# Semiparametric estimation, finite sample theory

Andreas Andresen

22.1.2011

## general setting

Let $\mathbb{Y} \sim I\!\!P$ for some probability measure $I\!\!P$ .

We choose the model $(I\!\!P_{\boldsymbol{\theta},\boldsymbol{\eta}})$ , where $(\boldsymbol{\theta},\boldsymbol{\eta}) \in \varTheta \subset I\!\!R^{p} \times I\!\!R^{p_1}$ .

$\boldsymbol{\theta}$ target, $\boldsymbol{\eta}$ nuisance.

Goal: inference on $\boldsymbol{\theta}$

Example in mind:

- an inverse problem with error in operator;
observed $\mathbb{Y}$ and $A_{\varepsilon_A}$ following $\mathbb{Y} = A_0\boldsymbol{\theta}_0 + \varepsilon_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta}_0$ and $A_0$ unknown;

## general setting

Model assumption (SPA):

$$\boldsymbol{Y} \sim I\!\!P \in \left( I\!\!P_{\boldsymbol{\theta},\boldsymbol{\eta}},\ \boldsymbol{\theta} \in \Theta, \boldsymbol{\eta} \in \Upsilon \right)$$

Log-likelihood:

$$L(\boldsymbol{\upsilon}) = \frac{dI\!\!P_{\boldsymbol{\theta},\boldsymbol{\eta}}}{d\boldsymbol{\mu}_0}(\boldsymbol{Y})$$

Profile MLE:

$$\widetilde{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \max_{\boldsymbol{\eta}} L(\boldsymbol{\theta}, \boldsymbol{\eta}) = \operatorname*{argmax}_{\boldsymbol{\theta}} \widetilde{L}(\boldsymbol{\theta})$$

$$\widetilde{L}(\boldsymbol{\theta}) = \max_{\boldsymbol{\eta}} L(\boldsymbol{\theta}, \boldsymbol{\eta})$$

## asymptotical results

Murphy, van der Vaart (2000), Kosorok (2005, 2008): Let $I\!P = I\!P_{\boldsymbol{\theta}^*, \boldsymbol{\eta}^*}$ . Then $\widetilde{\boldsymbol{\theta}}$ is

- root- $n$ consistent and normal
- semiparametrically efficient
- $2\widetilde{L}(\widetilde{\boldsymbol{\theta}}) - 2\widetilde{L}(\boldsymbol{\theta}^*) \xrightarrow{w} \chi^2_p$ , where $p = \dim(\Theta)$ .

Limitations:

- hard optimization problem, often unfeasible
- SPA is crucial but questionable
- large sample asymptotics

## linear semiparametric models: adaptivity

Consider the model:

$$\boldsymbol{Y} = \Psi^\top \boldsymbol{\theta}^* + \Phi^\top \boldsymbol{\eta}^* + \boldsymbol{\varepsilon},$$
$$I\!\!E\boldsymbol{\varepsilon} = 0, \ \operatorname{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n.$$

Trying to solve this we observe

$$\widetilde{\boldsymbol{\theta}} = \begin{pmatrix} I_p & 0 \end{pmatrix} \begin{pmatrix} \Psi\Psi^\top & \Psi\Phi^\top \\ \Phi\Psi^\top & \Phi\Phi^\top \end{pmatrix}^{-1} \begin{pmatrix} \Psi \\ \Phi \end{pmatrix} \boldsymbol{Y}$$

So if $\Psi\Phi^\top = 0$ the profile reads $\widetilde{\boldsymbol{\theta}} = (\Psi\Psi^\top)^{-1}\Psi\boldsymbol{Y}$ .

We set $\boldsymbol{\eta}^* = \widetilde{\boldsymbol{\eta}^*} - (\Phi\Phi^T)^{-1}\Phi\Psi^T\boldsymbol{\theta}^*$ then

$$\Psi^\top \boldsymbol{\theta}^* + \Phi^\top \boldsymbol{\eta}^* + \boldsymbol{\varepsilon} = \breve{\Psi}\boldsymbol{\theta}^* + \Phi^T\widetilde{\boldsymbol{\eta}^*} + \boldsymbol{\varepsilon}$$

with $\breve{\Psi}\Phi^T = 0$ and we are in the so called adaptive case.

# linear semiparametric models: the pMLE

**Theorem**

*The profile MLE $\widetilde{\boldsymbol{\theta}}$ reads as*

$$\widetilde{\boldsymbol{\theta}} = \left(\breve{\Psi}\breve{\Psi}^{\top}\right)^{-1}\breve{\Psi}\boldsymbol{Y},$$

$$\breve{\Psi} = \Psi - \Psi\Pi_{\boldsymbol{\eta}} = \Psi - \Psi\Phi^{\top}\left(\Phi\Phi^{\top}\right)^{-1}\Phi.$$

*Gauss-Markov: The quadratic risk is minimal in the class of all unbiased linear estimates of* $\boldsymbol{\theta}^{*}$

## linear semiparametric models: "Wilk's phenomenon"

We write

$$\widetilde{L}(\boldsymbol{\theta}) := \max_{\boldsymbol{\eta}} L(\boldsymbol{\theta}, \boldsymbol{\eta}).$$

**Theorem**

*Let the matrix* $D^2 = \sigma^{-2}\breve{\Psi}\breve{\Psi}^\top$ *be non-degenerated. It holds*

$$2\left\{\widetilde{L}(\widetilde{\boldsymbol{\theta}}) - \widetilde{L}(\boldsymbol{\theta}^*)\right\} = \|D\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\|^2 = \|\boldsymbol{\xi}\|^2,$$

$$\boldsymbol{\xi} = D\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right), \ \mathbb{E}\boldsymbol{\xi} = 0, \mathrm{Var}(\boldsymbol{\xi}) = I_p.$$

*If* $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$, *then* $\boldsymbol{\xi}$ *is standard normal in* $\mathbb{R}^p$ *and*

$$2\{\widetilde{L}(\widetilde{\boldsymbol{\theta}}) - \widetilde{L}(\boldsymbol{\theta}^*)\} \sim \chi_p^2 .$$

## general theory: basic idea

We want to extend the results for the linear model to the general case. But allow

- misspecification: $\mathbb{Y} \not\sim \mathbb{P}_{\boldsymbol{\theta}^*,\boldsymbol{\eta}^*}$
- finite sample size.

We aim at a "nearly" $\chi^2_p$-distributed random variable $\|\boldsymbol{\xi}\|^2$ such that we can find

- random sets $\mathcal{A}(\mathbb{Y},c)$ with

$$I\!P(\boldsymbol{\theta}^* \in A(\mathbb{Y},c)) \geq I\!P(\|\boldsymbol{\xi}\|^2 < c)$$

- sets $\mathcal{B}(c)$ with

$$I\!P(\widetilde{\boldsymbol{\theta}} \in B(c)) \geq I\!P(\|\boldsymbol{\xi}\|^2 < c)$$

Then we only have to use

$$I\!P(\|\boldsymbol{\xi}\|^2 > c) \lesssim e^{-(c-p)}$$

## the conditions

To do so we need a set of conditions. We write $\boldsymbol{\upsilon} = (\boldsymbol{\theta}, \boldsymbol{\eta})$
Define

$$\begin{aligned}
\boldsymbol{\upsilon}^* &:= \operatorname{argmax} I\!E[L(\boldsymbol{\upsilon})] \\
L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) &= L(\boldsymbol{\upsilon}) - L(\boldsymbol{\upsilon}^*) \\
\zeta(\boldsymbol{\upsilon}) &:= L(\boldsymbol{\upsilon}) - I\!E[L(\boldsymbol{\upsilon})] \\
\mathcal{D}_0^2 &:= -I\!E[\nabla^2 L(\boldsymbol{\upsilon}^*)] \\
\mathcal{V}_0^2 &= Var(\nabla\zeta(\boldsymbol{\upsilon}^*)) \\
\Theta_0(r) &:= \{\|\mathcal{V}_0(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\| < r\}
\end{aligned}$$

Then the conditions read:

- $$I\!\!E[\exp(\mu L)] < \infty, \text{, for some } \mu > 0$$

- $$I\!\!E[\exp(\mu \frac{\langle \nabla \zeta, \gamma \rangle}{\|\mathcal{V}_0 \gamma\|})] < \infty, \text{ for all } |\mu| \leq g$$

- $$I\!\!E[\exp(\mu \frac{\langle \nabla \zeta(\boldsymbol{\upsilon}) - \nabla \zeta(\boldsymbol{\upsilon}^*), \gamma \rangle}{\omega(r) \|\mathcal{V}_0 \gamma\|})] < \infty, \text{ for all } |\mu| \leq g \text{ and } \boldsymbol{\upsilon} \in \Theta_0(r)$$

- $$\nabla^2 L \approx -\mathcal{D}_0^2 \text{ the difference is quantified by } \delta(r)$$

## central theorem

For $\boldsymbol{\epsilon} = (\delta, \varrho)$ set $\mathcal{D}_{\boldsymbol{\epsilon}}^2 := (1-\delta)\mathcal{D}_0^2 + \varrho\mathcal{V}_0^2$, and for $\underline{\boldsymbol{\epsilon}} = -\boldsymbol{\epsilon}$ define $\mathcal{D}_{\underline{\boldsymbol{\epsilon}}}$ accordingly. Then define

$$\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) := \nabla L(\boldsymbol{\upsilon}^*)(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*) - \|\mathcal{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2/2$$

### Theorem

*Assume the conditions $(E), ..., (\mathcal{L}_0)$ hold. Let for some $\mathtt{r}$, the values $\varrho \geq \sqrt{3\nu_0}\,\omega(\mathtt{r})$ and $\delta \geq \delta(\mathtt{r})$ ensure $\mathcal{D}_{\boldsymbol{\epsilon}} \geq 0$. Then*

$$\mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}) \leq L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \leq \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) + \diamondsuit_{\boldsymbol{\epsilon}}(\mathtt{r}),$$

*where the random variables $\diamondsuit_{\boldsymbol{\epsilon}}(\mathtt{r}), \diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r})$ fulfill*

$$I\!P\big\{\varrho^{-1}\diamondsuit_{\boldsymbol{\epsilon}}(\mathtt{r}) \geq \mathfrak{z}(p^*, \mathtt{x})\big\} \leq \exp\big(-\mathtt{x}\big).$$

# semiparametric estimation: applying main theorem

Observe

$$\begin{aligned}
&\sup_{\boldsymbol{\upsilon}} \mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} \mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \mathit{\Delta}_{\boldsymbol{\epsilon},1}(\mathbf{r}) \\
&\quad \le \sup_{\boldsymbol{\upsilon}} L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \\
&\quad \le \sup_{\boldsymbol{\upsilon}} \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) + \mathit{\Delta}_{\boldsymbol{\epsilon},1}(\mathbf{r})
\end{aligned}$$

where the partial deficiency is given by

$$\mathit{\Delta}_{\boldsymbol{\epsilon},1}(\mathbf{r}) := \diamondsuit_{\boldsymbol{\epsilon}}(\mathbf{r}) + \diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathbf{r}) + \sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} \mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)$$

## using linear structure

As in the linear case we can do a variable substitution trick. We write

$$\mathcal{D}_{\boldsymbol{\epsilon}}^2 = \begin{pmatrix} D_{\boldsymbol{\epsilon}}^2 & A_{\boldsymbol{\epsilon}} \\ A_{\boldsymbol{\epsilon}}^\top & H_{\boldsymbol{\epsilon}}^2 \end{pmatrix},$$

set $\breve{D}_{\boldsymbol{\epsilon}}^2 = D_{\boldsymbol{\epsilon}}^2 - A_{\boldsymbol{\epsilon}} H_{\boldsymbol{\epsilon}}^{-2} A_{\boldsymbol{\epsilon}}^\top$ and obtain

$$2 \sup_{\boldsymbol{\upsilon}} \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - 2 \sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) = \|\breve{D}_{\boldsymbol{\epsilon}}^{-1} \breve{\zeta}_{\boldsymbol{\epsilon}}\|^2,$$

with

$$\breve{\zeta}_{\boldsymbol{\epsilon}} = (\nabla_{\boldsymbol{\theta}} + H_{\boldsymbol{\epsilon}}^{-2} A_{\boldsymbol{\epsilon}}^\top \nabla_{\boldsymbol{\eta}}) L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*).$$

We also see

$$\begin{aligned} 2 \sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - 2 \sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} \mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \\ = \|H_{\boldsymbol{\epsilon}}^{-1} \nabla_{\boldsymbol{\eta}} L\|^2 - \|H_{\underline{\boldsymbol{\epsilon}}}^{-1} \nabla_{\boldsymbol{\eta}} L\|^2 \end{aligned}$$

## concentration results

We write

$$\sup_{P_\theta \boldsymbol{\upsilon} = \boldsymbol{\theta}^*} L(\boldsymbol{\upsilon}) =: \widetilde{L}(\boldsymbol{\theta}^*).$$

Consider for some $c > 0$ the random set

$$\mathcal{A}(\mathbb{Y}, c) = \{\boldsymbol{\theta} : \widetilde{L}(\widetilde{\boldsymbol{\theta}}) - \widetilde{L}(\boldsymbol{\theta}) \leq c\}.$$

We have

### Corollary

*Suppose* $(E), \ldots, (\mathcal{L}_0)$ *hold. For any* $c > 0$, *it holds with* $\breve{\boldsymbol{\xi}}_{\boldsymbol{\epsilon}} = D_{\boldsymbol{\epsilon}}^{-1} \breve{\zeta}_{\boldsymbol{\epsilon}}$

$$I\!P\big\{\mathcal{A}(\mathbb{Y}, c) \not\ni \boldsymbol{\theta}^*, \, (\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\eta}}) \in \Theta_0(r)\big\} \leq I\!P\big\{\|\breve{\boldsymbol{\xi}}_{\boldsymbol{\epsilon}}\|^2 \geq 2c - 2\Delta_{\boldsymbol{\epsilon},1}(\mathbf{r})\big\}.$$

## concentration results

We define the total deficiency

$$\Delta^*_{\epsilon}(\mathbf{r}) = \diamondsuit_{\epsilon}(\mathbf{r}) + \diamondsuit_{\underline{\epsilon}}(\mathbf{r}) + \sup_{\boldsymbol{\upsilon}} \mathbb{L}_{\epsilon}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \sup_{\boldsymbol{\upsilon}} \mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) + \Delta_{\epsilon,1}(\mathbf{r}).$$

Consider for some $c > 0$ the set

$$\mathcal{B}(c) = \{\boldsymbol{\theta} : \|\breve{D}_{\epsilon}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \le c\}.$$

We have

**Corollary**

*Suppose $(E), ..., (\mathcal{L}_0)$ hold. For any $c > 0$, it holds with $\breve{\boldsymbol{\xi}}_{\epsilon} = \breve{D}_{\epsilon}^{-1}\breve{\zeta}_{\epsilon}$*

$$I\!P\{\mathcal{B}(c) \not\ni \widetilde{\boldsymbol{\theta}},\ (\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\eta}}) \in \Theta_0(r)\} \le I\!P\{\|\breve{\boldsymbol{\xi}}_{\epsilon}\|^2 \ge c - \sqrt{2\Delta^*_{\epsilon}(\mathbf{r})}\}.$$

## Making the probabilities small

So if $\Delta_{\epsilon}^{*}, \Delta_{\epsilon,1}(\mathtt{r})$ are neglectably small, we only have to

- controll the deviation of $\|\breve{\boldsymbol{\xi}}_{\epsilon}\|^2$ from its mean
- choose $c$ big enough
- choose $\mathtt{r}$ big enough to ensure a high probability of $\{(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\eta}}) \in \Theta_0(r)\}$

# large deviation

**Lemma**

*Suppose* $(E), ..., (\mathcal{L}_0)$ *hold. For* $x > 0$ *with* $r(x) > 0$ *such that*

$$\max_{\mu}\{-\log I\!\!E[\exp(\mu L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*))]\} \geq (1+s)t(\boldsymbol{\upsilon}) + \delta_\mu(\boldsymbol{\upsilon}) + \mathfrak{z}_1(x)$$

*on* $\Theta(r(x))^c$. *Then*

$$I\!\!P(\|\mathcal{V}_0(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*, \widetilde{\boldsymbol{\eta}} - \boldsymbol{\eta}^*)\| > r(x)) \leq e^{-x+1}.$$

# nearly Chi square distribution

**Lemma**

*Suppose* $(E), ..., (\mathcal{L}_0)$ *hold. Then there is a* $x_c \in \mathbb{R}$

$$\mathbb{P}(\|\breve{\boldsymbol{\xi}}_{\boldsymbol{\epsilon}}\|^2 > \mathfrak{z}_{\boldsymbol{\theta}}(x)) \leq 2e^{-x} + 8.4e^{-x_c}$$

*where* $\mathfrak{z}_{\boldsymbol{\theta}}(x) \cong \mathbb{E}[\|\breve{\boldsymbol{\xi}}_{\boldsymbol{\epsilon}}\|^2]$. *And*

$$\mathbb{P}(\|H_{\epsilon}^{-1}\nabla_{\boldsymbol{\eta}} L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\|^2 > \mathfrak{z}_{\boldsymbol{\eta}}(x)) \leq 2e^{-x} + 8.4e^{-x_c}$$

*where* $\mathfrak{z}_{\boldsymbol{\eta}}(x) \cong \mathbb{E}[\|H_{\epsilon}^{-1}\nabla_{\boldsymbol{\eta}} L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\|^2]$ *and* $x_c \cong g$.

## definition of regularity

Are we done?

Problem: the values above can not really be determined on this level of generality.

Further regularity conditions make it possible solve this.

Represent

$$\mathcal{D}_0^2 = \begin{pmatrix} D_0^2 & A_0 \\ A_0^T & H_0^2 \end{pmatrix}, \; \mathcal{V}_0^2 = \begin{pmatrix} V_0^2 & B_0 \\ B_0^T & Q_0^2 \end{pmatrix}$$

Then we impose with $\mathfrak{a}^2 < \frac{1-\delta}{\varrho}$

$$\|D_0^{-1} A_0 H_0^{-2} A_0^T D_0^{-1}\|_\infty \leq \gamma < 1, \qquad \mathfrak{a}^2 D_0^2 \geq V_0^2, \qquad \mathfrak{a}^2 H_0^2 \geq Q_0^2.$$

# implications of regularity

## Lemma

*Under the regularity conditions above and that*

$$\gamma_{\epsilon} := \gamma \frac{1 + \mathfrak{a}^2(\varrho + 2\delta\varrho) + \mathfrak{a}^4\varrho^2 + \delta^2 + \delta}{((1-\delta) - \varrho\mathfrak{a}^2)^2} < 1$$

*we have that* $\breve{D}_{\epsilon}$ *is invertable and the deficiency satisfies*

$$\begin{aligned} \Delta^*_{\epsilon}(\mathbf{r}) \le 2(\diamondsuit_{\epsilon}(\mathbf{r}) + \diamondsuit_{\underline{\epsilon}}(\mathbf{r})) + \alpha_{\boldsymbol{\eta},\epsilon}\|H_{\epsilon}^{-1}\nabla_{\boldsymbol{\eta}}L\|^2 \\ + \alpha_{\boldsymbol{\theta},\epsilon}\|\breve{D}_{\epsilon}^{-1}\breve{\boldsymbol{\zeta}}_{\epsilon}\|^2. \end{aligned}$$

*Where* $\alpha_{\boldsymbol{\theta},\epsilon}, \alpha_{\boldsymbol{\eta},\epsilon}$ *are of order* $\epsilon$.

$$\mathbb{E}[\|\breve{\boldsymbol{\xi}}_{\epsilon}\|^2] \le C(\mathfrak{a}, \epsilon)p$$

$$\mathbb{E}[\|H_{\epsilon}^{-1}\nabla_{\boldsymbol{\eta}}L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\|^2] \le \mathfrak{a}^2((1-\delta) - \varrho\mathfrak{a}^2))^{-1}p^*$$

## range of applicability

Now we collect setting $c(x) = \varrho\mathfrak{z}(p^*, \mathtt{x}) + \alpha_{\eta,\epsilon}\mathfrak{z}_{\boldsymbol{\eta}}(x) + \alpha_{\theta,\epsilon}\mathfrak{z}_{\boldsymbol{\theta}}(x)$

$$I\!P(\Delta^*_{\epsilon}(\mathbf{r}) > c(x)) \leq e^{-x} + 2e^{-x+1} + 16.8e^{-x_c}$$

We want that $c(x)/p$ is neglectable. Note that

$\varrho, \alpha_{\eta,\epsilon}, \alpha_{\eta,\epsilon}$ are of order $\omega(\mathbf{r})$.

So we need that $\omega(r(x))p^*/p$ is neglectably small.

In the i.i.d. case $\omega(r(x))$ is of order $n^{-\frac{1}{2}}p^* \log p^*$, which leads to

$$n^{-\frac{1}{2}}(p^*)^2 \log(p^*)/p, \text{ neglectable}$$

## an example: the model

Our model reads

$$\mathbb{Y} = A_0 \boldsymbol{\theta}_0 + \varepsilon_\theta,$$
$$A_{\varepsilon_A} := A_0 + \varepsilon_A,$$

where

- $\mathbb{E}[A_{\varepsilon_A}] =: A_0 \in \mathbb{R}^{n \times p}$
- $\mathbb{E}[\varepsilon_{\boldsymbol{\theta}}] = 0$, $\mathbb{E}[\varepsilon_A] = 0$
- $\mathbb{E}[(\varepsilon_\theta, \varepsilon_A)(\varepsilon_\theta, \varepsilon_A)^T] =: \Sigma$

$$\Sigma = \begin{pmatrix} \Sigma_\theta & \Sigma_{\theta,A} \\ \Sigma_{\theta,A}^T & \Sigma_A \end{pmatrix}$$

Aim: $(\boldsymbol{\theta})_0$
Estimator: $\widetilde{\boldsymbol{\theta}}$ maximizer of following functional

$$\max_A L(\boldsymbol{\theta}, A) = \max_A (-\|\mathbb{Y} - A\boldsymbol{\theta}\|^2/2 - \lambda tr[(A_{\varepsilon_A} - A)^T (A_{\varepsilon_A} - A)]/2).$$

## the elements

$$\langle \nabla\zeta(\boldsymbol{\theta}, A), (\boldsymbol{\theta}', A')\rangle_{I\!R^p \times I\!R^{n\times p}} = \langle A\boldsymbol{\theta}' + A'\boldsymbol{\theta}, \varepsilon_\theta\rangle_{I\!R^p} + \lambda tr((A')^T\varepsilon_A)$$

$$I\!E[\nabla\zeta(\boldsymbol{\theta}^*, A^*) \otimes \nabla\zeta(\boldsymbol{\theta}^*, A^*)] =: \mathcal{V}_0^2 = \begin{pmatrix} V_0^2 & B_0 \\ B_0^\top & Q_0^2 \end{pmatrix}$$

with

$$\begin{aligned} V_0^2 &= (A^*)^T\Sigma_\theta A^* \\ Q_0^2[A, A] &= (A\boldsymbol{\theta}^*)^T\Sigma_\theta(A\boldsymbol{\theta}^*) + \lambda^2\Sigma_A[A, A] + 2\lambda\Sigma_{\boldsymbol{\theta},A}[A\boldsymbol{\theta}^*, A] \\ B_0[\boldsymbol{\theta}, A] &= (A^*\boldsymbol{\theta})^T\Sigma_\theta A\boldsymbol{\theta}^* + \lambda\Sigma_{\boldsymbol{\theta},A}[A^*\boldsymbol{\theta}, A] \end{aligned}$$

## the elements

$$-D^2 \mathbb{E}[L(\boldsymbol{\theta}^*, A^*)] =: \mathcal{D}_0^2 = \begin{pmatrix} D_0^2 & A_0 \\ A_0^\top & H_0^2 \end{pmatrix}$$

with

$$\begin{aligned} D_0^2 &= (A^*)^T A^* \\ H_0^2[A, A] &= (A\boldsymbol{\theta}^*)^T A\boldsymbol{\theta}^* + \lambda tr(A^T A) \\ A_0[\boldsymbol{\theta}, A] &= \boldsymbol{\theta}^T (A^*)^T A\boldsymbol{\theta}^* - \langle \mathbb{E}[\mathbb{Y}] - A^*\boldsymbol{\theta}^*, A\boldsymbol{\theta}\rangle_{\mathbb{R}^p} \end{aligned}$$

## the conditions

### Lemma

*If* $A^*$ *has full rank,* $\Theta \subset I\!\!R^p \otimes I\!\!R^{n\times p}$ *is bounded and for any* $q \in [0, \frac{1}{2}]$ $(\varepsilon_\theta, \varepsilon_A)$ *satisfy with some* $\widetilde{g} > 0$ *for all* $|\mu| \le \widetilde{g}$

$$\sup_{\gamma \in I\!\!R^p,\ \|\gamma\|=1} \log I\!\!E[\exp\left\{\frac{C_\Theta \mu}{\lambda_{\min}(\mathcal{V}_0^2)^{1-q}} \langle \gamma, \varepsilon_\theta \rangle_{I\!\!R^p}\right\}] \le \nu_0^2 \mu^2/2$$

$$\sup_{\gamma \in I\!\!R^{n\times p},\ tr(\gamma^T\gamma)=1} \log I\!\!E[\exp\left\{\frac{\lambda \mu}{\lambda_{\min}(\mathcal{V}_0^2)^{1-q}}\ tr(\gamma^T \varepsilon_A)\right\}] \le \nu_0^2 \mu^2/2$$

*the conditions* $(E), ..., (\mathcal{L}_0)$ *are satisfied with*

$$g = \frac{1}{2}\lambda_{\min}(\mathcal{V}_0^2)^q \widetilde{g},\ \ \omega(r) \le \frac{2r}{\lambda_{\min}(\mathcal{V}_0^2)^{1-q}} \text{ and } \delta(r) \le \frac{C_{\Theta_0}}{\lambda_{\min}(\mathcal{D}_0^2)}.$$

## regularity

### Lemma

*If* $(1+\lambda)(\|\Sigma_\theta\|_\infty \vee \|\Sigma_A\|_\infty) < \frac{1-\delta}{\varrho}$ *then*

$$\mathfrak{a}^2 D_0^2 > V_0^2 \text{ and } \mathfrak{a}^2 H_0^2 > Q_0^2.$$

*with* $\mathfrak{a}^2 < \frac{1-\delta}{\varrho}$ .

### Lemma

*For every* $n \in \mathbb{N}$ *and* $\lambda > 0$ *there are suitalbe constants* $c_1 > 0, c_2 > 0$ *, that grow with* $n$ *such that if* $\|\mathbb{E}[\mathbb{Y}] - A^*\boldsymbol{\theta}^*\| < c_1, \cos(\frac{\varphi_0}{2}) < c_2$ *then*

$$\|D_\epsilon^{-1} A_\epsilon H_\epsilon^{-2} A_\epsilon^\top D_\epsilon^{-1}\|_\infty =: \gamma_\epsilon < 1,$$

*where*

$$\varphi_0 := \varphi(-((A^*)^T A^*)^{-\frac{1}{2}}\boldsymbol{\theta}^*, ((A^*)^T A^*)^{-\frac{1}{2}}(A^*)^T(\mathbb{E}[\mathbb{Y}] - A^*\boldsymbol{\theta}^*))$$

# still to do

- Calculate $r(x)$.
- identify range of applicability
- Specify optimal choice of $\lambda$ and $q$

Reference

V. Spokoiny.
Parametric estimation. finite sample theory.
2011.