NBER Summer Institute

What's New in Econometrics – Time Series

Lecture 4

July 14, 2008

Weak Instruments, Weak Identification, and Many Instruments, Part II

Outline

5) Classical IV regression IV: Estimation

Estimation is much harder than testing or confidence intervals

- Uniformly unbiased estimation is impossible (among estimators with support on the real line), uniformly in μ^2
- Estimation must be divorced from confidence intervals

<u>Partially robust estimators (with smaller bias/better MSE than TSLS):</u>

Remember *k*-class estimators?

$$\hat{\beta}(\underline{k}) = [\mathbf{Y}'(I - \underline{k}M_{\mathbf{Z}})\mathbf{Y}]^{-1}[\mathbf{Y}'(I - \underline{k}M_Z)\mathbf{y}]$$

TSLS: $\underline{k} = 1$,

LIML: $\underline{k} = \hat{\underline{k}}_{LIML}$ = smallest root of $\det(Y^{\perp\prime}Y^{\perp} - \underline{k}Y^{\perp\prime}M_Z Y^{\perp}) = 0$

Fuller: $\underline{k} = \hat{\underline{k}}_{LIML} - c/(T\text{–}k\text{–}\#included\ exog.)$, $c > 0$

Comparisons of *k*-class estimators

Anderson, Kunitomo, and Morimune (1986) – using second order theory
Hahn, Hausman, and Kuersteiner (2004) – using MC simulations

LIML

- median unbiased to second order
- HHK simulations – LIML exhibits very low median bias
- no moments exist! There can be extreme outliers
- LIML also can be shown to minimize the AR statistic:

$$\hat{\beta}^{LIML}\text{: } \min_{\beta} \text{AR}(\beta) = \frac{(\mathbf{y}-\mathbf{Y}\beta_0)'P_{\mathbf{Z}}(\mathbf{y}-\mathbf{Y}\beta_0)/k}{(\mathbf{y}-\mathbf{Y}\beta_0)'M_{\mathbf{Z}}(\mathbf{y}-\mathbf{Y}\beta_0)/(T-k)}$$

so LIML necessarily falls in the AR confidence set if it is nonempty

Comparisons of *k*-class estimators, ctd.

Fuller

- With c = 1, lowest RMSE to second order among a certain class (Rothenberg (1984))
- In simulation studies (m=1), Fuller performs very well with c = 1

Others

- (Jacknife TSLS; bias-adjusted TSLS) are dominated by Fuller, LIML

Summary and recommendations

- Under strong instruments, LIML, TSLS, *k*-class will all be close to each other.
- under weak instruments, TSLS has greatest bias and large MSE
- LIML has the advantage of minimizing AR – and thus always falling in the AR (and CLR) confidence set. LIML is a reasonable (good) choice as an alternative to TSLS.

What about the bootstrap or subsampling?

The bootstrap is often used to improve performance of estimators and tests through bias adjustment and approximating the sampling distribution.

A straightforward bootstrap algorithm for TSLS:

$$y_t = \beta Y_t + u_t$$

$$Y_t = \Pi' Z_t + v_t$$

i) Estimate β, Π by $\hat{\beta}^{TSLS}$, $\hat{\Pi}$

ii) Compute the residuals $\hat{u}_t$, $\hat{v}_t$

iii) Draw T "errors" and exogenous variables from $\{\hat{u}_t, \hat{v}_t, Z_t\}$, and construct bootstrap data $\tilde{y}_t$, $\tilde{Y}_t$ using $\hat{\beta}^{TSLS}$, $\hat{\Pi}$

iv) Compute TSLS estimator (and t-statistic, etc.) using bootstrap data

v) Repeat, and compute bias-adjustments and quantiles from the boostrap distribution, e.g. bias = bootstrap mean of $\hat{\beta}^{TSLS} - \hat{\beta}^{TSLS}$ using actual data

Bootstrap, ctd.

- Under strong instruments, this algorithm works (provides second-order improvements).
- Under weak instruments, this algorithm (or variants) does not even provide first-order valid inference
- The reason the bootstrap fails here is that $\hat{\Pi}$ is used to compute the bootstrap distribution. The true pdf depends on μ^2, say $f_{TSLS}(\hat{\beta}^{TSLS};\mu^2)$ (e.g. Rothenberg (1984 exposition above, or weak instrument asymptotics). By using $\hat{\Pi}$, μ^2 is estimated, say by $\hat{\mu}^2$. The bootstrap correctly estimates $f_{TSLS}(\hat{\beta}^{TSLS};\hat{\mu}^2)$, but $f_{TSLS}(\hat{\beta}^{TSLS};\hat{\mu}^2) \neq f_{TSLS}(\hat{\beta}^{TSLS};\mu^2)$ because $\hat{\mu}^2$ is not consistent for μ^2. ‘

Bootstrap, ctd.

- This is simply another aspect of the nuisance parameter problem in weak instruments. If we could estimate μ^2 consistently, the bootstrap would work – but we if so wouldn't need it anyway (at least to first order) since we would have operational first order approximating distributions!
- This story might be familiar – it is the same reason the bootstrap fails in the unit root model, and in the local-to-unity model, which led to Hansen's (1999) grid bootstrap, which has been shown to produce valid confidence intervals for the AR(1) coefficient by Mikusheva (2007).
- Failure of bootstrap in weak instruments is related to failure of Edgeworth expansion (uniformly in the strength of the instrument), see Hall (1992) in general, Moreira, Porter, and Suarez (2005a,b) in particular.

Bootstrap, ctd.

- One way to avoid this problem is to bootstrap test statistics with null distributions that do not depend on μ^2. Bootstrapping AR and LM *does* result in second order improvements, see Moreira, Porter, and Suarez (2005a,b).

What about subsampling?

Politis and Romano (1994), Politis, Romano and Wolf (1999)

Subsampling uses smaller samples of size *m* to estimate the parameters directly. If the CLT holds, the distribution of the subsample estimators, scaled by $\sqrt{m/T}$, approximates the distribution of the full-sample estimator.

A subsampling algorithm for TSLS:

(i) Choose subsample of size *m* and compute TSLS estimator

(ii) Repeat for all subsamples of size *m* (in cross-section, there are $\binom{T}{m}$ such subsamples; in time series, there are $T–m$)

(iii) Compute bias adjustments, quantiles, etc. from the rescaled empirical distribution of the subsample estimators.

Subsampling, ctd.

- Subsampling works in some cases in which bootstrap doesn't (Politis, Romano, and Wolf (1999))
- However, it doesn't work (doesn't provide first-order valid approximations to sampling distributions) with weak instruments (Andrews and Guggenberger (2007a,b)).
- The subsampling distribution estimates $f_{TSLS}(\hat{\beta}^{TSLS}; \mu_m^2)$, where μ_m^2 is the concentration parameter for m observations. But this is less (on average, by the factor m/T) than the concentration parameter for T observations, so the scaled subsample distribution does not estimate $f_{TSLS}(\hat{\beta}^{TSLS}; \mu_T^2)$.
- Subsampling can be size-corrected (in this case) but there is power loss relative to CLR; see Andrews and Guggenberger (2007b)

Outline

1) What is weak identification, and why do we care?
2) Classical IV regression I: Setup and asymptotics
3) Classical IV regression II: Detection of weak instruments
4) Classical IV regression III: hypothesis tests and confidence intervals
5) Classical IV regression IV: Estimation
6) **GMM I: Setup and asymptotics**
7) GMM II: Detection of weak identification
8) GMM III: Hypothesis tests and confidence intervals
9) GMM IV: Estimation
10) Many instruments

6) GMM I: Setup and asymptotics

GMM notation and estimator

GMM "error" term (G equations): $h(Y_t;\theta)$; θ_0 = true value

Errors times k instruments: $\phi_t(\theta) = \underset{G\times 1}{h(Y_t,\theta_0)} \otimes \underset{k\times 1}{Z_t}$

Moment conditions - k instruments: $E\phi_t(\theta) = E[\underset{G\times 1}{h(Y_t,\theta_0)} \otimes \underset{k\times 1}{Z_t}] = 0$

GMM objective function: $S_T(\theta) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]' W_T \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]$

GMM estimator: $\hat{\theta}$ minimizes $S_T(\theta)$

Linear GMM: $h(Y_t;\theta) = y_t - \theta' Y_t$

(linear GMM is the IV regression model, allowing for possible heteroskedasticity and/or serial correlation in the errors h)

Efficient GMM

Centered sample moments: $\Psi_T(\theta) = T^{-1/2}\sum_{t=1}^{T}\left(\phi_t(\theta) - E\phi_t(\theta)\right)$

Efficient (infeasible) GMM: $W_T = \Omega^{-1}$, $\Omega = E[\Psi_T(\theta)\Psi_T(\theta)'] = 2\pi S_{\phi_t(\theta)}(0)$

Feasible GMM

Estimator of Ω: $\hat{\Omega}(\theta)$ = HAC estimator of Ω $= \sum_{j=-S}^{S}\kappa_j\hat{\Gamma}_j(\theta)$,

where $\hat{\Gamma}_j(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\theta) - \overline{\phi_t(\theta)}\right)\left(\phi_{t-j}(\theta) - \overline{\phi_{t-j}(\theta)}\right)'$

$\{\kappa_j\}$ are kernel weights (e.g. Newey-West)

Feasible GMM variants

One-step W_T = fixed matrix (e.g. $W_T = I$)

Two-step efficient: $W_T^{(1)} = I$, $W_T^{(2)} = \hat{\Omega}(\hat{\theta}^{(1)})^{-1}$

Iterated: continue iterating, with $W_T^{(i+1)} = \hat{\Omega}(\hat{\theta}^{(i)})^{-1}$

CUE (Hansen, Heaton, Yaron 1996): $W_T = \hat{\Omega}(\theta)^{-1}$ (evaluate $\hat{\Omega}$ at every θ!)

Standard asymptotics

1) Establish consistency by showing the minimum of S_T will occur local to the true value θ_0: $\Pr[S_T(\theta) < S_T(\theta_o)] \to 0$ for $|\theta - \theta_0| > \varepsilon$
so by smoothness of the objective function, $\Pr[|\hat{\theta} - \theta_0| > \varepsilon] \to 0$

2) Establish normality by making quadratic approximation to S_T, based on consistency (which justifies dropping the higher order terms in the Taylor expansion):

$$S_T(\hat{\theta}) \approx S_T(\theta_0) + \sqrt{T}(\hat{\theta} - \theta_0)' \frac{1}{\sqrt{T}} \left.\frac{\partial S_T(\theta)}{\partial \theta}\right|_{\theta_0}$$

$$+ \tfrac{1}{2}\sqrt{T}(\hat{\theta} - \theta_0)' \left[\frac{1}{T} \left.\frac{\partial^2 S_T(\theta)}{\partial \theta \partial \theta'}\right|_{\theta_0} \right] \sqrt{T}(\hat{\theta} - \theta_0)$$

so $$\sqrt{T}(\hat{\theta} - \theta_0) \approx \left[\frac{1}{T} \left.\frac{\partial^2 S_T(\theta)}{\partial \theta \partial \theta'}\right|_{\theta_0} \right]^{-1} \frac{1}{\sqrt{T}} \left.\frac{\partial S_T(\theta)}{\partial \theta}\right|_{\theta_0}$$

If $W_T \xrightarrow{p} W$ (say), then

$$\left.\frac{1}{T}\frac{\partial^2 S_T(\theta)}{\partial\theta\partial\theta'}\right|_{\theta_0} \xrightarrow{p} DWD', \text{ where } D = E\left.\frac{\partial\phi_t(\theta)}{\partial\theta}\right|_{\theta_0}$$

$$\left.\frac{1}{\sqrt{T}}\frac{\partial S_T(\theta)}{\partial\theta}\right|_{\theta_0} \xrightarrow{d} N(0, DW\Omega W'D')$$

so

$$\sqrt{T}(\hat{\theta} - \theta_0) \approx \left[\left.\frac{1}{T}\frac{\partial^2 S_T(\theta)}{\partial\theta\partial\theta'}\right|_{\theta_0}\right]^{-1} \left.\frac{1}{\sqrt{T}}\frac{\partial S_T(\theta)}{\partial\theta}\right|_{\theta_0}$$

$$\xrightarrow{d} N(0, [DWD']^{-1}DW\Omega W'D'[DWD']^{-1})$$

<u>Feasible efficient GMM</u>

For two-step, iterated, and CUE, $W_T \xrightarrow{p} \Omega^{-1}$, so $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma)$

where $\Sigma = (\mathrm{D}\Omega^{-1}\mathrm{D}')^{-1}$

Estimator of variance matrix: $\hat{\Sigma} = [\hat{D}(\hat{\theta})\hat{\Omega}(\hat{\theta})\hat{D}(\hat{\theta})']^{-1}$

Weak identification in GMM – what goes wrong in the usual proof?

Digression:

- We will use the term "weak identification" because "weak instruments" is not precise in the nonlinear setting
- In the linear case, the strength of the instruments doesn't depend on θ
- In nonlinear GMM, the strength of the instruments can depend on θ: they can be weak for some departures $h(Y_t,\theta)$ - $h(Y_t,\theta_0)$, but strong for others

When identification is weak, there are 3 problems with the usual proof:

(a) The curvature, which reflects the amount of information, is small, so the maximum of S_T might not be close to θ_0.

(b) The curvature matrix is not well-approximated as nonrandom

(c) The linear term, $\left.\frac{\partial S_T(\theta)}{\partial \theta}\right|_{\theta_0}$, is not approximately normal with mean 0

Illustration: linear IV in the GMM framework

The TSLS objective function (two-step GMM) is exactly quadratic:

$$S(\theta) = (\mathbf{y} - \mathbf{Y}\theta)'P_{\mathbf{Z}}(\mathbf{y} - \mathbf{Y}\theta)$$

$$= [\mathbf{u} - \mathbf{Y}(\theta - \theta_0)]'P_{\mathbf{Z}}[\mathbf{u} - \mathbf{Y}(\theta - \theta_0)]$$

$$= \mathbf{u}'P_{\mathbf{Z}}\mathbf{u} + (2\mathbf{u}'P_{\mathbf{Z}}\mathbf{Y})(\theta - \theta_0) - ½(\theta - \theta_0)'(2\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y})(\theta - \theta_0)$$

or

$$S_T(\hat{\theta}) = S_T(\theta_0) + \sqrt{T}(\hat{\theta} - \theta_0)'\frac{1}{\sqrt{T}}\left.\frac{\partial S_T(\theta)}{\partial\theta}\right|_{\theta_0}$$

$$+ ½\sqrt{T}(\hat{\theta} - \theta_0)'\left[\frac{1}{T}\left.\frac{\partial^2 S_T(\theta)}{\partial\theta\partial\theta'}\right|_{\theta_0}\right]\sqrt{T}(\hat{\theta} - \theta_0)$$

where $S_T(\theta_0) = \mathbf{u}'P_{\mathbf{Z}}\mathbf{u}$

$$\frac{1}{\sqrt{T}}\left.\frac{\partial S_T(\theta)}{\partial\theta}\right|_{\theta_0} = 2\mathbf{u}'P_{\mathbf{Z}}\mathbf{Y}$$

$$\frac{1}{T}\left.\frac{\partial^2 S_T(\theta)}{\partial\theta\partial\theta'}\right|_{\theta_0} = 2\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y}$$

Illustration: linear IV in the GMM framework, ctd.

(a) The curvature is small (so estimator need not be local)

$$\left.\frac{1}{T}\frac{\partial^2 S_T(\theta)}{\partial\theta\partial\theta'}\right|_{\theta_0} = 2\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y}$$

$$= 2\frac{\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y}/k}{\mathbf{Y}'M_{\mathbf{Z}}\mathbf{Y}/(T-k)}\mathbf{Y}'M_{\mathbf{Z}}\mathbf{Y}/(T-k)$$

$$= 2kFs_v^2,$$

where F is the first-stage F and s_v^2 is the estimator of σ_v^2.

(b) The curvature is random – not well approximated by a constant

$$F/\mu^2 \to 1 \text{ as } \mu^2 \to \infty, \text{ but for small } \mu^2,\ F = \mu^2 + o_p(1)$$

(c) Under weak instrument asymptotics, the linear term is:

$$\left.\frac{1}{\sqrt{T}}\frac{\partial S_T(\theta)}{\partial\theta}\right|_{\theta_0} = 2\mathbf{u}'P_{\mathbf{Z}}\mathbf{Y} \xrightarrow{d} 2(\lambda + z_v)'z_u,$$

which has a mixture-of-normals distribution with a nonzero mean (recall the distribution of TSLS under weak instrument asymptotics)

Alternative asymptotics for weak identification

As in the linear case, we need asymptotics for GMM that are tractable; provides good approximations uniformly in strength of identification; and can be used to compare procedures.

Alternative approaches:

1. Finite sample – good luck
2. Edgeworth and related expansions – useful for developing partially robust procedures but won't cover complete range through unidentified case
3. Bootstrap & resampling – doesn't work in linear IV special case
4. Weak identification asymptotics – provide nesting (parameter sequence) that provides an approximation uniformly in strength of identification

Weak ID asymptotics in GMM

(Stock and Wright (2000))

Use local sequence (sequence of mean functions) to provide non-quadratic global approximation to $S_T(\theta)$:

$$S_T(\theta) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]' W_T \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]$$

Write

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta) &= T^{-1/2}\sum_{t=1}^{T}\left[\phi_t(\theta) - E\phi_t(\theta)\right] + T^{-1/2}\sum_{t=1}^{T}E\phi_t(\theta) \\ &= \Psi_T(\theta) + \sqrt{T}\,E\phi_t(\theta) \\ &= \Psi_T(\theta) + m_T(\theta) \end{aligned}$$

Weak ID asymptotics in GMM, ctd.

Applied to the linear IV regression model, this reorganization yields,

$$T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta) = T^{-1/2}\sum_{t=1}^{T}(y_t - \theta' Y_t)Z_t$$

$$= T^{-1/2}\sum_{t=1}^{T}\left(u_t - (\theta-\theta_0)' Y_t\right)Z_t$$

$$= T^{-1/2}\sum_{t=1}^{T}\zeta_t \; - E\left(T^{-1/2}\sum_{t=1}^{T}(\theta-\theta_0)' Y_t Z_t\right)$$

$$= \Psi_T(\theta) + m_T(\theta)$$

where $\zeta_t = u_t Z_t - \left[(\theta-\theta_0)' Y_t Z_t - E(\theta-\theta_0)' Y_t Z_t\right]$. Now:

- $\Psi_T(\theta) = T^{-1/2}\sum_{t=1}^{T}\zeta_t \xrightarrow{d}$ N(0, Ω) (because ζ_t is mean zero and i.i.d. – instrument strength doesn't enter this limit (subtracted out))
- The mean function $m_T(\theta)$ is a finite nonrandom (linear) function under the local nesting $\Pi = T^{-1/2}C$

Weak ID asymptotics in GMM, ctd.

$$T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta) = T^{-1/2}\sum_{t=1}^{T}\left[\phi_t(\theta)-E\phi_t(\theta)\right]+T^{-1/2}\sum_{t=1}^{T}E\phi_t(\theta) = \Psi_T(\theta) + m_T(\theta)$$

Suppose:

1. $m_T \overset{p}{\to} m$ uniformly in θ, where $m(\theta)$ is a limiting (finite continuous differentiable) function.

   This is the function extension of assuming $\Pi = T^{-1/2}C$

2. $\Psi_T(\bullet) \Rightarrow \Psi(\bullet)$, where $\Psi(\theta)$ is a Gaussian stochastic process on Θ with mean zero and covariance function $\Omega(\theta_1,\theta_2) = E\,\Psi(\theta_1)\,\Psi(\theta_2)'$

Weak ID asymptotics in GMM, ctd.

2. $\Psi_T \Rightarrow \Psi$, where $\Psi(\theta)$ is a Gaussian stochastic process on Θ with mean zero and covariance function $\Omega(\theta_1, \theta_2) = E\Psi(\theta_1)\Psi(\theta_2)'$

Digression on $\Psi_T \Rightarrow \Psi$:

Item #2 is an extension of the FCLT. Generally, the FCLT talks about convergence in distribution of a sequence of random functions, to a limiting function, which has a (limiting) distribution. In the time series FCLT introduced in Lecture 2, the function is indexed by s = τ/T $\in$ [0,1], and the limiting process has the covariance matrix of Brownian motion (it is Brownian motion). Here, the function is indexed by θ, and the limiting process has the covariance matrix $\Omega(\theta_1, \theta_2)$. The proof of the FCLT entailed proving:

Weak ID asymptotics in GMM, ctd.

i. *Convergence of finite dimensional distributions*. Here, this corresponds the joint distributions of $\Psi_T(\theta_1)$, $\Psi_T(\theta_2),\ldots,$ $\Psi_T(\theta_r)$. But $\Psi_T(\theta) = T^{-1/2}\sum_{t=1}^{T}\left[\phi_t(\theta) - E\phi_t(\theta)\right]$, so it is a weak (standard) assumption that $\Psi_T(\theta_1)$, $\Psi_T(\theta_2),\ldots,$ $\Psi_T(\theta_r)$ will converge jointly to a normal; the covariance matrix is filled out using $\Omega(\theta_1,\theta_2)$ (applied to all the points).

ii. *Tightness (or stochastic equicontinuity)*. That is, for θ_1 and θ_2 close, that $\Psi_T(\theta_1)$ and $\Psi_T(\theta_2)$ must be close (with high probability). This allows going from the function evaluated at finitely many points, to the function itself. Proving this is application specific (depends on $h(Y_t,\theta)$). Proof in the linear GMM case is in Stock and Wright (2000).

Weak ID asymptotics in GMM, ctd.

Back to main argument…

Under 1 and 2, $T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta) \Rightarrow \Psi(\theta) + m(\theta)$

3. $W_T(\theta) \xrightarrow{p} W(\theta)$ uniformly in θ, where $W(\theta)$ is psd, continuous in θ

Under 1, 2, and 3, $S_T(\theta) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]' W_T(\theta)\left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]$

$$\Rightarrow S(\theta) = [\psi(\theta) + m(\theta)]'W[\psi(\theta) + m(\theta)]$$

and

$$\hat{\theta} \Rightarrow \theta^*, \text{ where } \theta^* = \text{argmin } S(\theta)$$

Weak ID asymptotics in GMM, ctd.

$$\hat{\theta} \Rightarrow \theta^* = \text{argmin}\ \{S(\theta) = [\Psi(\theta) + m(\theta)]'W[\Psi(\theta) + m(\theta)]\}$$

Comments

- With $\phi_t(\theta) = (y_t - \theta' Y_t)Z_t$ and $W_T = (\mathbf{Z'Z}/T)^{-1}$, this yields the weak IV asymptotic distribution of TSLS obtained earlier.
- $S_T(\theta)$ is not well approximated by a quadratic (is not quadratic in the limit) with a nonrandom curvature matrix that gets large – instead, $S_T(\theta)$ is $O_p(1)$
- $\hat{\theta}$ is not consistent in this setup
- $\hat{\theta}$ has a nonstandard limiting distribution
- Standard errors of $\hat{\theta}$ aren't meaningful ($\pm 1.96SE$ isn't valid conf. int.)
- J-statistic doesn't have chi-squared distribution
- Well-identified elements of $\hat{\theta}$ have the usual limiting normal distributions, under the true values of the weakly identified elements
- Extensions and proofs are in Stock and Wright (2000)

Outline

1) What is weak identification, and why do we care?
2) Classical IV regression I: Setup and asymptotics
3) Classical IV regression II: Detection of weak instruments
4) Classical IV regression III: hypothesis tests and confidence intervals
5) Classical IV regression IV: Estimation
6) GMM I: Setup and asymptotics
7) **GMM II: Detection of weak identification**
8) GMM III: Hypothesis tests and confidence intervals
9) GMM IV: Estimation
10) Many instruments

7) GMM II: Detection of weak identification

This is an open area of research with no best solution. Some thoughts:

1. In linear GMM, the noncentrality parameter of the ***first-stage F*** and the concentration parameter are no longer the same thing if there is heteroskedasticity and/or serial correlation in $h(Y_t,\theta)$. With heteroskedasticity, the first-stage F still provides a reasonable guide (MC findings) but with serial correlation the first stage F isn't very reliable.

2. ***Wright (2003)*** provides a test for weak instruments, based on the extension of the Cragg-Donald (1993) using the estimated curvature of the objective function. The test is a test of non-identification (contrast with Stock-Yogo, testing whether μ^2 exceeds a critical cutoff; in

Wright (2003), the cutoff is taken to be $\mu^2 = 0$ in linear IV case). The test is conservative, which gives it low power against weak identification – a benefit in this instance. Important drawback is that it is only local (multiple peak problem).

3.***Some symptoms of weak identification*****:**

- CUE, two-step, and interated GMM converge to quite different values (see Hansen, Heaton, Yaron (1996) MC results)
- for two-step and iterated, the normalization matters
- multiple valleys in the CUE objective function
- Significant discrepancies between GMM-AR confidence sets (discussed below) and conventional Wald confidence sets

Outline

1) What is weak identification, and why do we care?
2) Classical IV regression I: Setup and asymptotics
3) Classical IV regression II: Detection of weak instruments
4) Classical IV regression III: hypothesis tests and confidence intervals
5) Classical IV regression IV: Estimation
6) GMM I: Setup and asymptotics
7) GMM II: Detection of weak identification
8) **GMM III: Hypothesis tests and confidence intervals**
9) GMM IV: Estimation
10) Many instruments

8) GMM III: Hypothesis tests and confidence intervals

Extensions of methods in linear IV:

<u>(1) The GMM-Anderson Rubin statistic</u>

(Kocherlakota (1990); Burnside (1994), Stock and Wright (2000)) The extension of the AR statistic to GMM is the CUE objective function evaluated at θ_0:

$$S_T^{CUE}(\theta_0) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_0)\right]'\hat{\Omega}(\theta_0)^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_0)\right]$$

$$\xrightarrow{d} \psi(\theta_0)'\Omega(\theta_0)^{-1}\Psi(\theta_0) \sim \chi_k^2$$

- Thus a valid test of H_0: $\theta = \theta_0$ can be undertaken by rejecting if $S_T(\theta_0)$ > 5% critical value of χ_k^2.

The GMM-Anderson Rubin statistic, ctd

- The statistic above tests all elements of θ. If some elements are strongly identified, they can be concentrated out (estimated under the null) for valid subset inference. Specifically, let $\theta = (\alpha, \beta)$, and let α be weakly identified and β be strongly identified. Fix α at the hypothesized value α_0 and let $\hat{\beta}^{GMM}$ be an efficient GMM estimator of β, at the given value of α_0. Then construct the CUE objective function, using the hypothesized value of α and the estimated value of β:

$$S_T^{CUE}(\alpha_0, \hat{\beta}^{GMM}) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\alpha_0, \hat{\beta}^{GMM})\right]'\hat{\Omega}(\alpha_0, \hat{\beta}^{GMM})^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\alpha_0, \hat{\beta}^{GMM})\right]$$

The statistic $S_T^{CUE}(\alpha_0, \hat{\beta}^{GMM})$ has a $\chi^2_{k-\dim(\beta)}$ distribution under H_0: $\alpha = \alpha_0$, and is a weak-identification robust test statistic for H_0: $\alpha = \alpha_0$.

GMM-Anderson-Rubin, ctd.

In the homoskedastic linear IV model, the GMM-AR statistic simplifies to the AR statistic (up to a degrees of freedom correction):

$$S_T^{CUE}(\theta_0) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_0)\right]'\hat{\Omega}(\theta_0)^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_0)\right]$$

$$= \left[T^{-1/2}\sum_{t=1}^{T}(y_t - \theta_0{}'Y_t)Z_t\right]'\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}s_v^2\right)^{-1}\left[T^{-1/2}\sum_{t=1}^{T}(y_t - \theta_0{}'Y_t)Z_t\right]$$

$$= \frac{(\mathbf{y}-\mathbf{Y}\theta_0)'P_{\mathbf{Z}}(\mathbf{y}-\mathbf{Y}\theta_0)}{(\mathbf{y}-\mathbf{Y}\theta_0)'M_{\mathbf{Z}}(\mathbf{y}-\mathbf{Y}\theta_0)/(T-k)} = k \times \text{AR}(\theta_0)$$

Comments:

- The statistic, $S_T^{CUE}(\theta_0)$, is called various things in the literature, including the *S*-statistic, the CUE objective function statistic, the nonlinear AR statistic, and the GMM-AR statistic. I think GMM-AR is the most descriptive and we will use that term here.

GMM-Anderson-Rubin, ctd.

- The GMM-AR statistic has the same issues of interpretation issues as the AR, specifically, the GMM-AR rejects because of endogenous instruments and/or incorrect θ
- With little information, the GMM-AR can fail to reject any values of θ (remember the Dufour (1997) critique of Wald tests)

(2) GMM-LM

Kleibergen (2005) – develops score statistic (based on CUE objective function – details of construction matter) that provides weak-identification valid hypothesis testing for sets of variables

(3) GMM-CLR

Andrews, Moreira, Stock (2006) – extension of CLR to linear GMM with a single included endogenous regressor, also see Kleibergen (2007). Very limited evidence on performance exists; also problem of dimension of conditioning vector

(4) Other methods

Guggenberger-Smith (2005) objective-function based tests based on Generalized Empirical Likelihood (GEL) objective function (Newey and Smith (2004)); Guggenberger-Smith (2008) generalize these to time series data. Performance is similar to CUE (asymptotically equivalent under weak instruments)

Confidence sets

- Fully-robust 95% confidence sets are obtained by inverting (are the acceptance region of) fully-robust 5% hypothesis tests
- Computation is by grid search in general: collect all the points θ which, when treated as the null, are not rejected by the GMM-AR statistic.
- Subsets by projection (see Kleibergen and Mavroeidis (2008) for an application of GMM-AR confidence sets and subsets)
- Valid tests must be unbounded (contain Θ) with finite probability with weak instruments

Bottom line recommendation

Work is under way in this area, but the best thing for now is to use the GMM-AR statistic to test $\theta = \theta_0$, and to invert the GMM-AR statistic to construct the GMM version of the AR confidence set. The GMM-AR statistic must in general be inverted by grid search. The GMM-AR confidence set, if nonempty, will contain the CUE estimator.

Outline

1) What is weak identification, and why do we care?
2) Classical IV regression I: Setup and asymptotics
3) Classical IV regression II: Detection of weak instruments
4) Classical IV regression III: hypothesis tests and confidence intervals
5) Classical IV regression IV: Estimation
6) GMM I: Setup and asymptotics
7) GMM II: Detection of weak identification
8) GMM III: Hypothesis tests and confidence intervals
9) **GMM IV: Estimation**
10) Many instruments

9) GMM IV: Estimation

- Impossibility of a (data-based) fully robust estimators are available – just just as in linear case
- The challenge is to find partially robust estimators – estimators that improve upon 2-step and iterated GMM (which perform terribly – just like TSLS)

(a) The continuous updating estimator (CUE)

Hansen, Heaton, Yaron (1996). The CUE minimizes,

$$S_T^{CUE}(\theta) = \left[T^{-1/2} \sum_{t=1}^{T} \phi_t(\theta) \right]' \hat{\Omega}(\theta)^{-1} \left[T^{-1/2} \sum_{t=1}^{T} \phi_t(\theta) \right]$$

Basic idea: "same θ in the numerator and the denominator".

CUE, ctd

Comments

- The CUE might seem arbitrary but actually it isn't. In fact, it was shown above that in the linear model with spherical errors, the CUE objective function *is* the AR statistic, $S_T^{CUE}(\theta) = \text{AR}(\theta)$. It was stated above (without proof) that LIML minimizes the AR statistic. So in the special case of linear GMM when there is no heteroskedasticity or serial correlation, the CUE estimator is LIML (asymptotically under weak instrument asymptotics if Ω is estimated).
- CUE will always be contained in the GMM-AR set
- The CUE seems to inherit median unbiasedness of LIML (MC result; for some theory see Hausman, Menzel, Lewis, and Newey (2007))
- CUE (like LIML) exhibits wide dispersion in MC studies (Guggenberger 2005)

(b) Other estimators

- Generalized empirical likelihood (GEL) family. Interestingly, GEL estimators are asymptotically equivalent to CUE under weak instrument asymptotics (Guggenberger and Smith (2005))
- Fuller-*k* type modifications explored in Hausman, Menzel, Lewis, and Newey (2007), with some simulation evidence.
- These alternative estimators are promising but preliminary and their properties, including the extent to which they are robust to weak instruments in practice, are not yet fully understood.

Example #3 (linear GMM): New Keynesian Phillips Curve

References: Gali and Gertler (1999); Mavroedis (2005), Nason and Smith (2007), Dufour, Khalath, and Kichian (2006), Kleibergen and Mavroeidis (2008)

Hybrid NKPC: $\pi_t = \lambda x_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \eta_t$

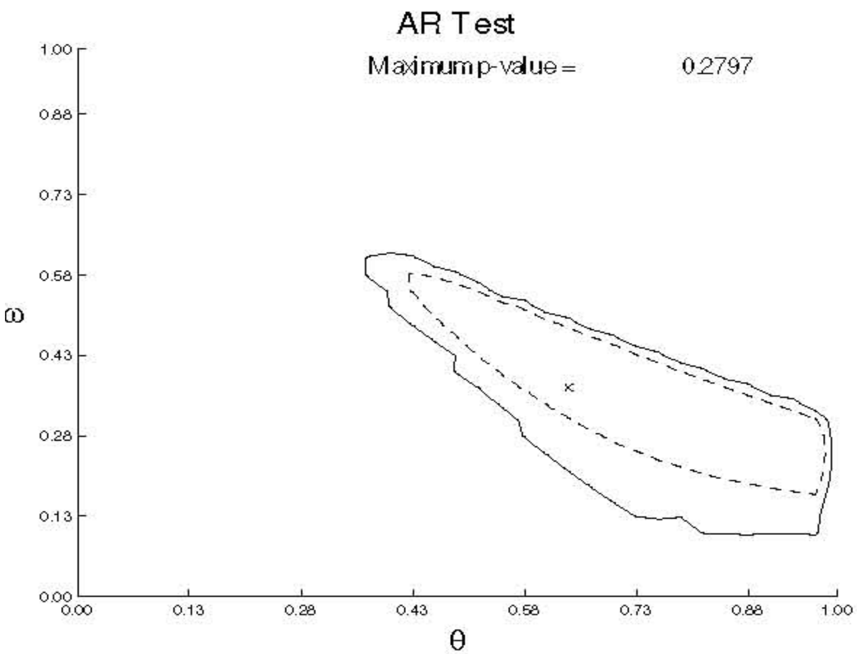
Rational expectations: $E_t(\pi_t - \lambda x_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}) = 0$

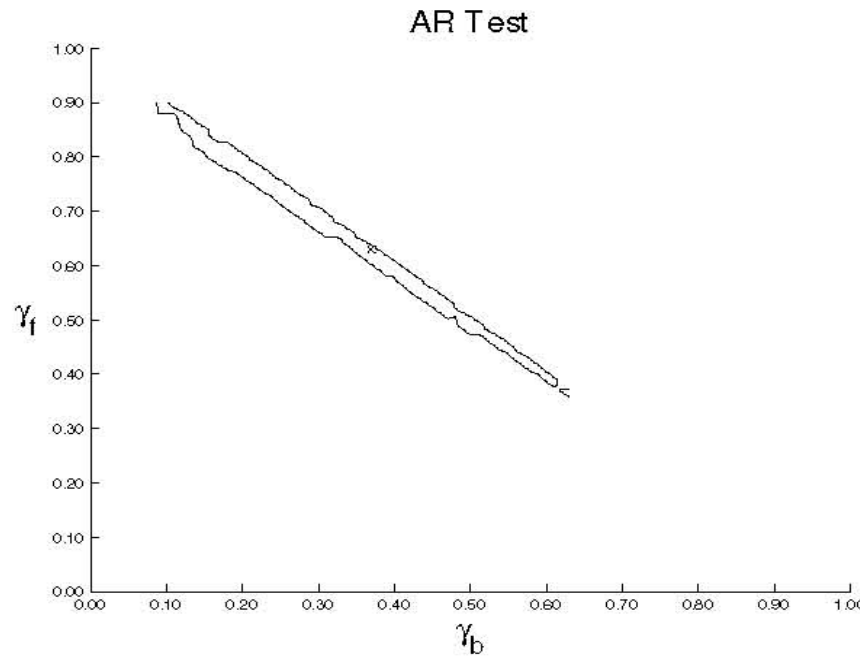
GMM moment condition: $E[(\pi_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1} - \lambda x_t) Z_t] = 0$

Instruments: $Z_t = \{\pi_{t-1}, x_{t-1}, \pi_{t-2}, x_{t-2}, \ldots\}$

$m = 2$, so AR sets are needed. Confidence intervals can be computed by projecting the sets to the axes – see the example from Kleibergen and Mavroeidis (2008) below:

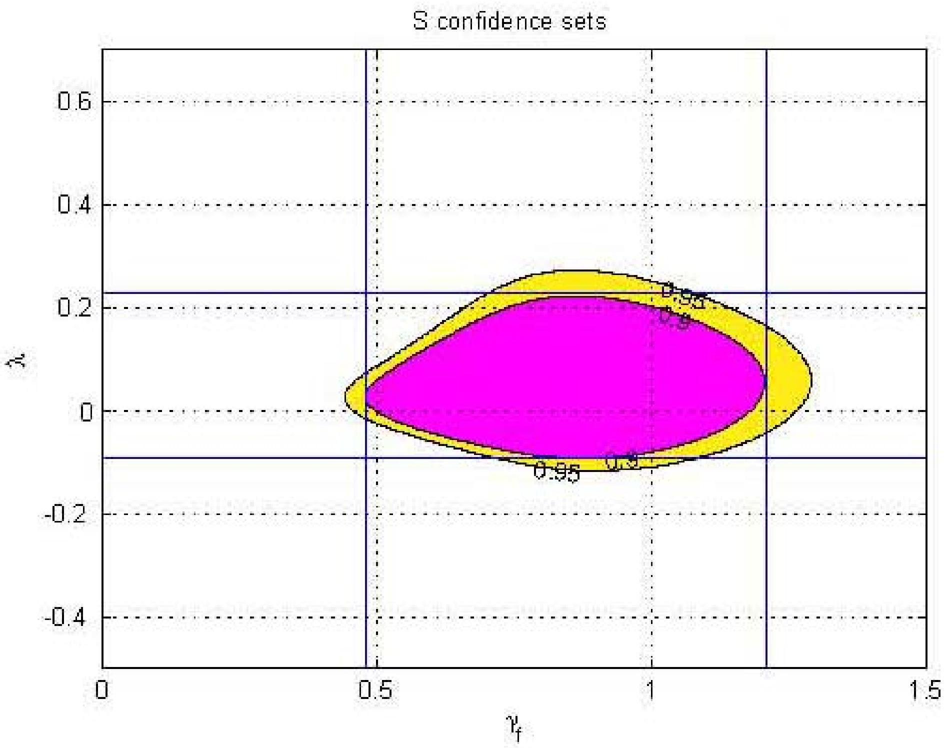
Anderson-Rubin confidence sets for NKPC parameters from Dufour, Khalath, and Kichian (2006) (2-dim confidence sets computed by grid search as nonrejection region of AR(β) test)

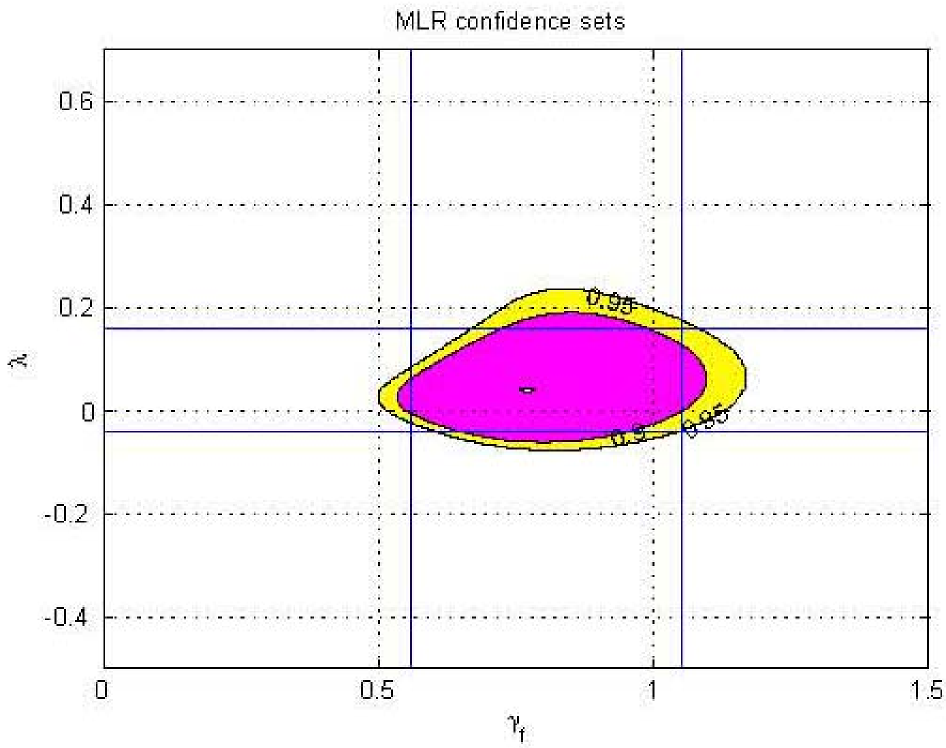




- Confidence set isn’t an ellipse (it could be disjoint!)
- Set is obtained by trying lots of values – over the grid

Kleibergen-Mavroedis confidence sets using *S*-sets (GMM-AR sets) (left) and also by inverting a bivariate linear GMM extension of the CLR statistic (right). Confidence intervals are computed by projecting the set onto the axes:





Example #4: Nonlinear CCAPM

Stock and Wright (2000), Neely, Roy, Whiteman (2001)

$$h(Y_t,\theta) = \delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \overset{G\times 1}{R_{t+1}} - \iota_G$$

where R_{t+1} is a $G\times 1$ vector of asset returns and ι_G is the G-vector of 1's.

GMM moment conditions (Hansen-Singleton (1982)):

Specific illustration (Stock and Wright (2005)) – AR sets

Annual data, 1871-1993

Two equations: stock returns & bond returns

Instruments: SR_{t-2}, BR_{t-2}, Δc_{t-2} (2nd lag because of temporal aggregation)

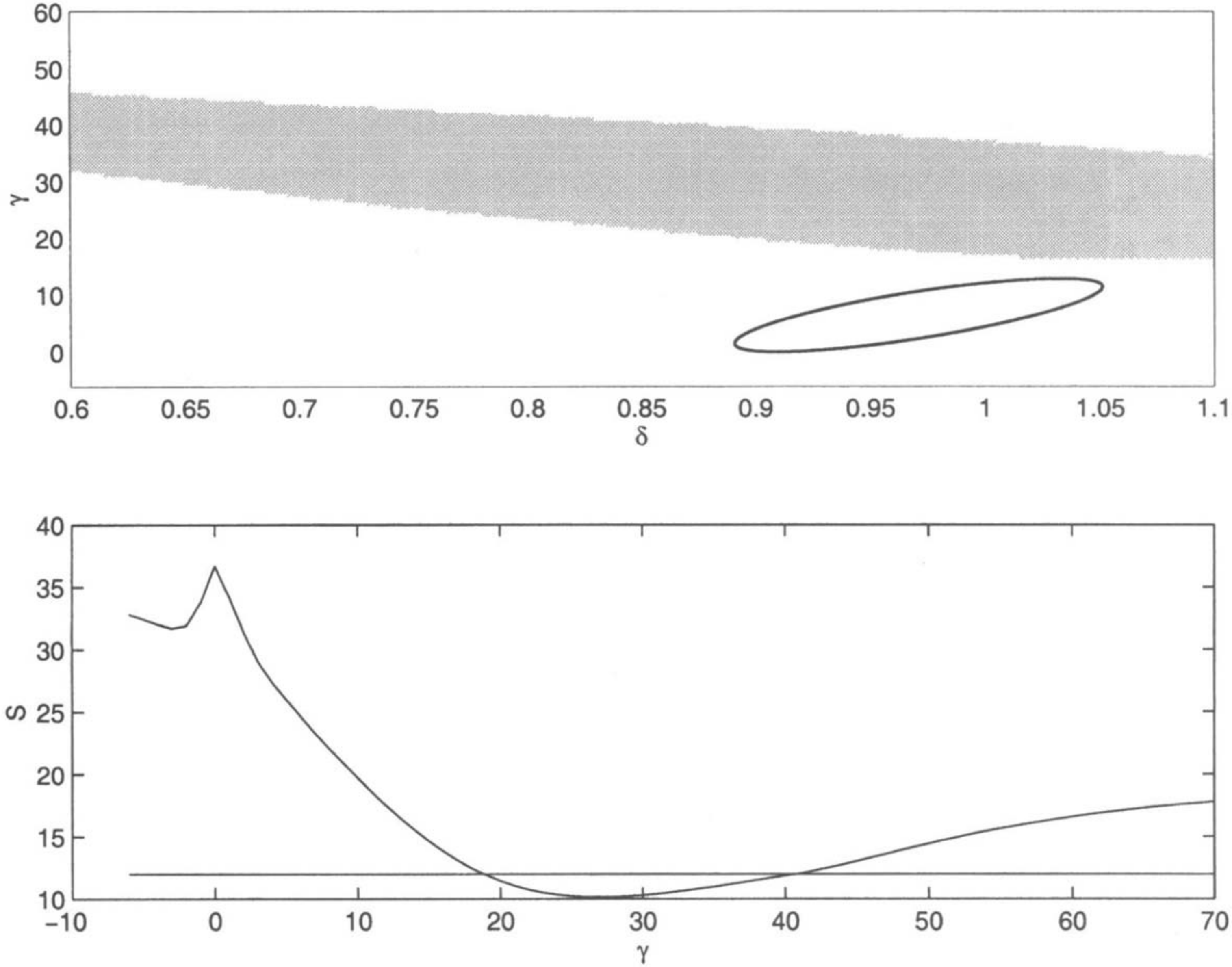


FIGURE 3.—Joint S-set and concentrated objective function: model CRRA-5. (a) Joint 90% S-set (shaded) and 90% GMM confidence ellipse for (γ, δ) (upper panel); (b) objective function concentrated with respect to δ (lower panel).

Outline

1) What is weak identification, and why do we care?
2) Classical IV regression I: Setup and asymptotics
3) Classical IV regression II: Detection of weak instruments
4) Classical IV regression III: hypothesis tests and confidence intervals
5) Classical IV regression IV: Estimation
6) GMM I: Setup and asymptotics
7) GMM II: Detection of weak identification
8) GMM III: Hypothesis tests and confidence intervals
9) GMM IV: Estimation
10) **Many instruments**

10) Many Instruments

<u>The appeal of using many instruments</u>

- Under standard IV asymptotics, more instruments means greater efficiency.
- This story is not very credible because
  (a) the instruments you are adding might well be weak (you already have used the first two lags, say) and
  (b) even if they are strong, this requires consistent estimation of increasingly many parameter to obtain the efficient projection – hence slow rates of growth of the number of instruments in efficient GMM literature.

Example of problems with many weak instruments – TSLS

Recall the TSLS weak instrument asymptotic limit:

$$\hat{\beta}^{TSLS} - \beta_0 \xrightarrow{d} \frac{(\lambda + z_v)' z_u}{(\lambda + z_v)'(\lambda + z_v)}$$

with the decomposition, $z_u = \delta z_v + \eta$. Suppose that k is large, and that $\lambda'\lambda/k \to \Lambda_\infty$ (one way to implement "many weak instrument asymptotics"). Then as $k \to \infty$,

$$\lambda' z_v/k \xrightarrow{p} 0 \text{ and } \lambda' z_u/k \xrightarrow{p} 0$$

$$z_v' z_v/k \xrightarrow{p} 1 \text{ and } z_v'\eta/k \xrightarrow{p} 0 \text{ (} z_v \text{ and } \eta \text{ are independent by construction)}$$

Putting these limits together, we have, as $k \to \infty$,

$$\frac{(\lambda + z_v)' z_u}{(\lambda + z_v)'(\lambda + z_v)} \xrightarrow{p} \frac{\delta}{1 + \Lambda_\infty}$$

In the limit that $\Lambda_\infty = 0$, TSLS is consistent for the *plim* of OLS!

Comments

- Strictly this calculation isn't right – it uses sequential asymptotics ($T \to \infty$, then $k \to \infty$). However the sequential asymptotics is justified under certain (restrictive) conditions on K/T (specifically, $k^4/T \to 0$)
- Typical conditions on k are $k^3/T \to 0$ (e.g. Newey and Windmeijer (2004))
- Many instruments can be turned into a blessing (if they are not too weak! They can't push the scaled concentration parameter to zero) by exploiting the additional convergence across instruments. This can lead to bias corrections and corrected standard errors. There is no single best method at this point but there is promising research, e.g. Newey and Windmeijer (2004), Chao and Swanson (2005), and Hansen, Hausman, and Newey (2006))

Comments, ctd.

- For testing, the AR, LM, and CLR are all valid under many instruments (again, slow rate: $k \to \infty$ but $k^3/T \to 0$) in the classical IV regression model; the CLR continues to be essentially most powerful (the power of the AR deteriorates substantially because of the large number of restrictions being tested)
- An important caveat in all of this is that the rates suggest that the number of instruments must be quite small compared to the number of observations. (The specific rate at which you can add instruments depends on their strength – the stronger the instruments, the more you can add; see the discussion in Hansen, Hausman, and Newey (2006) for example.) Consider the $k^3/T \to 0$ rate:

  with $T = 200$ and $k = 6$, $k^3/T = 1.08$.

  with $T = 329{,}509$ and $k = 178$, $k^3/T = 17$ (!)

Comments, ctd.

- There is interesting recent work on many instruments using ideas of dynamic factor models – postpone discussion of this until the discussion of DFMs. This is conceptually different (uses information in the instruments themselves to address the many instrument problem, without reference to Y)

Instrument selection

- Donald and Newey (2001) provide an information criterion instrument selection method in the classical linear IV model that applies when some instruments are strong (θ strongly identified) and others possibly weak. Problem with is that you need to know which are strong.
- Unaware of instrument selection methods that are appropriate when all instruments are possibly weak.

Final comments on many instruments

- Strong instruments: more instruments, more efficiency
- Weak instruments: more weak instruments, less reliable inference – more bias, size distortions (using standard estimators – two-step and iterated GMM)

- Don't be fooled by standard errors that get smaller as you add instruments. Remember the result that $\hat{\beta}^{TSLS} - \hat{\beta}^{OLS} \xrightarrow{p} 0$ as $k \to \infty$ (and $k^3/T \to 0$) when all but a few instruments are irrelevant.
- Some gains seem to be possible in theory (papers cited above) by exploiting the idea of many instruments but the theory is delicate: bias adjustments and size corrections that hold for rates such as $k \to \infty$ but $k^3/T \to 0$, but break down for k too large. Work needs to be done before these are ready for implementation
- For now, the best advice is to restrict attention to relatively few instruments, to use judgment selecting the strongest (recent lags, not distant ones), and to use relatively well understood.

Bottom line recommendations

- Weak instruments/weak identification comes up in a lot of applications
- In the linear case, it is helpful to check the first-stage F to see if weak instruments are plausibly a problem.
- TSLS and 2-step efficient GMM can give highly misleading estimates if instruments are weak.
- TSLS and 2-step GMM confidence intervals, constructed in the usual way (± 1.96 standard errors) are highly unreliable (can have very low true coverage rates) if instruments are weak.
- If you have weak instruments, the best thing to do is to get stronger instruments, but barring that you should use econometric procedures that are robust to weak instruments. Robust procedures give valid inference even if the instruments are weak.

Bottom line recommendations, ctd.

- In the linear case with m=1 and no serial correlation, the CLR and CLR confidence intervals are recommended. Estimation by LIML is preferred to TSLS, but LIML can deliver very large outliers. Fuller is also a plausible option (see above).
- In the general nonlinear GMM case, GMM-AR confidence sets are recommended, but care must be taken in interpreting these (see discussion above). If you must compute an estimator, CUE seems to be the best choice given the current state of knowledge.