Weierstraß-Institut für
Angewandte Analysis und Stochastik

Semidefinite optimization in IV Estimation with shape constraints

Vladimir Spokoiny
(joint with Joel Horowitz)

1 Motivation, Data and Problem

- IV problem
- Sieve approach

Observed: a sample from (Z, X, W).

Model

$$Y = f(X) + U, \qquad \mathbb{E}\left[U \mid W\right] = 0.$$

where
Y, an explained variable
X, an explanatory variable
W, an instrument.

Target: the regression function $f(\cdot)$.

Some references

Sieve estimation:

Newey and Powell (2003)

Blundell, Chen, Kristensen (2007)

Horowitz (2009)

Kernel:

Hall and Horowitz (2005)

Darolles, Florens, and Renault (2006).

Optimal rates:

Hall and Horowitz (2005)

Chen and Reiss (2007).

Let $\psi_1(x), \ldots, \psi_j(x), \ldots$ be a functional basis. Consider an approximation

$$f(x) = \theta_1 \psi_1(x) + \ldots + \theta_J \psi_J(x)$$

or in vector form

$$f(x) = \Psi(x)^\top \theta$$

with $\Psi(x) = \big(\psi_1(x), \ldots, \psi_J(x)\big)^\top \in \mathbb{R}^J$ and $\theta = (\theta_1, \ldots, \theta_J)^\top \in \mathbb{R}^J$.

An approximating model I:

$$Y = \Psi(X)^\top \theta_0 + U, \qquad \mathbb{E}\big[U \mid W\big] = 0.$$

A sieve approximation: for $\Psi(x) = \bigl(\psi_1(x), \ldots, \psi_J(x)\bigr)^\top \in I\!R^J$

$$Y = \Psi(X)^\top \theta_0 + U, \qquad I\!E\bigl[U \mid W\bigr] = 0.$$

Dual representation: for any function $\phi(W)$

$$I\!E\bigl[Y\phi(W)\bigr] = I\!E\bigl[\phi(W)\Psi(X)^\top\bigr]\theta_0$$

An approximating model II: for a collection of functions $\phi_1(w), \ldots, \phi_M(w)$, consider

$$I\!E\bigl[Y\Phi(W)\bigr] = I\!E\bigl[\Phi(W)\Psi(X)^\top\bigr]\theta_0 = A_0\theta_0$$

with

$$\begin{aligned} \Phi(w) &= (\phi_1(w), \ldots, \phi_M(w))^\top \in I\!R^M, \\ A_0 &= I\!E\bigl[\Phi(W)\Psi(X)^\top\bigr] \in I\!R^{M\times J}. \end{aligned}$$

Define

$$
\begin{aligned}
Z &= \mathbb{E}_n\big[Y\Phi(W)\big] &&= \tfrac{1}{n}\sum_i Y_i\Phi(W_i) &&\in \mathbb{R}^M,\\
\widetilde{A} &= \mathbb{E}_n\big[\Phi(W)\Psi(X)^\top\big] &&= \tfrac{1}{n}\sum_i \Phi(W_i)\Psi(X_i)^\top &&\in \mathbb{R}^{M\times J},\\
\varepsilon &= \mathbb{E}_n\big[\Phi(W)U\big] &&= \tfrac{1}{n}\sum_i \Phi(W_i)U_i &&\in \mathbb{R}^M.
\end{aligned}
$$

The IV problem translates to

$$Z = \widetilde{A}\theta_0 + \varepsilon.$$

Here $\widetilde{A}$ is a sample counterpart of the unknown operator $A_0 = \mathbb{E}\big[\Phi(W)\Psi(X)^\top\big]$, and ε is the error vector.

Problem for the analysis: $\widetilde{A}$ is random and correlated with Z and ε.

2 Procedure for a given operator

- Regularization
- Estimation via semidefinite programming
- Oracle bounds
- Incorporating the shape constraints

A partial model for A given (probably misspecified):

$$Z = A\theta + \varepsilon.$$

Least squares contrast

$$L(\theta) = \|Z - A\theta\|^2$$

LSE of θ_0:

$$\widetilde{\theta} = \left(A^*A\right)^{-1} A^*Z$$

where A^* is the conjugated to A and A^*A is a $J \times J$ matrix.

Problem: A smooth and $(A^*A)^{-1}$ unbounded.

A need in a regularization.

- Galerkin (spectral) cut-off: take a relatively small J.
- Tikhonov (ridge regression): use $(A^*A+\lambda I)^{-1}$ in place of $(A^*A)^{-1}$:

$$\widetilde{\theta}_\lambda = \left(A^*A+\lambda I\right)^{-1} A^*Z$$

- (Sobolev) smoothness constraint: given a $J \times J$ matrix S, optimize over the class of θ with $\|S\theta\| \le L$.
- Roughness penalty: given S, define

$$\widetilde{\theta}_{\lambda,S} = \operatorname*{arginf}_{\theta}\left\{\|Z-A\theta\|^2+\lambda\|S\theta\|^2\right\} = (A^*A+\lambda S^2)^{-1}A^*Z.$$

The regularized estimate:

$$\widetilde{\theta}_{\lambda,S} = \underset{\theta}{\operatorname{arginf}}\left\{ \|Z - A\theta\|^2 + \lambda \|S\theta\|^2 \right\} = (A^*A + \lambda S^2)^{-1} A^* Z.$$

The use of regularization introduces some estimation bias: $\widetilde{\theta}_{\lambda,S}$ effectively estimates the vector

$$\theta_{0,\lambda}(A) \stackrel{\text{def}}{=} \underset{\theta}{\operatorname{arginf}} L_\lambda(\theta, A) = (A^*A + \lambda S^2)^{-1} A^* \mathbb{E} Z.$$

If $\mathbb{E} Z \approx f = A\theta_0$, then

$$\theta_{0,\lambda}(A) \approx (A^*A + \lambda S^2)^{-1} A^* A \theta_0$$

Problem: a proper choice of the degree of regularization/smoothness.

- Classical (ML-LS) approach: minimize the statistical loss $\|Z - A\theta\|^2$ under smoothness (roughness) constraints $\|S\theta\| \le L$.

- Dual approach (statistical uncertainty principle): minimize roughness under the model constraint that the residuals $Z - A\theta$ are indistinguishable from the noise with zero mean.

Leads to the problem

$$\widetilde{\theta} \stackrel{\text{def}}{=} \operatorname*{arginf}_{\theta} \|S\theta\| \qquad \text{subject to} \qquad \text{“no signal” in } Z - A\theta.$$

Let Z be supposed to follow $Z = A\theta_0 + \varepsilon$, where $\mathrm{Var}(\varepsilon_m) = \sigma^2$ is given (possibly misspecified).

Given θ, test the hypothesis of “no signal in $Z - A\theta$” by a componentwise check:

$$\|Z - A\theta\|_\infty \le \sigma\mathfrak{z}.$$

A proper choice of the critical value $\mathfrak{z}$ is $\sqrt{(2+\varepsilon)\log M}$.

Level dependent c.v.'s are possible, $\mathfrak{z} = (\mathfrak{z}_1, \ldots, \mathfrak{z}_M)^\top$; cf. Hoffmann and Reiss (2007).

Estimate θ_0 by the constrained minimization:

$$\widetilde{\theta} \stackrel{\text{def}}{=} \operatorname*{arginf}_{\theta} \|S\theta\| \qquad \text{subject to} \qquad \|Z - A\theta\|_{\infty} \leq \sigma\mathfrak{z}$$

A convex problem with a quadratic objective function.

The other measures of smoothness/complexity can be used:

- dimensionality $\|\theta\|_0$; cf. hard thresholding, Donoho and Johnstone (1995).

- sparsity penalty $\|\theta\|_1$, cf. the Dantzig selector, Candes and Tao (2007).

Oracle accuracy

"Oracle" - optimal choice of the tuning parameters (degree of smoothness) for the model.

Adaptive "oracle" results: the procedure mimics the oracle, i.e., it is nearly as good as the oracle.

Alternatively: the procedure performs at least as good as one with any particular choice of the tuning parameter.

Oracle results automatically imply the rate optimality.

Let θ_0 be a "good" choice, that is,

$$Z - A\theta_0 \approx \varepsilon$$

and $\|S\theta_0\| \le L$.

Implies that the constraints $\|Z - A\theta_0\|_\infty \le \sigma\mathfrak{z}$ are fulfilled (at least with a high probab - choice of $\mathfrak{z}$!).

The solution $\widetilde{\theta}$ exists and fulfills $\|Z - A\widetilde{\theta}\|_\infty \le \sigma\mathfrak{z}$ and $\|S\widetilde{\theta}\| \le L$.

Putting together yields

$$\|A(\widetilde{\theta} - \theta_0)\|_\infty \le 2\sigma\mathfrak{z}, \qquad \|S\widetilde{\theta}\| \le L,$$

meaning that $\widetilde{\theta}$ is in a proper vicinity of θ_0.

Implies the correct rate of estimation of θ_0 within the $\mathfrak{z}$-factor.

Let $Z = A_0\theta_0 + \varepsilon$ be the true model.

The use of a wrong operator A in place of A_0 leads to the change of the target:

$$\theta(A) = \operatorname*{arginf}_{\theta} \|A\theta - A_0\theta_0\|.$$

Requirements on A:

$\|A\theta(A) - A_0\theta_0\|$ is not too big and the smoothness condition $\|S\theta(A)\| \le L$ is (nearly) preserved:

$$\|A_0\theta_0 - A\theta(A)\|_\infty \lesssim \sigma\mathfrak{z}, \qquad \|S\theta(A)\| \lesssim L. \tag{1}$$

Then the “oracle” result continues to apply.

The same arguments for σ. Can be misspecified provided that the bound (1) is (nearly) preserved.

Many shape constraints can be represented as a family of linear inequalities. Examples:

- positivity $f \geq 0$,
- monotonicity $f' \geq 0$,
- convexity $f'' \geq 0$.

Kernel characterization: for a family of kernels $K_{h,t}$, it holds

$$K_{h,t} * f \geq 0.$$

Here h is a bandwidth, t, a location. cf. Dümbgen and Sp. (2001).

Translation into the parameter space:

$$\widetilde{K}_{h,t} \theta \geq 0$$

where $\widetilde{K}_{h,t} = K_{h,t} \Psi$.

Let $\psi_1, \ldots, \psi_J$ be a set of linear splines with equidistant knots (a recommended choice), and let $\theta_1, \ldots, \theta_J$ be the corresponding slope coefficients.

- positivity: $c_0 + \theta_1 + \ldots + \theta_j \geq 0$.
- Monotonicity: $\theta_j \geq 0,\ j = 1, \ldots, J$.
- Concavity: $\theta_{j-1} - 2\theta_j + \theta_{j+1} \geq 0$.

Add the translated shape constraints into the problem:

$$\widetilde{\theta} \stackrel{\text{def}}{=} \operatorname*{arginf}_{\theta} \|S\theta\| \qquad \text{subject to} \qquad \|Z - A\theta\|_\infty \leq \sigma_{\mathfrak{z}}, \quad \inf_{h,t} \widetilde{K}_{h,t}\theta \geq 0.$$

Still a feasible convex optimization problem.

The oracle bounds apply provided that there is a good candidate θ_0 fulfilling the constraints.

3 A unknown. Semiparametric approach

Consider the operator A as the nuisance parameter for which a pilot $\widetilde{A}$ is available.

A partial model for A given:

$$Z = A\theta + \varepsilon.$$

Partial least square contrast

$$\widetilde{\theta}(A) = \operatorname*{argmin}_{\theta} \|Z - A\theta\|^2.$$

- Plug-in: The use of $\widetilde{A}$ instead of A.
- Profile: optimize the joint contrast w.r.t. θ and A:

$$\widetilde{\theta} = \operatorname*{argmin}_{\theta} \min_{A} \left\{ \|Z - A\theta\|^2 + \|\widetilde{A} - A\|^2 \right\}.$$

Regularized problem: minimize

$$L(\theta, A) = \|Z - A\theta\|^2 + \|\widetilde{A} - A\|^2$$

subject to $\|S\theta\| \leq L$.

A regularization in the operator is possible; cf. Hoffmann and Reiss (2007). However, in many cases an undersmoothing (taking a proper truncation point in the definition of A) is sufficient.

The plug-in method (the use $\widetilde{A}$ instead of A) is possible for getting the optimal rate, but one does not reach semiparametric efficiency bound.

Classical (ML-LS) approach: minimizing the statistical loss $\|Z - A\theta\|^2 + \|\widetilde{A} - A\|^2$ under smoothness (roughness) constraints on θ and probably on A.

Dual approach: minimize roughness under the constraints that $Z - A\theta$ and $\widetilde{A} - A$ are indistinguishable from the noise.

Leads to the problem

$$\begin{aligned} &\text{minimize } \left\{ \|S\theta\| + \|S_1 A\| \right\} \\ &\text{s.t.} \quad \|Z - A\theta\|_\infty \le \mathfrak{z}\sigma, \ \|\widetilde{A} - A\|_\infty \le \mathfrak{z}_1\sigma. \end{aligned}$$

Pros: relaxation in A relative to the plug-in method.

Cons: the constraint $\|Z - A\theta\|_\infty \le \mathfrak{z}\sigma$ is not convex.

Consider again the shape constraints in the form $\widetilde{K}_{h,t}\theta \geq 0$.

Leads to the problem

$$\begin{aligned} &\text{minimize } \{\|S\theta\| + \|S_1 A\|\} \\ &\text{s.t.} \quad \|Z - A\theta\|_\infty \leq \mathfrak{z}\sigma,\ \|\widetilde{A} - A\|_\infty \leq \mathfrak{z}_1\sigma, \quad \inf_{h,t} \widetilde{K}_{h,t}\theta \geq 0. \end{aligned}$$

Numerical solution via alternating:

- start with $A = \widetilde{A}$.
- solve the partial (in θ) problem with a fixed A (convex o.p.).
- fix the obtained θ and solve w.r.t. A (convex o.p.)
- iterate until convergence

The oracle bounds apply under the condition that $\widetilde{A}$ concentrates in the set of admissible values

$$\|A_0\theta_0 - A\theta(A)\|_\infty \lesssim \sigma_{\mathfrak{z}}, \qquad \|S\theta(A)\| \lesssim L.$$

Requires some minimal smoothness of A_0.