Efficient Instrumental Variables Estimation for Autoregressive Models with Conditional Heteroskedasticity

Guido Kuersteiner

Econometric Theory 2002

Overview

- How can we use the information in past observations to construct an *efficient* IV estimator?
- Can we do better than assuming homoskedasticity and using a Gaussian ML estimator?
- line of attack (common to a lot of papers in this literature):
  1. construct optimal IV estimator that is not feasible
  2. then construct feasible estimator that has the same asymptotic distribution as the infeasible estimator

Setup

- $\phi_0(L) y_t = \varepsilon_t$
- $\phi_0(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$
- $\phi_0(L)$ has all roots outside the unit circle
- ε_t is strictly stationary and ergodic
- $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$
- $\mathrm{E}(\varepsilon_t^2) = \sigma^2 > 0$
- $\alpha_s = E(\varepsilon_t^2 \varepsilon_{t-s}^2) = \sigma(s) + \sigma^4 < \infty$
- $E(\varepsilon_t^2 \varepsilon_{t-s} \varepsilon_{t-r}) = 0 \forall r \neq s$
- $\sum |s| |\sigma(s)| = B < \infty$

IV estimation

- $z_t \in \mathbb{R}^p$ is $\mathcal{F}_{t-1}$ measurable, square integrable, strictly stationary and ergodic
- moment condition:
$$E[\phi_0(L) y_t z_t] = 0 \tag{1}$$
- IV estimator in matrix notation
$$\widetilde{\phi} = (Z' Y_{-1})^{-1} Z' Y \tag{2}$$
- how do we pick z_t if we want to minimize the variance of $\widetilde{\phi}$?
- restrict z_t to be linear in past observations

Some more notation

- $\phi_0^{-1}(L) = \sum_{j=0}^{\infty} \psi_j L^j$
- $b_j' = (\psi_{j-1} \ldots \psi_{j-p})$
- $P_m' = [b_1 \ldots b_m]$
- let $a_j \in \mathbb{R}^p$ be a sequence of vector valued constants
- $\Omega_m = diag(\alpha_1 \ldots \alpha_m)$

The optimal, but infeasible IV estimator

- $z_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j}$
- $$var(\sqrt{n}\widetilde{\phi}) = \lim_{m\to\infty} \sigma^{-4}(P_m' A_m)^{-1} A_m' \Omega_m A_m (A_m' P_m)^{-1} \quad (3)$$
- the minimal variance is given by

$$\Xi = \lim_{m\to\infty} = \sigma^{-4}(P_m' \Omega_m^{-1} P_m)^{-1} \quad (4)$$

- $l_{\psi_0}(L) = \sum_{j=1}^{\infty} b_j/\alpha_j L^j$
- the optimal instruments are then given by $z_t = l_\psi(L)\varepsilon_t$

Why infeasible?

- optimal instruments depend on infinite history of obervables (equivalently, disturbances)
- additionally, the filters $\phi_0(L)$ and $l_{\psi_0}(L)$ are unknown. So in contrast to standard IV we have an estimation error in the instruments as well.

Constructing a feasible estimator

- frequency domain techniques
- turns convolutions (i.e. filters) into products of frequency domain objects
- avoids direct calculation of instruments
- Assuming we only have finite data, but knowing the true filter we can form an estimator of the following form:

$$\widetilde{\phi}(h_0)_{FD} = [\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) h_0^x(\phi_0, \lambda_j)]^{-1} [\sum_{j=1}^{n-1} I_{n,yy} h_0(\phi_0, \lambda_j)] \quad (5)$$

- under the given assumptions we have

$$\widetilde{\phi}(h_0)_{FD} - \widetilde{\phi} = o_p(n^{-1/2}) \tag{6}$$

and

$$\sqrt{n}(\widetilde{\phi}(h_0)_{FD} - \phi_0) \Rightarrow N(0, \Xi) \tag{7}$$

constructing the feasible IV estimator

1. use a consistent but inefficient method (here: OLS) to get initial parameter estimates $\widehat{\phi}$
2. construct sample analogues of filters based on the initial parameter estimate
3. plug those into the formula for the frequency domain estimator to get $\widetilde{\phi}(\widehat{h})_{FD}$
4. then one can show

$$\widetilde{\phi}(\widehat{h})_{FD} - \widetilde{\phi} = o_p(n^{-1/2}) \tag{8}$$

5. Ξ can be consistently estimated using the initial parameter estimates

Asymptotic gains

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{9}$$

$$\varepsilon_t = u_t(.1 + \gamma_1 \varepsilon_{t-1}^2)^{1/2} \tag{10}$$

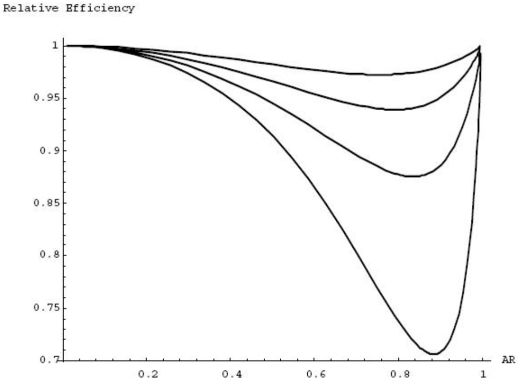

Figure 3.1: *Asymptotic efficiency of OLS relative to the IV estimator as a function of the parameter* ϕ. *Generating mechanisms considered are from bottom to top:* $\gamma_1 = .5$, $\gamma_1 = .4$, $\gamma_1 = .3$ *and* $\gamma_1 = .2$.

Finite sample performance

- IV estimator does about as well as OLS in finite samples (250 observations for the example above), gets the upper hand as sample size increases