Parametric and Semiparametric Efficiency

Marc Hallin

Université Libre de Bruxelles

London, December 11, 2015

- LAN is a parametric theory, leading to parametrically efficiency under density f
- but the parametrically efficient tests based on $\mathbf{\Delta}_{\boldsymbol{\theta}}^{(n)}$ in general are not valid under $g \neq f$; nor do the parametrically efficient (one-step) estimators remain $\boldsymbol{\nu}(n)$-(root-n) consistent under $g \neq f$

Outline

Parametric and semiparametric efficiency

- 1. Parametric efficiency at f
- 2. Semiparametric efficiency at f

Outline

Parametric and semiparametric efficiency

- 1. Parametric efficiency at f
- 2. Semiparametric efficiency at f

Semiparametric models

- So far, we have been dealing with (sequences of) parametric models, of the form

$$\mathcal{E}_f^{(n)} := \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}_f^{(n)} := \left\{ \mathrm{P}_{\boldsymbol{\theta};f}^{(n)} \middle| \boldsymbol{\theta} \in \mathbb{R}^K \right\}\right)$$

- from now on, such parametric models are embedded into larger models: we assume that the observation $\mathbf{X}^{(n)}$ is generated by

$$\mathcal{E}^{(n)} := \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} := \left\{ \mathrm{P}_{\boldsymbol{\theta};g}^{(n)} \middle| \boldsymbol{\theta} \in \mathbb{R}^K, \ g \in \mathcal{F} \right\}\right)$$

where $\boldsymbol{\theta}$ is a finite-dimensional parameter of interest, and $g \in \mathcal{F}$ an infinite-dimensional nuisance (here, some unknown noise or innovation density). Such models are called semiparametric models

- throughout, the fixed-g parametric submodels

$$\mathcal{E}_g^{(n)} := \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}_g^{(n)} := \left\{ \mathrm{P}_{\boldsymbol{\theta};g}^{(n)} \middle| \boldsymbol{\theta} \in \mathbb{R}^K \right\}\right)$$

  are assumed to be ULAN, with (parametric) central sequence $\boldsymbol{\Delta}_{\boldsymbol{\theta};g}^{(n)}$, and (parametric) information matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta};g}$; it is also assumed that there exist *residuals* $Z_t^{(n)}(\boldsymbol{\theta})$ such that $\mathbf{X}^{(n)} \sim \mathrm{P}_{\boldsymbol{\theta};g}^{(n)}$ iff the $Z_t^{(n)}(\boldsymbol{\theta})$'s are, for instance, i.i.d., with density g (other concepts of white noise are possible)

- the fixed-$\boldsymbol{\theta}$ submodels

$$\mathcal{E}_{\boldsymbol{\theta}}^{(n)} := \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}_{\boldsymbol{\theta}}^{(n)} := \left\{ \mathrm{P}_{\boldsymbol{\theta};g}^{(n)} \middle| \; g \in \mathcal{F} \right\}\right)$$

  are nonparametric models indexed by an "infinite-dimensional parameter" $g \in \mathcal{F}$

- parametric efficiency (at f) is defined within the $\mathcal{E}_f^{(n)}$ submodels, and is entirely characterized by the information matrix $\mathbf{\Gamma}_{\boldsymbol{\theta};f}$: this matrix defines the noncentrality parameters in the power of parametrically optimal (at f) tests; its inverse is also the asymptotic covariance matrix of the parametrically optimal (at f) estimators of $\boldsymbol{\theta}$
- however, the implementation of parametrically efficient procedures was based on the fact that, under the parametric local experiments of $\mathcal{E}_f^{(n)}$,

$$\mathbf{\Delta}_{\boldsymbol{\theta};f}^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{\Gamma}_{\boldsymbol{\theta};f}\boldsymbol{\tau},\ \mathbf{\Gamma}_{\boldsymbol{\theta};f}) \quad \boldsymbol{\tau} \in \mathbb{R}^K;$$

  which does not hold under the semiparametric local experiments of $\mathcal{E}^{(n)}$ anymore
- in particular, $\mathbf{\Delta}_{\boldsymbol{\theta};f}^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{\Gamma}_{\boldsymbol{\theta};f}\boldsymbol{\tau},\ \mathbf{\Gamma}_{\boldsymbol{\theta};f})$ in general is not true anymore under $\mathrm{P}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau};g}^{(n)}$, $g \neq f$

the cost of going semiparametric

hence, parametrically efficient procedures, as a rule, are no longer valid, and parametric efficiency cannot be reached, in the semiparametric model: the fact of not knowing g in general has a cost

Can we characterize that cost?

parametric nuisances

Recall that, in case $\boldsymbol{\theta}$ splits into a parameter of interest $\boldsymbol{\theta}_1$ and a (parametric) nuisance $\boldsymbol{\theta}_2$,

- optimal (efficient) inference on $\boldsymbol{\theta}_1$ can be based on the residual $\boldsymbol{\Delta}_1^{(n)^\perp} := \boldsymbol{\Delta}_1^{(n)} - \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Delta}_2^{(n)}$; of the regression of the $\boldsymbol{\theta}_1$-part $\boldsymbol{\Delta}_1^{(n)}$ of the central sequence $\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} = (\boldsymbol{\Delta}_{\boldsymbol{\theta};1}^{(n)\prime}, \boldsymbol{\Delta}_{\boldsymbol{\theta};2}^{(n)\prime})'$ with respect to the $\boldsymbol{\theta}_2$-part $\boldsymbol{\Delta}_2^{(n)}$, with (under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)(\boldsymbol{\tau}',_1,\boldsymbol{\tau}'_2)'}$)

$$\boldsymbol{\Delta}_1^{(n)^\perp} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\boldsymbol{\Gamma}_{11}^{\perp}\boldsymbol{\tau}_1,\, \boldsymbol{\Gamma}_{11}^{\perp}\right), \qquad \boldsymbol{\Gamma}_{11}^{\perp} := \boldsymbol{\Gamma}_{11} - \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Gamma}'_{12};$$

- the matrix $\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Gamma}'_{12}$ is the (asymptotic) information loss due to nonspecification of $\boldsymbol{\theta}_2$
- this regression actually is projecting $\boldsymbol{\Delta}_1^{(n)}$ onto the L_2 space orthogonal to $\boldsymbol{\Delta}_2^{(n)}$, thus neutralizing the impact of a local perturbation of $\boldsymbol{\theta}_2$

tangents

Here the nuisance g ranges over an infinite-dimensional space $\mathcal{F}$; but this space $\mathcal{F}$ contains parametric subspaces. Consider "paths" $q := \boldsymbol{\eta} \mapsto q(\boldsymbol{\eta}) \in \mathcal{F}$, where $\boldsymbol{\eta} \in (-1,1)^K$, $q(\mathbf{0}) = g$, such that the parametric model

$$\mathcal{E}_q^{(n)} := \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}_q^{(n)} := \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}, q(\boldsymbol{\eta})}\}\right)$$

be LAN at $(\boldsymbol{\theta}', \mathbf{0}')$ with respect to $(\boldsymbol{\theta}', \boldsymbol{\eta}')$, with central sequence $\begin{pmatrix} \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};g} \\ \mathbf{H}^{(n)}_{q;\boldsymbol{\theta};g} \end{pmatrix}$ and information matrix $\begin{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\theta};g} & \mathbf{C}'_{q;\boldsymbol{\theta};g} \\ \mathbf{C}_{q;\boldsymbol{\theta};g} & \boldsymbol{\Gamma}^{\mathbf{H}}_{\boldsymbol{\theta};g} \end{pmatrix}$ ($\boldsymbol{\eta}$ here plays the role of a parametric nuisance)

- denote by $\mathcal{Q}$ the family of all such paths

- efficient inference on $\boldsymbol{\theta}$ in that model thus should be based on the residual of the regression of the $\boldsymbol{\theta}$-part $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};g}$ of the central sequence with respect to the $\boldsymbol{\eta}$-part $\mathbf{H}^{(n)}_{q;\boldsymbol{\theta};g}$

$$\boldsymbol{\Delta}^{(n)q}_{\boldsymbol{\theta};g} := \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};g} - \mathbf{C}'_{q;\boldsymbol{\theta};g}(\boldsymbol{\Gamma}^{\mathbf{H}}_{\boldsymbol{\theta};g})^{-1}\mathbf{H}^{(n)}_{q;\boldsymbol{\theta};g} \quad \text{with} \quad \boldsymbol{\Gamma}^{q}_{\boldsymbol{\theta};g} := \boldsymbol{\Gamma}_{\boldsymbol{\theta};g} - \mathbf{C}'_{q;\boldsymbol{\theta};g}(\boldsymbol{\Gamma}^{\mathbf{H}}_{\boldsymbol{\theta};g})^{-1}\mathbf{C}_{q;\boldsymbol{\theta};g}$$

- which is the projection of $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};g}$ onto the L_2 space orthogonal to $\mathbf{H}^{(n)}_{q;\boldsymbol{\theta};g}$, neutralizing the impact on $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};g}$ of a local perturbation $q(n^{-1/2}\boldsymbol{\tau}_2)$ of $g =: q(\mathbf{0})$ along $q(\boldsymbol{\eta})$
- $\boldsymbol{\Delta}^{(n)q}_{\boldsymbol{\theta};g} \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{\Gamma}^{q}_{\boldsymbol{\theta};g}\boldsymbol{\tau}_1,\ \boldsymbol{\Gamma}^{q}_{\boldsymbol{\theta};g})$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}_1;q(n^{-1/2}\boldsymbol{\tau}_2);g}$,

least favorable submodel

- Assume that a q_{lf} exists such that $\mathbf{\Delta}^{(n)q_{lf}}_{\boldsymbol{\theta};g}$ is asymptotically L_2-orthogonal to $\mathbf{H}^{(n)}_{q;\boldsymbol{\theta};g}$ for any q: clearly, $\mathbf{\Gamma}^{q_{lf}}_{\boldsymbol{\theta};g}$ would be minimal among all $\mathbf{\Gamma}^{q}_{\boldsymbol{\theta};g}$'s, and q_{lf} would provide the least favorable local perturbation of g, and $\mathbf{\Delta}^{(n)q_{lf}}_{\boldsymbol{\theta};g}$ the 'maximal" part of the original central sequence insensitive to all such local perturbations (the collection of which is called the tangent space)
- semiparametric efficiency (at $\boldsymbol{\theta}$ and g) is characterized by the residual information matrix $\mathbf{\Gamma}^{*}_{\boldsymbol{\theta};g} := \mathbf{\Gamma}^{q_{lf}}_{\boldsymbol{\theta};g}$
- in the parametric model $\mathcal{E}_g$, this semiparametric efficiency can be achieved by considering the semiparametrically efficient central sequence

$$\mathbf{\Delta}^{(n)*}_{\boldsymbol{\theta};g} := \mathbf{\Delta}^{(n)q_{lf}}_{\boldsymbol{\theta};g} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{\Gamma}^{*}_{\boldsymbol{\theta};g}\boldsymbol{\tau}_1,\ \mathbf{\Gamma}^{*}_{\boldsymbol{\theta};g}) \quad \text{under } \mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}_1;q(n^{-1/2}\boldsymbol{\tau}_2);g}$$

summing up

semiparametric efficiency (at given f and $\boldsymbol{\theta}$) is characterized by the Gaussian shift model

$$\boldsymbol{\Delta}^* \sim \mathcal{N}\left(\boldsymbol{\Gamma}^*_{\boldsymbol{\theta};f}\boldsymbol{\tau}, \boldsymbol{\Gamma}^*_{\boldsymbol{\theta};f}\right), \ \tau \in \mathbb{R}^m$$

hence by the semiparametrically efficient (at f) information matrix $\boldsymbol{\Gamma}^*_{\boldsymbol{\theta};f}$ (noncentrality parameters for optimal tests; asymptotic covariance matrices of optimal estimators)

- if $\boldsymbol{\Gamma}^*_{\boldsymbol{\theta};f} = \boldsymbol{\Gamma}_{\boldsymbol{\theta};f}$: the model is "adaptive" at f
- in general, $\boldsymbol{\Gamma}^*_{\boldsymbol{\theta};f} < \boldsymbol{\Gamma}_{\boldsymbol{\theta};f}$: the cost of not knowing the "true" density, at f, is strictly positive

- BUT ...

many problems remain unsolved!

- how can we construct such least favorable models? how can we compute $\mathbf{\Delta}^{(n)*}_{\boldsymbol{\theta};f}$ and $\mathbf{\Gamma}^{*}_{\boldsymbol{\theta};f}$?
- assuming that $\mathbf{\Delta}^{(n)*}_{\boldsymbol{\theta};f}$ can be computed, its asymptotic distribution is known under $\mathcal{E}^{(n)}_{f}$, but not, in general, under the semiparametric model $\mathcal{E}^{(n)}$ (under $\mathcal{E}^{(n)}_{g}$, $g \neq f$, such asymptotic distribution even may not exist!) ... hence, inference based on $\mathbf{\Delta}^{(n)*}_{\boldsymbol{\theta};f}$ is valid under density f only
- ... inference based on $\mathbf{\Delta}^{(n)*}_{\boldsymbol{\theta};\hat{g}^{(n)}}$, where $\hat{g}^{(n)}$ is such that $\mathbf{\Delta}^{(n)*}_{\boldsymbol{\theta};\hat{g}^{(n)}} - \mathbf{\Delta}^{(n)*}_{\boldsymbol{\theta};g} = o_{\mathrm{P}}(1)$ at "all" g (nonparametric estimation $\hat{g}^{(n)}$ of g, additional regularity assumptions, slow convergence, sample splitting, and other niceties ...) ...