin{pmatrix} Y_t \\ Y_t^2 \end{pmatrix} \middle| m(X_t, \theta) = z \right)^{-1} z. \tag{38}$$

Quantile Regression

- Generalized Piecewise Linear (GPL) class of loss functions:

$$\rho\big(Y_t, q_\alpha(X_t,\theta)\big) = \big(\mathbb{1}_{\{Y_t \le q_\alpha(X_t,\theta)\}} - \alpha\big) g_t(q_\alpha(X_t,\theta)) - \mathbb{1}_{\{Y_t \le q_\alpha(X_t,\theta)\}} g_t(Y_t) + a(Y_t)$$

  where g_t are strictly increasing functions.

- The *associated* identification functions are given by

$$\psi_g(Y_t, X_t, \theta) = \nabla_\theta q_\alpha(X_t,\theta) g_t'(q_\alpha(X_t,\theta)) \big(\mathbb{1}_{\{Y_t \le q_\alpha(X_t,\theta)\}} - \alpha\big) \quad (39)$$

$$= A(X_t) \big(\mathbb{1}_{\{Y_t \le q_\alpha(X_t,\theta)\}} - \alpha\big). \quad (40)$$

- Efficient choice (Z-estimator):

$$A_C^*(X_t) = C \cdot \frac{1}{\alpha(1-\alpha)} \nabla_\theta q_\alpha(X_t,\theta_0) f_t\big(q_\alpha(X_t,\theta_0)\big) \quad (41)$$

- Efficient choice (M-estimator):

$$g_t(z) = F_t(z) \qquad \text{and thus} \qquad g_t'\big(q_\alpha(X_t,\theta_0)\big) = f_t\big(q_\alpha(X_t,\theta_0)\big). \quad (42)$$

A Double Quantile Regression I

- We jointly model the $\alpha, \beta \in (0,1)$, $\alpha < \beta$ quantiles through the joint model
$$m(X_t, \theta) = \begin{pmatrix} q_\alpha(X_t, \theta^\alpha) & q_\beta(X_t, \theta^\beta) \end{pmatrix}. \tag{43}$$
- Multivariate GPL class of loss functions:
$$\begin{aligned} &\rho\big(Y_t, q_\alpha(X_t, \theta^\alpha), q_\beta(X_t, \theta^\beta)\big) \\ &= \big(\mathbb{1}_{\{Y_t \le q_\alpha(X_t,\theta^\alpha)\}} - \alpha\big) g_{1,t}(q_\alpha(X_t, \theta^\alpha)) - \mathbb{1}_{\{Y_t \le q_\alpha(X_t,\theta^\alpha)\}} g_{1,t}(Y_t) \\ &+ \big(\mathbb{1}_{\{Y_t \le q_\beta(X_t,\theta^\beta)\}} - \beta\big) g_{2,t}(q_\beta(X_t, \theta^\beta)) - \mathbb{1}_{\{Y_t \le q_\beta(X_t,\theta^\beta)\}} g_{2,t}(Y_t) + a(Y_t) \end{aligned}$$
where $g_{1,t}$ and $g_{2,t}$ are strictly increasing and smooth functions.
- The *associated* identification functions are given by
$$\begin{aligned} &\psi_{g_1,g_2}\big(Y_t, q_\alpha(X_t, \theta^\alpha), q_\beta(X_t, \theta^\beta)\big) \\ &= \nabla q_\alpha(X_t, \theta^\alpha) g'_{1,t}(q_\alpha(X_t, \theta^\alpha)) \big(\mathbb{1}_{\{Y_t \le q_\alpha(X_t,\theta^\alpha)\}} - \alpha\big) \\ &+ \nabla q_\beta(X_t, \theta^\beta) g'_{2,t}(q_\beta(X_t, \theta^\beta)) \big(\mathbb{1}_{\{Y_t \le q_\beta(X_t,\theta^\beta)\}} - \beta\big) \\ &= A_{g_1,g_2}(X_t) \begin{pmatrix} \mathbb{1}_{\{Y_t \le q_\alpha(X_t,\theta^\alpha)\}} - \alpha \\ \mathbb{1}_{\{Y_t \le q_\beta(X_t,\theta^\beta)\}} - \beta \end{pmatrix}. \end{aligned}$$

A Double Quantile Regression II

- Thus

$$A_{g_1,g_2}(X_t,\theta_0) = \begin{pmatrix} \nabla_\theta q_\alpha(X_t,\theta_0^\alpha)^\top g'_{1,t}\Big(q_\alpha(X_t,\theta_0^\alpha)\Big) & 0 \\ 0 & \nabla_\theta q_\beta(X_t,\theta_0^\beta)^\top g'_{2,t}\Big(q_\beta(X_t,\theta_0^\beta)\Big) \end{pmatrix}$$

- Efficient choice:

$$A^*_C(X_t,\theta_0) = C \cdot \begin{pmatrix} \nabla_\theta q_\alpha(X_t,\theta_0^\alpha)^\top f_t\Big(q_\alpha(X_t,\theta_0^\alpha)\Big) & 0 \\ 0 & \nabla_\theta q_\beta(X_t,\theta_0^\beta)^\top f_t\Big(q_\beta(X_t,\theta_0^\beta)\Big) \end{pmatrix} \cdot \begin{pmatrix} \alpha(1-\alpha) & \alpha(1-\beta) \\ \alpha(1-\beta) & \beta(1-\beta) \end{pmatrix}^{-1}$$

- This looks like $g_{1,t}(\cdot) = g_{2,t}(\cdot) = f_t(\cdot)$ would attain the efficiency bound.

A Double Quantile Regression III

Theorem 7

Assume that

(DQR1) *the support of the pushforward measure of* $\nabla q_\alpha(X_t, \theta_0^\alpha)$ *contains at least* $k_1 + 1$ *different values* $v_1, \ldots, v_{k_1+1}$*, such that any subset of cardinality* k_1 *of* $\{v_1, \ldots, v_{k_1+1}\}$ *is linearly independent. Equivalently, the support of the pushforward measure of* $\nabla q_\beta(X_t, \theta_0^\beta)$ *contains at least* $k_2 + 1$ *such values.*

(DQR2) *The ratio* $\frac{f_t(q_\alpha(X_t, \theta_0^\alpha))}{f_t(q_\beta(X_t, \theta_0^\beta))}$ *is not constant almost surely.*

Then, $A_C^*(X_t) \neq A_{g_1,g_2}(X_t)$ *with positive probability for some* $t = 1, \ldots, T$*. Thus, the M-estimator cannot attain the efficiency bound theoretically.*

A Joint Regression for the quantile and ES I

- We jointly model the α-quantile and α-ES through the joint model

$$m(X_t, \theta) = \Big(q_\alpha(X_t, \theta^q), \quad e_\alpha(X_t, \theta^e) \Big) . \tag{44}$$

- FZ class of loss functions:

$$\begin{aligned} &\rho\Big(Y_t, q_\alpha(X_t, \theta^q), e_\alpha(X_t, \theta^e)\Big) \\ &= \Big(\mathbb{1}_{\{Y_t \le q_\alpha(X_t, \theta^q)\}} - \alpha\Big) g_t(q_\alpha(X_t, \theta^q)) - \mathbb{1}_{\{Y_t \le q_\alpha(X_t, \theta^e)\}} g_t(Y_t) \\ &\quad + \phi_t'(e_\alpha(X_t, \theta^e)) \left(e_\alpha(X_t, \theta^e) - q_\alpha(X_t, \theta^q) + \frac{(q_\alpha(X_t, \theta^q) - Y_t)\mathbb{1}_{\{Y_t \le q_\alpha(X_t, \theta^q)\}}}{\alpha} \right) \\ &\quad - \phi_t(e_\alpha(X_t, \theta^e)) + a(Y_t), \end{aligned}$$

where g_t are strictly increasing and ϕ_t strictly increasing and strictly convex.

- The *associated* identification functions are given by

$$\begin{aligned} &\psi_{g,\phi}\Big(Y_t, q_\alpha(X_t, \theta^q), e_\alpha(X_t, \theta^e)\Big) \\ &= \nabla q_\alpha(X_t, \theta^q) \left(g_t'(q_\alpha(X_t, \theta^q)) + \frac{\phi_t'(e_\alpha(X_t, \theta^e))}{\alpha} \right) \Big(\mathbb{1}_{\{Y_t \le q_\alpha(X_t, \theta^q)\}} - \alpha\Big) \\ &+ \nabla e_\alpha(X_t, \theta^e) \phi_t''(e_\alpha(X_t, \theta^e)) \left(e_\alpha(X_t, \theta^e) - q_\alpha(X_t, \theta^q) + \frac{(q_\alpha(X_t, \theta^q) - Y_t)\mathbb{1}_{\{Y_t \le q_\alpha(X_t, \theta^q)\}}}{\alpha} \right) \end{aligned}$$

A Joint Regression for the quantile and ES II

- Efficient choice $A^*_C(X_t) = C \cdot D(X_t)^\top \cdot S(X_t)^{-1}$, where

$$D(X_t) = \begin{pmatrix} \nabla q_\alpha(X_t, \theta_0^q) f_t(q_\alpha(X_t, \theta_0)^q) & 0 \\ 0 & \nabla e_\alpha(X_t, \theta_0^e) \end{pmatrix} \quad \text{and} \tag{45}$$

$$S(X_t) = \begin{pmatrix} \alpha(1-\alpha) & (1-\alpha)\Big(q_\alpha(X_t, \theta_0) - e_\alpha(X_t, \theta_0)\Big) \\ (1-\alpha)\Big(q_\alpha(X_t, \theta_0) - e_\alpha(X_t, \theta_0)\Big) & S_{22} \end{pmatrix}, \tag{46}$$

$$S_{22} = \frac{1}{\alpha} \operatorname{Var}_t \left(Y_t - q_\alpha(X_t, \theta_0^q) \,\middle|\, Y_t \le q_\alpha(X_t, \theta_0^q)\right) + \frac{1-\alpha}{\alpha}\Big(e_\alpha(X_t, \theta_0^e) - q_\alpha(X_t, \theta_0^q)\Big)^2 \tag{47}$$

Theorem 8

Assume that

(QESR1) *the support of the pushforward measure of* $\nabla q_\alpha(X_t, \theta_0^q)$ *contains at least* $k_1 + 1$ *different values* $v_1, \dots, v_{k_1+1}$*, such that any subset of cardinality* k_1 *of* $\{v_1, \dots, v_{k_1+1}\}$ *is linearly independent. Equivalently, the support of the pushforward measure of* $\nabla e_\alpha(X_t, \theta_0^e)$ *contains at least* $k_2 + 1$ *such values.*

(QESR2) *At least one of the following equalities does not hold almost surely:*

$$\mathbb{E}_t\left[\left(q_\alpha(X_t, \theta_0^q) - Y_t\right)^2 \mathbb{1}_{\{Y_t \le q_\alpha(X_t, \theta_0^q)\}}\right] = \left(\frac{1-\alpha}{\alpha} - \alpha^2\right)\left(q_\alpha(X_t, \theta_0^q) - e_\alpha(X_t, \theta_0^e)\right)^2$$

$$\frac{1}{\alpha^2}\mathbb{E}_t\left[\left(q_\alpha(X_t, \theta_0^q) - Y_t\right)^2 \mathbb{1}_{\{Y_t \le q_\alpha(X_t, \theta_0^q)\}}\right] + \left(e_\alpha(X_t, \theta_0^e) - q_\alpha(X_t, \theta_0^q)\right)^2$$

$$= \phi_t''(e_\alpha(X_t, \theta_0^q))^{-1}$$

$$(1-\alpha)\left(\alpha g_t'(q_\alpha(X_t, \theta_0^q)) + \phi_t'(e_\alpha(X_t, \theta_0^e))\right) = f_t(q_\alpha(X_t, \theta_0^q))$$

Then, $A_C^*(X_t) \neq A_{g,\phi}(X_t)$ *with positive probability for some* $t = 1, \dots, T$*. Thus, the M-estimator cannot attain the efficiency bound theoretically.*

DQR Simulation Setup, iid errors

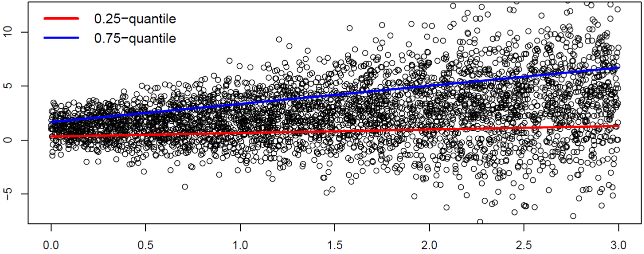
$$X_t \sim \Big(1, U[0,3]\Big) \qquad u_t \sim \mathcal{N}(0,1), \tag{48}$$

$$Y_t = X_t^\top \gamma_0 + (X_t^\top \eta_0) u_t \qquad (\gamma_0, \eta_0) = (1,1,1,1) \tag{49}$$

$$q_\alpha(X_t, \theta_0^\alpha) = X_t^\top \Big(\gamma_0 + \eta_0 z_\alpha\Big) \qquad q_\beta(X_t, \theta_0^\beta) = X_t^\top \Big(\gamma_0 + \eta_0 z_\beta\Big) \tag{50}$$

- We need that $\frac{f_t(q_\alpha(X_t,\theta_0^\alpha))}{f_t(q_\beta(X_t,\theta_0^\beta))}$ is not deterministic
- In location-scale models with $u_t \sim iid$:

$$\frac{f_t(q_\alpha(X_t, \theta_0^\alpha))}{f_t(q_\beta(X_t, \theta_0^\beta))} = \frac{f_{u_t}(z_\alpha)}{f_{u_t}(z_\beta)} = \frac{f_u(z_\alpha)}{f_u(z_\beta)} = \text{constant} \tag{51}$$



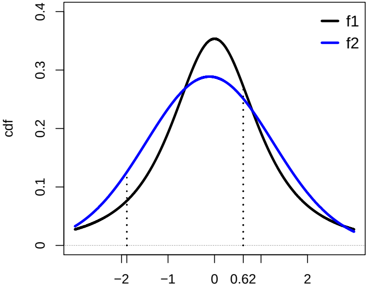
DQR Simulation Setup, non-iid errors

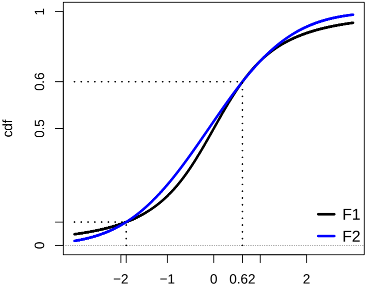
$$u_t \sim t_{\nu_t}(\mu_t, \sigma_t), \qquad \nu_t = 2 + t/T(100-2) \tag{52}$$

$$\sigma_t = \frac{Q_\alpha(t_{\nu_0}) - Q_\beta(t_{\nu_0})}{Q_\alpha(t_\nu) - Q_\beta(t_\nu)} \qquad \mu_t = Q_\beta(t_{\nu_0}) - \sigma_\nu Q_\beta(t_\nu) \tag{53}$$

$$\implies q_{0.1}(u_t) = -1.88 \qquad q_{0.6}(u_t) = 0.62 \tag{54}$$

$$\implies \frac{f_1(q_{0.1}(u_1))}{f_1(q_{0.6}(u_1))} = 0.28 \qquad \frac{f_T(q_{0.1}(u_T))}{f_T(q_{0.62}(u_T))} = 0.50 \tag{55}$$





DQR Simulation Results, non-iid errors

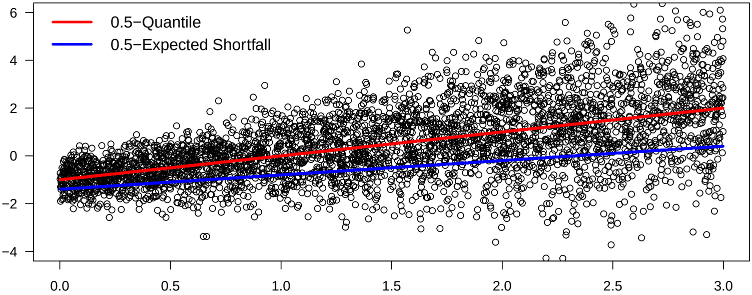
Estimation Method	$u_t \sim t_\nu$			
	θ_1	θ_2	θ_3	θ_4
	True Asymptotic SD			
M-est	2.4351	2.1294	5.4898	4.7882
Z-est	2.4351	2.1294	5.4898	4.7882
M-est eff.	2.1829	1.9124	5.1889	4.5350
Z-est p.eff	2.1829	1.9124	5.1889	4.5350
Z-est	2.1694	1.9005	5.1569	4.5070
	Estimated Asymptotic SD			
M-est	2.4344	2.1288	5.4806	4.7906
Z-est	2.4344	2.1288	5.4806	4.7906
M-est eff.	2.1829	1.9125	5.1794	4.5367
Z-est p.eff	2.1833	1.9127	5.1804	4.5370
Z-est	2.1693	1.9006	5.1474	4.5085
	Empirical SD			
M-est	2.3575	2.0358	5.3285	4.6281
Z-est	2.3559	2.0347	5.3201	4.6231
M-est eff.	2.1002	1.8103	5.0785	4.4149
Z-est p.eff	2.1022	1.8064	5.1095	4.4414
Z-est	2.1271	1.8247	5.2245	4.5301

QESR Simulation Setup, iid errors

$$X_t \sim \big(1, U[0,3]\big) \qquad u_t \sim \mathcal{N}(0,1), \tag{56}$$

$$Y_t = X_t^\top \gamma_0 + (X_t^\top \eta_0) u_t \qquad (\gamma_0, \eta_0) = (-1, 1, 0.5, 0.5) \tag{57}$$

$$q_\alpha(X_t, \theta_0^q) = X_t^\top \big(\gamma_0 + \eta_0 z_\alpha\big) \qquad e_\alpha(X_t, \theta_0^e) = X_t^\top \big(\gamma_0 + \eta_0 \xi_\alpha\big) \tag{58}$$



QESR, Simulation Results:

	g	ϕ'	True Asymptotic SD			
M-est	$g(z) = z$	$\phi'(z) = F_{log}(z)$	5.9924	4.5612	5.5715	6.4873
Z-est	$g(z) = z$	$\phi'(z) = F_{log}(z)$	5.9924	4.5612	5.5715	6.4873
M-est	$g(z) = F_t(z)$	$\phi'(z) = F_{log}(z)$	5.4118	4.2179	5.5715	6.4873
Z-est	$g(z) = z$	$\phi'(z) = -1/z$	5.7265	4.4176	5.0074	3.8683
Z-est eff			5.3484	4.1389	4.9828	3.8560

	g	ϕ'	Estimated Asymptotic SD			
M-est	$g(z) = z$	$\phi'(z) = F_{log}(z)$	6.0181	4.5629	5.5894	6.5037
Z-est	$g(z) = z$	$\phi'(z) = F_{log}(z)$	6.0181	4.5630	5.5895	6.5023
M-est	$g(z) = F_t(z)$	$\phi'(z) = F_{log}(z)$	5.4352	4.2195	5.5837	6.5060
Z-est	$g(z) = z$	$\phi'(z) = -1/z$	5.7509	4.4186	5.0422	3.8707
Z-est eff			5.3753	4.1420	5.0020	3.8554

	g	ϕ'	Empirical SD			
M-est	$g(z) = z$	$\phi'(z) = F_{log}(z)$	5.7764	4.5348	5.4280	6.3952
Z-est	$g(z) = z$	$\phi'(z) = F_{log}(z)$	5.7875	4.5308	5.4273	6.3953
M-est	$g(z) = F_t(z)$	$\phi'(z) = F_{log}(z)$	5.2376	4.1857	5.4237	6.4019
Z-est	$g(z) = z$	$\phi'(z) = -1/z$	6.1330	4.6506	9.5877	5.6558
Z-est eff			5.6945	4.3757	5.2574	4.1161

Conclusions

- For univariate models: (**M-Estimator**)$\widehat{=}$(**Z-Estimator**).
- However, for multivariate models: (**M-Estimator**) $\subseteq$ (**Z-Estimator**)
- For univariate models, the efficiency bound can be reached by both, the M- and Z-estimator.
- For multivariate models, there are examples where the efficiency bound **cannot** be reached by the M-estimator.
  - This depends on the richness of the class of strictly consistent loss functions.

References I

Königsberger, K. (2004). *Analysis 2*. Springer-Verlag, Berlin Heidelberg New York, 5th edition.

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier.

Newey, W. K. (1993). 16 efficient estimation of models with conditional moment restrictions. In *Econometrics*, volume 11 of *Handbook of Statistics*, pages 419 – 454. Elsevier.

Mean Regression

- Bregman class of loss functions:

$$\rho\big(Y, m_1(X,\theta)\big) = -\phi(m_1(X,\theta)) + \phi'(m_1(X,\theta))\big(m_1(X,\theta) - Y\big) + a(Y), \tag{59}$$

  where ϕ is a strictly convex function.
- The *associated* identification functions are given by

$$\psi(Y, X, \theta) = \nabla_\theta m_1(X,\theta)\phi''(m_1(X,\theta))\big(m_1(X,\theta) - Y\big) \tag{60}$$

$$= A(X,\theta)\big(m_1(X,\theta) - Y\big). \tag{61}$$

- Efficient choice:

$$A_C^*(X,\theta_0) = C \cdot \nabla_\theta m_1(X,\theta_0)\frac{1}{\mathrm{Var}(Y|X)} \tag{62}$$

  and thus $C = I$ and

$$\phi''(z) = \frac{1}{\mathrm{Var}\left(Y \middle| m_1(X,\theta) = z\right)} \tag{63}$$

- Here, we need the additional condition that $\mathrm{Var}(Y|X)$ is almost surely uniquely determined by the value of $m_1(X,\theta_0) = \mathbb{E}[Y|X]$.